The Henryk Niewodniczański
Institute of Nuclear Physics
Polish Academy of Sciences

Study of the Bose–Einstein correlations in proton-proton and proton-lead collisions at LHCb

Author:
Bartosz Małecki

Supervisor:
dr hab. inż. Marcin Kucharczyk,
prof. IFJ PAN

Thesis submitted for the degree of
Doctor of Philosophy in Physics

https://doi.org/10.48733/msd.bm.18.02.2022

Kraków, November 2021

Abstract

This thesis describes a study of the Bose–Einstein correlations for same-sign charged pions originating from proton-proton and proton-lead collisions recorded in the LHCb experiment at $\sqrt{s} = 7\,\text{TeV}$ centre-of-mass energy and $\sqrt{s_{\text{NN}}} = 5.02\,\text{TeV}$ centre-of-mass energy per nucleon. Both measurements are the first of this type performed in the forward region at LHC energies. The proton-proton (proton-lead) dataset used in the analysis was recorded in 2011 (2013) and corresponds to an integrated luminosity of $1.0\,\text{fb}^{-1}$ ($1.6\,\text{nb}^{-1}$). Correlation parameters are determined for different regions of charged-particle multiplicity. It is observed that the correlation radius (the intercept parameter) increases (decreases) with the charged-particle multiplicity, which is consistent with observations from other experiments at the LHC in the central rapidity region. The measured correlation radii scale linearly with the cube root of the charged-particle multiplicity. Such a behaviour is compatible with predictions based on the hydrodynamic models. Moreover, hints for a dependence of the correlation radius on pseudorapidity are observed.

The thesis also reports the developments in the simulation of the RICH detectors for the upgrade of the LHCb experiment. The main focus of this work is the preparation for the Run3 data-taking period. The developed tools are used to assess the performance of the RICH system in the Run3 conditions with different detector configurations. Conducted studies indicate that the upgraded RICH detectors well meet their nominal performance expectations.

Streszczenie

Niniejsza rozprawa opisuje wyniki badań korelacji Bosego–Einsteina dla jednoimiennie naładowanych pionów pochodzących ze zderzeń proton-proton oraz proton-ołów zarejestrowanych w eksperymencie LHCb przy energii odpowiednio $\sqrt{s} = 7\,\text{TeV}$ w układzie środka masy oraz $\sqrt{s_{\text{NN}}} = 5.02\,\text{TeV}$ w układzie środka masy przypadającej na nukleon. Obydwa pomiary są pierwszymi tego typu wykonanymi w obszarze do przodu przy energiach LHC. Wykorzystane w analizie dane proton-proton (proton-ołów) zostały zebrane w 2011 (2013) roku i odpowiadają scałkowanej świetlności $1.0\,\text{fb}^{-1}$ ($1.6\,\text{nb}^{-1}$). Parametry korelacji są zmierzone dla różnych przedziałów krotności cząstek naładowanych. Wyznaczone promienie korelacji (parametry przecięcia) rosną (maleją) z krotnością cząstek naładowanych, co jest zgodne z obserwacjami z innych eksperymentów przy LHC w centralnym obszarze pseudopospieszności. Zmierzone promienie korelacji skalują się liniowo z pierwiastkiem sześciennym z krotności cząstek naładowanych. Takie zachowanie jest zgodne z przewidywaniami bazującymi na modelach hydrodynamicznych. Ponadto, uzyskane wyniki wskazują na zależność promienia korelacji od pseudopospieszności.

W rozprawie przedstawiono również rozwój symulacji detektorów RICH dla zmodernizowanego eksperymentu LHCb. Głównym celem tych prac jest przygotowanie do okresu zbierania danych Run3. Opracowane rozwiązania są wykorzystane do oceny wydajności systemu RICH w warunkach panujących w trakcie Run3 przy różnych konfiguracjach detektorów. Przeprowadzone badania wskazują, że zmodernizowane detektory dobrze spełniają pierwotne założenia co do ich wydajności.

Acknowledgements

I would like to express my gratitude to my supervisor, dr hab. inż. Marcin Kucharczyk, for his guidance and help in the preparation of this thesis. In addition, I would like to thank prof. dr hab. Mariusz Witek for the opportunity of joining the LHCb IFJ PAN group in the first place. I am also grateful to Dr. Carmelo D'Ambrosio and Dr. Sajan Easo, who were the supervisors of my *Doctoral Student Program* at CERN. I would like to thank them and other members of the LHCb RICH group for many interesting discussions on the RICH system and for the opportunity of contributing to this exciting project. I would also like to express my gratitude to Richa Nevatia for proofreading this thesis.

This study has been supported by KNOW (PhD scholarship) and the National Science Centre (grant numbers 2013/11/B/ST2/03829 and 2018/29/N/ST2/01641). Part of this research has been done using the PLGrid Infrastructure.

Contents

Chapter 1

Introduction

Multiparticle production in relativistic particle and nuclear collisions is a basic, yet a complex process in the field of the high-energy physics. It is a multistage phenomenon which cannot by fully described by calculations based on first principles, *e.g.* using the perturbative *quantum chromodynamics* (QCD) theory [1], and many aspects of its characterisation rely on the introduction of phenomenological models. For this reason, it is a field where significant developments are often driven by experimental observations that provide an input and allow for the addition of constraints on such models. It has been recognised in the high-energy physics community that the so-called *small systems* – such as proton-proton (pp) and proton-ion (pA) collisions – are particularly interesting in these studies, as their lifetimes are significantly shorter than the ones in the heavy-ion (AA) collisions, thus presenting a better experimental insight into the early system dynamics and the initial geometry [2–5].

One of the observables that provide valuable information on the system evolution is the HBT radius. It is named after two astronomers, Hanbury-Brown and Twiss, who introduced a novel method into the radio astronomy – intensity interferometry – to measure the angular dimensions of distant stars [6–8]. A similar approach was first employed in particle physics by Goldhaber *et al.* in 1959 [9,10]. The observed radii provide information on the particle-emitting source geometry [11] at the moment of the kinetic freeze-out. The HBT interferometry allows measuring momentum correlations between identical particles originating from the same particle-emitting source. These correlations are quantum in nature and emerge from statistics describing the given particle system. In the case of bosons, which obey the Bose–Einstein statistics, an enhanced production of identical particles with a small relative momentum is observed and this effect is referred to as the Bose–Einstein correlations (BEC). For fermions, which are described by the Fermi–Dirac statistics, the effect is opposite and is named as the Fermi–Dirac correlations (FDC). Effectively, the correlations measure scales that are referred to as *lengths of homogeneity* [12, 13], which correspond to a limited region of the particle-emitting source surface (in opposition to the full source size). These values can be related to the geometrical size of the particle-emitting source [2, 11].

The current thesis presents a study of the two-particle Bose–Einstein correlations for same-sign charged pions in two types of small systems in the *Large Hadron Collider beauty* (LHCb) experiment [14, 15]. The analysis employs pp and pPb collisions recorded at $\sqrt{s} = 7\,\text{TeV}$ centre-of-mass energy and $\sqrt{s_{\text{NN}}} = 5.02\,\text{TeV}$ centre-of-mass energy per nucleon, respectively. The study is performed in bins of the charged-particle multiplicity.

The LHCb detector acceptance ($2.0 < \eta < 5.0$) is unique among other experiments at the *Large Hadron Collider* (LHC) [16] and delivers results that are complementary to the measurements performed in the central rapidity region. It provides an exclusive input on the particle production process in the forward direction at LHC energies for the development of the theoretical models. Furthermore, a direct comparison of the results obtained by a single experiment in two different systems allows the establishment of a strong reference on the potential variations between the pp and the pPb collisions. The analysis of the BEC effect in the pp system described in this thesis has been published by the LHCb experiment [17], and the publication procedure for the pPb study is ongoing [18]. Furthermore, the Bose–Einstein and some other types of two-particle correlations at LHCb have been discussed in a number of articles by the author of this thesis [19–22].

The author has also been actively involved in the development of the future *Ring-Imaging Cherenkov* (RICH) detectors at LHCb. This activity has been carried out in terms of the *Doctoral Student Program* at CERN (*European Organization for Nuclear Research*) [23] within a project named 'LHCb RICH Upgrade: Photon detector installation and software development'. As part of this effort, the author has participated in the development of the simulation of the upgraded RICH detectors, and the progress on this matter has been reported in internal experiment documentation. The developed tools have been used to evaluate the expected performance of the modernised system in different detector configurations, which has been employed, for example, in the investigations described in Ref. [24].

The thesis is organised as follows. An overview of both the theoretical and experimental insights into the HBT measurements as well as a broader physical context of the present study is given in Chap. 2. The current LHCb experiment and its planned upgrades are described in Chap. 3. A special focus is given to the development of the modernised RICH detectors. The study of the BEC effect in the two collision system types is reported in Chap. 4, which is followed by a discussion of the analysis results in Chap. 5. A summary of the studies performed within this thesis is given in Chap. 6.

Chapter 2

Theory overview and previous experimental results

Experimental data is crucial in providing constraints for the development of the phenomenological models describing the process of multiparticle production. One of the prominent examples of the interplay between the theoretical and experimental insights into this phenomenon has been given by the discovery of a strongly-coupled *quark-gluon plasma* (QGP) [25] in heavy-ion collisions at the *Relativistic Heavy Ion Collider* (RHIC) [26–29], which was later confirmed at the LHC [30, 31]. The observed collective behaviour[1] of the produced particles, together with some other signatures – such as the jet quenching in large systems [34, 35], has been interpreted as a creation of a new state of hot matter (in opposition to cold matter in a weakly-coupled gas of hadrons). The evolution of the quark-gluon plasma is found to be described well in terms of the relativistic viscous hydrodynamics [36] which shows that the new medium has properties of a nearly *perfect* fluid [37] (it refers to the extremely low viscosity of QGP, which is close to the theoretical bounds on this parameter for any fluid [38, 39]). The fact that the quark-gluon plasma is created in the heavy-ion collisions supported the historical distinction between the latter and the small systems, where the application of a hydrodynamic description was thought to be not justified.

Another surprise in the field arrived with the observation of collective phenomena similar to those in the heavy-ion collisions also in the small systems – in high-multiplicity pp [40–43] and pPb [44–47] collisions at LHC and in the dAu (deuteron-gold) system [48] at RHIC. Up to now, many of the signatures that are commonly associated with the QGP formation in heavy-ion collisions have also been confirmed in the small systems [49]. It is worth noting, that at this point *e.g.* no jet quenching has been observed in the small systems, but it is expected that this effect would be significantly less prominent than in the AA collisions, due to the shorter lifetimes of the potentially created medium (studies in this area are ongoing [33, 49]). The increasing number of similarities between the various systems has sparked an active debate on whether the detected collective phenomena can be of the same physical origin [32, 33] and, if they are, what are the minimal conditions for the QGP formation.

A word on the limits of validity for the relativistic fluid dynamics is in order here. In

[1] Following Refs. [32, 33], the *collectivity* is meant here as a correlated behaviour in the particle production (*e.g.* a preferred direction, or similar momenta of the emitted particles), regardless of its physical origins.

a classical view, hydrodynamics requires a medium that is at thermal equilibrium, or at least close to it, which is not the case in the highly inhomogeneous small systems. For this reason, despite the early considerations on using the hydrodynamic description in the particle collisions [50], this approach was never widely favoured in the field, where models based on the QCD calculations prevailed [33]. However, it was understood that the criterion of the thermal equilibrium is not easily satisfied even in the heavy-ion collisions [51, 52]. Recent developments in the relativistic viscous fluid dynamics model allowed for the determination that this description can be applied even to systems that are far from equilibrium, as long as the so-called nonhydrodynamic modes are dominated by the hydrodynamic ones (the moment when this condition begins to be satisfied in the system evolution is referred to as the *hydrodynamisation* time – see Sect. 2.2.1) [5, 51]. It was found that for a fluid with the QGP properties (nearly inviscid fluid), the above criterion corresponds to a droplet of a size of 0.15 fm [52, 53], which does not make it unreasonable to apply the hydrodynamic picture to the *pp* collisions (comparing to the proton size of ~ 0.86 fm) or even to the e^+e^- system [54]. It should be noted that this reasoning refers only to the theoretical limits of the relativistic fluid dynamics description, leaving the question on whether the QGP is actually formed in the small systems for the interpretation of the experimental results.

The discussion on the origins of the collective phenomena in the small and the large systems is focused on the distinction whether the observed collectivity emerges at the level of final-state momentum correlations or if it is already imprinted in the initial-state momentum anisotropy. The former predictions are often explained to arise in the process of hydrodynamic evolution, which in principle propagates the initial-state spatial correlations into the final-state momentum anisotropy [5, 33]. The momentum correlations already in the initial state are usually reproduced within a *Color Glass Condensate* (CGC) framework [55, 56], which is a QCD-based effective field theory where a phenomenon of gluon saturation is assumed [57]. While other approaches are also discussed, they do not seem to describe the data as universally as the approaches mentioned above [32]. It has been appreciated by the theoretical and experimental community that the discrimination between the competing models requires simultaneous comparisons to a variety of experimental data [11], due to a typically large number of free parameters in such descriptions[2]. The current hydrodynamic models have been shown to be extremely successful in describing the data, being able to properly reproduce the dynamics of the *pp*, the *p*Pb and the PbPb collisions with a single set of input parameters [62] (however, without strong conclusions on the QGP formation in all of those systems). The CGC framework is also able to predict some of the patterns recognised in the experimental results, nevertheless, certain data features remain a challenge for this approach [33, 54]. One of those aspects has been shown by a study from RHIC which emphasized the critical impact of the initial geometry on the observed collective behaviour [63–67]. The mentioned series of analyses employed three types of collisions – *p*Au, *d*Au and ^{3}HeAu – to engineer intrinsically different initial geometries (dominantly circular, elliptical and triangular, respectively). It has been found that the patterns observed in the data cannot be reproduced by the CGC framework without the hydrodynamic evolution of the system and that they are naturally emerging within the expanding medium picture. Those observations strongly support

[2]It should be noted that the experimental data is also used to put constraints on the mentioned model parameters. An example of the hydrodynamic framework can be given here, where results on a number of observables are used simultaneously to determine the properties of the evolving fluid [58–61].

the creation of QGP in those three types of collisions.

In the long-established debate between the two frameworks, there is no doubt that the initial-state momentum correlations are also present, but the question is at what multiplicity levels they would prevail the final-state phenomena from the, apparently, hydrodynamic evolution [68]. Ultimately, both approaches could be combined to achieve the highest precision. A recent study of the small systems employing the CGC framework to provide detailed information on the initial condition (comprising not only the initial geometry, but also the early system dynamics) for the subsequent hydrodynamic expansion acknowledges that the hydrodynamic picture dominates the qualitative behaviour of the data [69]. However, the specifics of the initial dynamics is stated to be crucial for quantitative comparisons, and the increasing role of the initial-state phenomena at lower multiplicities is confirmed. The described developments represent a change in the paradigm of the particle production modelling, where, with the increasing number of similarities observed between the small and the large systems, their historical distinction into different physics categories may be not so well founded [5].

As already mentioned, the HBT studies are one of the measurements that can be employed to put constraints on the phenomenological models described above and to gain further insight into the process of multiparticle production. The current chapter provides an introduction to the theoretical concepts of the BEC effect (see Sect. 2.1) and describes methods of incorporating this phenomenon into the high-energy physics simulation frameworks (see Sect. 2.2). An overview of the experimental results in the HBT measurements is given in Sect. 2.3.

2.1 Introduction to BEC

As already mentioned, the BEC and the FDC effects emerge from the quantum statistics describing the given particle system, namely due to the symmetrisation (antisymmetrisation) of the wavefunction representing a system of bosons (fermions). The resulting correlations can provide information on the geometry of the particle-emitting source at a femtometer scale[3]. The correlations have a complex structure and in general case, their relation to the underlying source distribution is highly nontrivial (or even not analytically known) [2, 70]. The study described in the current thesis is based on the assumption of static, spherically-symmetric sources that can be characterised by univariate distributions, hence the detailed parts of the theory discussion are narrowed down to such a case. This class of sources is commonly used in the HBT analyses, since, as shown in the further sections, the measured correlations in this case can be relatively simply related to the effective size of the particle-emitting source.

2.1.1 Quantum correlations

Let us begin with introducing a generic form of the two-particle quantum correlations description, which can be conveniently done using a Wigner-function formalism [71–73]. In such a representation, a source function $S(x, k)$ is a covariant Wigner transform of

[3]Thus, the HBT measurements are also often referred to as *femtoscopy* (although the latter has a broader meaning, encompassing also, among others, studies of strong interactions [11]).

the source density matrix and corresponds to a classical probability that a particle of momentum k is emitted at point x, where both x and k are four-vectors [2].

Then, a single- and a two-particle invariant momentum distribution of the produced particles – $N_1(k)$ and $N_2(k_1, k_2)$, respectively – can be expressed as [2]

$$N_1(k) = \int \mathrm{d}^4 x S(x,k) \ , \tag{2.1}$$

$$N_2(k_1,k_2) = \int \mathrm{d}^4 x_1 \mathrm{d}^4 x_2 S(x_1,k_1) S(x_2,k_2) \left| \Psi_{k_1,k_2}(x_1,x_2) \right|^2 , \tag{2.2}$$

where $\Psi_{k_1,k_2}(x_1,x_2)$ is a wavefunction describing the system of two particles. A two-particle correlation function is defined as

$$C_2(k_1,k_2) \equiv \frac{N_2(k_1,k_2)}{N_1(k_1) N_1(k_2)} = \frac{\int \mathrm{d}^4 x_1 \mathrm{d}^4 x_2 S(x_1,k_1) S(x_2,k_2) \left| \Psi_{k_1,k_2}(x_1,x_2) \right|^2}{\int \mathrm{d}^4 x_1 S(x_1,k_1) \int \mathrm{d}^4 x_2 S(x_2,k_2)} \ . \tag{2.3}$$

Assuming a fully incoherent (chaotic) emission and that there are no final-state interactions between the particles in a pair (or that they are under experimental control), a plane-wave approximation can be used to construct a symmetric wavefunction for a bosonic system

$$\Psi_{k_1,k_2}(x_1,x_2) = \frac{1}{\sqrt{2}} \left[e^{ik_1 x_1 + ik_2 x_2} + e^{ik_1 x_2 + ik_2 x_1} \right] \ . \tag{2.4}$$

After combining Eqs. (2.3) and (2.4), the correlation function can be expressed in terms of the relative momentum of the particles in a pair $q = k_1 - k_2$ as [2]

$$C_{2,\mathrm{BEC}}(k_1,k_2) = 1 + \frac{\left| \tilde{S}(q,K) \right|^2}{\tilde{S}(0,k_1) \tilde{S}(0,k_2)} \approx 1 + \frac{\left| \tilde{S}(q,K) \right|^2}{\left| \tilde{S}(0,K) \right|^2} \ , \tag{2.5}$$

where $\tilde{S}(q,K)$ is the Fourier transform of the source function $S(x,K)$

$$\tilde{S}(q,K) = \int \mathrm{d}^4 x e^{iqx} S(x,K) \tag{2.6}$$

and K corresponds to the pair mean momentum

$$K = \frac{1}{2}(k_1 + k_2) \ . \tag{2.7}$$

In practice, the correlation function is usually studied using a Lorentz-invariant Q variable, which is analogous to the relative momentum of the two particles in a pair defined above

$$Q \equiv \sqrt{-q^2} = \sqrt{-(k_1 - k_2)^2} \ . \tag{2.8}$$

The form of Eq. (2.5) provides a straightforward relation between the observable correlation function and the assumed (known) source function distribution.

2.1.2 Intercept parameter λ

Apart from the correlation radius, studying the BEC effect allows for a determination of the so-called intercept parameter λ, which corresponds to the extrapolated value of the correlation function at $Q = 0\,\mathrm{GeV}$ [2][4]. This observable can be interpreted within the core-halo model [74, 75] which assumes that the particle emission can take place in a central *core* or in an extended *halo* originating from long-lived resonance decays. If the decay widths of the latter are below the experimental Q resolution, then these contributions are not resolvable within the measured correlation function [76]. However, their effect can be accounted for in the core-halo picture, by assuming that the source function is split into two separate contributions from the mentioned regions, $S_\mathrm{c}(x,k)$ and $S_\mathrm{h}(x,k)$, which leads to

$$S(x,k) = S_\mathrm{c}(x,k) + S_\mathrm{h}(x,k) \ . \tag{2.9}$$

Then, Eq. (2.5) takes a slightly modified form

$$C_{2,\mathrm{BEC}}(k_1,k_2) = 1 + \lambda(K)\,\frac{\left|\tilde{S}_\mathrm{c}(q,K)\right|^2}{\tilde{S}_\mathrm{c}(0,k_1)\tilde{S}_\mathrm{c}(0,k_2)} \approx 1 + \lambda(K)\,\frac{\left|\tilde{S}_\mathrm{c}(q,K)\right|^2}{\left|\tilde{S}_\mathrm{c}(0,K)\right|^2} \ , \tag{2.10}$$

where the intercept parameter $\lambda(K)$ indicates the fraction of particles originating from the core

$$\lambda(K) = \left(\frac{N_\mathrm{c}(K)}{N_\mathrm{c}(K) + N_\mathrm{h}(K)}\right)^2 \ . \tag{2.11}$$

In reality, the value of the intercept parameter can be modified by various experimental effects (such as misreconstructed pairs of nonidentical particles, or imperfections in the description of the final-state interactions – see Sect. 4.1), hence this quantity is also often referred to as a *correlation strength*. However, despite its sensitivity to the experimental methods, it can carry valuable physics information – not only on the fraction of the unresolved halo contribution, but also on the level of (in)coherence of the particle emission in the core [2, 77].

2.1.3 Static sources

Let us now consider a simple example of static, spherically-symmetric sources to illustrate the basic concepts of the correlation function and its relation to the underlying source distribution. In such a case, the source function can be simplified to factorised univariate terms for the spatial and the momentum components [76]

$$S(x,k) = f(x)g(k) \ , \tag{2.12}$$

where the normalisation of the spatial distribution is set to

$$\int \mathrm{d}x f(x) = 1 \tag{2.13}$$

[4] If not indicated otherwise, natural units with $\hbar = c = 1$ are used throughout this thesis. In particular, the measured R values are initially given in GeV^{-1}. Conversion to femtometers is done using relation $R[\,\mathrm{fm}] = 0.197 \times R[\,\mathrm{GeV}^{-1}]$.

and both x and k are from now on one-dimensional variables (in opposition to the generic case of x, k four-vectors discussed in Sect. 2.1.1). Then, the single-particle spectrum is given by

$$N_1(k) = \int \mathrm{d}x f(x) g(k) = g(k) \ . \tag{2.14}$$

Following the reasoning of Sect. 2.1.1, one arrives at

$$C_{2,\mathrm{BEC}}(k_1, k_2) = 1 + \left|\tilde{f}(q)\right|^2 \ , \tag{2.15}$$

where

$$\tilde{f}(q) = \int \mathrm{d}x e^{iqx} f(x) \ , \tag{2.16}$$

which means that the correlation function directly measures the absolute value squared of the Fourier transform of the source spatial distribution [76].

Assuming a univariate Gaussian distribution of the source density, the correlation function becomes

$$C_{2,\mathrm{BEC}}(Q) = 1 + e^{-R^2Q^2} \ , \tag{2.17}$$

where R stands for the correlation radius, which is defined as the width of the source density distribution. Another commonly used parametrisation is the Lorentzian (or Cauchy) source distribution, which results in an exponential form of the correlation function

$$C_{2,\mathrm{BEC}}(Q) = 1 + e^{-|RQ|} \ . \tag{2.18}$$

Both of the above parametrisations allow obtaining a characteristic scale of the source spatial distribution, which can be interpreted as the effective size of the particle-emitting source at the kinetic freeze-out (see Sect. 2.2.1) [2,70]. They are often employed in the HBT studies, depending on which one better describes the relevant data, to allow for an interpretation of the measured correlation radii and for comparisons between various experiments. However, neither of these distributions describe the shape of the correlation peak with high precision, hence more elaborate parametrisations are also used (see Sect. 2.1.4).

2.1.4 Lévy-type (static) sources

A more general class of sources (with respect to the simple cases discussed in Sect. 2.1.3) can be described by the so-called symmetric Lévy-stable distributions [76]. In the case of static, univariate sources, the Lévy-type correlation function can be expressed as

$$C_{2,\mathrm{BEC}}(Q) = 1 + e^{-|RQ|^{\alpha_\mathrm{L}}} \ , \tag{2.19}$$

where α_L is a parameter that can take values in range of $0 < \alpha_\mathrm{L} \leqslant 2$ and is referred to as a Lévy index of stability. This type of parametrisation usually provides better quality of the fits to the data than the simple cases considered in Sect. 2.1.3, however, the interpretation of the measured correlation radius is not straightforward in this case. It is worth noting that for two specific values of the α_L parameter of unity and two, the exponential and Gaussian parametrisations are reproduced.

It has been found in a study employing the Lévy parametrisation [78] that indeed the measured index of stability values are neither consistent with the Gaussian nor

the exponential scenario, with the results at the level of $\alpha_{\mathrm{L}} \sim 1.2$. Some of the potential explanations on processes leading to the observation of Lévy-type sources, including QCD critical phenomena, jet fragmentation and an anomalous diffusion in expanding systems, are discussed in Refs. [79–82].

2.2 Modelling of the BEC effect

The focus of the current section is to explain how the BEC effect can be introduced into the simulation of high-energy physics collisions. An example is given in Sect. 2.2.2 based on the implementation of this phenomenon in the PYTHIA generator [83, 84]. It is preceded by a description of how the particle collisions are modelled in general (see Sect. 2.2.1). The latter is illustrated using an example of the Björken model [85], which is commonly employed in the description of relativistic heavy-ion collisions. Keeping in mind the general discussion on the differences between the small and the large systems mentioned at the beginning of Chap. 2, the applicability of a similar picture in the *pp* collisions seems to be gaining an increasing support in the field [33]. The intention of introducing this model is to provide a general idea on the characteristic processes in the particle production, not to state whether the hydrodynamic expansion is present in (all) small systems or not. It should also be noted that the various stages of the system evolution can be described in different ways by competitive phenomenological models, while only some example implementations are mentioned in the following. The usage of the phenomenological models is necessary due to the nonperturbative nature of the simulated processes, which makes it not feasible to describe them using the first principles of the perturbative QCD theory. A detailed review on the available models and on the particle production simulation can be found, for example, in Refs. [5, 32].

2.2.1 Particle production model

As already mentioned, the relativistic nuclear collisions are usually described using the Björken model [33, 85], which is illustrated in Fig. 2.1. The first stage of the system evolution is the actual nuclei collision, where the incoming projectiles are highly Lorentz-contracted due to their relativistic velocities. There are several models that allow the prediction of the distribution of the energy deposit in this process [32]. The state of the system (*e.g.* its geometrical size) at which the hydrodynamic evolution is introduced, is referred to as the *initial condition* and constitutes a critical input for the relativistic fluid dynamics calculations. One of the methods that allow the reconstruction of this initial state is the Glauber model [86]. It is a purely geometrical approach, where an interaction between two individual nucleons in the colliding nuclei is declared if some defined areas (based on the interaction cross-section) around those nucleons overlap. This picture is quite successful in the description of heavy-ion collisions, where the initial geometry is dominated by the nucleon positions in the colliding nuclei. The subnucleonic structure can be reproduced in an extension of the Glauber model, where so-called constituent quarks are introduced within the individual nucleons [62, 87]. While the Glauber model, even with its extensions, is still a simple geometrical method (in particular, it does not imply any system evolution at the initial stage, providing a direct input for the hydrodynamic simulation), it has been successfully applied in the description of even the small systems like *pp* and

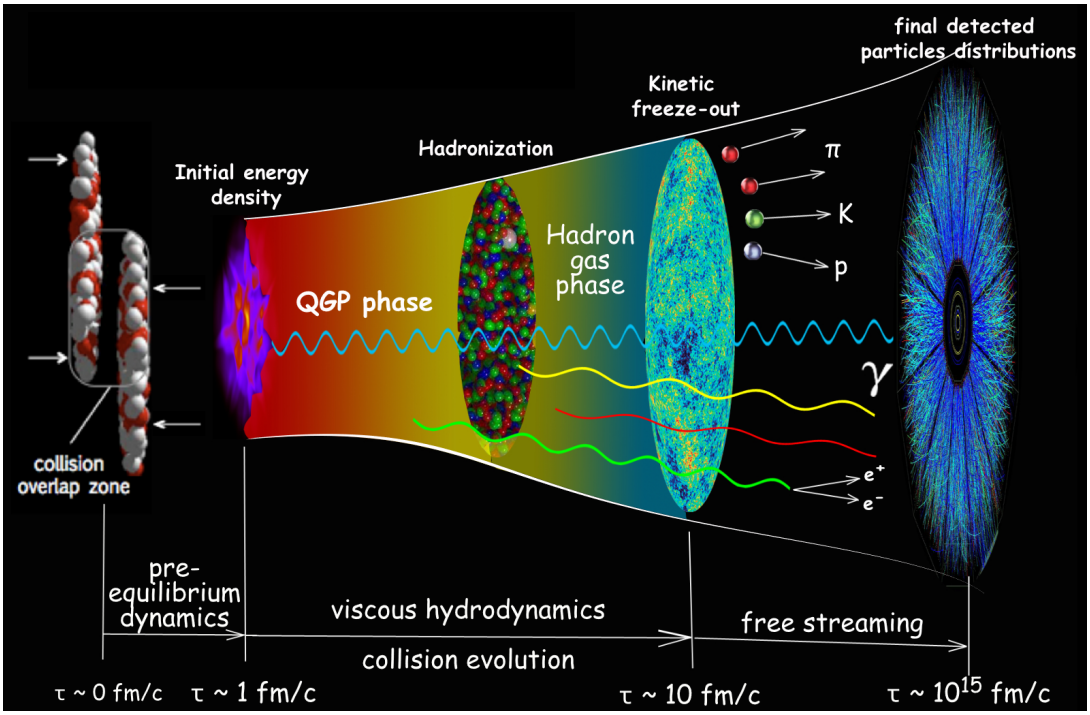

Figure 2.1: Schematic view of a relativistic nuclear collision. Various stages of the system evolution are indicated, such as the initial dynamics up to the hydrodynamisation time, the subsequent hydrodynamic expansion of the QGP phase and the hadronisation process. The hadronic scatterings continue up to the kinetic freeze-out phase, which is the point from which the particles propagate freely towards the detector sensors [89].

pA collisions, where the subnucleonic structure can be more relevant for the subsequent evolution [32, 62, 67]. An alternative approach is presented in the IP-Glasma model [88], which is based on the QCD calculations within the CGC framework [55, 56]. This solution allows the procurement of information on the initial condition including basic dynamics of the system at its earliest stages. It provides a detailed description of the initial state that can be propagated to the subsequent hydrodynamic simulation [3, 69].

The system evolution begins to be described in terms of the relativistic viscous hydrodynamics[5] when a *hydrodynamisation* time is achieved [5]. Those calculations are based on a QCD-motivated equation of state of the underlying fluid with properties constrained with the usage of the experimental data. The QGP medium cools down during the expansion until the point where a crossover transition takes place [90, 91] and the hot matter is confined into hadrons (this process is referred to as the hadronisation and is usually described in terms of the Cooper–Frye mechanism [92]). The created hadrons still experience inelastic and elastic scatterings, which terminate at the moment of the so-called chemical and kinetic freeze-out, respectively. This process is usually modelled via a hadron cascade simulation (*e.g.* Ref. [93]), where the individual hadrons propagate in straight

[5]Historically, the common understanding was that the system needs to be close to a thermal equilibrium to apply the hydrodynamic description, and the evolution before this moment was referred to as the preequilibrium dynamics. However, as mentioned at the beginning of Chap. 2, the developments in the relativistic hydrodynamics (and its understanding) allow its application even to systems that are far from equilibrium [5, 51], and the hydrodynamisation and *prehydrodynamisation evolution* terms are introduced. Such a nomenclature is used uniformly in the current thesis.

lines, unless they come close enough to each other to interact (where the relevant distance is motivated by the scattering cross-section of the given hadron species). After the kinetic freeze-out, the hadrons propagate freely with properties that can be observed in detectors.

2.2.2 BEC effect in the PYTHIA generator

The PYTHIA generator [83,84], which is employed in the current *pp* study (see Sects. 3.1.5 and 4.2.1), is used as an example of implementing the BEC effect in the high-energy physics simulation. This particular framework is based on a Lund string model [94], where the system evolution up to the hadronisation stage (with no hydrodynamic expansion) is governed by the fragmentation of the so-called colour strings between the underlying quarks and antiquarks (partons). The potential energy accumulated in the string is growing with the increasing distance in the quark-antiquark pair. If the energy is high enough, the string can break up and create a new pair. This process is continued as long as the energy of the strings allows for the creation of further pairs. The final hadrons are formed from the partons that arise from the partitions of the adjacent strings.

The BEC effect is introduced into the simulation after the phase of string fragmentation and modelling decays of short-lived resonances. At this stage, the Q values are calculated for each pair of identical particles in the simulated event. The BEC effect is introduced based on an assumed parametrisation of the correlation function (corresponding to the Gaussian or exponential shape – see Sect. 2.1.3). A shift in the momenta of the paired particles needed to reproduce a required parametrisation is determined. Once the shifts for all pairs in an event are calculated, momentum of each individual particle is modified according to the sum of the shifts determined for this particle. This process does not conserve the total energy in an event, and several different algorithms are available in PYTHIA to correct for this effect. When this procedure is finalised, the simulation continues with modelling decays of long-lived particles.

As shown above, reproducing the BEC effect in the simulation is not straightforward. The HBT correlations are introduced with various assumptions, such as the one on the parametrisation of the BEC effect. This picture can be further complicated by taking into account the final-state interactions in the studied particle pairs (see Sect. 4.1.2). The limitations mentioned here should be kept in mind while employing simulations including the BEC effect. For these reasons, as indicated in Sect. 4.2.1, the BEC modelling is not activated in the simulation used in the current studies.

2.3 Experimental results overview

The BEC and the FDC effects have been studied across many collision systems with a wide spectrum of energies at *e.g.* AGS [95], ISR [96–98], SPS [99–101], LEP [102–117], RHIC [78, 118–129], LHC and others [130, 131][6]. The sizes of the studied systems vary tremendously, from e^+e^- collisions at LEP to (mainly) AuAu collisions at RHIC and PbPb beams at LHC. At the LHC alone, the HBT effect has been investigated in *pp* [17,132–144], *p*Pb [134,140,141,143,145–147] and PbPb [134,140,141,148–150] systems.

[6]The names of some of the facilities mentioned in the text are explained in the following: AGS (*Alternating Gradient Synchrotron*), ISR (*Intersecting Storage Rings*), SPS (*Super Proton Synchrotron*), LEP (*Large Electron-Positron collider*).

The mentioned studies involved many generations of accelerators and experiments, resulting in the development of a mature community in the field and a common paradigm in the performed analyses. This allowed for a recognition of various systematic patterns in the correlation parameters, sometimes strikingly similar across many different collision systems [11]. Some of the observed trends are described in the following paragraphs. They include studies of the correlation parameters dependence on the charged-particle multiplicity N_{ch}, the average pair transverse momentum k_{T} ($k_{\mathrm{T}} = \frac{1}{2}|\vec{p}_{\mathrm{T}_1} + \vec{p}_{\mathrm{T}_2}|$, where $\vec{p}_{\mathrm{T}_1}$ and $\vec{p}_{\mathrm{T}_2}$ are the transverse momenta of the individual particles in a pair) and the rest mass (species) of the investigated hadrons.

2.3.1 Correlation radius *vs.* charged-particle multiplicity

It was observed already by the AFS experiment[7] [151] at ISR that the measured correlation radii increase with the charged-particle multiplicity in the collision [96]. This dependence has been scrutinised in many studies since then. Various variables that reflect the same behaviour are employed in the individual analyses. Two of them are defined here for the purpose of the further discussion: the mean charged-particle multiplicity $\langle N_{\mathrm{ch}} \rangle$ and the average local charged-particle multiplicity density (in pseudorapidity[8]) $\langle \mathrm{d}N_{\mathrm{ch}}/\mathrm{d}\eta \rangle$ (but it can also be *e.g.* the number of nucleons participating in a collision in the case of the heavy-ion studies). The correlation radii seem to scale universally with the cube root of the charged-particle multiplicity [11]. The data from systems spanning from *pp*, through *p*Pb, to AA collisions is shown in Fig. 2.2, where a slightly different slope of this relation between the three system groups mentioned above is visible. It is also clearly pronounced (especially for the AA collisions) that the correlation radii do not depend strongly on the type and energy of the colliding nuclei, as long as the comparison is done at similar multiplicity values. This suggests that the BEC effect shows (at most) a weak dependence on the collision energy, and the apparent increase of the measured radii at higher energies stems mainly from the fact that more energetic collisions can produce higher charged-particle multiplicities [11, 135]. The mentioned feature has also been distinctively illustrated *e.g.* by the CMS experiment for *pp* collisions, by showing that the results obtained at $\sqrt{s} = 7\,\mathrm{TeV}$ and $\sqrt{s} = 13\,\mathrm{TeV}$ energies are in good agreement (see Fig. 2.3 (left)).

The universal scaling mentioned above can indicate similarities in the evolution of different collision systems (in the case of the AA collisions, this behaviour is interpreted as a signature of the hydrodynamic expansion [36]). As mentioned at the beginning of Chap. 2, there is an active ongoing debate on whether the hydrodynamic description can also be applied to the small systems. The said scaling may not be so clear especially in the *pp* system, where the signs of a potential saturation of the correlation radius at high multiplicities in $\sqrt{s} = 7\,\mathrm{TeV}$ collisions have been reported by ATLAS [144]. However, the uncertainties of this measurement are quite significant in the high-multiplicity region. A more recent study for *pp* collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ by CMS [135] seems to be in agreement with the ATLAS results (within the uncertainties), but indicates that the potential saturation may be present at multiplicities much higher than the ones suggested by ATLAS (see Fig. 2.3 (right)).

[7] AFS – *Axial Field Spectrometer*.

[8] Pseudorapidity is defined as $\eta = -\ln(\tan\frac{\theta}{2})$, where θ is the angle between the beam and the particle momentum.

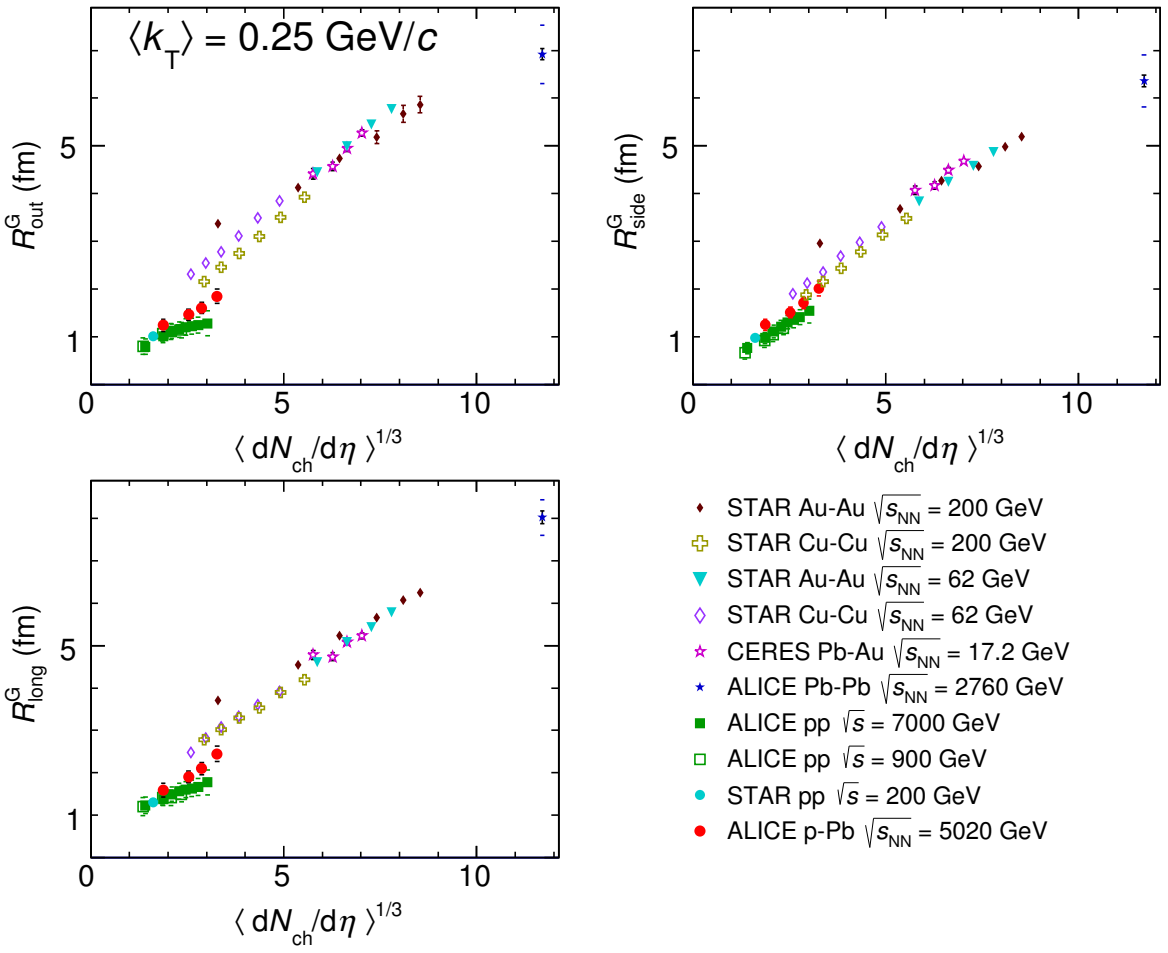

Figure 2.2: Comparison of the correlation radii, as a function of $\langle \mathrm{d}N_{\mathrm{ch}}/\mathrm{d}\eta \rangle^{1/3}$, determined for various collision systems and energies by CERES [101], STAR [121, 122, 125], PHENIX [126] and ALICE [137, 145] experiments. The presented radii are measured in multidimensional HBT analyses where the correlations are decomposed into different directions as indicated by the subscripts (see *e.g.* Ref. [145] for details on the multidimensional HBT studies). The figure is taken from Ref. [145].

Opportunities for direct comparisons between the measured correlation radii and predictions of the theoretical models are often limited, due to the fact that the relevant calculations are not always available for the given collision energy, experimental acceptance, *etc.* Furthermore, the scales predicted by the theoretical models may not always correspond to the experimentally measurable HBT radii, but rather to the geometrical size of the particle-emitting source at some stage of the system evolution [11]. However, comparing the qualitative trends can still provide valuable information on the system dynamics. In particular, predictions on the correlation radius dependence on multiplicity both in the hydrodynamic [3, 152–154] and CGC [3, 155, 156] framework are available also for the small systems. The former predict the correlation radius scaling with $\langle \mathrm{d}N_{\mathrm{ch}}/\mathrm{d}\eta \rangle^{1/3}$, while the latter conclude that the radii should increase up to some multiplicity values, with a saturation reached above that level. Both of these scenarios are tested by the CMS experiment in the *pp* collisions study at $\sqrt{s} = 13\,\mathrm{TeV}$ [135] (see Fig. 2.4). The radii

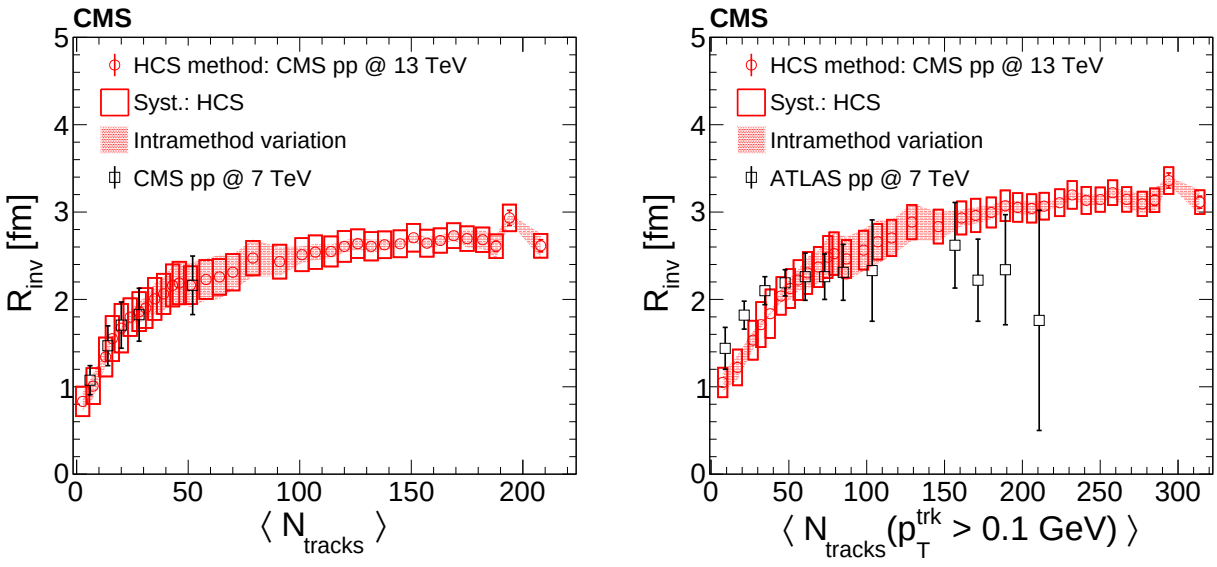

Figure 2.3: Correlation radii as a function of the average charged-particle multiplicity in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ measured by CMS [135] (red circles) compared to the results for pp collisions at $\sqrt{s} = 7\,\text{TeV}$ (black squares) from (left) CMS [134] and (right) ATLAS [144]. The error bars in the CMS case represent systematic uncertainties (statistical uncertainties are smaller than the marker size) and, in the ATLAS case, statistical and systematic uncertainties added in quadrature. The CMS data points in the right figure are adjusted to take into account the differences between both the analysis procedures and the determination of the charged-particle multiplicity in the two experiment acceptances, to be compatible with the ATLAS results and to allow a direct comparison of the measurements [135].

predicted within the CGC framework alone (with no subsequent hydrodynamic evolution) are significantly below the CMS data points. On the other hand, the qualitative CGC-motivated behaviour with a potential saturation at high multiplicities might be visible in the data. At the moment, the data does not allow the exclusion of any of the two scenarios, however such studies show the potential of the BEC measurements in providing constraints for the development of the theoretical models.

2.3.2 Transverse momentum (mass) dependence

Another interesting trend is related to the correlation radius dependence on the average pair transverse momentum (or transverse mass, $m_\text{T} = \sqrt{m^2 + {k_\text{T}}^2}$, where m is the rest mass of a single particle in the pair), which has been first observed for pion pairs (see *e.g.* Ref. [157]). An example of this relation in pp collisions at $\sqrt{s} = 7\,\text{TeV}$ is shown in Fig. 2.5 (left), where the correlation radii decrease with increasing k_T. Such a behaviour can be interpreted as a signature of the hydrodynamic expansion of the system [153, 154]. This effect is least prominent at the low multiplicity values, where the role of the hydrodynamic evolution is expected to be limited (or simply nonexistent). It has also been observed that the correlation radii become smaller with an increasing mass of the studied hadron species, which was a conclusion driven mainly by the excellent LEP measurements performed for a number of different types of hadron pairs such as pions [102], kaons [107]

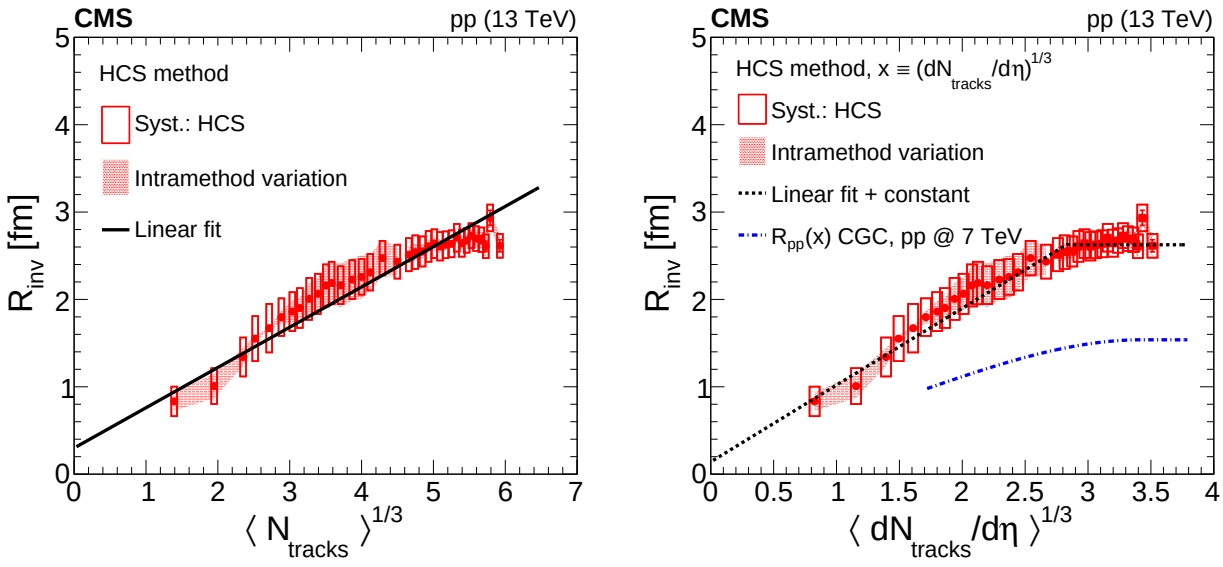

Figure 2.4: Testing the behaviour of the correlation radii predicted by some of the theoretical models performed by the CMS experiment in pp collisions at $\sqrt{s} = 13\,\text{TeV}$. (left) The correlation radius values are shown as a function of the cube root of the average charged-particle multiplicity, together with a linear fit illustrating the scaling expected from hydrodynamics (see *e.g.* Refs. [153, 154]). (right) The correlation radius dependence on the cube root of the average charged-particle multiplicity density is shown together with the predictions from the CGC framework. The linear plus constant function (dashed black lines) is shown to illustrate the qualitative behaviour suggested by the CGC (the matching point of the two lines is a result of a fit). The dot-dashed blue line is the direct CGC-based prediction for pp collisions at $\sqrt{s} = 7\,\text{TeV}$ [156]. Only statistical uncertainties are considered [135].

and lambdas [114]. Several theoretical models aiming to describe this dependence on the hadron mass (or transverse mass) were proposed [158–160]. However, a question raised by some of the authors [2] was whether the observed trend originated in the particle mass, or rather in the transverse mass, thus also being related to the previously-mentioned k_{T} dependence for pion sources. Due to the issues raised above, the usage of some of the mentioned models that aim to explain the correlation radius dependence based mainly on the particle species (and, therefore, mass) would be arguable [11].

Indeed, it has been shown *e.g.* by the ALICE collaboration that an approximate scaling of the correlation radius with m_{T} for different particle species can be observed in PbPb collisions (see Fig. 2.5 (right), where the results are shown in different centrality classes[9]). It is interpreted as a sign of the collective flow in such systems, when the evolution is largely independent of the scattering cross-section of the given particle type, but is driven mainly by the particle mass in a common flow field [11, 161]. It is not feasible to verify the statement on the universal m_{T} dependence in the LEP data, since the limited statistics only allowed for the determination of single parameters for the particles other than pions (with no distinction for different k_{T} or m_{T} bins). This kind of study (*i.e.*,

[9]Centrality is a measure of the impact parameter of a collision and can be related to the mean charged-particle multiplicity density, with the lower centrality values corresponding to the higher multiplicities.

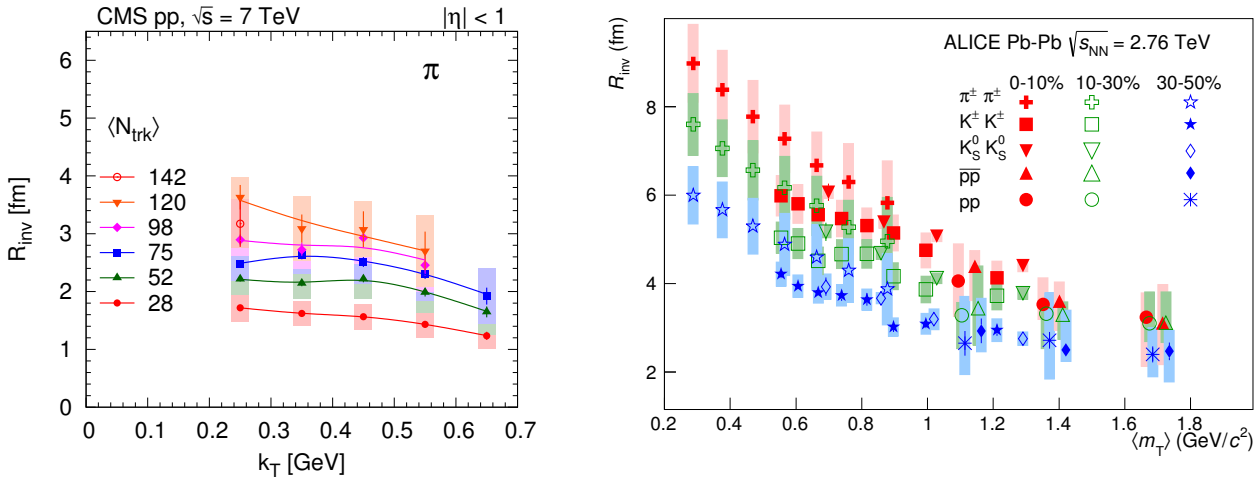

Figure 2.5: (left) k_{T} dependence of the correlation radius for several bins of charged-particle multiplicity measured in pp collisions at $\sqrt{s} = 7\,\mathrm{TeV}$ by CMS. Lines are drawn to guide the eye [134]. (right) Approximate scaling of the correlation radius with m_{T} for various particle pair types observed by the ALICE experiment in PbPb collisions. The results are shown in three centrality classes. Statistical (thin lines) and systematic (boxes) uncertainties are indicated [149].

performing the HBT measurements in multiple k_{T} or m_{T} bins for particles other than pions) has been done for the first time in pp collisions by the ALICE experiment [138] and was followed by an impressive series of analyses for other hadrons [139, 142, 143]. It seems that the approximate scaling with m_{T} can also be observed in the pp collisions, however, similar to the case of the multiplicity dependence, this picture is not as clear as in the AA collisions. In particular, at the lowest multiplicities the trend of the correlation radius decreasing with higher k_{T} tends to be reversed, which could point to the case where the hydrodynamic evolution is significantly limited (or not present at all) and the system dynamics becomes dominated by other processes. As indicated by the authors of these studies, further experimental and theoretical developments would be beneficial to better understand these phenomena [138, 139].

2.3.3 Particle pair rapidity

The last observation on the behaviour of the correlation parameters presented in this section is related to their sensitivity to the rapidity of the investigated particle pairs[10]. Although it is not as commonly studied as the multiplicity and k_{T} relations, results on this issue have been presented for pion pairs *e.g.* by the ATLAS experiment in pPb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$ (see Fig. 2.6). The observed radii display a slight, but nonnegligible dependence on the particle pair rapidity. It should be noted, however, that these interesting results are restricted to the central rapidity range. Performing similar measurements at LHCb could provide information on the particle production in the forward region, far from the acceptance limits of the other experiments at the LHC, thus presenting a unique input for the development of the theoretical models.

[10]The pair rapidity $y^*_{\pi\pi}$ is calculated relative to the nucleon-nucleon centre-of-mass frame, assuming that both particles in a pair have the mass of a pion [147].

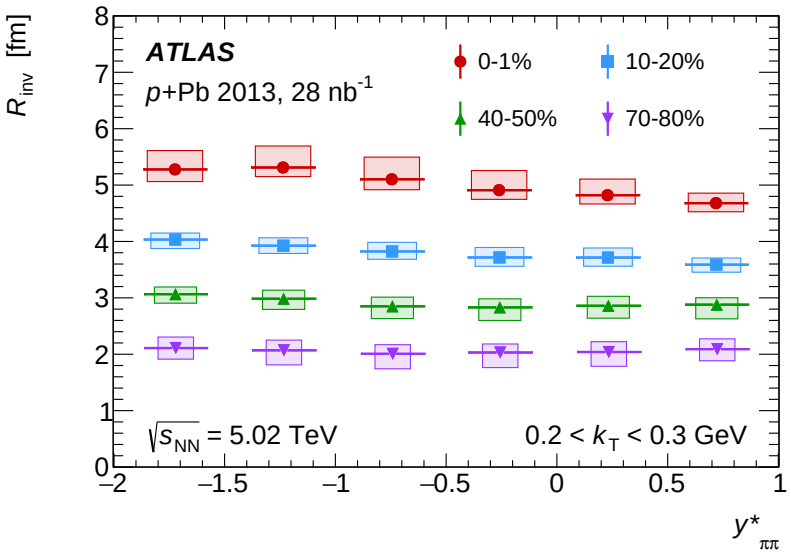

Figure 2.6: Correlation radii measured by the ATLAS experiment in the BEC study for same-sign charged pion pairs in pPb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$, as a function of pair rapidity. Four nonadjacent centrality intervals are shown. The vertical size of each box represents the systematic uncertainties, while the statistical ones are shown with vertical lines. The horizontal positions of the points are the average rapidity values in each interval and the horizontal lines indicate the standard deviation of the rapidity. The widths of the boxes differ among centrality intervals for visual clarity [147].

Chapter 3

LHCb experiment

The data used in this study was collected by the LHCb experiment [14, 15]. It is located at the CERN laboratory, which spans over the border between France and Switzerland near Geneva. The laboratory currently houses the most powerful particle accelerator in the world, the LHC machine [16]. This apparatus allows colliding protons at energies equivalent to 14 TeV in the pp centre-of-mass frame ($\sqrt{s} = 14\,\text{TeV}$). While the LHC is mainly operating with proton beams, for example lead ions can also be accelerated there, providing pPb and PbPb collisions. The full accelerator chain at CERN is complex and includes previous generations of machines. An overview of the accelerators and experiments located at CERN is shown in Fig. 3.1. Protons that are injected into the LHC are first preaccelerated to 50 MeV at the linear collider (LINAC2) [162] and then to 1.4 GeV at the *Proton Synchrotron Booster* (PSB) [163]. Subsequently, the protons are transferred to the *Proton Synchrotron* (PS) [164], where their energy increases to 26 GeV, and to the *Super Proton Synchrotron* (SPS) [165], where they are accelerated to 450 GeV. At this point, the beam is passed to the LHC ring. Some of the machines in the described process not only participate in preaccelerating the protons, but also still house active physics experiments.

The major experiments located at the LHC ring are: ATLAS (*A Toroidal LHC ApparatuS*) [167], CMS (*Compact Muon Solenoid*) [168], ALICE (*A Large Ion Colliding Experiment*) [169] and LHCb. The first two are general-purpose detectors (GPD), designed mainly for the Higgs boson (discovered in these experiments in 2012) and the New Physics searches. The ALICE experiment is dedicated for heavy-ion physics and can study the behaviour of matter in conditions close to those from the beginning of the Universe's existence (*e.g.* properties of the quark-gluon plasma).

The LHCb experiment was primarily designed for the study of particles containing b or c quarks, although now it also performs well as a general-purpose detector in the forward region, *e.g.* for heavy-ion and fixed-target collisions [170]. It has a unique acceptance as compared to the other mentioned experiments, which is dictated by the properties of the $b\bar{b}$ quark pairs production at the LHC conditions (see Sect. 3.1).

The LHCb detector is described in more detail in Sect. 3.1. The following Sect. 3.2 provides an overview of the planned experiment upgrades. A special focus is given to the description of the LHCb RICH system in Sect. 3.3, including the author's contributions to the simulation of the future RICH detectors.

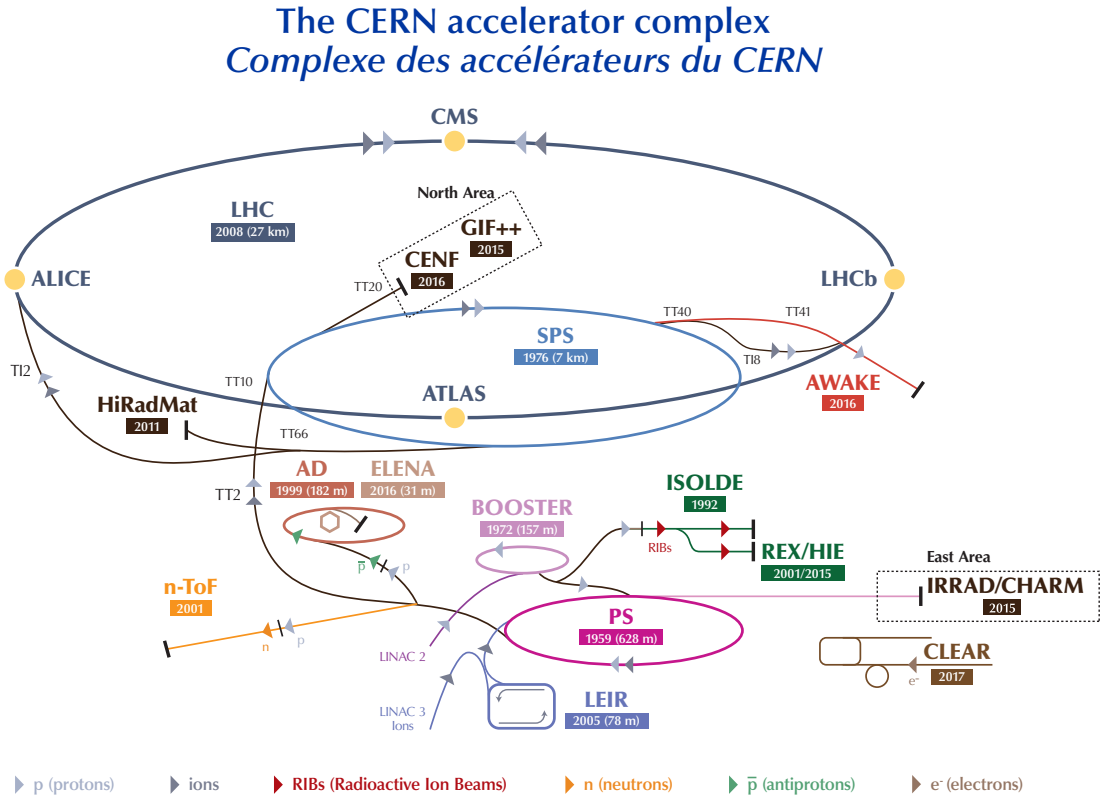

Figure 3.1: LHC is the last ring (dark blue line) in a complex chain of particle accelerators. The smaller machines are used in a chain to help boost the particles to their final energies and provide beams to a whole set of physics experiments [166].

3.1 LHCb detector

The LHCb detector is a single-arm spectrometer with an angular coverage of $10 < \theta < 300\,(250)$ mrad in the bending (nonbending) plane, which corresponds to the pseudorapidity range of $2.0 < \eta < 5.0$. As mentioned in the previous paragraph, this is an acceptance region which is unique among the other experiments at the LHC, and it is motivated by the properties of the $b\bar{b}$ quark pairs production. Gluon fusion is the dominant mechanism contributing to this phenomenon at the LHC. In this process, a significant asymmetry between the momenta of the inbound partons is present in the laboratory frame. As a result, the produced pair's centre-of-mass energy is boosted in the direction of the higher-momentum parton. It leads to the production of both quarks in the $b\bar{b}$ pair along the same direction in the forward or in the backward region. In the case of pp collisions at $\sqrt{s} = 14\,\text{TeV}$, 24% of all $b\bar{b}$ pairs are produced within the LHCb detector acceptance (see Fig. 3.2).

Details on the data-taking conditions in the LHCb experiment are reported in

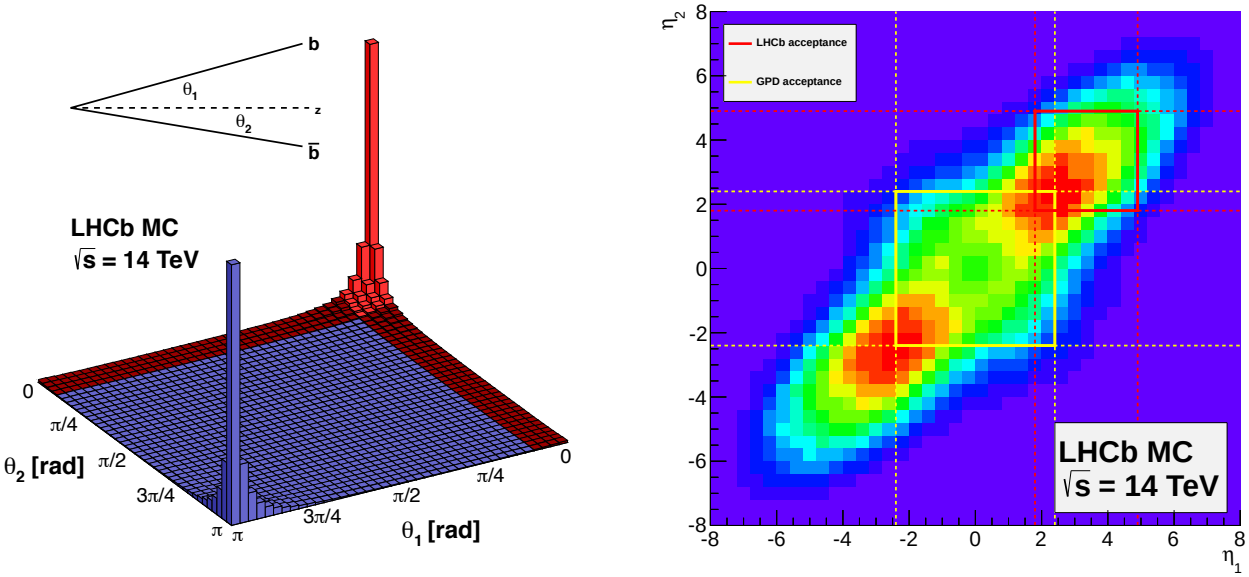

Figure 3.2: Production of the $b\bar{b}$ quark pairs (left) within the LHCb acceptance (here defined as $1.8 < \eta < 4.9$ – the region marked in red) and (right) comparing the LHCb and a general purpose detector acceptance (assumed to be $|\eta| < 2.4$) [171]. Both plots are produced using the simulation with the PYTHIA8.1 generator [83].

Sect. 3.1.1. An overview of the LHCb detector[1] is shown in Fig. 3.3 and a description of its individual components is provided in the following sections. The most distinctive features of the LHCb spectrometer are: a very precise track and primary vertex (PV) reconstruction as well as an outstanding performance of the particle identification (PID). The two systems that are dedicated to provide such a functionality are described in more detail in Sect. 3.1.2 and Sect. 3.1.3, respectively. Another important component of the LHCb experiment is the trigger, which allows for the selection of events relevant for a rich spectrum of physics studies (see Sect. 3.1.4). An overview of the LHCb simulation and software framework is given in Sect. 3.1.5.

3.1.1 Data-taking conditions

The LHCb experiment recorded its first pp collisions in 2009. The data-taking period that took place in years 2010–2013 (2015–2018) is referred to as *Run1* (*Run2*). The maximum centre-of-mass energy of the pp collisions achieved in Run1 (Run2) amounted to $\sqrt{s} = 8\,\text{TeV}$ ($\sqrt{s} = 13\,\text{TeV}$).

Two parameters of the recorded collisions are especially important for the detector operation. The first one is an instantaneous luminosity[2] $\mathcal{L}$ and another one is an average

[1]The LHCb coordinate system is right-handed, with the z-axis pointing along the beam axis, y-axis in the vertical direction, and x-axis in the horizontal direction. The (x, z)-plane is the bending plane of the dipole magnet.

[2]An instantaneous luminosity corresponds to the rate of collisions and can be expressed as $\mathcal{L} = \frac{1}{\sigma}\frac{\mathrm{d}N}{\mathrm{d}t}$, where σ denotes the cross-section of the given process and $\frac{\mathrm{d}N}{\mathrm{d}t}$ designates the number of observed events within a certain time period (in this case, both values refer to the proton-proton scattering). The value

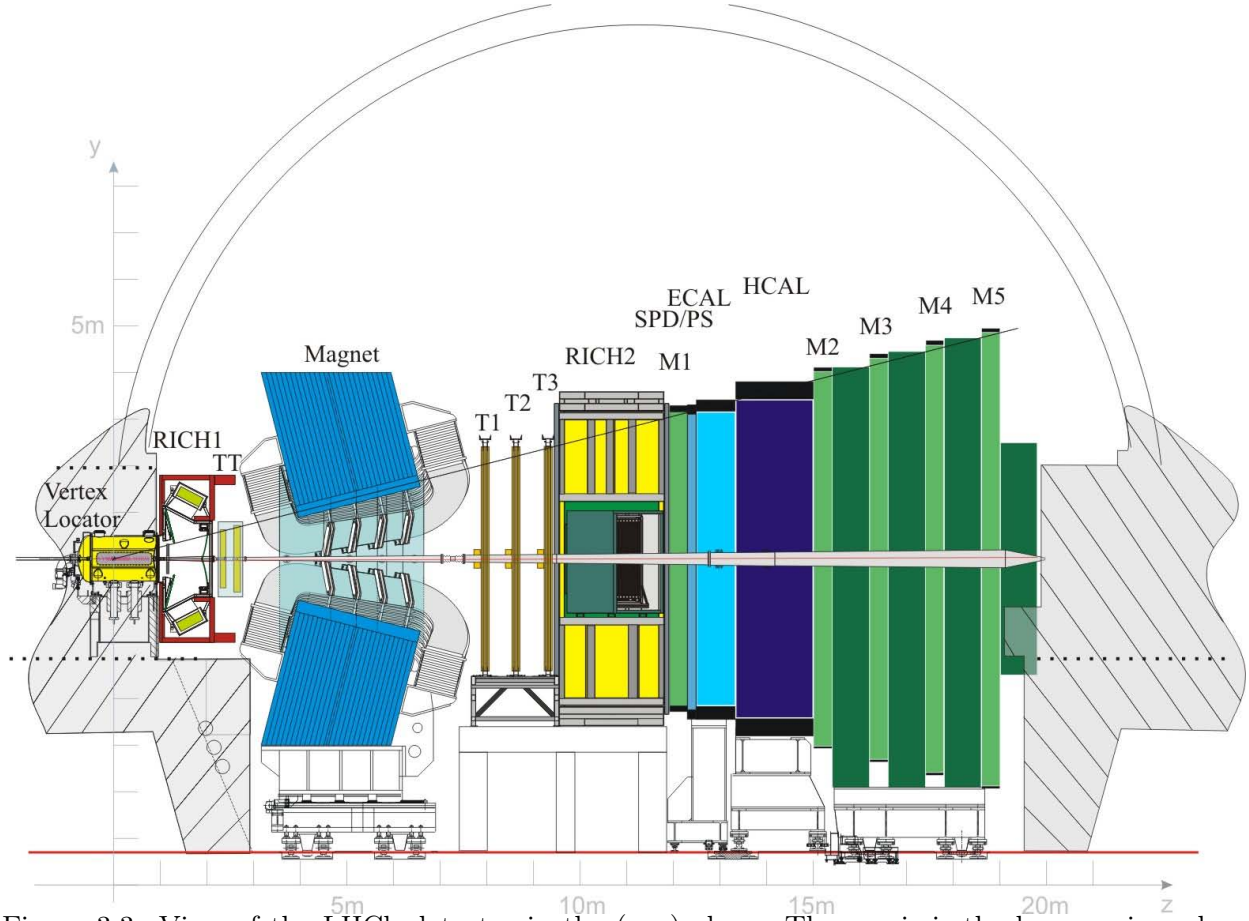

Figure 3.3: View of the LHCb detector in the (y, z)-plane. The z-axis is the beam axis and the x-axis is perpendicular to the page surface [172].

number of visible interactions per bunch crossing μ_{vis} (*pileup*)[3]. The number of interactions per bunch crossing is given by a Poisson distribution with the mean value proportional to the instantaneous luminosity. Already in Run1, the majority of the data was collected at the instantaneous luminosity level of $\mathcal{L} = 4 \times 10^{32}\,\text{cm}^{-2}\,\text{s}^{-1}$, which was twice the nominal value envisioned in the detector design (see Fig. 3.4). The original assumptions on the average number of visible interactions per bunch crossing ($\mu_{\text{vis}} = 0.7$) were exceeded significantly as well, amounting to ~ 0.5–2.5 (1.1) during the Run1 (Run2) period [15,174]. Despite these challenging conditions, the performance of the LHCb detector was in good agreement with the original assumptions during the whole activity period.

The *pp* collisions recorded in the LHCb experiment during Run1 and Run2 correspond to an integrated luminosity of 3 and $6\,\text{fb}^{-1}$, respectively. Other types of collisions were also registered to extend the reach of the LHCb physics programme, including samples of PbPb collisions at $\sqrt{s_{\text{NN}}} = 5.02\,\text{TeV}$ centre-of-mass energy per nucleon, equivalent to an integrated luminosity of $\sim 220\,\mu\text{b}^{-1}$. Another studied system was *p*Pb, with collisions recorded at $\sqrt{s_{\text{NN}}} = 5.02\,\text{TeV}$ and $\sqrt{s_{\text{NN}}} = 8.16\,\text{TeV}$ in 2013 and 2016, respectively (combined integrated luminosity of those datasets is at a level of $35\,\text{nb}^{-1}$). The LHCb detector

of $\mathcal{L}$ in an accelerator is dictated by the properties of the beams, such as the revolution frequency, beams geometry, colliding angle, *etc.* [16].

[3]A visible interaction in the LHCb experiment is defined as a PV reconstructed with at least five VELO tracks (see Sect. 3.1.2) [173].

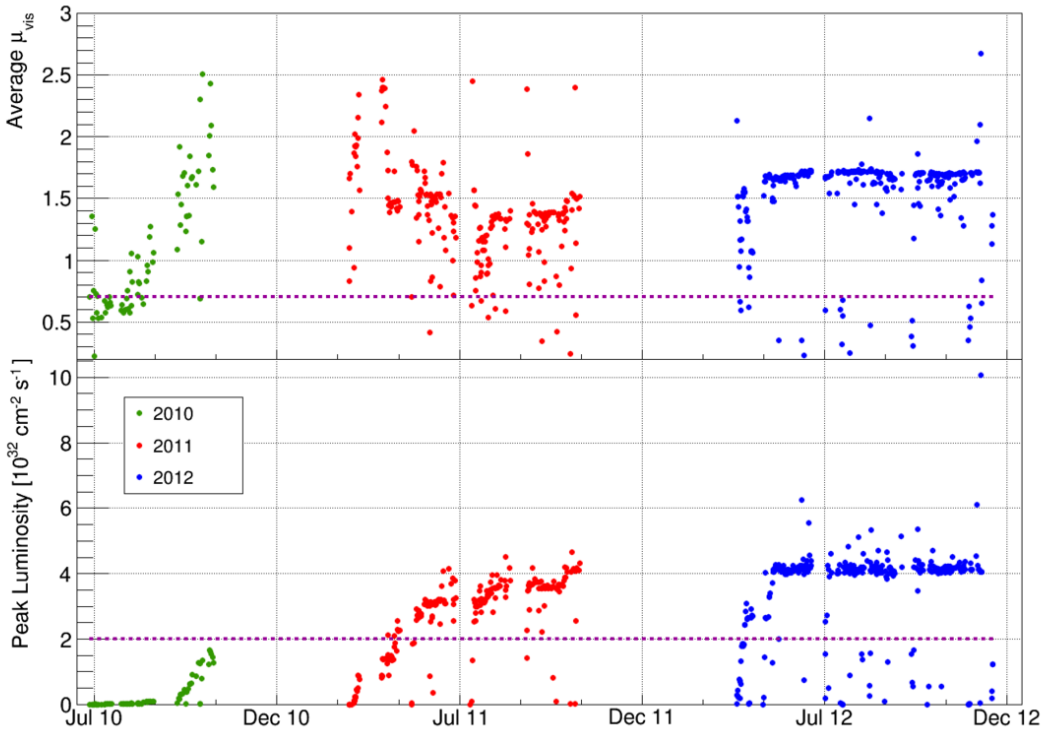

Figure 3.4: (top) Average number of visible interactions per bunch crossing and (bottom) instantaneous luminosity at the LHCb interaction point in the period 2010–2012. The dotted lines show the design values [15].

is also capable of operating in a fixed-target mode, thanks to the SMOG (*System for Measuring Overlap with Gas*) device [170]. It was used in the Run2 period to register pHe, pNe and pAr collisions at different energies ranging up to $\sqrt{s_{\mathrm{NN}}} = 110\,\mathrm{GeV}$.

The LHCb experiment follows various operational policies to reduce systematic uncertainties related to the data-taking conditions. One of them is the *luminosity-levelling* procedure, which allows for maintaining steady instantaneous luminosity (and detector occupancy) values during the given LHC fill. A stability within 5% is achieved by adjusting the overlap of the colliding beams in the course of the operation. Another policy is to periodically reverse the LHCb magnet polarity (see Sect. 3.1.2) throughout the data taking. Trajectories of charged particles are curved by the magnetic field, which can in principle lead to detection asymmetries. By systematically reversing the magnet polarity, this effect is significantly reduced. The positively (negatively) charged particles are bent in the horizontal plane towards the LHC centre in the mode with the magnetic field pointing upwards (downwards), which is referred to as the *MagUp* (*MagDown*) configuration.

3.1.2 Track and vertex reconstruction

The tracking system consists of a vertex locator VELO (which is a silicon-strip detector, surrounding the interaction region) and four tracking stations located at different positions with respect to a dipole magnet with a bending power of 4 Tm. A *Tracker Turicensis* detector (TT) is placed upstream of the magnet, while three T stations (T1–T3) are located downstream. The sensors used in TT and the inner parts of the T stations (*Inner*

Tracker – IT) are constructed using silicon microstrips, while the regions of the T stations that are further away from the beam pipe (*Outer Tracker* – OT), are covered by detectors built from straw drift-tubes. Combining the information from the tracking detectors located upstream and downstream of the magnet allows us to measure the deflection of charged particles in the magnetic field, and determine their momenta.

The VELO detector (see Fig. 3.5) is a crucial component of the tracking system, since it is located around the interaction point, which allows for a very precise reconstruction of primary and secondary vertices. It is particularly important for identifying decays of hadrons containing b or c quarks, based on their relatively long flight distance. The VELO detector consists of 42 modules, each containing a pair of (R, ϕ) sensors (which are designed to measure the radial and polar coordinates of the traversing particle). At the time of data taking, when the beam is stable, the modules are placed at a very close distance of 8.2 mm from the beam. However, during the beam injection, those elements are retracted by 29 mm in the horizontal direction, to avoid potential detector damage. Moreover, the beam pipe in the VELO region is removed, to reduce the number of the material interactions before reaching the detector sensors. The VELO stations are placed in a vacuum vessel and are separated from the beam only by a thin aluminium foil (*RF foil*). The VELO detector allows for the reconstruction of primary vertices with a spatial resolution of 13 μm in the x and y directions, and 71 μm along the z-axis (performance for vertices reconstructed from 25 tracks).

Both the TT and the IT (see Fig. 3.6) detectors are constructed using silicon microstrips, and they are also referred to as the *Silicon Tracker* (ST). The active areas of these systems correspond to approximately 8 m^2 and 4 m^2, respectively, and the single-hit resolution amounts to ~ 50 μm in each of them. The *Outer Tracker* is a drift-tube gas detector, which uses a mixture of argon (70%), CO_2 (28.5%) and O_2 (1.5%). Such a gas composition allows us to provide a spatial resolution of 200 μm and a drift time below 50 ns. Each of the tracking stations consists of four detection layers, which are arranged in a $x - u - v - x$ geometry. The relevant modules in the x-layers are oriented vertically, while the ones in the u and the v layers are tilted by $\pm 5^\circ$ with respect to the vertical direction. Such an arrangement provides a more accurate spatial information for the track reconstruction, thus also improving the resolution of the particle momentum measurement.

The LHCb track reconstruction algorithms use the information on hits in the various tracking detectors to determine the trajectories of charged particles. Depending on the combination of detectors that have hits associated with the track candidate, the following track types are defined (see Fig. 3.7):

- *long tracks* have associated hits both in the VELO and the T stations (usually also in TT, but this is not required); since they traverse the full tracking system, they have the most accurate information on the momentum and are the most important tracks for the physics studies,

- *VELO tracks* are usually related to particles travelling in the backward direction (negative z values in the LHCb coordinate system) or at large angles with respect to the beam axis, which produces hits only in the VELO detector; such tracks are useful in the primary vertex reconstruction as well as for the determination of the PV charged-particle multiplicity,

- *upstream tracks* are the tracks with momenta too low to pass the whole magnet

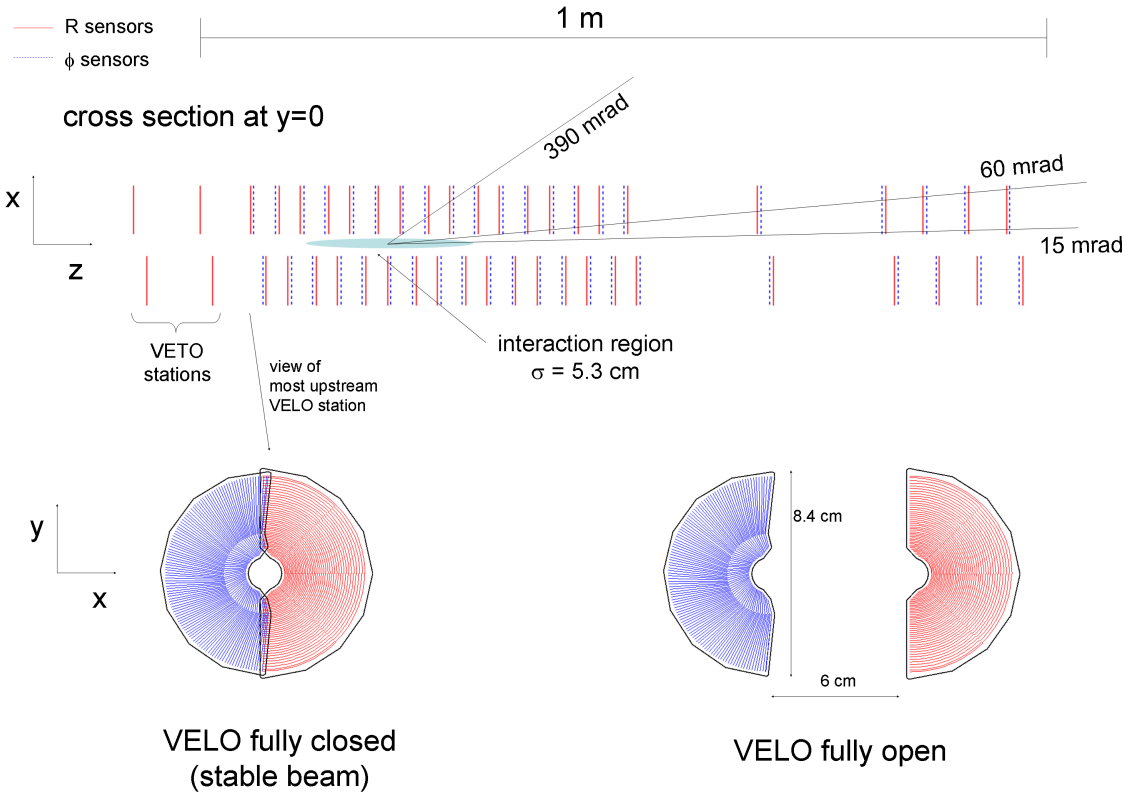

Figure 3.5: Cross-section in the (x, z)-plane of the VELO silicon sensors, at $y = 0$, with the detector in the fully closed position. The front face of the first modules is also illustrated in both the closed and open positions [14].

region, thus leaving hits only in the VELO and the TT stations; yet, they travel through the RICH1 detector and can be used for the background estimation in the RICH1 PID algorithms,

- *downstream tracks* are reconstructed from hits present only in the TT and the T stations; those tracks are used in the reconstruction of long-lived particles, which decay outside of the region covered by the VELO detector,

- *T tracks* have hits only in the T stations; usually they originate from secondary interactions, but still can be useful for the PID algorithms of the RICH2 detector.

In the case of the long tracks, the reconstruction efficiency is at a level of 96% in the momentum range of 5–200 GeV (which corresponds to the phase-space region covered by the LHCb detector). In the analyses described in this thesis, the long tracks are employed to study the Bose–Einstein correlations, while the VELO tracks are used to determine the charged-particle multiplicity in a primary vertex.

Due to imperfections of the reconstruction algorithms, some of the reconstructed tracks do not correspond to a real charged particle. Two types of the misreconstructed tracks are defined below:

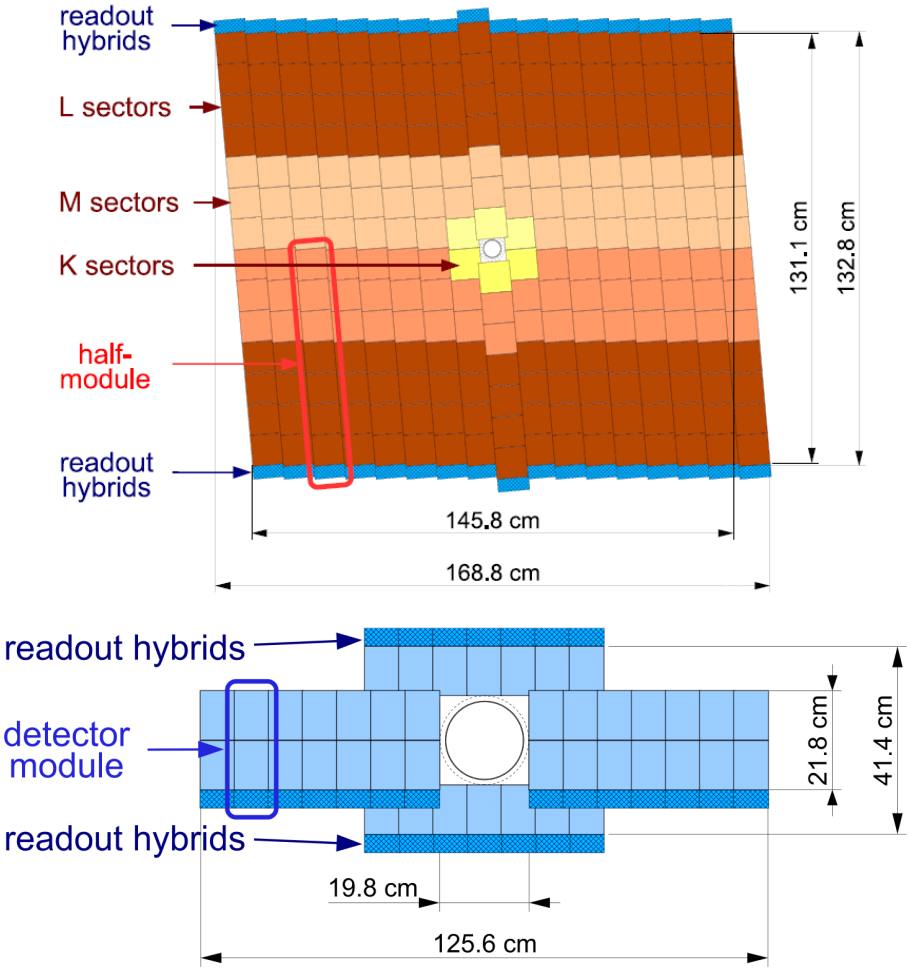

Figure 3.6: (top) Layout of a TT detection layer and (bottom) a diagram of the x detection layer in the second IT station [14].

- *fake* (*ghost*) tracks are the ones which do not correspond to any particle trajectory, but they are reconstructed from a number of unrelated hits,
- *clone* tracks are multiple tracks reconstructed from hits which were deposited by a single charged-particle.

The rate of the misreconstructed tracks is significantly reduced by the use of a neural network classifier [175] that combines information from the track fit, its kinematics and the number of measured and expected hits in the tracking stations.

The track reconstruction system is capable of determining the momentum of charged particles with a relative uncertainty $\Delta p/p$ varying between 0.5% for low momenta to 1.0% for momenta at the level of 200 GeV. The system also allows for the measurement of the minimum distance of a track to its associated primary vertex, the *impact parameter* (IP), with a resolution of $(15 + 29/p_{\mathrm{T}})$ µm, where p_{T} (given in GeV) is the particle transverse momentum with respect to the beam direction. The *associated PV* is defined as the PV that best fits to the flight direction of the particle.

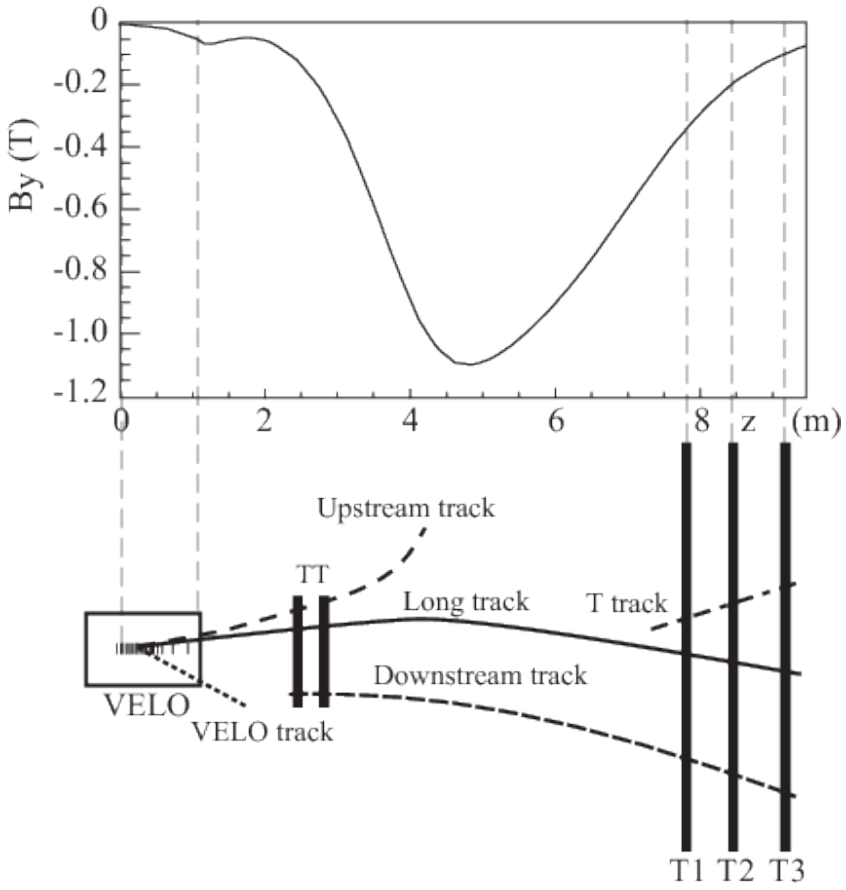

Figure 3.7: Schematic illustration of the various track types: long, upstream, downstream, VELO and T tracks. For reference, the main B-field component (B_y) is plotted above as a function of the z coordinate [14].

3.1.3 Particle identification

The particle identification is based on information from a number of different subdetectors. Two RICH detectors (RICH1, RICH2) allow us to distinguish the different species of charged hadrons. A system of calorimeters (comprising an electromagnetic – ECAL – and a hadronic one – HCAL) with scintillating-pads and preshower detectors (SPD/PS) provides discrimination between photons, electrons and hadrons. Muons are identified using a set of stations (M1–M5), consisting of iron layers and multiwire proportional chambers.

The RICH detectors (see Fig. 3.8) are particularly important for the BEC studies, since they allow us to distinguish among different species of charged hadrons (see more details on the RICH detectors and their development in Sect. 3.3). The RICH1 detector covers the full LHCb acceptance between 25 and 300 (250) mrad in the bending (nonbending) plane. It employs silica aerogel[4] and CF_4 gas as radiators with refractive indices

[4]The aerogel radiator (which was intended to improve the PID capabilities in the momentum region below 10 GeV) was removed from the detector before the Run2 data-taking period due to its relatively low photon yield and an observed limited effectiveness in a high-luminosity environment of LHC. As described in Ref. [176], the aerogel removal resulted in a higher photon yield in the C_4F_{10} gas, where previously part of the emitted Cherenkov photons was blocked from reaching the detection area. This aspect, together with a reduced material budget within the detector acceptance, outweighed the loss

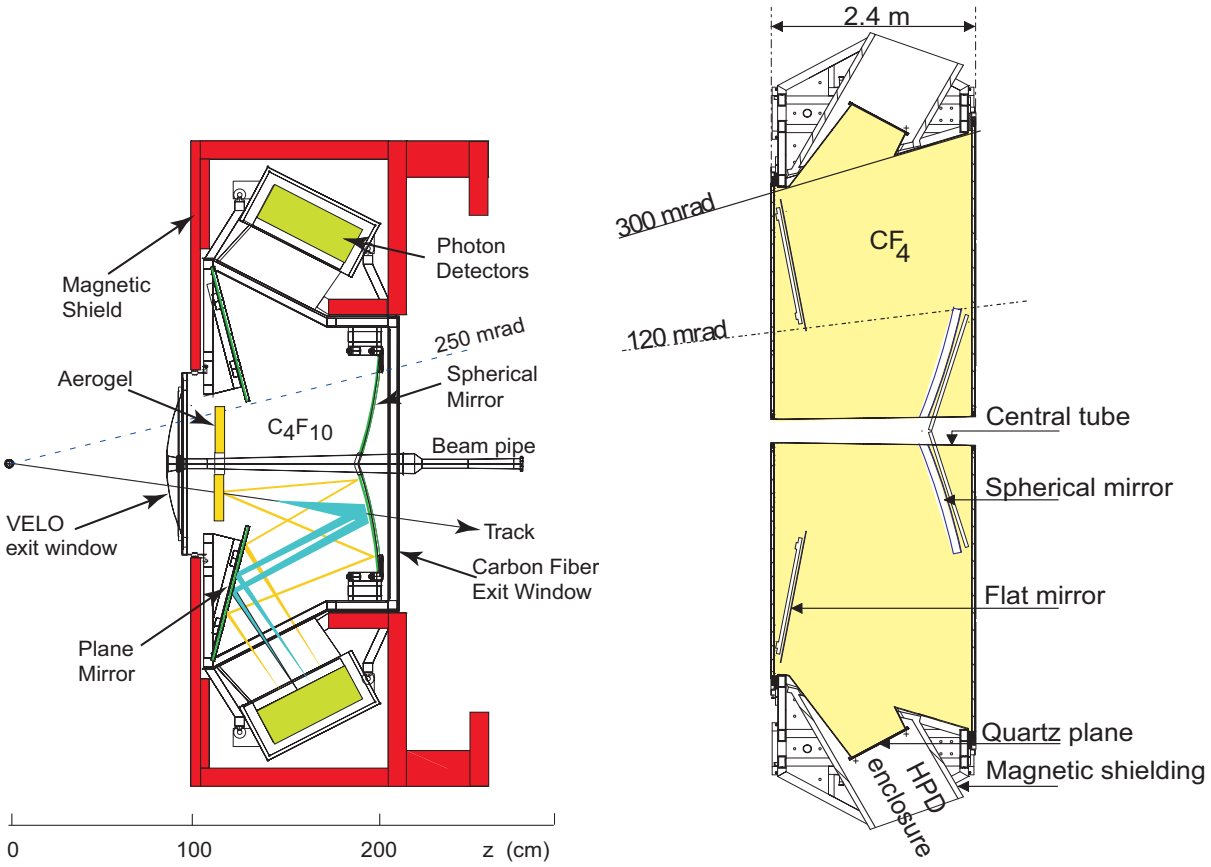

Figure 3.8: (left) Side-view schematic layout of the RICH1 detector and (right) top-view schematic layout of the RICH2 detector [14].

of 1.03 and 1.0014, respectively, and allows for the identification of charged particles in the low momentum range (2–60 GeV). The high-momentum region between 15–100 GeV is addressed by the RICH2 detector. It is filled with C_4F_{10} gas as a radiator (refractive index equal to 1.0005) and has a limited angular coverage of 15–120 (100) mrad in the horizontal (vertical) direction, which is motivated by the region in which most of the high-momentum particles are produced. The Cherenkov photons are emitted in the radiator at an angle depending on the velocity of the traversing charged-particle (see Fig. 3.9) and they are collected using a *Hybrid Photon Detector* (HPD) technology (see Fig. 3.10). Combined information from the RICH system allows for the identification of charged particles (primarily hadrons – π, K and p, but also an input for the e and μ identification is provided) in the momentum range of 2–100 GeV.

The calorimeter system at LHCb contributes to the particle identification by distinguishing photons, electrons and hadrons (including the neutral ones, such as π^0). It also measures their energies and positions, and provides significant input for the hardware trigger system, by indicating candidates with high transverse energies (see Sect. 3.1.4 for details). The ECAL detector constitutes interchanging lead and scintillating layers and extends over a region corresponding to $X_0 = 25$ radiation lengths. The overall size of

of the PID information from the aerogel radiator and an overall improvement in the PID performance during Run2 was achieved.

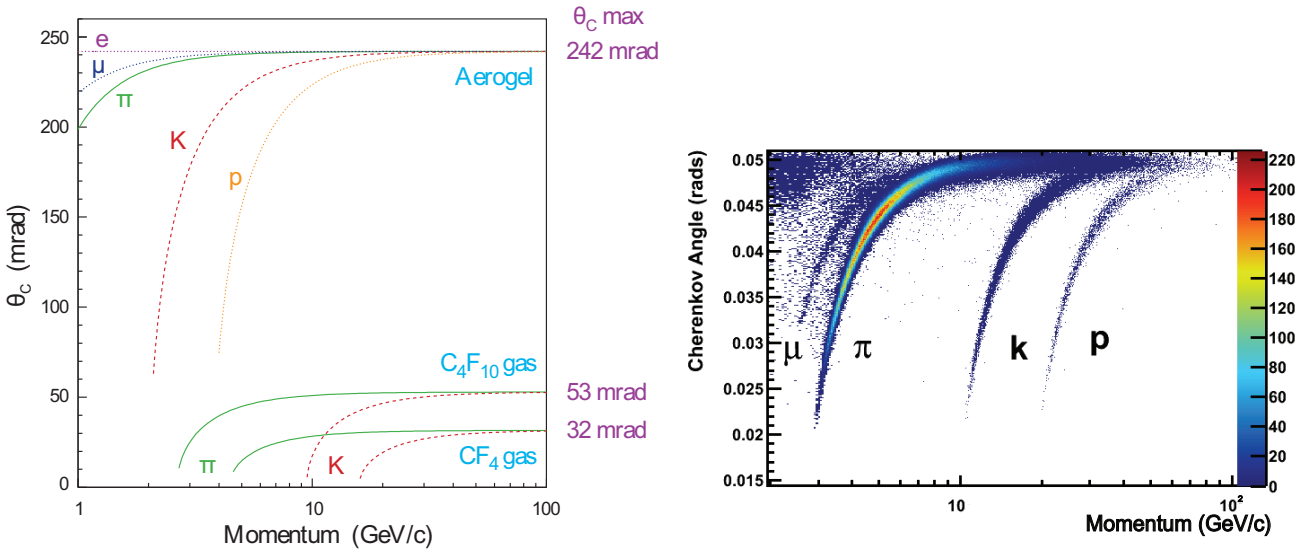

Figure 3.9: (left) Expected Cherenkov angle versus particle momentum for the RICH radiators [14] and (right) reconstructed Cherenkov angle as a function of track momentum in the C_4F_{10} radiator. The Cherenkov bands for muons, pions, kaons and protons are clearly visible [177].

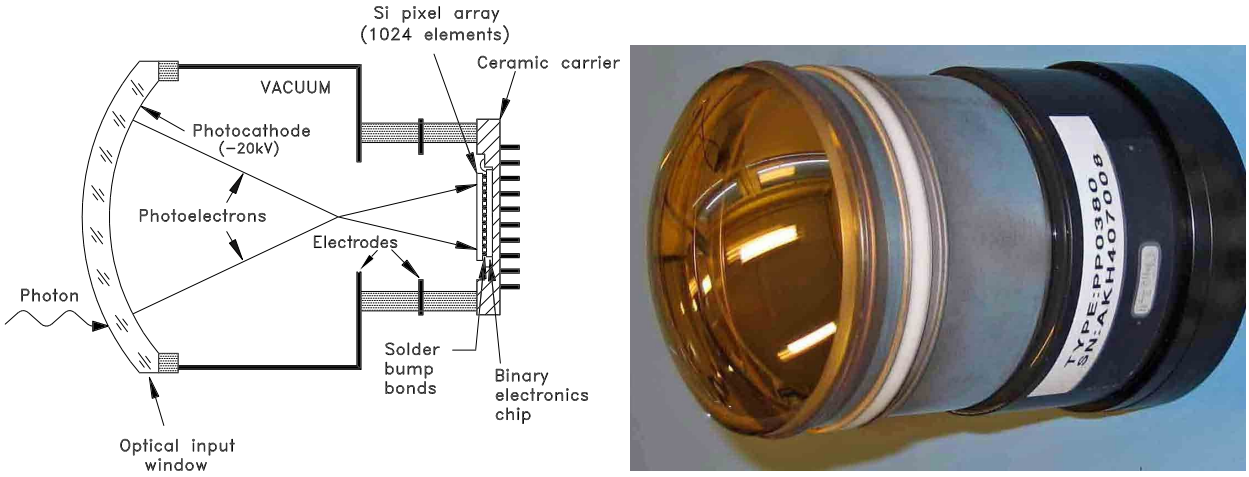

Figure 3.10: (left) Schematic view and (right) a photograph of the HPD [14].

ECAL is chosen to match the acceptance of the tracking system, with a limited coverage of $\theta > 25\,\text{mrad}$ in the beam-pipe region. The hadronic calorimeter is composed of iron and scintillating tiles. Its thickness corresponds to $\lambda_{\text{int}} = 5.6$ nuclear interaction lengths. The relative energy resolutions of the calorimeters are

$$\left(\frac{\sigma_E}{E}\right)_{\text{ECAL}} = 1\% \oplus \frac{10\%}{\sqrt{E}} \;\; ; \;\; \left(\frac{\sigma_E}{E}\right)_{\text{HCAL}} = 9\% \oplus \frac{69\%}{\sqrt{E}} \;\; ,$$

where E is given in GeV and the operator $\oplus$ denotes that the contributions should be added in quadrature.

The muon system allows for the identification of muons and is one of the key components of the LHCb hardware trigger. The total area covered by the detector is $435\,\text{m}^2$. The stations M2–M5 are separated by absorbers that consist of 80 cm thick iron layers,

which, together with the calorimeter region, corresponds to $\lambda_{\text{int}} = 20$ nuclear interaction lengths. The first muon station is located upstream of the calorimeter system to improve the resolution of the p_{T} measurement, which is used in the hardware-based trigger (see Sect. 3.1.4).

As already mentioned, all of those systems contribute to the particle identification at LHCb. In each case, log-likelihood functions are constructed for the observed signatures and different particle hypotheses [15] (see Sect. 3.3.1 for more details on this procedure in the RICH system). Then, the identification is based on the requirement that the difference between the log-likelihood functions (DLL) for given hypotheses be greater than a certain level – for example, a $\text{DLL}_{K\pi} > 0$ prerequisite can be imposed to differentiate a kaon from a pion. In this case, a more strict requirement (such as $\text{DLL}_{K\pi} > 5$) leads to a decreased rate of pions misidentified as kaons, with a reduced efficiency of kaon identification (fraction of all true kaons properly identified as kaons).

Information from different detectors is combined to maximise the PID performance. In a basic technique, the log-likelihood functions from different systems are added linearly and used in a similar manner as described above. A more refined approach based on a dedicated neural network [178] implemented within a TMVA (multivariate analysis) toolkit [179] is also employed, where the log-likelihood functions are only one of the inputs in the procedure, and correlations between the different detector systems are taken into account. This modified approach results in producing single numbers reflecting the probability that a particle is of a given type, labelled as *e.g.* $ProbNN$(pion).

The combined information from the PID system provides an excellent performance of the particle identification procedure. As an example, kaons can be identified at an efficiency level of 95% with a 5% probability of misidentifying a pion for a kaon. Also, for the muon identification with an efficiency of 97%, the probability of the pion misidentification amounts to less than 3%.

3.1.4 Trigger and online data-processing

During data taking, events are selected using a trigger, which consists of a hardware (*Level-0* – L0) and a software (*High Level Trigger* – HLT) stage. The goal of the trigger is to select events that are potentially valuable for the given type of analysis, and to reduce the amount of data to a level which can be saved to a long-term storage. A scheme of the LHCb trigger system (as in 2012) is presented in Fig. 3.11.

The hardware trigger at LHCb is based on the information from the calorimeters and the muon system, and accepts events with signatures of high values of the transverse momentum or energy. It allows for the reduction of the event rate from the initial 40 MHz bunch-crossing rate of the LHC accelerator to approximately 1 MHz. It corresponds to a rate at which the complete detector can be read out.

The number of events is further reduced by the software trigger, which consists of two levels. The first one (HLT1) uses part of the available subdetector information that permits, for example, the identification of tracks that are significantly displaced from the primary vertex, which can be a signature of decays of long-lived particles that contain b or c quarks. The second level (HLT2) performs a full event reconstruction and applies a mixture of both inclusive and exclusive selections. The software trigger adopts mostly the same reconstruction algorithms that are used in the offline processing, with certain simplifications introduced in order to optimise the computation time. As an example, for

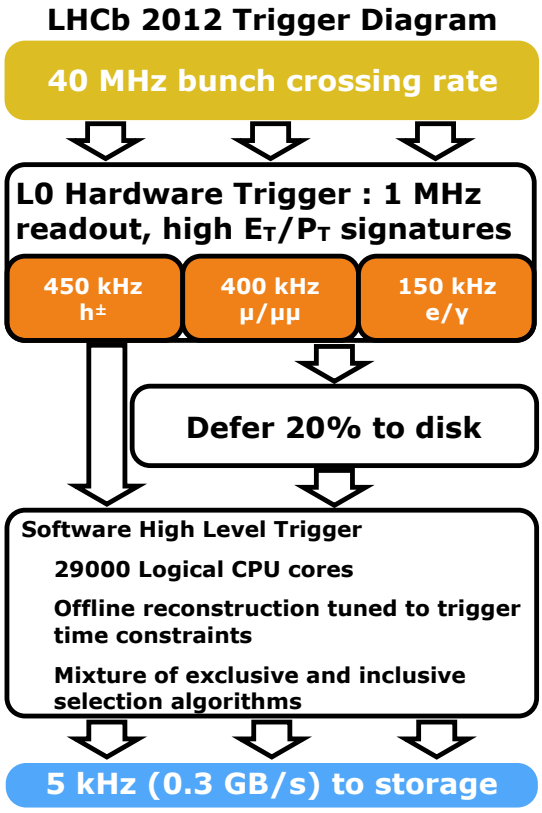

Figure 3.11: Diagram of the online data-processing flow at LHCb in 2012 [180].

the pp collisions collected in 2011 (2012), HLT2 was taking into account only the particles with $p_{\mathrm{T}} > 500\,(300)$ MeV. During the operational period in 2012, the data rate for the final storage was being reduced by the trigger to a level of 2–5 kHz.

Part of the events that are accepted by the hardware trigger are deferred to a temporary storage and are processed at a later stage, when no collisions are provided by the LHC machine. It allows for the optimisation of the use of available computing resources and the improvement of the trigger efficiency, by running more detailed reconstruction algorithms, *e.g.* with less strict requirements on the particles' transverse momenta.

Events accepted by the trigger are further processed offline, employing the full event reconstruction. This procedure is performed periodically, to benefit from potential developments in the reconstruction algorithms and improvements in the detector alignment and calibration. Offline event selections are applied to the reconstructed data to provide information required for different types of physics analyses. This process, in which more analysis-specific physical quantities are also added to the samples, is referred to as *stripping* in the LHCb experiment. In the BEC studies, a *minimum-bias* stripping line is used, which corresponds to a very loose selection requirements, close to the ones applied by the hardware trigger.

3.1.5 Simulation and software

The LHCb software applications are based on experiment-independent ROOT [181] and GAUDI [182] frameworks. The structure of the LHCb software and data-processing scheme

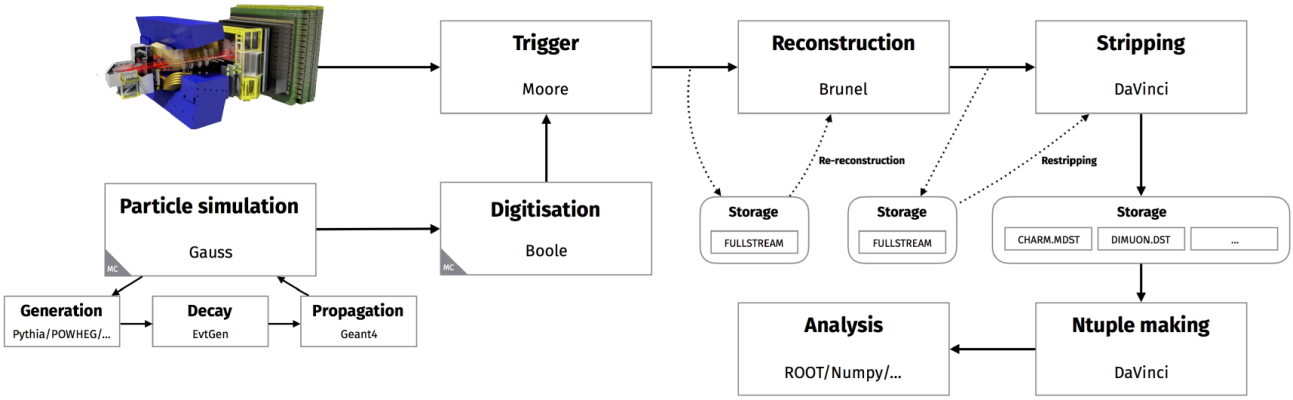

Figure 3.12: LHCb data-flow and data-processing applications [183].

is shown in Fig. 3.12.

A full detector simulation is required to understand various detector acceptance effects as well as an impact of the chosen selection requirements on the physics results. In the LHCb simulation, the pp collisions are produced using the PYTHIA [83, 84] generator with an experiment-specific configuration [184]. Alternative generators are adopted in special cases, such as *e.g.* EPOS [185] for the pPb collisions. Decays of unstable particles are modelled by EVTGEN [186] and the final-state radiation is described using PHOTOS [187]. The GEANT4 toolkit [188] is used to simulate interactions of the generated particles with the detector, within a GAUSS application [189]. Response of the detector electronics (including effects such as electronic noise and channel cross-talk) is modelled by a BOOLE project [190] (this part of the simulation is referred to as *digitisation*). The MOORE application [191] is an implementation of the HLT software trigger. Event reconstruction, including the particle identification algorithms, is run within a BRUNEL project [192], for both the data and simulated samples. The DAVINCI package [193] performs tasks related to data selection (*e.g.* stripping) and provides higher-level information for the physics analyses.

3.2 Experiment upgrades

The full potential of the LHC machine can be exploited in an approach which is referred to as the intensity frontier. Increasing the instantaneous luminosity in the accelerator allows for the collection of more data for the physics studies in a shorter time and also provides many interesting challenges for the development of the detector technology. An upgrade of the machine is planned to a *High-Luminosity LHC* (HL-LHC) [194], where collisions at an instantaneous luminosity of $\mathcal{L} = 2 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ will be provided to the LHCb experiment, employing the luminosity-levelling procedure (see Sect. 3.1.1). The current LHCb detector, despite its exceptional operation in Run1 and Run2, will not be able to maintain its performance in such conditions. It is, among others, due to the maximum rate at which the data can be collected. Experiment upgrades are planned up to the HL-LHC era, which will allow us to increase the precision of many interesting physics measurements to an unprecedented level, especially in the heavy-flavour studies [195].

The HL-LHC accelerator is scheduled to become operational around the year 2025 and

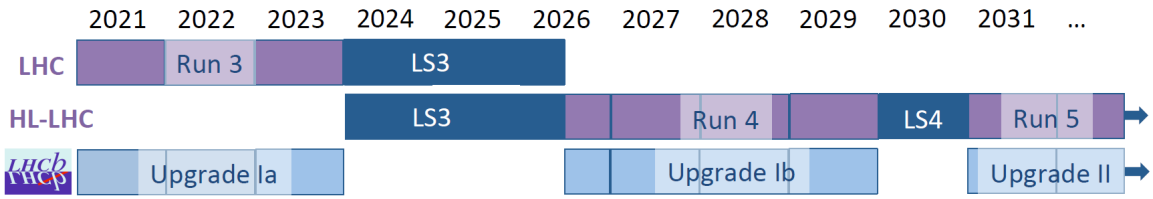

Figure 3.13: Timeline of accelerator and experiment operations over the decade 2021–2031. The periods of data-taking and long shutdowns, where the upgrades are performed, are indicated for the LHC, the HL-LHC and LHCb [195].

be used till the late '30s. A timeline of the operations and the planned upgrades to both the machine and the LHCb experiment is presented in Fig. 3.13. The data-taking periods are referred to as *Runs* (*e.g.* Run3), while the nonoperational years are designated *long shutdowns* (*e.g.* LS3), where most of the work related to the upgrades is performed.

In both Run3 and Run4, the instantaneous luminosity of the collisions in the LHCb experiment will be maintained at a level of $\mathcal{L} = 2 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, which will be increased to $\mathcal{L} = 2 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ in Run5. These numbers amount to an increase in the luminosity with respect to the values from the end of Run2 at the order of five and fifty, respectively, and to similar changes in the average number of visible interactions per bunch crossing (from $\mu_{\mathrm{vis}} = 1.1$ in Run2 to the value of $\mu_{\mathrm{vis}} = 5.2$ during Run3 and Run4 [196], and $\mu_{\mathrm{vis}} = 56$ in Run5 [195]). The nominal design beam energies of the LHC will also be achieved as of Run3, which corresponds to the energy of 14 TeV in the centre-of-mass frame. The experiment upgrades will allow for the collection of data equivalent to a total integrated luminosity of $50\,\mathrm{fb}^{-1}$ by the end of Run4 and of $300\,\mathrm{fb}^{-1}$ when Run5 is completed (compared to the value of $9\,\mathrm{fb}^{-1}$ recorded in Run1 and Run2). The expected levels of the integrated and instantaneous luminosity at LHCb during the periods mentioned above are shown in Fig. 3.14.

The LHCb experiment upgrades are designed to cope with the increased amount of data produced in the collisions, while maintaining (or even improving) the current detector performance and possibly extending its physics coverage by adding new subsystems or upgrading the existing ones. The first upgrade (Upgrade I) [197,198] will be commissioned for Run3 (this phase is also referred to as Upgrade Ia) and its consolidation is planned before Run4 (Upgrade Ib). The Upgrade II will take place in the HL-LHC era and it will provide further interesting prospects for the heavy-flavour physics studies [195]. More details on each of the upgrades are given in Sects. 3.2.1, 3.2.2 and 3.2.3, focusing mainly on the differences with respect to the previous system.

3.2.1 Upgrade Ia

The increased instantaneous luminosity in Upgrade I will lead to major modernisations in the LHCb experiment, *e.g.*, for the sensor technologies and for the readout electronics of individual subsystems. In total, more than 90% of the active detector channels will be replaced in that period.

A distinctive feature of Upgrade Ia is the ability to read out the complete detector at the LHC bunch-crossing rate of 40 MHz [199]. A new, fully software-based trigger system will also be developed (with no hardware level at all), which will provide flexibility in

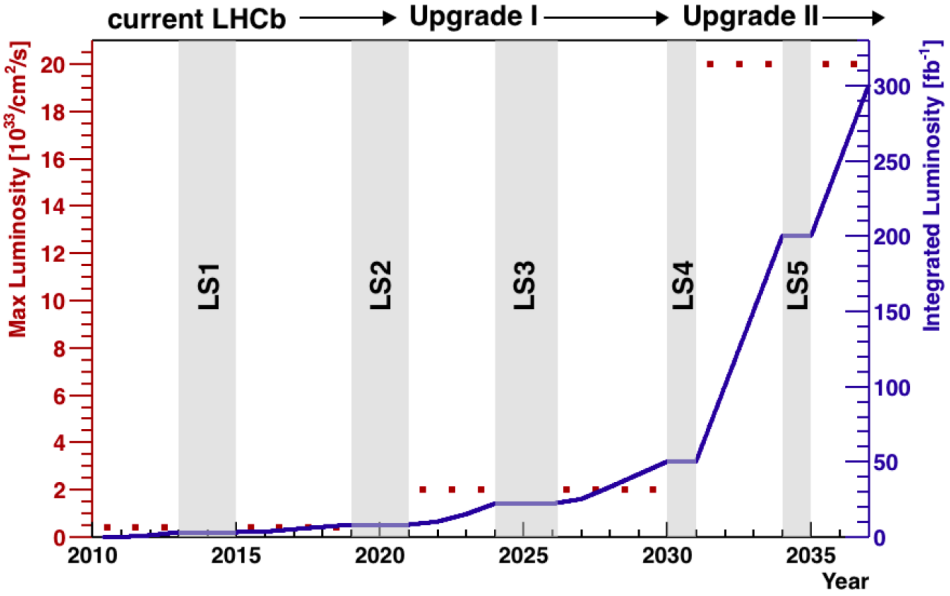

Figure 3.14: Luminosity projections for the original LHCb, Upgrade I and Upgrade II experiments as a function of time. The red points and the left scale indicate the anticipated instantaneous luminosity during each period, with the blue line and right scale indicating the integrated luminosity accumulated [195].

the application of selection requirements on the data. With the instantaneous luminosity increased by a factor of five and the trigger efficiency doubled for numerous physics channels, the annual yield in many studies will be enhanced by an order of magnitude with respect to the current detector. An overview of the LHCb spectrometer in this period is presented in Fig. 3.15 and the modifications with respect to the original experiment are described in the following sections.

3.2.1.1 Track and vertex reconstruction

The various components of the tracking system will be improved [196, 200], which will allow us to maintain the resolution of the momentum measurement below $\Delta p/p = 1.0\%$ for momenta up to 200 GeV. A hybrid-pixel technology with a pitch of 55 µm will be used instead of the silicon-strip sensors in VELO. This will help perform the vertex reconstruction in the Run3 conditions with an increased pileup. The TT detector will be replaced by a new silicon *Upstream Tracker* (UT) and a *Scintillating Fibre* (SciFi) detector will be installed in the three downstream tracking stations. As a result of a lower material budget in the new tracking modules, and therefore less multiple scattering interactions, the momentum resolution is expected to improve by about 10%–20% as compared to the current LHCb experiment.

3.2.1.2 Particle identification

The PID system will also be modernised significantly [201]. The RICH upgrades involve changes to the photodetectors and electronics as well as to some parts of the mechanical design and optics. In particular the HPD sensors (which are limited by their embedded 1 MHz readout electronics) will be replaced by *Multianode Photomultiplier Tubes* (MaPMTs)

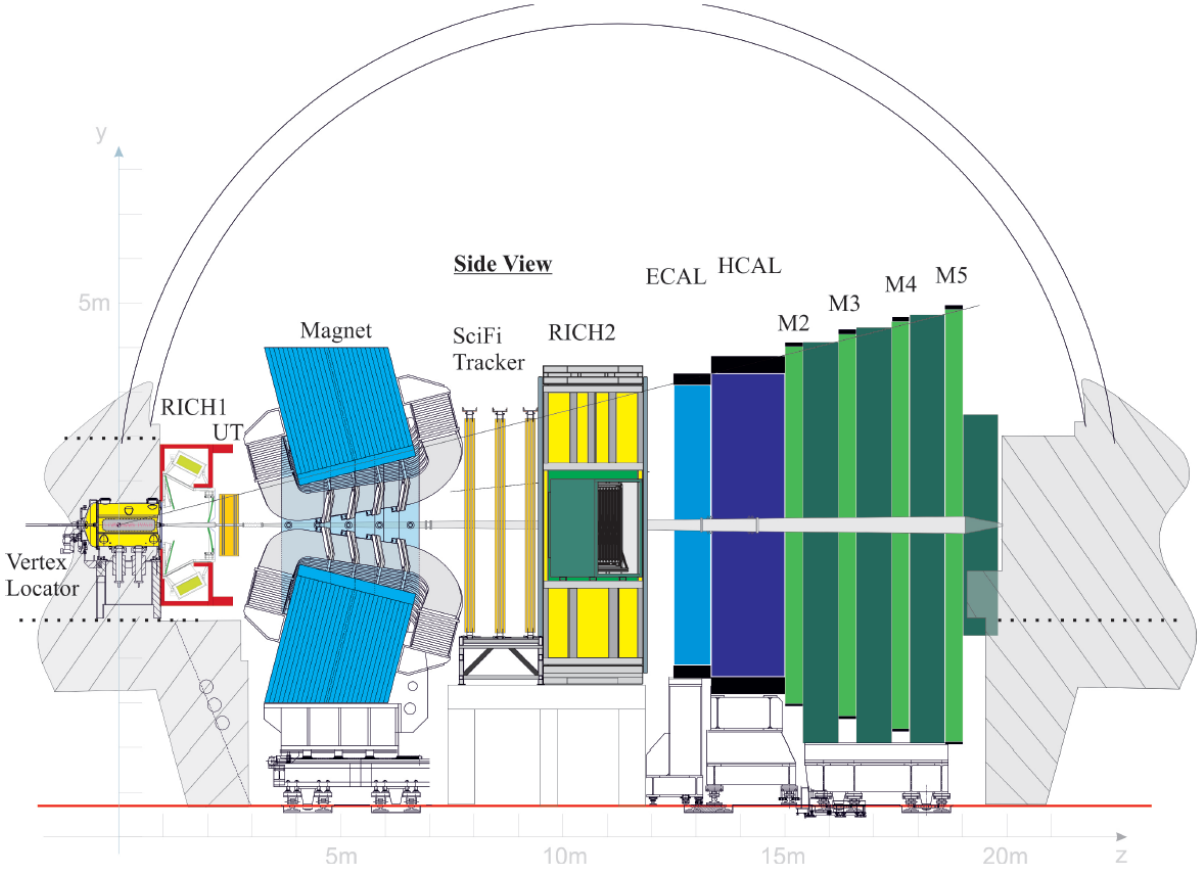

Figure 3.15: Schematic view of the LHCb Upgrade I detector [196].

with custom CLARO readout chips (see Sect. 3.3.3.2). These developments are described in more detail in Sect. 3.3.3. The M1 station of the muon system will be substituted by a neutron shield, to provide the SciFi photon detectors a protection for backsplash from the electromagnetic calorimeter. In the current LHCb experiment, this station was mainly used to enhance the p_{T} measurement for the hardware trigger, which will not be relevant for the completely software-based system.

3.2.1.3 Trigger and online data-processing

The increased amount of data and the new trigger implementation will also be taken into account in the updated computing model [183, 202]. The data-processing scheme for Run3 is shown in Fig. 3.16. The Upgrade Ia approach is based on solutions which were already implemented and successfully tested in Run2. The software trigger will be divided into two stages. A partial event reconstruction will be performed at the HLT1 phase and the data will be saved to buffer disks at a rate of approximately 1 MHz. This information will allow us to perform a real-time calibration and alignment of the detector elements, which will be used by HLT2. The latter will perform a full event reconstruction with a quality equivalent to the offline one and will apply the relevant trigger selections. This will mean that analyses can be performed directly on the data provided by the trigger and no additional offline reconstruction will be necessary. The HLT2 phase will run asynchronously, which will allow for the optimal use of available computing resources (*e.g.* by executing in the periods where no collisions are recorded). One of the important features of this new online system

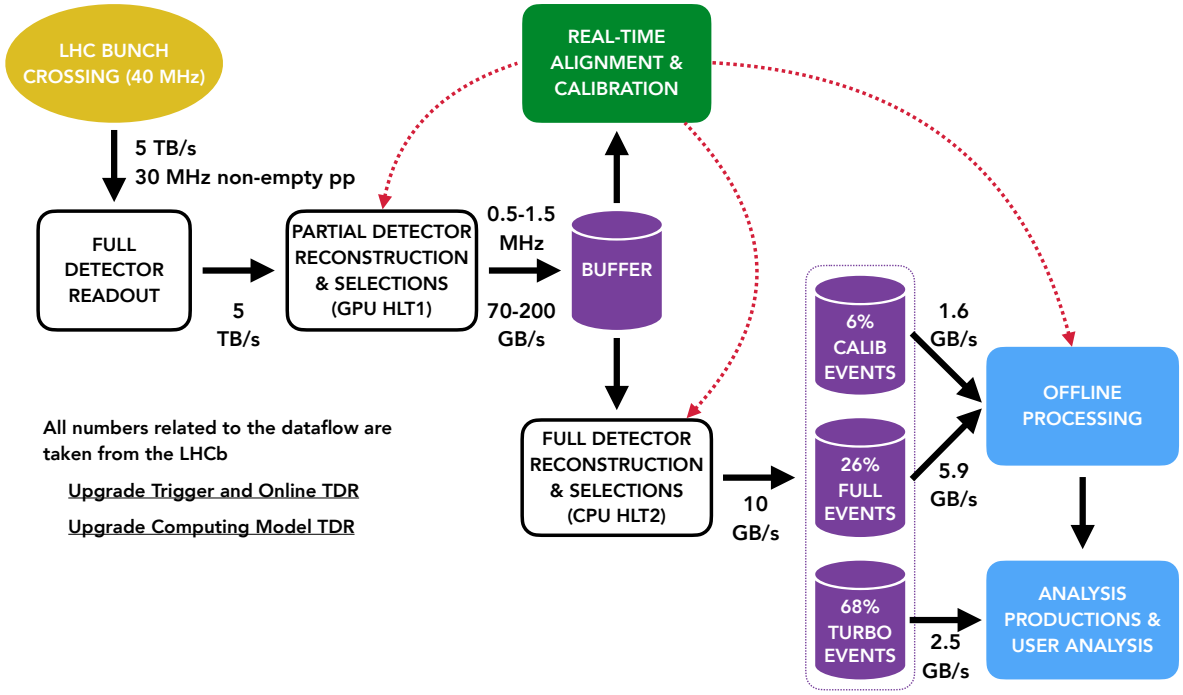

Figure 3.16: Diagram of the online data-processing flow for the upgraded LHCb detector, focusing on the real-time aspects [180].

is that the initial data volume will be sent by radiation-hard optical links from the detector front-end electronics to a processing farm located in a data-centre building on the surface. The radiation-free environment will allow the computing to benefit from commercially available technologies. It will permit to facilitate the most cost-effective solutions for real-time data processing and will be crucial not only in the current Upgrade Ia, but also in the Upgrade II period.

3.2.2 Upgrade Ib

The Upgrade Ib phase is considered as a consolidation of the previous detector modernisations. However, due to the extended long-shutdown period, it is also a good opportunity for some modest detector improvements and preparations to Upgrade II. Two particularly interesting developments are planned for this stage, which are related to entirely new subdetectors that will expand the physics coverage of the LHCb experiment. The first extension will be made in the tracking system, by installing additional magnet stations. These scintillating-fibre detectors will be placed at the internal surfaces of the dipole magnet (see Fig. 3.17) and will expand the experiment physics reach to low momentum tracks (such as low momentum electrons used in studies of the photon scattering processes [203]). The second improvement involves a novel TORCH (*Time Of internally Reflected CHerenkov light*) module [204], which is in principle envisaged for the Upgrade II phase, but will aim for installation to be already done in the Upgrade Ib period (see more details in Sect. 3.2.3.2). The RICH detectors will also take advantage of this opportunity to implement a redesigned, fast electronics chain. While this will already provide an improvement in the PID performance for Run4, the time-resolved sensors will be absolutely crucial for RICH detectors in Upgrade II. More details on the usage of

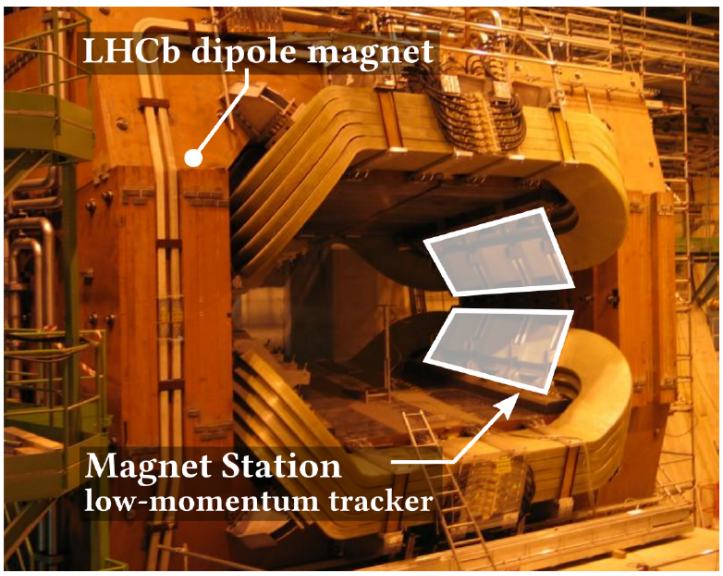

Figure 3.17: LHCb dipole magnet, with the white outline indicating the area to be covered by the magnet station. A symmetrical module will cover the opposite face of the magnet [195].

the time information in the RICH system can be found in Ref. [205]. The Upgrade Ib period will also allow some of the subdetectors (ECAL and SciFi) to replace their inner regions due to the expected high radiation damage. In the case of SciFi, populating silicon sensors in the regions closest to the beam pipe will also be an initial step in modernising the tracking system for Upgrade II (see more details on *Mighty Tracker* in Sect. 3.2.3.1).

3.2.3 Upgrade II

At the Upgrade II [195] conditions, each bunch-crossing is expected to provide $\mu_{\text{vis}} = 56$ visible interactions, which corresponds to around 2500 particles produced in the LHCb acceptance. It poses an extreme challenge for a successful reconstruction of the recorded tracks. Those demands will be addressed by increasing the detector granularity and reducing the amount of material interactions (*e.g.* by removing the RF foil in VELO), but it is the precision timing that will play a crucial role in that period. The overall design of the LHCb detector in the Upgrade II phase is presented in Fig. 3.18 and the modernisations with respect to the Upgrade I version are detailed in the following sections.

3.2.3.1 Track and vertex reconstruction

Some of the planned improvements of the tracking system in Upgrade II are already mentioned in the previous paragraphs, but here a more comprehensive overview is given. The VELO system will be upgraded to a 4D hybrid-pixel detector, with an increased granularity and additional time information. The RF foil will also be removed to reduce the material budget between the interaction point and the VELO sensors. Precise time information at the level of the vertex reconstruction will be crucial in lowering the combinatorics between the VELO tracks and the tracks from other parts of the tracking system, which will be highly beneficial in terms of the necessary computing resources. The reconstructed vertex time will also be useful for other subdetectors that will handle the timing

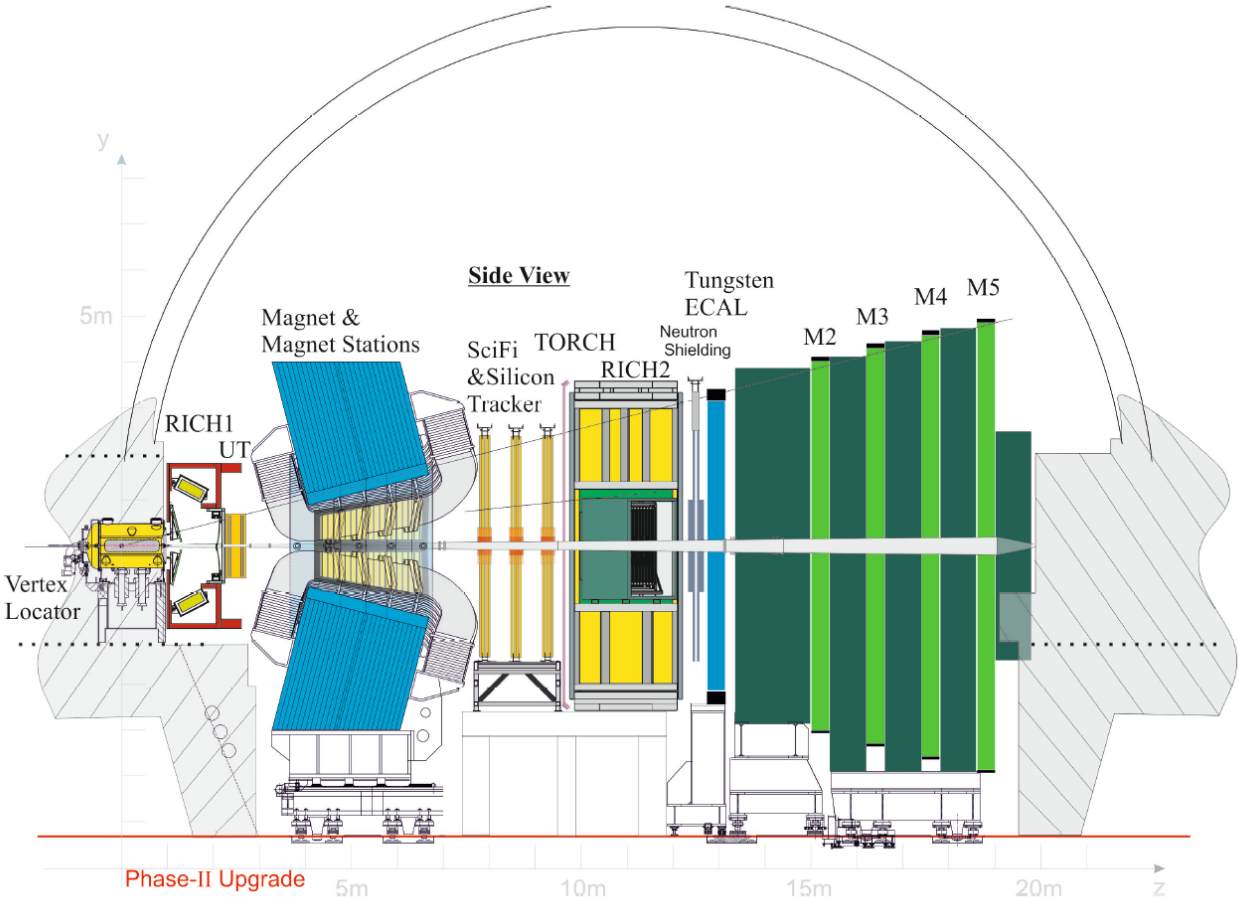

Figure 3.18: Schematic side-view of the Upgrade II detector [195].

information, *e.g.* TORCH and RICH in the particle identification. High-occupancy levels require an increase in the granularity of the downstream tracking stations as well. The inner and middle regions are planned to be replaced with silicon sensors, and this project, combined with the outer SciFi parts, is referred to as the Mighty Tracker (see Fig. 3.19). As mentioned in Sect. 3.2.2, the inner parts of this new silicon tracker will be already installed in Upgrade Ib. For similar reasons as in the VELO case, there are ongoing studies on implementing the time information in at least a part of the tracking detectors.

3.2.3.2 Particle identification

The particle identification system will need to undergo significant modernisations to maintain (or even improve) its excellent performance in Upgrade II. Similar to the tracking detectors, it will be mainly related to increasing the granularity and introducing subnanosecond timing, which will allow us to combine the signals with the relevant vertices, and thus effectively reduce the combinatorics in the PID algorithms.

The increased occupancies will require higher granularity of the RICH photodetectors, together with the subnanosecond time resolution. Several sensor technologies are being investigated, including HPDs, MaPMTs, *microchannel plates* (MCPs) and *silicon photomultipliers* (SiPMs). An overall improvement of the Cherenkov angle resolution

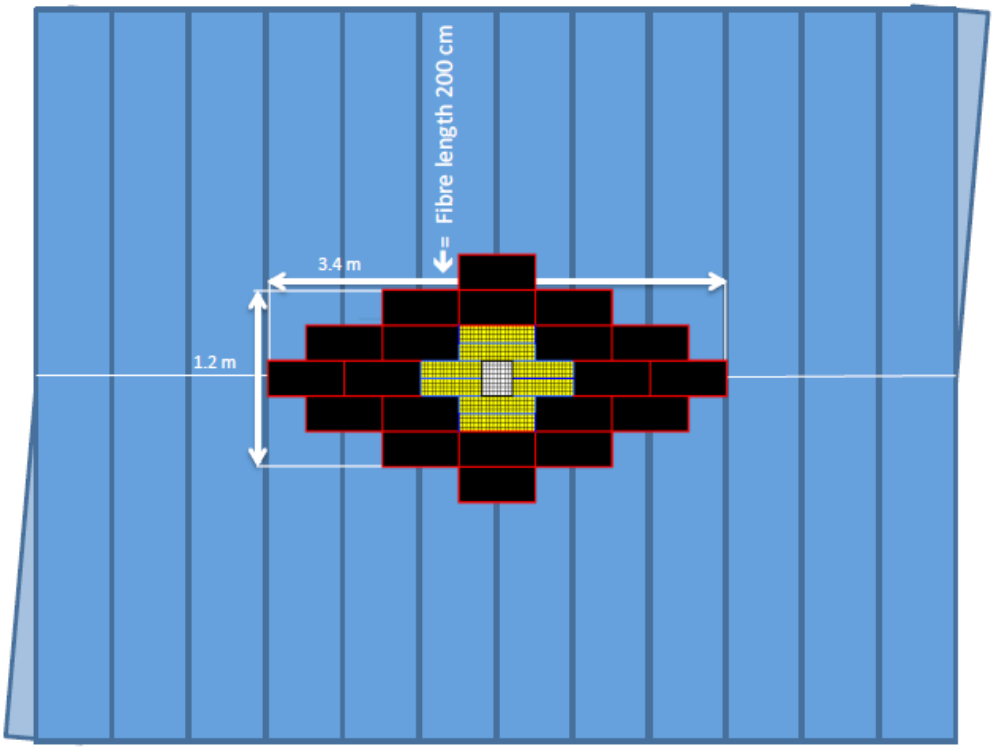

Figure 3.19: Proposed layout of the Mighty Tracker station with scintillating fibres (blue), silicon middle (black) and silicon inner tracker (yellow) [195].

by a factor at a level of three is required (with respect to Upgrade I values), which will involve redesigning the optics of the RICH detectors. Some of these aspects are described in more detail in Sect. 3.3.

An interesting extension of the PID capabilities of the LHCb experiment can be achieved by the installation of a novel TORCH detector. This apparatus provides the time-of-flight information by collecting internally-reflected Cherenkov photons produced in a thin quartz plate (see Fig. 3.20). It will allow for the positive identification of low-momentum protons as well as kaons in the momentum region below 10 GeV, which is not possible with the current RICH system. The TORCH modules will be installed upstream of the RICH2 detector, possibly already in the Upgrade Ib phase.

As mentioned in Sect. 3.2.2, the inner part of ECAL will need to be already replaced in Upgrade Ib, due to the expected radiation damage. It corresponds to the central detector band with a vertical size of approximately 1.5 m. Even higher occupancies and radiation doses are expected in Upgrade II, which is why the calorimeter granularity will be improved, and the radiation hardness will be an important factor in the technology choice. The upgraded ECAL will also be capable of fast timing resolution, which will allow for the reduction of the combinatoric background, for example in the reconstruction of the neutral π^0 mesons.

Similar to some other subdetectors, new modules with increased granularity will be installed in the inner parts of the muon system. An important factor will also be an additional shielding from the flux of particles arriving from the upstream region. This will be achieved by replacing HCAL (which was mostly used to provide fast information for

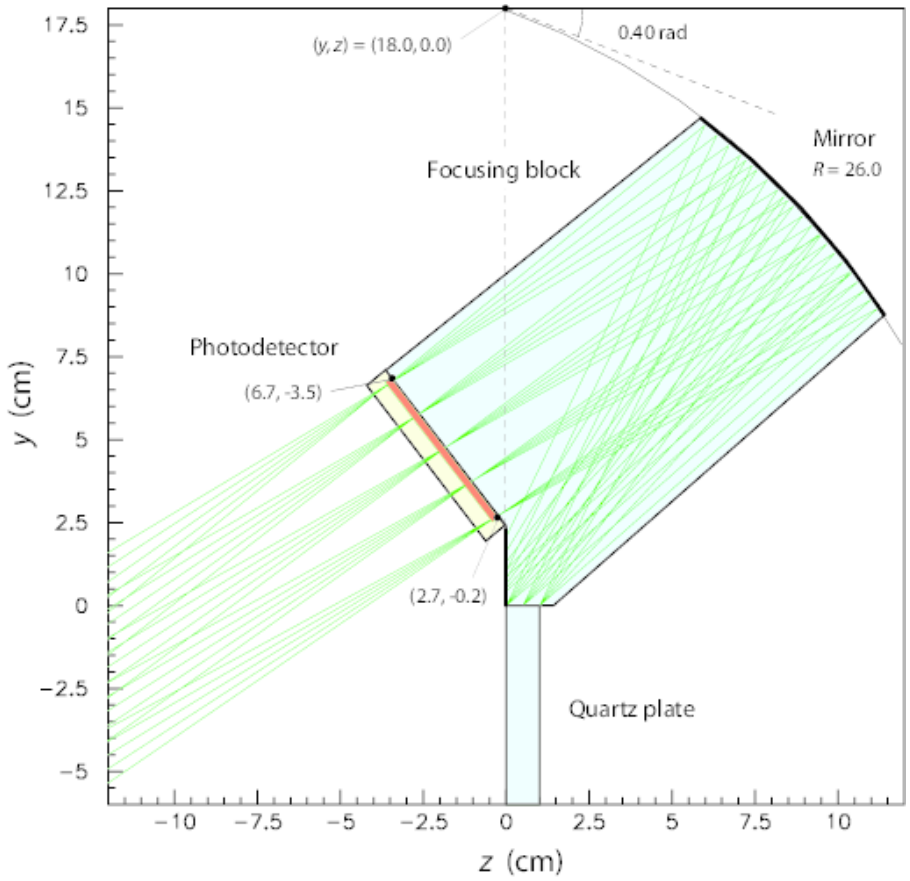

Figure 3.20: Cross-section of the focusing element of the TORCH detector. The focusing of photons is shown for five illustrative angles between 450 and 850 mrad, emerging at different points across the edge of the plate [204].

the hardware trigger, and is not so relevant for Upgrade I and Upgrade II) with a 1.7 m thick layer of iron, which is a volume corresponding to four additional interaction lengths.

3.2.3.3 Trigger and online data-processing

The high number of interactions per bunch-crossing in Upgrade II will lead to a shift in the paradigm of the trigger system. In the previous periods the event selection is based in many cases on inclusive topological selections (such as a displaced vertex, which can be a signature of a decaying heavy-flavour hadron). In the Upgrade II conditions, multiple interactions of this kind can happen in a single bunch-crossing, so such selections will not be effective in reducing the data rate. An alternative method will be based on the *pileup suppression*, which means that the events will be reconstructed more globally, but the hits not associated with the given primary interaction will be discarded at an early stage of this process, effectively limiting the relevant combinatorics. Same as in Upgrade I, the data-processing farm located on the surface (in a radiation-free environment) will allow us to benefit from the most effective commercial computing solutions available at the time of Upgrade II.

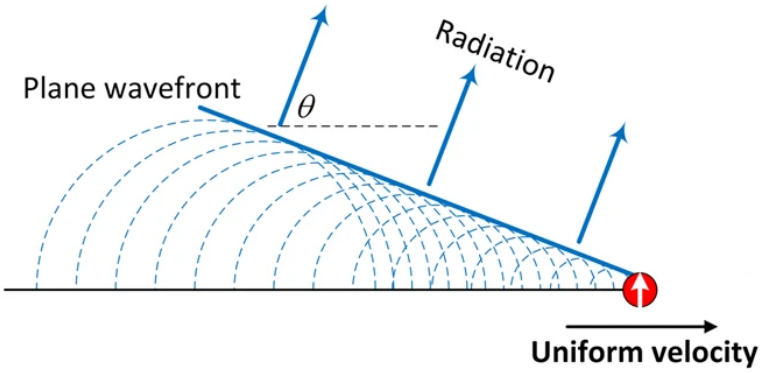

Figure 3.21: Diagram illustrating the Cherenkov radiation emitted by a charged particle moving at a constant velocity [206].

3.3 LHCb RICH detectors – performance and upgrade

While the LHCb RICH detectors are already described in Sects. 3.1.3, 3.2.1.2, 3.2.2 and 3.2.3.2, these sections mainly aim to characterise the role of these modules in the overall PID system of the current LHCb experiment and of its upgrades. The present section is focused on the development of the RICH detectors themselves, illustrating basic concepts of such detectors, and the measures taken or considered to meet the PID requirements of the LHCb experiment upgrades. The basic concepts are explained using an example of the current LHCb detector [14,15,176,177] in Sect. 3.3.1, while Sect. 3.3.2 describes a general approach to improving some aspects of the RICH performance in future upgrades. An overview of the modifications applied to the RICH system in the Upgrade Ia period [201] is provided in Sect. 3.3.3. The author of this thesis has been actively engaged in the preparation of the simulation of the upgraded RICH detectors, and his contribution is described in more detail in Sect. 3.3.4.

3.3.1 Basic concepts of the RICH detectors

The RICH detectors in the LHCb experiment provide an identification of long-lived charged hadrons, which is particularly important in the heavy-flavour physics studies, to discriminate the decays with similar topologies. A mechanism employed in this kind of devices is the Cherenkov radiation, which occurs when a charged particle traverses a medium with a velocity higher than the speed of light in the given material. The Cherenkov photons are emitted at a characteristic angle θ_C, forming a cone-like structure (see Fig. 3.21). The value of the Cherenkov angle is described by

$$\cos(\theta_C) = \frac{1}{n\beta} , \tag{3.1}$$

where n is the refractive index of the medium and $\beta = v/c$ is the ratio of the charged particle velocity to the speed of light in the vacuum. Thus, the measurement of the Cherenkov angle provides information on the particle velocity, which, combined with the particle momentum measured by the tracking system, allows us to identify a particle by its mass.

It is worth noting at this point that for particles with $\beta \to 1$, the Cherenkov angle approaches its maximum value ${\theta_C}^{\max}$ in the given medium and the discrimination between

different particle species is not possible. Such tracks are referred to as *saturated* ones, and the saturation is reached by different particle species at different velocity values (see Fig. 3.9). Hence, the choice of the radiators is essential to ensure the required discrimination properties across a specified momentum range. In the case of the LHCb RICH detectors, two fluorocarbon gases are used as radiators – C_4F_{10} and CF_4 for RICH1 and RICH2, respectively[5]. These gases are chosen due to their relatively narrow Cherenkov angle range (their refractive indices are close to unity) and low chromatic dispersion (the latter is relevant for the resolution of the measured Cherenkov angle – see the following paragraphs for more details).

The concept of the ring-imaging Cherenkov detectors was proposed in Ref. [207]. Devices of this type make use of spherical mirrors in order to image the cone-like structure of the emitted Cherenkov photons into a ring shape on a dedicated detection plane. Measuring the radius of this ring allows us to obtain the value of the Cherenkov angle. In the LHCb RICH detectors, two types of mirrors are used to place the detection plane outside of the spectrometer acceptance as well as to meet the space constraints in the LHCb experiment. Slightly tilted spherical mirrors allow for the deflection of the photons away from the beam direction. Secondary flat mirrors are installed to redirect the photons towards the detection plane. The optical system is designed in such a way that the detection plane position is optimised with respect to the focal plane of the spherical mirrors, which allows focused images of the Cherenkov rings to be obtained on the photon detector planes (see Fig. 3.8 and Fig. 3.22).

A resolution of the Cherenkov angle measurement is the key factor that defines the performance of the RICH detectors. The main contributions to the overall single-photon resolution σ_θ are listed below:

- *chromatic error* ($\sigma_{\text{chromatic}}$) – related to the dispersion in the radiator; the Cherenkov photons produced by the same charged particle are emitted at slightly different angles, as a result of the refractive index variations with the photon wavelength,

- *emission point error* (σ_{emission}) – due to optical aberrations of the mirror system, the photons that are created at various points along the particle trajectory arrive at different positions on the detection plane; the detection plane can also be displaced with respect to the actual focal plane, *e.g.* due to engineering constraints, which further contributes to this type of error,

- *pixel size error* (σ_{pixel}) – related to the granularity of the photon detectors, which needs to be optimised with respect to the detector requirements.

Assuming that all of the Cherenkov photons in the ring are produced by a single charged particle, the total Cherenkov angle resolution per (saturated) track $\Delta\theta_{\text{C}}$ is defined as

$$\Delta\theta_{\text{C}} = \frac{\sigma_\theta}{\sqrt{N_{\text{ph}}}} \oplus C_{\text{tracking}} \ , \tag{3.2}$$

where N_{ph} corresponds to the average number of detected Cherenkov photons in the ring (*photon yield*), the C_{tracking} component refers to the tracking resolution, and the $\oplus$ symbol denotes a quadratic sum.

[5]Initially, an aerogel radiator was also employed in RICH1, but it was removed due to its limited effectiveness before the Run2 operations – see Sect. 3.1.3 for more details.

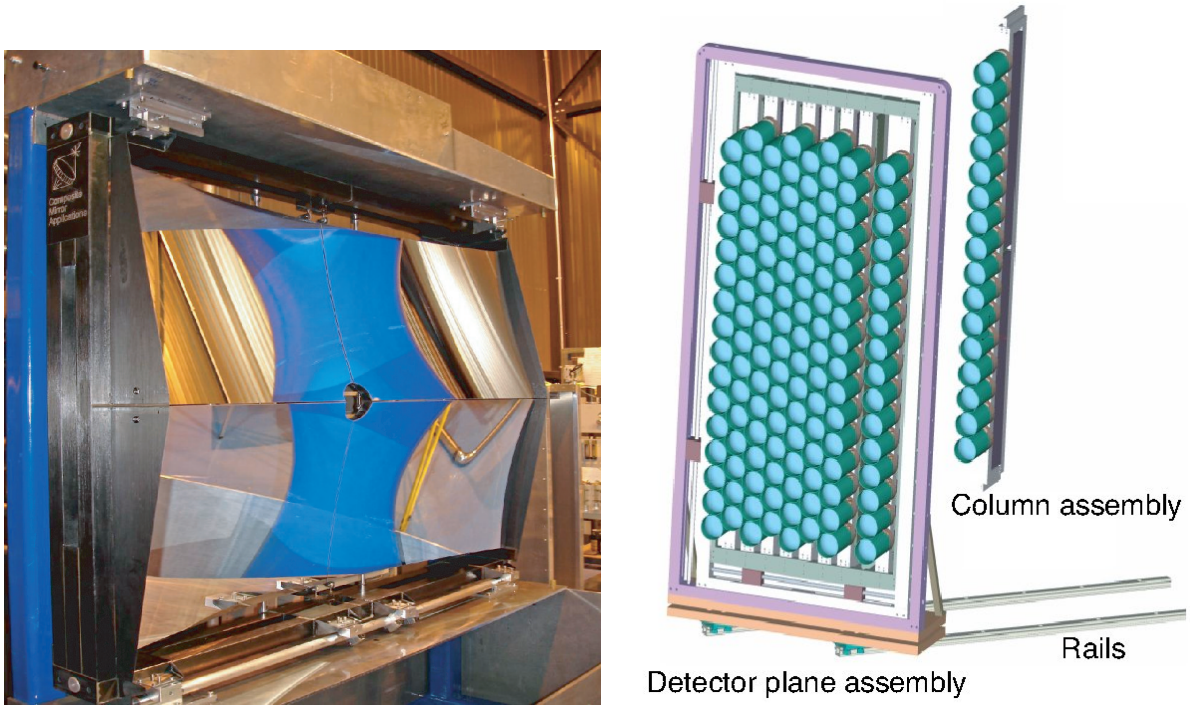

Figure 3.22: (left) Spherical-mirror array of the current RICH1 mounted onto the support frame viewed from the front and (right) the current RICH2 photon detector plane with the HPD tubes visible [14].

As indicated in Eq. (3.2), a higher photon yield improves the resolution of the Cherenkov angle measurement. The optical and photon detection systems need to be designed in a way that optimises their working range with respect to the Cherenkov photons spectrum (this also refers to the chromatic error reduction – see Sect. 3.3.2 for a more detailed discussion on this issue). One of the methods to achieve this is to apply dedicated coatings to the mirrors, in order to obtain the required reflectivity profile in the function of photon energy (or wavelength) (see Fig. 3.23 (left)). Furthermore, the choice of the photon detector technology is essential due to different *quantum efficiency* (QE) spectra of the detector photocathodes (see *e.g.* Fig. 3.23 (right)). This quantity is defined as

$$\mathrm{QE} = \frac{N_{pe}}{N_{\gamma}} \ , \tag{3.3}$$

where N_{γ} corresponds to the number of Cherenkov photons reaching the photocathode and N_{pe} denotes the number of created photoelectrons. The final single-photon detection efficiency is a product of QE and the inefficiencies of the detection system (related *e.g.* to the photon detector geometry, the charge amplification process and the digitisation of the photodetector response).

A typical event in RICH1 presenting photon hits on the detection plane and the corresponding ring candidates is shown in Fig. 3.24. As can be seen in the figure, many potential rings are overlapped in the event, and a pattern recognition algorithm is employed [208] to obtain the PID variables (see Sect. 3.1.3 for details). A hypothesis that each of the tracks is created by a pion is used in the first step of this procedure. Since the track momentum

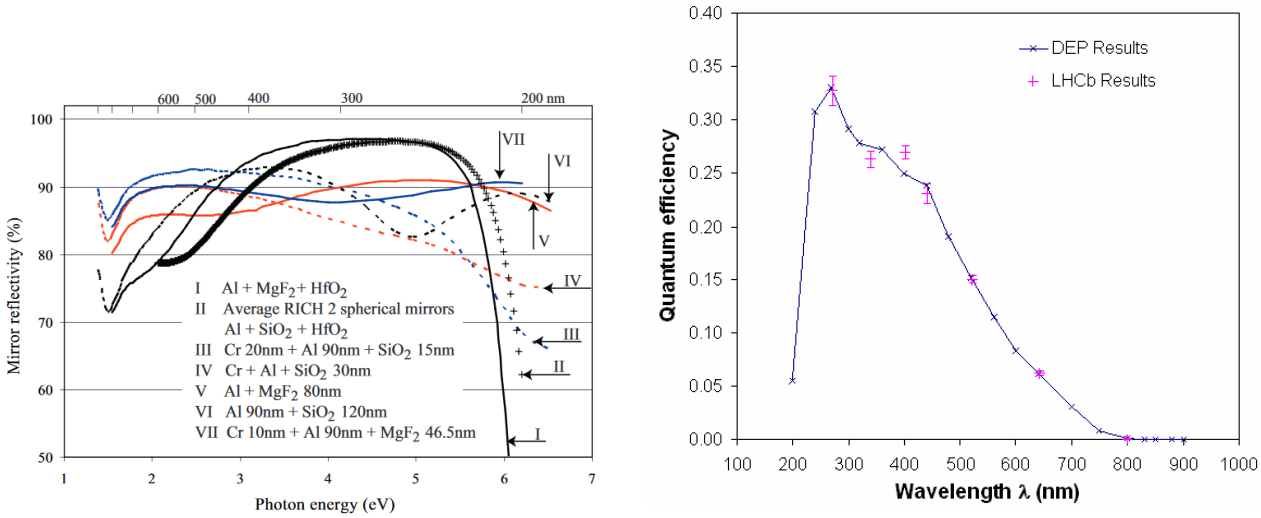

Figure 3.23: (left) Reflectivity of several mirror coatings on glass as a function of the photon energy with the angle of incidence of 30° and (right) QE measurement for one of the best HPDs in RICH [14].

and trajectory is known from the tracking system, it is possible to reconstruct the expected rings for the given mass hypothesis. The reconstructed rings are then compared with the observed hits and a global likelihood is determined. The likelihood value is optimised by changing the mass hypothesis for a single track at once, until the maximum value is obtained.

An exemplary efficiency of the PID algorithms during Run1 as a function of track momentum is shown in Fig. 3.25. As can be seen in the figure, a more strict requirement on the $\mathrm{DLL}_{K\pi}$ variable not only results in lower rates of misidentified pions, but also in a diminished kaon identification efficiency. The PID performance in that period is also illustrated in an alternative way, by presenting the fraction of misidentified pions as a function of efficiency of the kaon identification in Fig. 3.26. In this case, each data point corresponds to a different requirement on the $\mathrm{DLL}_{K\pi}$ variable, and the values are averaged over all track momenta. In general, it is observed that the PID performance degrades with increasing track multiplicity in an event.

3.3.2 Strategy for the RICH developments

The performance of the current LHCb RICH detectors in the upgrades conditions is restricted by a number of factors, such as the contributions to the single-photon resolution described in Sect. 3.3.1. Another limitation is related to the maximum occupancy at which the pattern recognition algorithms can perform effectively. A general approach to improving each of these factors in the future developments of the RICH detectors is outlined in Ref. [209] and summarised in the current section.

An evident solution for reducing the channel occupancy, and to improve the pixel size error, is to increase the detector granularity. It should be noted, however, that for an efficient performance of the pattern recognition algorithms, the following relation must be satisfied

$$\sigma_\theta f \lesssim A_{\mathrm{pixel}} \ , \tag{3.4}$$

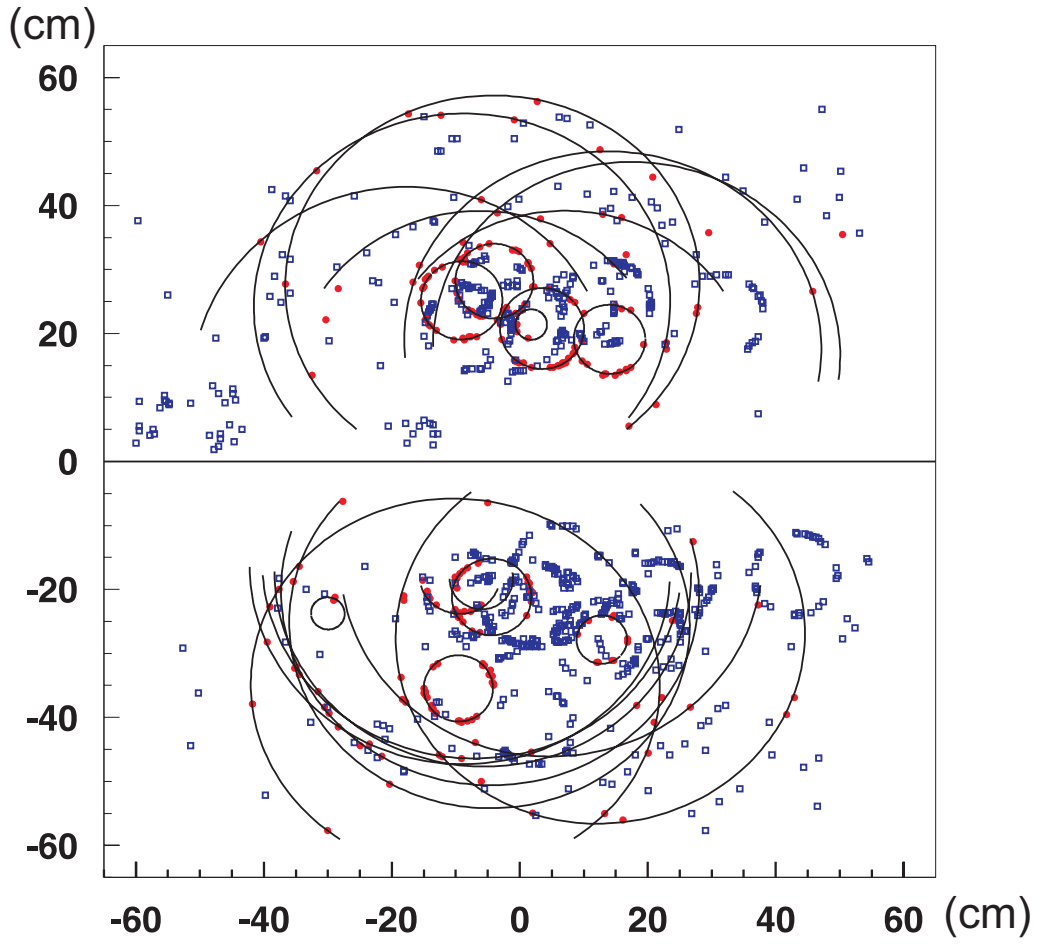

Figure 3.24: Display of a typical LHCb event in RICH1 [14].

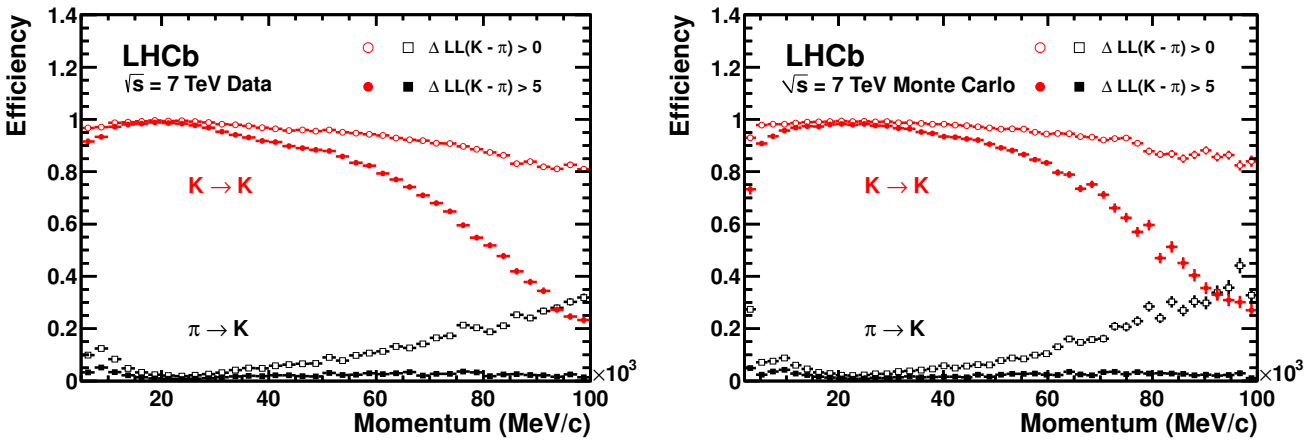

Figure 3.25: Kaon identification efficiency (red circles) and pion misidentification rate (black squares) as a function of track momentum (left) measured using the data and (right) from the simulation. Two different requirements on the $\mathrm{DLL}_{K\pi}$ value are imposed, resulting in the open and the filled marker distributions [177].

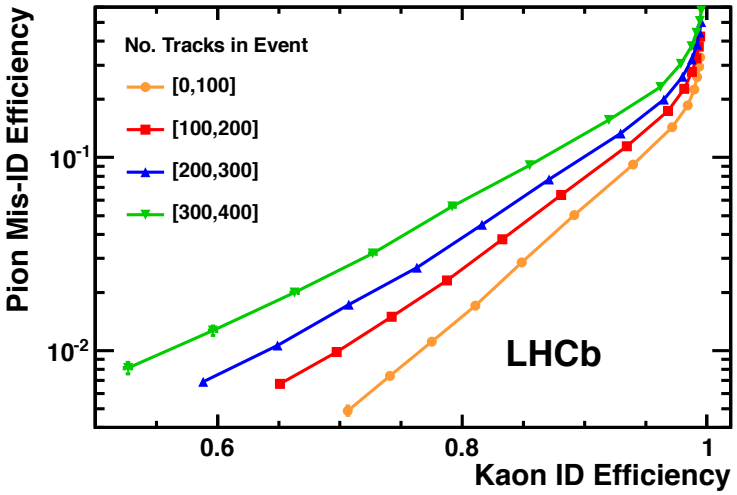

Figure 3.26: Pion misidentification rate as a function of kaon identification efficiency in the current LHCb experiment, as evaluated using the 7 TeV data, for different ranges of track multiplicity in an event [177].

where f denotes the focal length of the spherical mirrors and A_{pixel} corresponds to the area of the photon detector pixel. The above condition represents a situation when the Cherenkov rings have widths comparable to (not exceeding) a single pixel size.

The emission point error emerges from the optical aberrations of the RICH mirrors system. They can be related to the limitations of the mirrors manufacturing and to their positioning in the detector. The tilt of the spherical mirrors in the RICH system is necessary to redirect the photons to the flat mirrors, yet it inevitably contributes to the optical aberrations. The current flat mirrors need to be placed outside of the LHCb detector acceptance, due to their significant material budget that would affect the tracking performance. The emission point error can be improved *e.g.* by increasing the distance between the spherical and flat mirrors – thus also reducing the spherical mirrors tilt (this approach is employed in the Upgrade I phase) – or by using lightweight flat mirrors that could be placed within the spectrometer acceptance (see Fig. 3.27). The latter solution is envisaged for the Upgrade II period. The development of the lightweight composite mirrors is driven by the LHCb RICH group and their installation would significantly reduce the system aberrations during that phase. The emission point error can also be improved by increasing the focal length (and therefore the radius of curvature) of the spherical mirrors, which makes them less susceptible to the optical aberrations.

The chromatic error is related to the refractive index variations (with photon wavelength) in the RICH radiators. The fluorocarbon gases used in the current system are selected due to their relatively low dispersion, hence an improvement by the means of optimising the used radiators does not seem feasible. It is clear however, that the chromatic error is most prominent for the near-UV photons and it can be limited by selecting photon detectors (and also by applying customised coatings to the optical components) that are more sensitive in the wavelength region shifted towards the red. Furthermore, to compensate for the loss of the photons from the near-UV region and to keep a high yield of the Cherenkov photons, the overall quantum efficiency of the chosen photon detector technology needs to be respectively high.

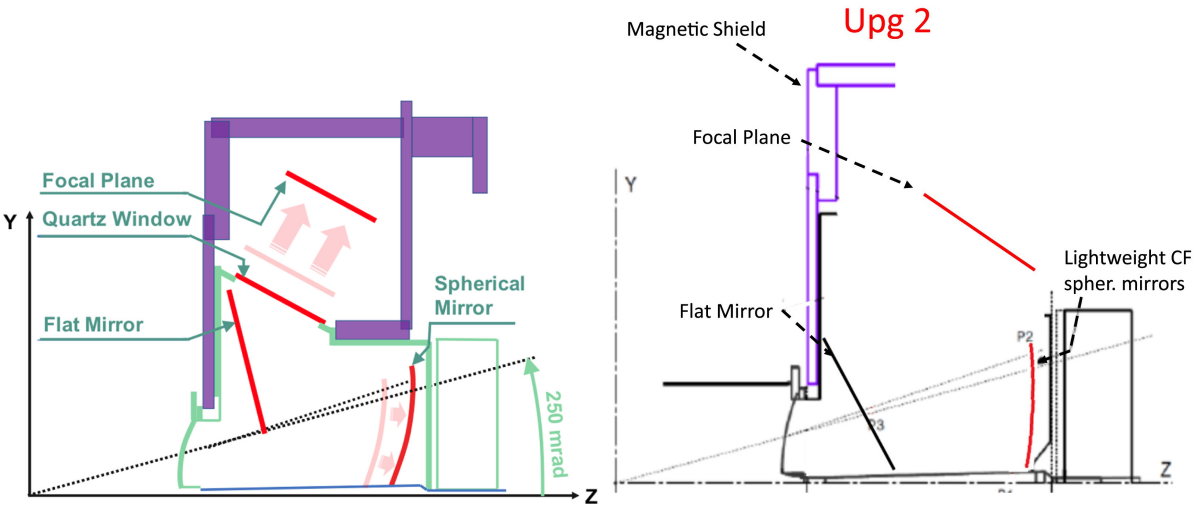

Figure 3.27: (left) Schematic drawing of the mechanical and optical system modifications for the RICH1 detector during Upgrade I compared to the current experiment [210] and (right) a proposed optical geometry for RICH1 in the Upgrade II period [209].

Preliminary detector designs have been proposed and studied for the LHCb upgrades, implementing the approaches described in this section to maintain or improve the RICH PID capabilities in the respective upgrade conditions. A comparison of the simulated performance of the RICH detectors in the mentioned upgrades is shown in Tab. 3.1. The table illustrates the evident advancements in the LHCb RICH design, where the various components of the performance are addressed, and significantly improved, in the consecutive detector versions. It also reflects the successful contribution of the RICH detectors to the LHCb experiment and its upgrades, with the Upgrade Ia detector already being installed, and the ongoing R&D activities on the detector technologies feasible for the Upgrade II construction.

3.3.3 Modifications in the Upgrade Ia phase

As already discussed in Sect. 3.2.1, the main challenges for the experiment operation in the Run3 conditions are related to the increased detector occupancies and the requirement of the full spectrometer readout at the 40 MHz rate. In the case of the RICH system, the increased occupancy would lead to a more significant overlap between the Cherenkov rings which would degrade the performance of the pattern recognition algorithms. The present section describes modifications designed for the Upgrade Ia RICH detectors to enhance the RICH performance. Excellent PID capabilities of the RICH detector were observed in Run1 and Run2. The mentioned modernisations are envisaged to maintain these capabilities also in the more demanding Run3 conditions. Adaptations to the optical system that aim to reduce the occupancy and to improve the total Cherenkov angle resolution are presented in Sect. 3.3.3.1. The following Sect. 3.3.3.2 describes the new photon detection chain that addresses the need of the 40 MHz readout rate and also mitigates the pixel error contribution to the overall single-photon resolution.

Table 3.1: Preliminary simulated performance of the various versions of the upgraded RICH detectors. The relevant quantities are also shown for the current RICH1 detector for reference [209].

Radiator		C_4F_{10}		CF_4	
Detector Version	RICH1 Current (HPD)	RICH1 UPG1	**RICH1 UPG2**	RICH2 UPG1	**RICH2 UPG2**
Average Photoelectron Yield	30	40	**60–30**	22	**30**
Single Photon Errors (mrad)					
Chromatic	0.84	0.58	**0.24–0.12**	0.31	**0.1**
Pixel	0.9	0.44	**0.15**	0.20	**0.07**
Emission Point	0.8	0.37	**0.1**	0.27	**0.05**
Overall	1.47	0.82	**0.3–0.2**	0.46	**0.13**

3.3.3.1 Optics

Compared to the current LHCb experiment conditions, a two-fold increase in the average channel occupancy is expected in the Upgrade I phase [209]. This would clearly deteriorate the PID performance – especially in RICH1, where the peak hit occupancy values would be at the level of 60%. To compensate for this effect, the whole optical system in RICH1 has been redesigned to increase the area corresponding to the photon ring images by a factor of two, thus smearing the photon hits over a larger surface on the photon detection plane and reducing the peak occupancy. This is also possible due to the lack of the aerogel radiator, which was removed from the current RICH detectors before the Run2 operations (see Sect. 3.1.3 for more details). The aerogel removal significantly limits the maximum sizes of the rings in RICH1 and allows for the usage of the full area of the photon detector planes for the detection of rings originating from the C_4F_{10} radiator. The magnification of the Cherenkov rings is achieved by increasing the focal length of the RICH1 spherical mirrors by a factor of $\sqrt{2}$ (which corresponds to installing larger mirrors with the radius of curvature close to 3.7 m, instead of 2.7 m). Both the spherical mirrors and the detection plane positions are also moved back with respect to their current placements and the mechanical structure is adapted accordingly (see Fig. 3.27 (left)). A slightly longer distance between the flat and the spherical mirrors allows for a smaller tilt of the latter, which, together with the increased curvature radius of the spherical mirrors, reduces the emission point error of the system (see Sect. 3.3.2). A longer path in the gas radiator for the charged tracks results in a higher yield of the Cherenkov photons and thus enhances the total Cherenkov angle resolution (see Tab. 3.1).

The components of the RICH1 optical system are covered with dedicated coatings to improve their optical properties [211]. This refers to an antireflective layer on both sides of the gas exit quartz-windows (see Fig. 3.27 (left)), which is applied to enhance the photon transmission in the quartz windows. Similarly, both the spherical and the flat mirrors

are covered with coatings that improve their reflectivity. The adaptations described in this paragraph provide an excellent performance of the RICH1 mirrors system, with the reflectivity values higher than 90% in the wavelength region from 275 nm up to 500 nm. The coatings also shift the working range of the RICH1 optical system towards the red, which is beneficial for reducing the detector chromatic error (see Sect. 3.3.2).

3.3.3.2 Photon detection chain

As mentioned in Sect. 3.2.1.2, the readout rate of the current RICH system is limited to 1 MHz by the embedded readout electronics of the HPD sensors. To allow for operating the detector at the LHC bunch-crossing rate of 40 MHz, the photon detection chain has been redesigned, with the HPDs replaced by MaPMT tubes [212, 213] that are read out by custom-made CLARO chips [214, 215].

The MaPMT tubes allow for a single-photon detection in the RICH system. This is achieved by amplifying the photoelectrons created at the photocathode in an array of dynode chains and collecting the created photocurrent in one of the respective anodes (pixels). Two types of MaPMTs are employed: a 1-inch R13742 (R-type) and a 2-inch R13743 (H-type) model. Both devices are customised versions of the original Hamamatsu[6] tubes from R11265 and R12699 series, respectively, with tightened specifications to meet the requirements of the RICH production [212, 213]. The tubes are equipped with super-bialkali photocathodes deposited on UV-glass windows. A typical quantum efficiency of the photocathodes, which is an average of a number of MaPMT measurements performed by the LHCb RICH group at CERN, is shown in Fig. 3.28. Each tube has 8×8 pixels with the size of 2.8 mm or 6 mm (for the R-type and the H-type ones, respectively) and nominal gains at the order of 10^6 for a 1 kV bias voltage, with a pixel-to-pixel gain variation at the level of $1 : 3$. The H-type devices are populated in the outer regions of the RICH2 detector (see Fig. 3.29 (left)), where the hit rate is expected to be relatively low. In total, 3100 (450) units of the R-type (H-type) MaPMTs are produced and their characteristics have been tested in a *Photon Detector Quality Assurance* (PDQA) procedure [213].

The MaPMT tubes are read out asynchronously by the 8-channel CLARO ASIC chips [214, 215]. These devices incorporate a charge-sensitive amplifier and a discriminator with a configurable threshold, which allows for the transformation of the analogue MaPMT outputs to digital signals. The recovery time of CLARO is shorter than 25 ns, which allows for an operation at the 40 MHz rate. The chip design is also optimised for relatively low power consumption at the level of 1 mW per detector channel and a radiation-hardness sufficient for the Upgrade I conditions.

The CLARO outputs are captured asynchronously by an FPGA-based *Photon Detector Module Digital Board* (PDMDB), which processes the digital signals, synchronises them to the LHC clock and sends them in the proper format to the LHCb *data acquisition* (DAQ) system using optical links. The FPGA boards are adjusted to the Upgrade I conditions in terms of the radiation hardness and the usage of their logic resources is kept at the minimum level to reduce the probability of the potential electronics upsets. The installed FPGAs allow us to sample the CLARO response at a 320 MHz rate, producing 8-bit time-structure patterns for each 25 ns period (see Fig. 3.30 (left)). The accepted patterns can be provided by a programmable lookup table, effectively enabling the implementation of various readout

[6]Hamamatsu Photonics (`https://www.hamamatsu.com`).

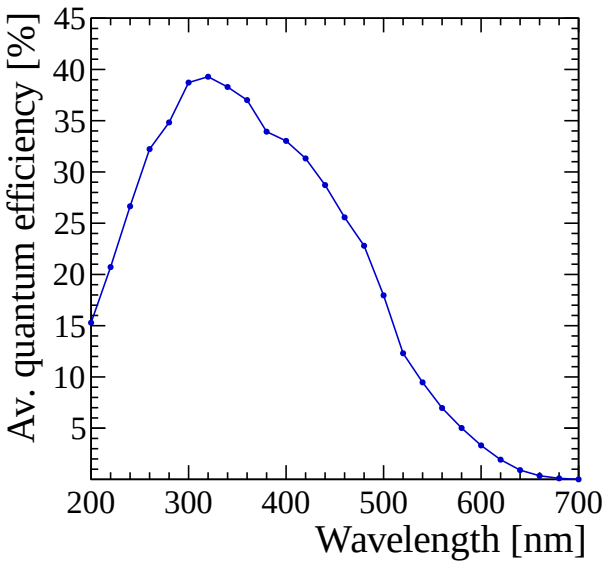

Figure 3.28: Average photocathode quantum efficiency for a number of RICH MaPMTs measured at CERN [216].

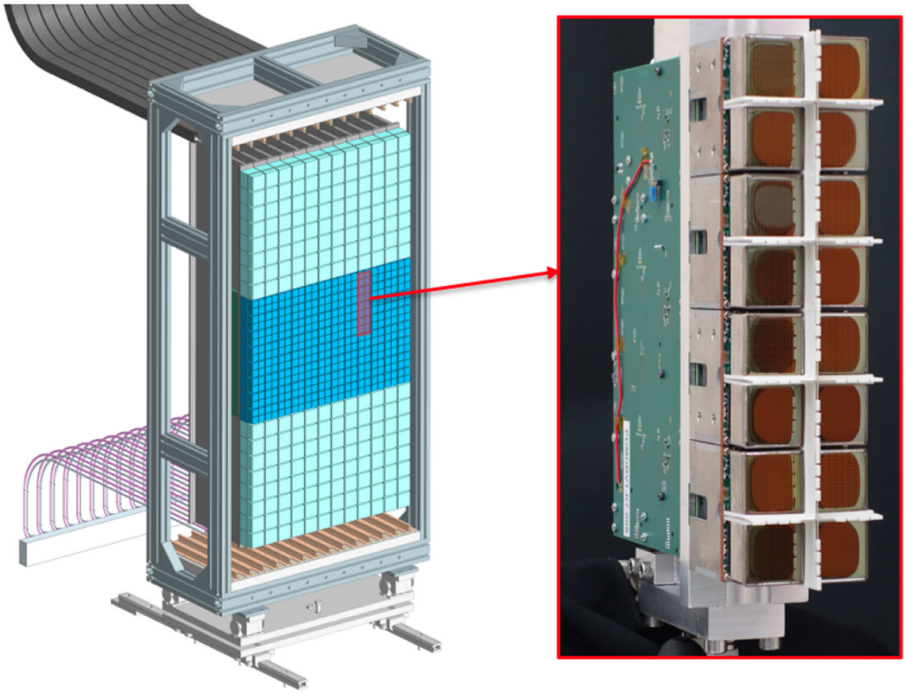

Figure 3.29: (left) Model of the RICH2 photodetector plane and (right) a picture of a photon detector module (with magnetic shields for RICH1 only) [210].

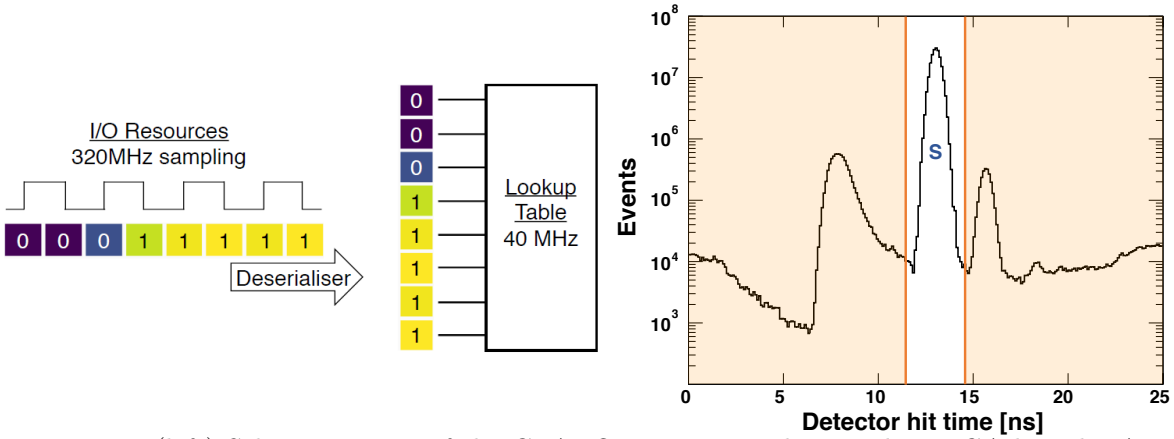

Figure 3.30: (left) Schematic view of the CLARO output sampling in the FPGA boards. A programmable lookup table of the accepted patterns allows for a flexible readout configuration. (right) Simulated photon detector hit time for RICH1. The signal peak (S), background contributions and a potential time gate implemented in the front-end electronics are shown [205].

configuration scenarios. This feature will be used to introduce a time gate around the signal peak in front-end electronics, which can help to significantly reduce the contribution of the out-of-time noise and background sources. The concept of the time gate is illustrated in Fig. 3.30 (right). Depending on the final gate configuration, up to an eight-fold reduction in the detection of the noise sources that are uniformly distributed in time (such as the dark counts or a signal-induced noise – see Sect. 3.3.4.3) could be achieved.

The design of the photon detection chain is modular and the basic component of this scheme is referred to as an *Elementary Cell* (EC)[7] [217, 218]. Each EC consists of four R-type MaPMT tubes mounted on a baseboard that transfers the signals to the *Front-End Boards* (FEBs) housing the CLARO chips (see Fig. 3.31). There are four FEB boards in an EC and a single one contains eight CLARO chips. The FEBs are connected to a backboard that provides an interface to the PDMDBs. Due to the sensitivity of the MaPMT photon detection efficiency to the magnetic field, the ECs in RICH1 are also equipped with cross-shaped magnetic shields (in RICH2 the intensity of the magnetic field is significantly lower).

The next level of modularity is represented by *Photon Detector Modules* (PDMs). They comprise four elementary cells (see Fig. 3.29 (right)) connected to two PDMDB boards. The modules are organised into mechanical structures referred to as *columns*, which provide common supplies (*e.g.* cooling, low and high voltage, monitoring devices) and a support, also allowing for an easy removal of the modules for the maintenance purposes. The final photon detection planes consist of columns – eleven (twelve) per panel in RICH1 (RICH2) – each of them containing six modules (see Fig. 3.32). The outermost EC on each side of the RICH1 columns – where the hit rate is expected to be extremely low – is not mounted, to simplify the mechanical handling and the maintenance of the columns.

[7] In the main text, the structure of the photon detection chain is described, if not stated otherwise, for the components containing only the R-type MaPMTs (the two types of MaPMTs are not mixed within an elementary cell or a module). The description for the H-type components is given below. An H-type EC contains only a single H-type MaPMT, which corresponds to fewer channels as compared to the R-type case. Therefore, the number of other components in the chain is reduced to optimise the use of resources – it is found that the operation with two FEBs per H-type EC and a single PDMDB board for an H-type module is optimal.

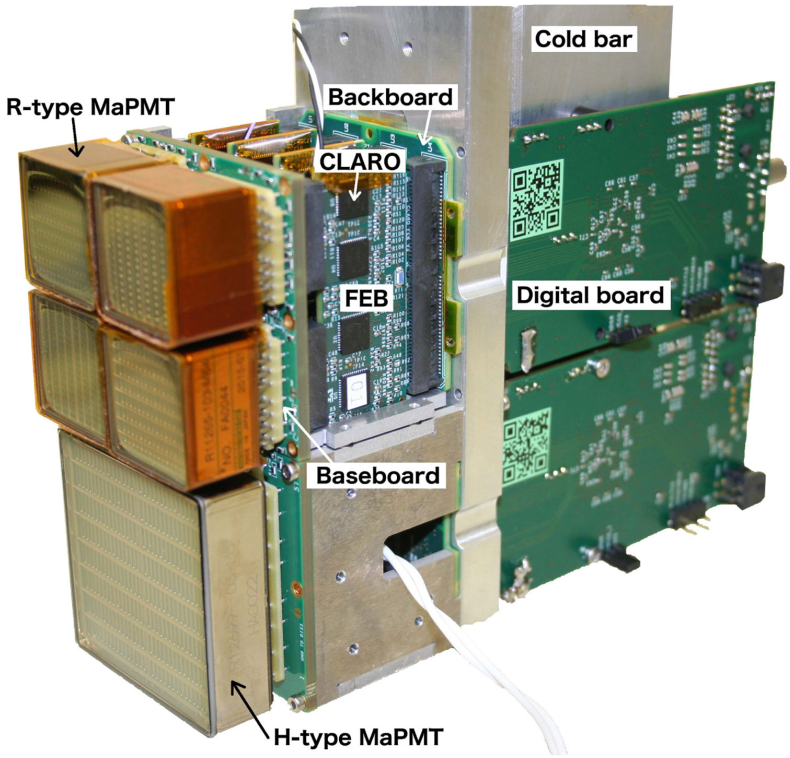

Figure 3.31: Picture of the R-type (top) and the H-type (bottom) elementary cells with the PDMDB boards mounted [219].

For the same reasons, one additional EC on each side is not installed in the column that is furthest away from the beam pipe in both of the RICH1 panels (which means there are only two ECs mounted in the outermost modules of these columns). While the RICH1 columns are constructed only from the R-type PDMs, the RICH2 ones are assembled with four H-type and two R-type modules. The H-type PDMs are placed in the outer regions of the detection planes (see Fig. 3.29 (left)), where the hit rate is expected to be significantly lower than in their central parts.

3.3.4 Simulation of the Upgrade Ia detector

A full detector simulation is essential for evaluating the performance and operating conditions of the constructed detector. The RICH detectors are simulated within the standard LHCb software and simulation framework described in Sect. 3.1.5. The RICH simulation [221] comprises the creation of the Cherenkov photons, their transport through the detector optical system (including their reflections and refraction at the optical boundaries) and the conversion to photoelectrons in the RICH photon detectors to create the final *photon hits*. The RICH geometry, specific properties of its optical components and of the photon detectors (such as the quantum efficiency) are taken into account in this first simulation stage, which is performed within the GAUSS application. The BOOLE project allows for the modelling of the digitisation process of the detector response, taking into account the detailed information on the pixel gains, time properties and

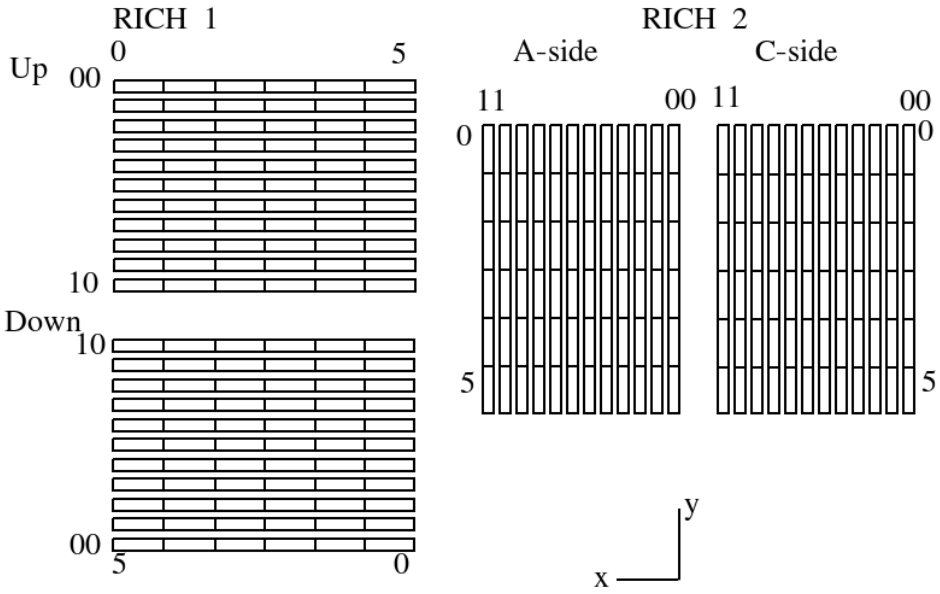

Figure 3.32: Layout of the RICH panels, together with the applied module and column numbering schemes. Only the extreme numbers are shown, the others are in integer steps between them [220].

electronics characteristics[8]. An accurate modelling of the photon detectors response and the readout electronics processing is necessary for a realistic detector simulation, in particular in terms of the noise and timing studies, which can be already useful in the Upgrade Ia phase, but will be a necessary element in the research on the future upgrades (see Sects. 3.2.2 and 3.2.3).

The author of this thesis has been actively involved in the development of the RICH detectors simulation for the Upgrade Ia phase. These advancements consider various stages of the detector modelling, from the detector description, through the digitisation of its response, to evaluating the final performance of the system. An overview of the developments in the detector description, especially with respect to the nominal Upgrade Ia simulation studies reported in Ref. [201], is given in Sect. 3.3.4.1. The following Sect. 3.3.4.2 presents general information on the digitisation chain for the photon hits. In particular, the author has contributed to the simulation of an unexpected *signal-induced noise* (SIN) observed in the MaPMT tubes (see Sect. 3.3.4.3 for details). Assessing the effects of this additional noise on the RICH operation has been an important element of the Run3 preparations. Another aspect is related to introducing a detailed time-response simulation, which enables the testing of various configurations of the detector readout, relevant also for reducing the impact of noise (see Sect. 3.3.4.4). Finally, the author has developed a unified methodology for studying the RICH performance using the simulation monitoring. The described contributions have been realised in close collaboration with the LHCb RICH group, especially in terms of realistic modelling of the detector based on its measured or expected properties. All of the mentioned components have been used to determine an anticipated baseline performance of the upgraded RICH system as well as to investigate the impact of different detector configurations on its operation, *e.g.* in the context of noise

[8]In the described studies, the *hit occupancy* refers to the number of the photon hits (photoelectrons registered in the MaPMT pixels). The *detected occupancy* corresponds to the values observed after the full digitisation process in Boole.

reduction (see Sects. 3.3.4.5 and 3.3.4.6). While the described developments are mainly focused on the Run3 preparations, some aspects of the simulation (such as the modelling of a specific detector noise or the methods developed for the RICH performance studies) can also be adapted for the research on the future upgrades.

3.3.4.1 Detector description

Description of the final RICH geometry, corresponding to the actual arrangement of the photon detectors and modules in the RICH panels as described in Sect. 3.3.3.2, is introduced within a relevant LHCb database[9]. It includes enforcing a uniform numbering scheme for the various components of the photon detection chain in the simulation. The main purpose of this scheme is to assign a unique global number to each MaPMT following rules that allow for a straightforward identification of the corresponding PDM, elementary cell and the position of the tube in the latter. A mapping for the readout of the individual pixels within the MaPMTs is also provided. The numbering schemes used in the simulation have been developed in collaboration with the members of the LHCb RICH group and follow the conventions used in the hardware installation, which are shown in Figs. 3.32 and 3.33.

Introducing the final detector geometry and numbering scheme allows for the creation of corresponding structures in the simulation database[10]. This database contains information on the readout configuration which is used in the simulation, including the mappings mentioned in the previous paragraph. It also enables storing conditions with the properties of the individual photon detectors and channels (see more details in Sect. 3.3.4.2).

The new numbering scheme is integrated within the LHCb software, which allows a full detector simulation to study the RICH performance and monitor its changes with the potential developments in material and geometry description emerging during the detector construction. The mentioned mappings are one of the elements used in the LHCb reconstruction to decode the information on the detector hits from the real data format as well as to produce events in the real data format using the simulation (*encoding*).

One of the developments in the full detector simulation is related to updating properties of the RICH detector components that evolved significantly since the nominal Upgrade Ia study reported in Ref. [201]. It mainly involves the improvements to the RICH1 optical system obtained with the dedicated mirror and gas exit quartz-window coatings described in Sect. 3.3.3.2 as well as choosing a slightly different version of the MaPMT tubes for the detector construction than what would be used in the nominal scenario. The final MaPMTs have an improved quantum efficiency, with the peak values reaching almost 40% for the photon wavelength around 320 nm (see Fig. 3.28). The quantum efficiency is also enhanced in the near-UV region with respect to the tube version from the nominal scenario. Those developments have a significant influence on the number of detected photons in the RICH system. The effect is more prominent for RICH1, where the overall performance of the optical system is also improved. More details on this study are given in the following paragraph. It is worth noting at this point, that the mentioned features can still slightly evolve as the detector installation is ongoing and the knowledge on the performance and alignment of some of its components is continuously improving.

[9] *Detector description database* (DDDB).
[10] *Simulation conditions database* (SIMCOND).

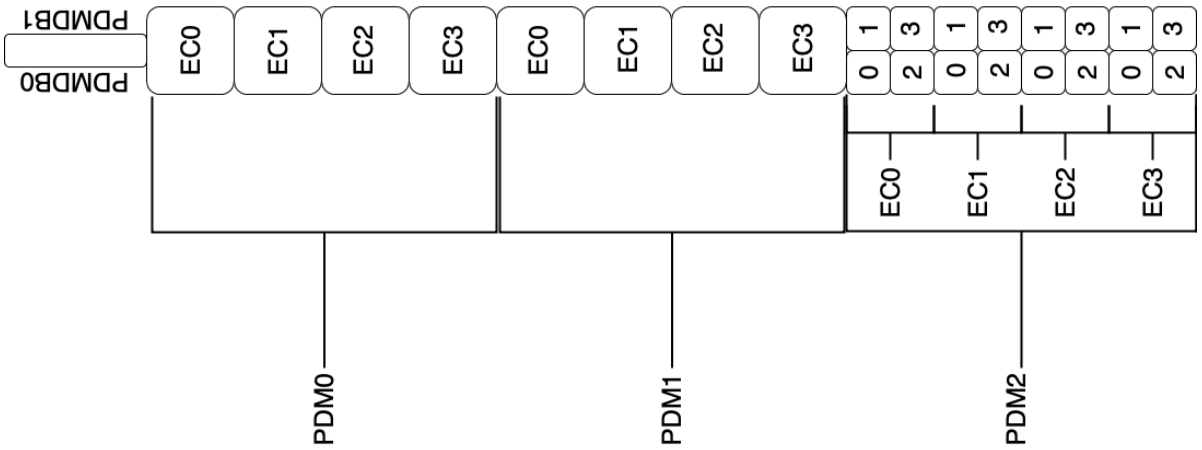

Figure 3.33: Schematic view of the numbering schemes within a RICH1 column. Only the first three modules are shown [222].

However, any potential further updates to the specified properties are expected to be minor with respect to the ones described in the current paragraph.

The effect of the mentioned updates to the RICH properties on the simulation results is shown in Figs. 3.34 and 3.35. A significant increase in the pixel hit occupancy is observed in both RICH1 and RICH2, together with an improvement in the PID performance. As mentioned before, the described developments in the detector design contribute to the increase of the number of detected photons in the RICH system, which is beneficial for the RICH performance, but it may lead to a more challenging operation of the detector. The increase in the occupancy is dominated by additional photons from the near-UV range[11], which is related to the higher sensitivity of the quantum efficiency in that region. The results presented in this paragraph aim only to illustrate the changes in the detector occupancy and PID efficiency observed after this particular step in the simulation development (see Sect. 3.3.4.5 for a detailed study and discussion on the RICH performance, including the updates described in the current section).

3.3.4.2 Digitisation of the detector response

One of the key differences in the Upgrade Ia digitisation with respect to the current RICH detectors using the HPDs is that the individual detector channel properties need to be taken into account, not only to improve the simulation accuracy, but also to account for a nonnegligible MaPMT pixel-to-pixel gain variation (see Sect. 3.3.3.2). The relevant properties are stored in dedicated structures within the simulation database and include information on pixel gains, CLARO thresholds and the input needed for the SIN simulation (see Sect. 3.3.4.3). The mentioned features depend on a number of factors, such as the MaPMT bias voltage, the detector calibration and the photon detectors ageing. The simulation database needs to be updated following changes in these aspects. Ultimately, when the detector installation is finished, these conditions will be populated with the data obtained during the various stages of the commissioning process and quality

[11] This refers to wavelength values below 300 nm and is denoted in this study as the *short-wavelength* region.

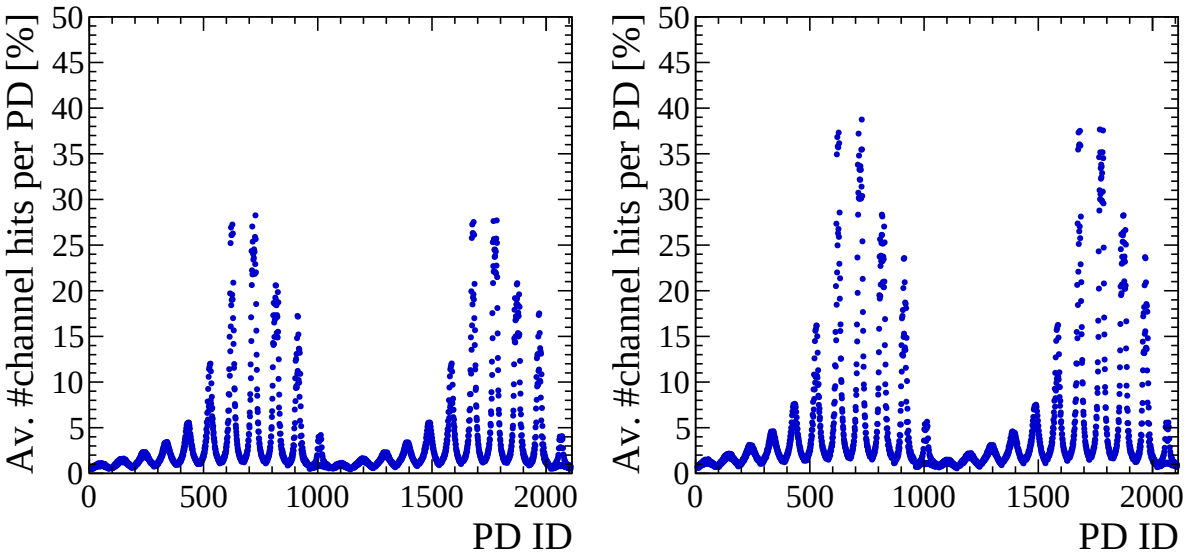

Figure 3.34: Average pixel hit occupancy for different MaPMTs in RICH1 (left) before and (right) after the detector updates described in the text. The values are obtained using the LHCb simulation.

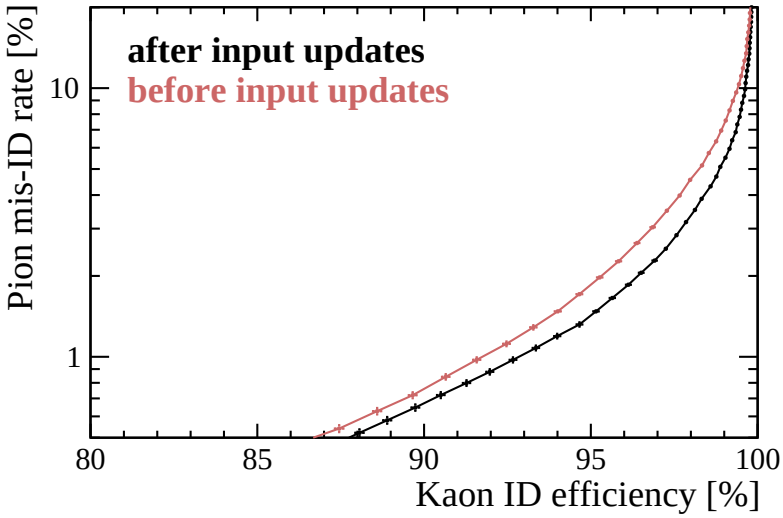

Figure 3.35: Comparison of the PID performance obtained from the LHCb simulation before (red) and after (black) the detector updates described in the text. Both curves illustrate the probability of misidentifying a pion as a kaon, for the given efficiency of the kaon identification.

Table 3.2: Realistic channel properties used in the simulation for different MaPMT bias voltage values, based on a representative subset of the commissioning data [223].

Bias voltage [V]	**Gain mean [Me^-]**	**Gain width [Me^-]**	**Threshold [Me^-]**
900	0.9	0.20	0.125
950	1.4	0.35	0.125
1000	2.0	0.55	0.125

assurance procedures (see Sect. 3.3.3.2) [213, 218]. Currently, the values used in the simulation aim to reproduce a realistic detector configuration in Run3 with a 900 V MaPMT bias voltage (see Tab. 3.2), based on a representative part of the commissioning data corresponding to approximately 70 000 detector channels [223]. The simulation database also contains information on the time-related properties of the CLARO chips and MaPMT tubes, which is used in the detailed time simulation (see Sect. 3.3.4.4).

The digitisation stage starts from the information on the photoelectrons registered in the given MaPMT pixel. The processing of each such deposit can be described as follows. In the first step, a random number of electrons is assigned to the given pixel based on the information on gain obtained from the simulation database. The pixel gain is modelled as a Gaussian distribution with the measured mean and width values. Then, the number of electrons is compared with the threshold set in the CLARO discriminator and only the visible deposits are processed further. This mechanism allows for the processing of any kind of deposits, regardless of their origin (*e.g.* the signal, but also the background, noise or spillover[12]), that contribute to the given channel and to study their influence on the final CLARO output signal. Additional effects like the electronics noise, the dark-count rate or the channel cross-talk are expected to be relatively low and negligible with respect to the SIN contribution, hence they are not taken into account in the digitisation process.

The digital CLARO response is sampled by FPGAs in the PDMDB, which is emulated in the simulation (see Sect. 3.3.4.4 for details). The 25 ns signal time-window (with adjustable starting point) is sampled into an 8-bit structure corresponding to 3.125 ns time slots, which is compared with the configurable list of patterns that are accepted for sending to the LHCb DAQ system. This implementation allows for the study of different options for the readout configuration during Run3, in particular the various scenarios for the signal time-gate (see Sect. 3.3.3.2) envisaged to optimise the noise and background reduction, and also to investigate the effects of the spillover hits. Since an optimal readout configuration is still under study (see Sect. 3.3.4.6 for more details), a simple level detection is applied by default, requiring a signal in any of the 3.125 ns time slots.

3.3.4.3 SIN modelling

The SIN effect has been observed in the MaPMT tubes during the last year of the Run2 data-taking period, when a prototype Upgrade Ia photon detector module was installed behind the RICH2 detection plane in order to study its performance in a realistic operational environment. The unexpected noise pulses seemed to arrive in a certain delay

[12] *Spillover* hits refer to the signals in the detector originating from the collisions that are adjacent to the main event. In the LHCb simulation, two minimum-bias collisions before and after the main event can be studied by default.

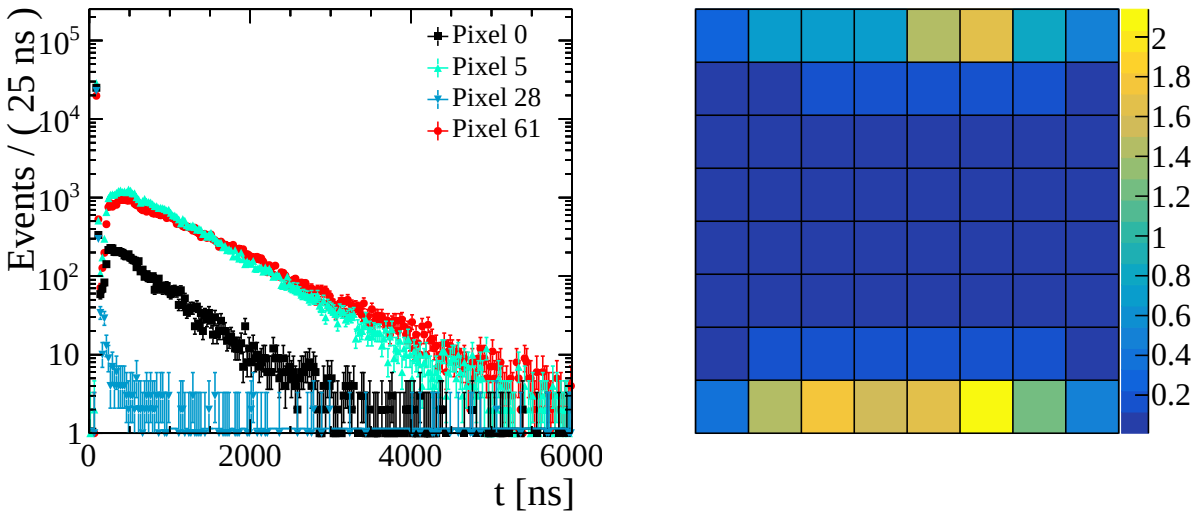

Figure 3.36: (left) Typical SIN time-spectra in a logarithmic scale for four different pixels in a MaPMT tube. The peaks that are closest to zero correspond to the signal hits. (right) Example map of the expected number of SIN pulses in the MaPMT pixels [24].

with respect to the anticipated signal arrival time. Since then, the effect has been studied thoroughly during the quality assurance measurements of the detector components as well as in dedicated setups to improve its understanding and to assess its influence on the detector performance. The results of these investigations, including the developed model of the SIN effect and its potential mitigation strategies, are detailed in Ref. [24]. It also reports studies of the detector performance in the presence of SIN in various detector configurations (see Sect. 3.3.4.6), which are done employing the combined contributions to the RICH simulation described in Sect. 3.3.4.

The SIN pulses are detected during a period of several microseconds after the signal hit. Their typical spectrum is reported in Fig. 3.36 (left) for four different pixels within the same MaPMT tube. As shown, the SIN intensity varies significantly between different pixels. The effect can be quantified in terms of the expected number of SIN pulses μ_{SIN}, which is derived from the SIN time spectra as the ratio of all pulses observed after the signal peak to the number of counts in the signal slot. It is observed that the SIN intensity is different between the MaPMT pixels and also between the tubes. However, a common characteristic pattern is visible in the affected MaPMTs. An example map of the expected number of SIN pulses in the MaPMT pixels is shown in Fig. 3.36 (right) – it is clearly visible that the additional noise is present mainly in two edge pixel rows. The SIN intensity is also observed to increase with the applied MaPMT bias voltage.

Due to the relatively long duration of the SIN pulses, in an environment with a high signal rate the pileup of the SIN pulses may lead to an increase in the observed pixel hit occupancy. Dedicated high-rate measurements employing laser pulses were taken to investigate this effect. The developed SIN model allows for the determination of the expected detection probability within a 25 ns time-slot including the presence of SIN as [24]

$$P_{\mathrm{SIN}} = 1 - \bar{P}_{\mathrm{Ckv}} e^{-\mu_{\mathrm{SIN}} P_{\mathrm{Ckv}}} \ , \tag{3.5}$$

where $P_{\mathrm{Ckv}} = 1 - \bar{P}_{\mathrm{Ckv}}$ corresponds to the pixel hit occupancy without the SIN pulses, which can be obtained using the simulation (the average occupancy values are the same within all 25 ns time-slots).

The SIN effect is introduced to the LHCb simulation at the digitisation stage to assess its impact on the RICH performance and on the overall PID efficiency. Since the SIN pulses are visible up to the order of microseconds after the signal event, an event-by-event simulation is not feasible, and a parametrised approach based on the SIN model is taken. The SIN hits are added randomly in individual detector pixels with a probability of x_{SIN}. This quantity fulfils the relationship $P_{\mathrm{SIN}} = P_{\mathrm{Ckv}} + x_{\mathrm{SIN}}$ (the two contributions to occupancy are considered independent) which, together with Eq. (3.5), leads to

$$x_{\mathrm{SIN}} = (1 - P_{\mathrm{Ckv}}) \left(1 - e^{-\mu_{\mathrm{SIN}} P_{\mathrm{Ckv}}}\right) \ . \tag{3.6}$$

The response of the RICH detector is modelled using a full time-simulation, and the SIN hits are also introduced independently in the two time-windows before the signal one. The acquisition time (see Sect. 3.3.4.4) of each of the SIN pulses is assigned randomly within the given 25 ns slot. The SIN hits are processed as any other deposit types, which allows us to study the influence of SIN on the final CLARO time-response and the overall readout performance (see Sect. 3.3.4.2). In particular, the simulation reproduces a scenario in which a signal hit is preceded by a SIN pulse, which can lead to a failure to register a genuine Cherenkov photon.

The μ_{SIN} values for each detector pixel and the simulated average pixel hit occupancy that are necessary for the SIN simulation are stored in the simulation database. While in the final Upgrade Ia simulation database the various channel properties will be filled using the full commissioning dataset (see Sect. 3.3.4.2), estimating the impact of SIN on the PID performance requires reproducing a realistic detector configuration with the limited data available. Based on the results obtained during the quality assurance measurements, seven representative classes of MaPMTs are defined, with respect to their observed SIN intensity, and μ_{SIN} maps are produced for each class (see Fig. 3.36 (right)) [223]. Similar to the other channel properties, the values obtained with 900 V MaPMT bias voltage are used by default. The MaPMTs in the simulation are populated on the photon detector planes based on their assigned class and the simulated minimum-bias hit occupancy. The ones least affected by SIN are placed in the highest-occupancy regions.

The characteristic patterns related to the simulated SIN deposits (increased occupancy in the edge rows of some of the MaPMTs) are clearly visible in the map of the detected channel occupancy for RICH2 shown in Fig. 3.37. As can be expected, the PID performance is observed to decrease slightly due to this extra source of noise (see Fig. 3.38). Initial studies indicate that the presence of SIN should not have a significant influence on the physics performance in the Run3 period and further strategies are explored to mitigate this effect. The most straightforward approach is that the MaPMTs will be operated with the lowest possible bias voltage, since the number of SIN pulses increases with the latter. A batch of mechanically modified MaPMT tubes with a largely suppressed SIN intensity will also be procured. These devices will be populated in the areas with the highest-occupancy in the RICH1 detector. Another mitigation strategy is related to the usage of a signal time-gate mentioned in Sect. 3.3.3.2. More detailed discussion on the SIN effects in different detector configurations is given in Sect. 3.3.4.6.

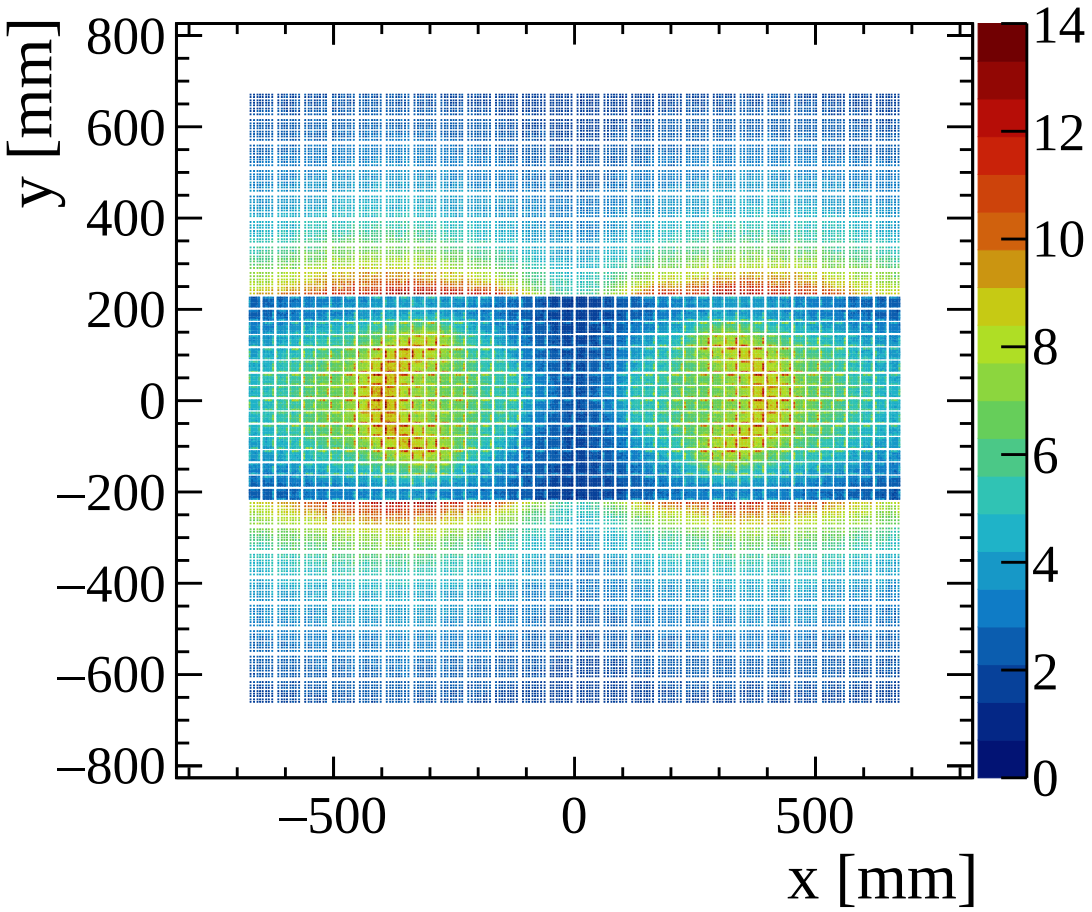

Figure 3.37: Map of the detected channel occupancy in RICH2 obtained from the LHCb simulation with the SIN noise. The occupancy values are given as a percentage.

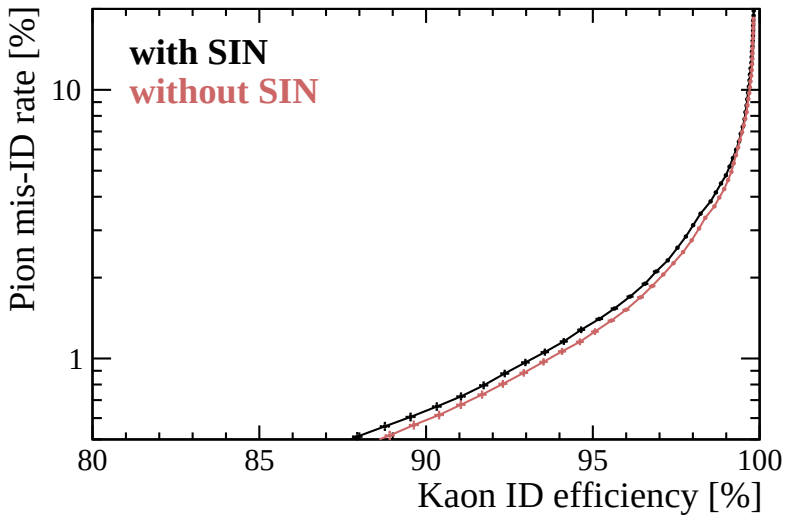

Figure 3.38: PID performance with (black) and without (red) the SIN effect enabled in the simulation. The curves correspond to a rate of misidentifying a pion as a kaon, as a function of the efficiency of kaon identification.

Table 3.3: Transit time of the MaPMT tubes used in RICH. The information is based on the manufacturer data sheets [224].

MaPMT type	Transit time mean [ns]	Transit time spread (FWHM) [ns]
R-type	5.10	0.35
H-type	5.30	0.28

3.3.4.4 Detailed time simulation

Employing the hit time information can be essential in reducing the detector occupancy and combinatorial background in the future upgrades conditions (see *e.g.* Sect. 3.2.3.2). A detailed discussion on the usage of the time information in the LHCb RICH detectors is given in Ref. [205]. Accurate information on the hits time and their time-processing in the photon detection chain is needed to simulate and study different time-gate scenarios for the Run3 operations (see Sect. 3.3.3.2).

Two time-related quantities are particularly necessary to simulate the final detector response. It is the absolute time from the bunch crossing to the CLARO output signal latching in the PDMDB FPGA boards (*acquisition time* – TOA) and the period for which that signal remains visible (*time-over-threshold* – TOT)[13]. These two values allow for the emulation of the 8-bit sampled pattern in the digital boards by testing the state of the CLARO output in each of the corresponding time slots.

Information on the absolute time from the bunch crossing to the photoelectron creation in the MaPMT photocathode can be obtained from the full event simulation of the charged tracks and the Cherenkov photons paths in the detector using the Gauss application. This simulation also takes into account the intrinsic time spread of the initial collision primary vertex (see Ref. [205]). Two additional steps in the time-processing of the RICH signals must be considered: the transit time through the MaPMT tubes and the CLARO chip delay from the input signal to reaching the output level that is latched by the FPGA boards. The first component is based on the manufacturer data sheets for the MaPMTs and the transit time is modelled using a Gaussian distribution with the provided mean and width values (see Tab. 3.3). The CLARO delay as well as the output time-over-threshold depends on the input signal and the threshold set in the CLARO discriminator. Both of these quantities have been measured in detail for a number of channels by the LHCb RICH group and the results of this study are shown in Fig. 3.39 (top). Interpolation between the curves is performed to assign the proper TOT and CLARO delay during the digitisation to the deposits with the simulated input signal and the relevant discriminator threshold. The visible channel-to-channel variation in the CLARO response delay will be corrected for during the detector operation using a built-in input-delay feature of the FPGA boards. Since the CLARO input signal is a random number depending on the MaPMT pixel gain, the delay in the given channel has a certain spread, which is referred to as the CLARO *time-walk* (see Fig. 3.39 (bottom)). It constitutes the dominant contribution to the overall span of the RICH detectors time-response.

By taking into account the components mentioned above, it is possible to accurately reconstruct the acquisition time and time-over-threshold for various deposits, including

[13]The FPGA boards latch when the output signal of the CLARO discriminator reaches 50% of its amplitude. The time-over-threshold refers to the time when the signal stays above the mentioned value.

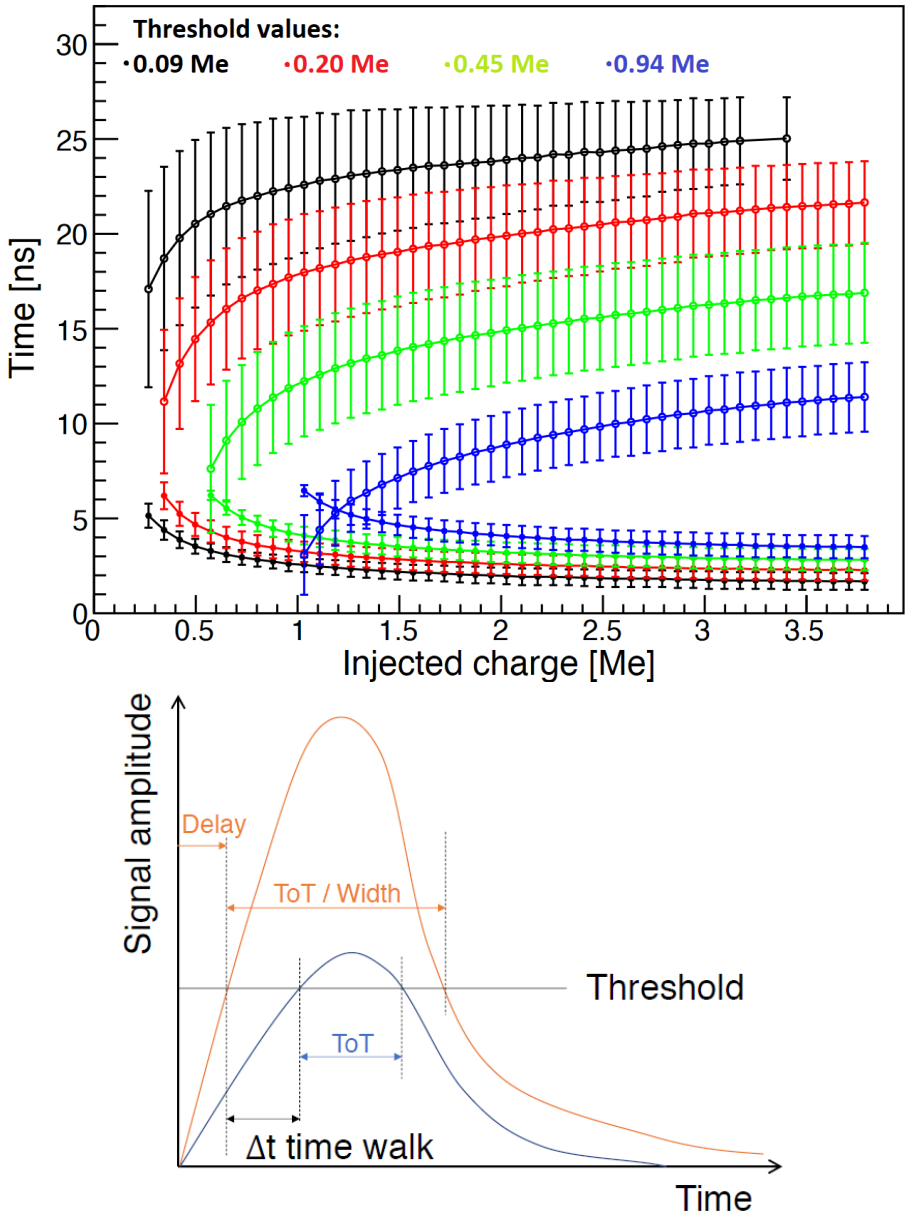

Figure 3.39: (top) Time-over-threshold (open points, upper distributions) and delay (filled points, lower distributions) of the CLARO output pulse versus the input signal amplitude, for different threshold settings. The error bars correspond to the measured channel-to-channel variations [225]. (bottom) Time walk arises due to the variation of the MaPMT signal amplitudes at a fixed CLARO threshold [205].

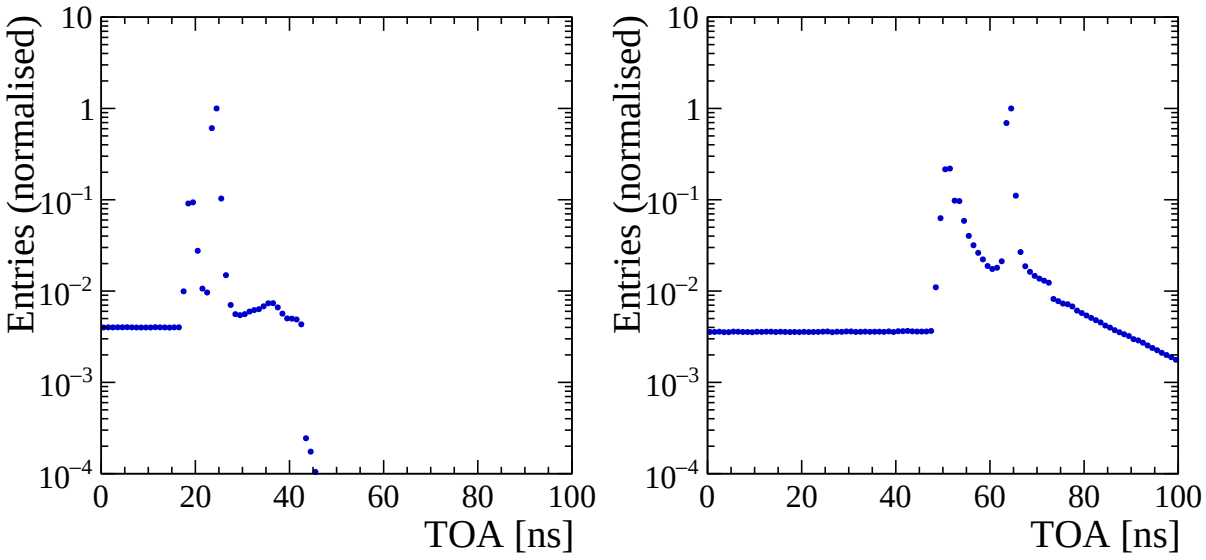

Figure 3.40: Acquisition time of deposits in the detector channels for (left) RICH1 and (right) RICH2, counting from the bunch crossing. The values are obtained using the LHCb simulation and include the simulated SIN hits.

the ones related to the noise or background, and to study their effects on the final sampled CLARO output signal (see Sect. 3.3.4.2), which can be useful for finding an optimal readout configuration for the Run3 period. The TOA distribution for both RICH1 and RICH2 is shown in Fig. 3.40. The most prominent peaks in those distributions correspond to the signal Cherenkov photons. The less pronounced structures arriving before the signal are related to the background originating from charged particles and photons travelling directly from the interaction point towards the detection planes. Rises that are significantly less evident can also be observed after the signal regions, which is associated with Cherenkov photons undergoing multiple reflections on their way through the optical system. The SIN effect is distributed uniformly in time and is best visible as flat areas preceding the previously mentioned contributions. The SIN deposits are added in the signal region and also before it, to study the influence of the noise TOT on the signal time-window (see Sect. 3.3.4.3 for more details). The fact that SIN is simulated in a limited time range is the reason for sharp transitions visible at approximately 43 and 73 ns in RICH1 and RICH2, respectively.

3.3.4.5 Baseline performance studies

The LHCb simulation framework, including the developments described in the previous sections, is employed to evaluate the performance of the RICH detectors in the Upgrade Ia phase. This simulation version represents the most recent details of the RICH geometry, optics, and properties of the photon detectors and electronics. The characteristics of the individual detector channels are accounted for as well – including pixel gains, CLARO thresholds and, in particular, the SIN effect. As mentioned in the previous sections, the individual channel properties are based on the limited dataset acquired during different phases of the detector commissioning [223] and aim to represent the realistic operational

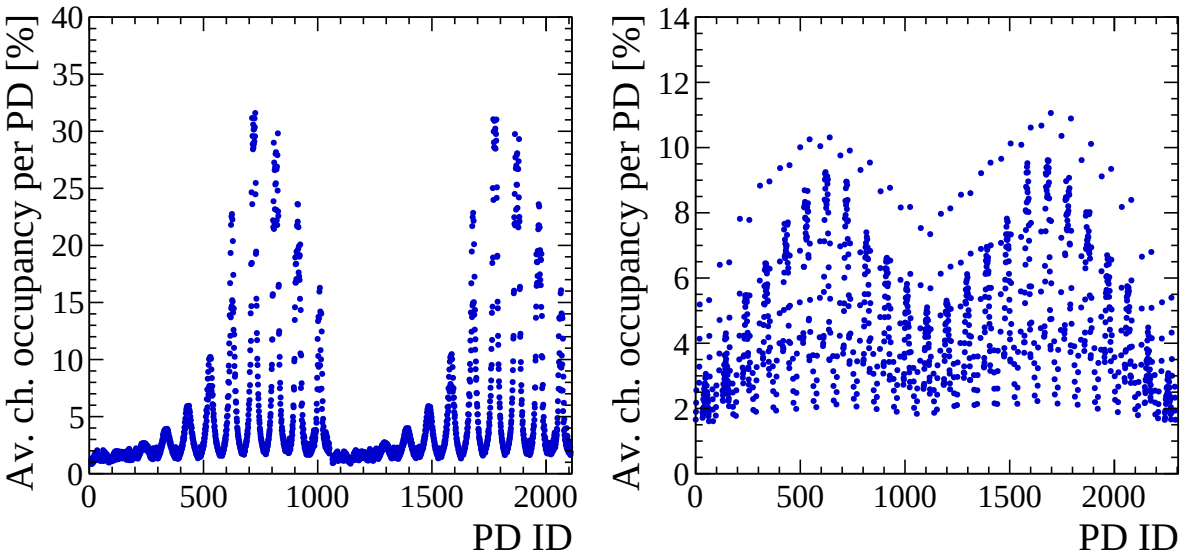

Figure 3.41: Average detected channel-occupancy obtained using the LHCb simulation. The values are shown as a function of a MaPMT identifier for the (left) RICH1 and (right) RICH2 detector.

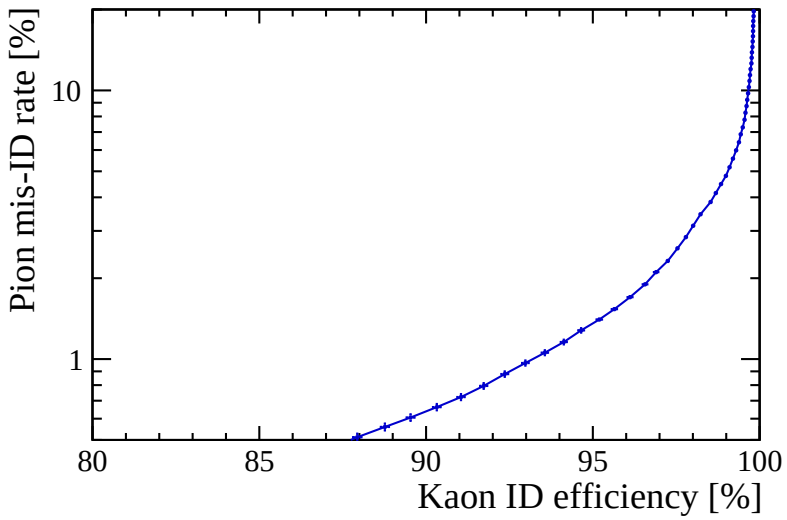

Figure 3.42: Typical performance of the kaon identification in the Run3 period. The curve is obtained using the LHCb simulation as described in the text.

conditions and calibration of the RICH detectors during the Run3 period (the values measured with a 900 V MaPMT bias voltage are used as a baseline scenario hereafter). The performance study is done employing $B_s^0 \to \phi\phi$ events representing a common signal type. The detected occupancy in the RICH channels obtained in the current investigations is reported in Fig. 3.41 and a typical PID performance of the kaon identification is shown in Fig. 3.42.

A unified method for evaluating the various contributions to the RICH detectors performance (see Sect. 3.3.1) using the simulation has been developed to compare dif-

ferent detector configurations. Information on the single-photon resolution is obtained using a simplified reconstruction within the monitoring of the GAUSS application. By default, in this simplified approach, it is assumed that a Cherenkov photon originates from a location in the middle of the particle trajectory in the radiator and reaches the centre of the corresponding detector pixel. The overall single-photon resolution σ_θ is determined by investigating the resulting distribution. The individual components of this aspect (the chromatic, the emission point and the pixel errors) are evaluated by modifying the input to the process mentioned above. The pixel error, for example, is acquired by running the procedure with a genuine hit position in the pixel and comparing the resulting Cherenkov angle to the one obtained with the default input. An expected photon yield is also determined using the simulation, by studying the distribution of the number of photon hits per saturated tracks. The total Cherenkov angle resolution for a track $\Delta\theta_{\mathrm{C}}$ is calculated following Eq. (3.2). In the current studies, a tracking resolution of $C_{\mathrm{tracking}} = 0.35\,\mathrm{mrad}$ is assumed.

The investigations of the photon yield (${N_{\mathrm{ph}}}^{\mathrm{optimal}}$) and the Cherenkov angle resolution are performed using a particle gun. To ensure that the studied tracks are saturated and have a minor curvature due to the presence of the magnetic field, the particle gun is required to provide 80 GeV muons. Furthermore, the research is conducted in a limited acceptance region which corresponds to the region where an optimal performance of the RICH system is expected. It is equivalent to requiring tracks with the polar angle in the range of 90–180 mrad (40–90 mrad) for the RICH1 (RICH2) detector. The use of the particle gun is driven by a better control of the above track prerequisites. Nevertheless, the same results on the RICH performance are reproduced with the signal events given that the tracks with a quality corresponding to the one from the particle gun configuration are selected. Typical photon yields (${N_{\mathrm{ph}}}^{\mathrm{typical}}$) determined using the signal events without the strict track selection are also provided. These values are lower than the optimal yields, mainly due to some of the photons hitting the beam pipe for tracks in the acceptance region close to the beam.

Results of the performance studies described above are summarised in Tab. 3.4, where *nominal* values reported in Ref. [201] are also indicated for reference. Regardless of the fact that the essential elements of the simplified reconstruction in the GAUSS monitoring are the same as in the cited *nominal* case, the lack of a well-defined and consistent method for determining the individual parameters (*e.g.* in terms of the track quality requirements, or performing fits to the obtained distributions) limits the feasibility of a direct comparison to the other discussed scenarios. This refers in particular to a nontrivial evaluation of the contribution due to the chromatic error. Determination of the chromatic error has been reviewed in the current studies, where the individual components of the RICH performance are obtained using the unified approach mentioned at the beginning of the present section. This development provides a universal benchmark that allows for the monitoring of changes in each contribution to the performance for different detector configurations, as done in the paragraph below and in Sect. 3.3.4.6.

The visible differences between the *recent* and the *previous* scenarios are mainly related to the evolutions in the detector design since the *nominal* study reported in Ref. [201] (which are taken into account in the *recent* scenario, but not in the *previous* one – see Sect. 3.3.4.1 for details). As stated before, those developments lead to a significant surge in the detector occupancy, mostly due to photons from the short-wavelength spectrum. The present study shows an increase in the chromatic error for both RICH1 and RICH2,

Table 3.4: Expected performance of the RICH detectors in the Upgrade Ia phase obtained using the LHCb simulation. Values corresponding to the most recent simulation version described in the text (*recent*) are compared to the ones before the updates detailed in Sect. 3.3.4.1 (*previous*). The *nominal* design values [201, 209] are shown for reference (as indicated in the text, a direct comparison between the *nominal* and the other scenarios may not be feasible). For RICH2, the performance is reported only for the inner detector regions where the R-type MaPMTs are installed.

	Photon yield		**Cherenkov angle resolution [mrad]**				
	$N_{\mathrm{ph}}{}^{\mathrm{optimal}}$	$N_{\mathrm{ph}}{}^{\mathrm{typical}}$	$\sigma_{\mathrm{chromatic}}$	$\sigma_{\mathrm{emission}}$	σ_{pixel}	σ_{θ}	$\Delta\theta_{\mathrm{C}}$
RICH1 (*nominal*)	40	-	0.58	0.37	0.44	0.82	0.37
RICH1 (*previous*)	43	-	0.41	0.36	0.50	0.73	0.37
RICH1 (*recent*)	63	59	0.52	0.36	0.50	0.81	0.36
RICH2 (*nominal*)	22	-	0.31	0.27	0.20	0.46	0.36
RICH2 (*previous*)	24	-	0.26	0.32	0.22	0.44	0.36
RICH2 (*recent*)	34	30	0.34	0.32	0.22	0.52	0.36

which is related to the presence of additional short-wavelength photons originating from an increased QE sensitivity in that wavelength region (see Sect. 3.3.2). The emission point and the pixel errors are not affected by the mentioned updates, while the chromatic contribution leads to a slightly worse overall single-photon resolution. However, this effect is more than compensated by the significantly higher photon yield. In the end, the cited developments in the detector construction result in a slight improvement in the total Cherenkov angle resolution and in the overall PID efficiency (see Fig. 3.35), however, at the cost of increased detector occupancy (see Fig. 3.34). Keeping in mind the limited possibility of direct comparisons to the *nominal* studies on individual contributions to the RICH performance, the present investigations indicate that the Upgrade Ia RICH detectors well satisfy their *nominal* requirements on the total resolution of the Cherenkov angle measurement and that they will allow us to maintain the excellent PID capabilities in the Run3 conditions.

3.3.4.6 Preliminary investigations on the readout configuration

While Sect. 3.3.4.5 provides details on the expected baseline performance of the RICH detectors, the present section focuses on studies of more refined readout configurations. These investigations are considered preliminary, as they are based on a complete set of channel properties acquired at various stages of the commissioning process [223], taking into account the expected positions of MaPMTs in the detector (in the case of RICH1, the installation is ongoing at the moment this thesis is being written). However, some essential aspects (*e.g.* the CLARO threshold calibration in RICH1) are not yet finalised, which can have a nonnegligible influence on the simulation results. For this reason, the baseline performance study reported in Sect. 3.3.4.5 is based on the previously-defined realistic detector configuration derived from the data on a limited number of channels (see Sect. 3.3.4.2). Furthermore, the current studies involve more refined readout configuration scenarios (including edge detection and time-gate implementation) to illustrate the related potential gains in the detector performance. Despite the fact that the scenarios proposed below are well-motivated and realistic, it is not yet known whether such configurations

will be feasible or optimal in data taking (which depends, among others, on the final detector calibration and on the ultimate level of observed CLARO time-walk – see Sect. 3.3.4.4). Hence, the presented results are not considered a baseline scenario, but illustrate the expected improvement in performance with different detector configurations.

The current section intends to demonstrate research feasible with the developments in the RICH simulation described previously. It is done using an example of investigations performed to assess the outcome of various SIN-reduction measures which are reported also in Ref. [24]. In these studies, two different configurations are considered, both including the SIN and the spillover effects with a full time-simulation of the detector response. A nominal scenario corresponds to a situation when no dedicated procedures are employed to limit the SIN effect. In this instance, the MaPMTs are operated at a 1000 V bias voltage and the readout is configured to imitate a level detection in a defined 3.125 ns time slot after the signal peak (which is chosen to reduce the impact of spillover hits from the event preceding the one under question). An optimised scenario reflects a configuration implementing some of the SIN-mitigation measures mentioned in Sect. 3.3.4.3. In this case, the MaPMTs are operated at the lowest considered bias voltage of 900 V. Furthermore, a dedicated procedure is applied to optimise the positions of MaPMTs in RICH1 (where the regions with highest occupancy in the RICH system are present, which could potentially lead to strong SIN effects) with respect to both SIN levels and pixel gains. Additionally, a simple edge detection with a 9.375 ns time gate (see Sect. 3.3.3.2) around the signal region is emulated using the mechanism described in Sect. 3.3.4.2, where the time gate is expected to reduce contributions due to the random noise and background.

The effects of SIN on the RICH occupancy are illustrated in Fig. 3.43. As can be seen, the additional SIN deposits in the nominal scenario described above lead to a considerable increase in the detected occupancy. The points with values near unity originate from the H-type MaPMTs, where the observed level of SIN is minor. The occupancy increase is substantially mitigated in the optimised configuration, which is essential for an optimal detector operation – in this case, the rise in the occupancy is maintained below the level of 10% in all MaPMTs. As can be expected, the PID performance is also degraded due to the SIN contribution as illustrated in Fig. 3.44. However, taking into account the notable effects of SIN on the occupancy in the nominal scenario, the PID algorithms appear to be rather robust in the presence of random noise such as SIN. The presented results indicate that the employed SIN-mitagation strategies, including the time gate around the signal peak, are remarkably efficient, leading to an almost complete alleviation of the SIN effects in the PID context (see the reference performance with no SIN simulation).

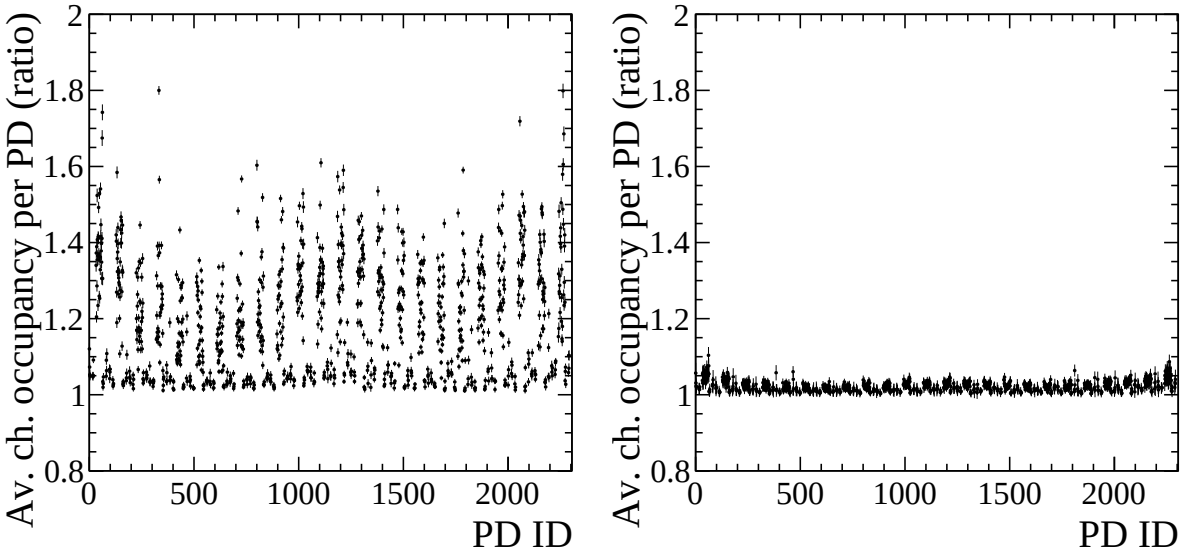

Figure 3.43: Ratio of the average detected channel-occupancy per MaPMT in RICH2 from the LHCb simulation with and without SIN using the (left) nominal and (right) optimised detector configuration described in the text [24].

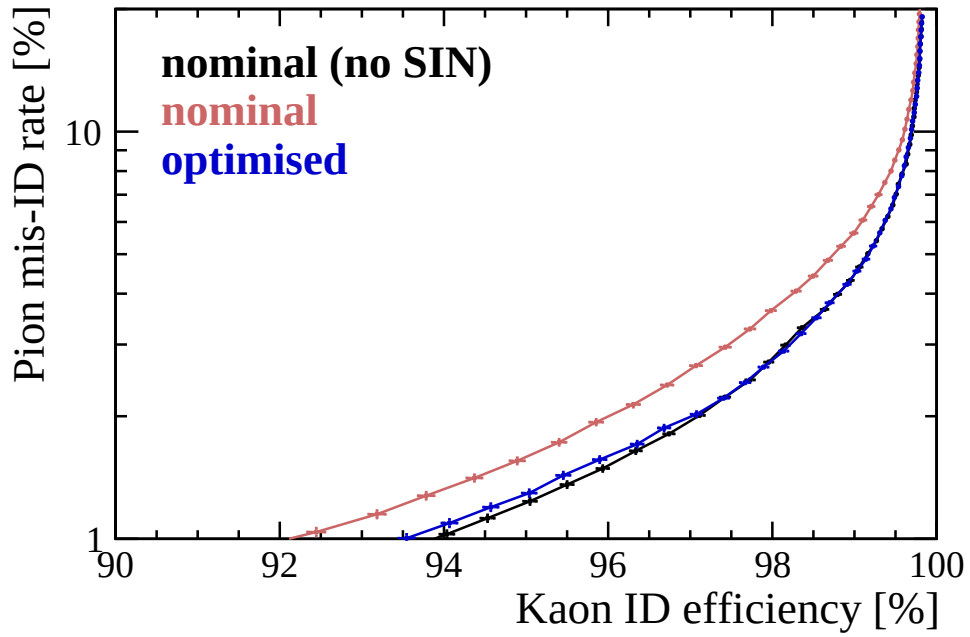

Figure 3.44: Comparison of the PID performance obtained from the LHCb simulation including SIN with the nominal (red) and optimised (blue) scenarios as stated in the text. A case corresponding to the nominal configuration without simulating the SIN effect (black) is also shown for reference. The curves illustrate the probability of misidentifying a pion as a kaon, as a function of the efficiency of kaon identification [24].

Chapter 4

Analysis of the Bose–Einstein correlations in pp and pPb collisions

In this chapter, two separate analyses of the Bose–Einstein correlations for same-sign charged pions in the LHCb experiment are described. The first one is performed for pp collisions and the second one in the pPb system (for the pPb case, results are preliminary). Both studies share a similar approach to the analysis method, data-processing and event selection, which is outlined in Sects. 4.1, 4.2 and 4.3, respectively. Analysis results are described separately for each of these systems in Sects. 4.4 and 4.5.

4.1 Methodology of the BEC studies

While Sect. 2.1 provides the fundamental concepts of the multiparticle production and quantum correlations, this section aims to describe the experimental aspects of the BEC measurements in the context of proton-proton and proton-lead collision systems.

4.1.1 Correlation function

The study of the Bose–Einstein correlations is commonly performed using the Lorentz-invariant variable Q, which is defined in Eq. (2.8). It corresponds to the proximity of two particles in the four-momenta phase-space [226]. A two-particle correlation function is constructed as a ratio of the Q distributions for signal and reference pairs

$$C_2(Q) = \left(\frac{N^{ref}}{N^{sig}}\right)\left(\frac{\mathrm{d}N^{sig}(Q)\,/\,\mathrm{d}Q}{\mathrm{d}N^{ref}(Q)\,/\,\mathrm{d}Q}\right) \ , \tag{4.1}$$

where N^{sig} and N^{ref} correspond to the number of signal and reference pairs, respectively, obtained from an integral of the relevant Q distributions. The signal pairs are pairs of selected prompt, same-sign charged (SS) pions that originate from the same primary vertex, which corresponds to the conditions where the Bose–Einstein correlations can be observed. The reference distribution should reproduce the kinematics and various effects present in the signal one, except for the BEC enhancement. In the first order, the correlation function is, by its construction, independent of effects related to the detector occupancy and material budget as well as to those related to the single-particle detection efficiency.

The reference sample can be constructed in different ways, however none of them is able to perfectly reflect the signal distribution properties [227, 228]. One of the employed

methods is to use pairs of opposite-sign charged (OS) particles originating from the same primary vertex. While this approach allows us to include correlations emerging from kinematics and conservation laws in the given event, the shape of such a distribution is heavily distorted by structures related to two-body resonance decays. Another way to construct the reference sample is to create pairs from particles originating from different primary vertices within the same event. However, such a solution is dependent on the accuracy of the tracks assignment to the PVs of their origin and also on the additional effects related to the track and vertex reconstruction.

In this study, the reference sample is constructed by creating pairs of particles originating from different events, where, by definition, the BEC effect is not present (a so-called *event-mixing* method). The reference pairs are processed in the same way as the signal ones (which means identical selection and particle-identification requirements), to ensure that the signal distribution is reproduced as closely as possible. Additional requirements may be added to combine particles originating from events with similar properties (*e.g.* a comparable charged-particle multiplicity in the given PV or position of the primary vertex in the detector), to further improve the agreement between the samples.

As mentioned above, the reference sample is not capable of perfectly reproducing the signal distribution and hence various distortions in the correlation function are expected. More details on these contributions and on methods to take them into account in the analysis are given in the following sections.

4.1.2 Final-state interactions

The shape of the correlation function is affected, among others, by final-state interactions between the produced hadrons. The effect of the strong interactions in the case of pions is relatively small (see *e.g.* Ref. [229]) and is usually neglected in the BEC studies. However, Coulomb interactions between the charged pions need to be taken into account in the analysis.

The same-sign charged and the opposite-sign charged pairs are impacted in different ways by the Coulomb effects. For the SS ones, a repulsive interaction leads to a decrease in the correlation function, which is mostly prominent for low Q-values. In the case of OS pairs, this effect is reversed and an enhancement is observed.

A general term for the Coulomb interaction $K(Q)$ is given by [230, 231]

$$K(Q) = \int \mathrm{d}^3\vec{r}\, f(\vec{r}) \left|\Psi_C(\vec{k},\vec{r})\right|^2 , \tag{4.2}$$

where $f(\vec{r})$ corresponds to the particle-emitting source distribution and $\Psi_C(\vec{k},\vec{r})$ is the wave-function for two Coulomb-interacting particles with the relative space and momentum separation given by $\vec{r}$ and $\vec{k}$, respectively. It can be used to correct the correlation function that describes the pure BEC effect – $C_{2,\mathrm{BEC}}(Q)$ – for the final-state Coulomb interactions

$$C_{2,\mathrm{BEC+Coulomb}}(Q) = K(Q)\, C_{2,\mathrm{BEC}}(Q) . \tag{4.3}$$

It should be noted that the $C_{2,\mathrm{BEC}}(Q)$ term above is not equivalent to the experimentally measured correlation function defined in Eq. (4.1).

For point-like sources, the Coulomb interaction term in Eq. (4.2) becomes equivalent to a so-called Gamov factor [232]

$$K^{\mathrm{SS}}_{Gamov}(\zeta) = \frac{2\pi\zeta}{e^{2\pi\zeta} - 1} \ , \ K^{\mathrm{OS}}_{Gamov}(\zeta) = \frac{2\pi\zeta}{1 - e^{-2\pi\zeta}} \ , \tag{4.4}$$

where $\zeta = \alpha m \,/\, Q$, α is the fine-structure constant and m is the particle rest mass. In the case of pions (sources of relatively small sizes), no significant differences are expected between the general form of the Coulomb correction and the Gamov factor, and the latter is often used to simplify the fit procedure.

Calculating the integral in Eq. (4.2) in its full form for extended (not point-like) sources is time-consuming and requires a numerical, iterative approach. It is a common solution to store the results of such calculations in a lookup table that is later used in the full correlation function fit [78, 100]. Another approach is to parametrise the results stored in such a table to further accelerate the fit procedure [134, 233]. Alternative methods for the Coulomb correction are also investigated [234].

A particularly interesting parametrisation of the $K(Q)$ term is given by

$$K(Q) = K_{Gamov}(Q) \left(1 + \frac{\alpha\pi m R_{\mathrm{eff}}}{1.26 + Q R_{\mathrm{eff}}}\right) \ , \tag{4.5}$$

where R_{eff} corresponds to the effective size of the particle-emitting source (see Sect. 2.1) and is provided in femtometers. The additional term with R_{eff} constitutes a correction for the Gamov factor that enables for a more precise characterisation of the Coulomb interactions for extended sources. This parametrisation was developed by the CMS experiment [134] and is valid for the Lévy-type sources (see Sect. 2.1.4) with the index of stability α_{L} equal to unity.

4.1.3 Nonfemtoscopic background

As mentioned in the previous sections, the correlation function shape is distorted by the presence of various nonfemtoscopic effects. There is no strict, theory-motivated description of such contributions, and different strategies can be applied to take them into account in the analysis (either by significantly reducing these effects or by providing an ad hoc background parametrisation that fits the data well - see Sects. 4.1.3.1 and 4.1.3.2).

One such nonfemtoscopic effect is related to long-range correlations that originate from the energy-momentum conservation in the event and which may not be well reproduced in the reference sample. These correlations are present in the full Q range and are the easiest to observe at the high-Q values ($Q > 1\,\mathrm{GeV}$), far from the BEC-signal region. Although different parametrisations can be employed, a simple term linear in Q is usually optimal (in terms of both accuracy and fit stability) to characterise this contribution and therefore commonly used (see Sects. 4.4.1 and 4.5.1.2). Effectively, the long-range correlations provide a reference baseline for other effects in the correlation function.

Another major input to the nonfemtoscopic background is referred to as *cluster contribution* (following the naming convention proposed by the CMS experiment in Refs. [134, 135]). It involves effects due to minijets fragmentation and multibody resonance decays, that are both most prominent in the low-Q range ($Q < 0.5$–$1.0\,\mathrm{GeV}$). This contribution is in general less straightforward to account for than the one from long-range

correlations, due to a more complicated shape and its overlap with the BEC signal. However, in some studies it is found that it can be successfully described using relatively simple parametrisations, *e.g.* employing a Gaussian function [134, 135, 145].

Constructing a correlation function for the opposite-sign charged pairs can be of particular use in the background studies, since similar effects can be expected for both the SS and the OS pairs. Care must be taken when investigating the opposite-sign charged pairs, due to structures related to two-body resonance decays, arising in the correlation function. Regions of Q which are affected by resonances are usually removed from the OS correlation functions.

4.1.3.1 Double-ratio method

The double-ratio approach allows us to correct the correlation function for the nonfemtoscopic background using the simulation. It also helps to reduce the effects due to potential imperfections of the reference sample. A double-ratio function is a ratio of the correlation functions for the data and the simulation

$$r_d(Q) = \frac{C_2^{\mathrm{DATA}}(Q)}{C_2^{\mathrm{MC}}(Q)} \,, \tag{4.6}$$

where the BEC effect should not be present in the simulation. In both cases, the correlation function is obtained using exactly the same method (meaning the same selection requirements, construction of the reference sample, *etc.*). Due to that, structures originating from phenomena properly modelled in the simulation, are removed from the initial correlation function for the data and in an ideal case only the pure BEC signal should be visible. This method allows for the reduction of contributions from the nonfemtoscopic background, often to a negligible level. However, it requires good agreement between the simulation and the data, and the choice of the simulation conditions (*e.g.* the Monte Carlo generator and its tuning) is often the main contribution to the systematic uncertainty. On the other hand, other sources of the systematic uncertainty can be reduced in this method, since the double-ratio function is, to a high extent, insensitive to the effects related to the optimisation of the selection criteria.

As mentioned in Sect. 4.1.3, the opposite-sign charged pairs can be useful in studying the background contributions. In the case of the double-ratio method, a double-ratio distribution for the OS pairs is a good indicator of the agreement between the simulation and the data – a flat distribution close to unity suggests that the various effects in the data (*e.g.* the long-range correlations and the cluster contribution) are well reproduced in the simulated sample.

4.1.3.2 Cluster subtraction technique

An alternative approach (as opposed to the double-ratio method described in Sect. 4.1.3.1) requires us to study and parametrise the background effects, in particular the nontrivial cluster contribution. The method described in this paragraph, referred to as the *cluster subtraction* (CS) method, was developed by the CMS experiment [134, 135] and represents a fully data-driven approach.

In the CS method, the cluster contribution is studied using the opposite-sign charged pairs and the shape parameters of the chosen function for the background description

(*e.g.* the width of a Gaussian function) are determined from the OS fits. As mentioned in Sect. 4.1.3, similar effects due to the background are expected in both the OS and the SS pairs, however with different amplitudes. It is related to the conservation of electric charge in both components of the cluster contribution (minijets fragmentation and multibody resonance decays), which results in more prominent correlations in the case of OS pairs. Therefore, the background shape parameters in the SS fits are fixed to the values determined by studying the OS pairs, with an additional scaling parameter that is introduced to take into account the different amplitudes of the cluster contribution in both pair types. The specifics of this method are dependent on the aspects of particular analysis, such as the function chosen for the background parametrisation (see Sect. 4.5.1 and also *e.g.* Ref. [134]).

A similar approach was developed by the ATLAS collaboration [147]. It also uses the OS pairs from the data to study the background contribution, however the simulation is used in this case to understand and map the differences in the various background parameters between the OS and the SS samples. It makes this method dependent on the simulation accuracy, but on the other hand allows for the reduction of the number of arbitrary parameters used for this kind of mapping in the pure CS method. Hence, this approach is referred to as a *hybrid cluster subtraction* (HCS) method.

4.1.4 Fitting the correlation function

As described in detail in Sect. 2.1.4, the BEC effect for Lévy-type sources can be parametrised as (see Eq. (2.19))

$$C_{2,\mathrm{BEC}}(Q) = 1 + e^{-|RQ|^{\alpha_\mathrm{L}}} \ .$$

This generic form usually provides the best description of the data (in terms of the fit quality), however the interpretation of the measured correlation radii is not straightforward. In the current study, the parametrisation with the α_L parameter fixed to unity is used, since it usually provides better fit quality than the parametrisation with the assumed Gaussian source-distribution (see Sect. 2.1.4). This form allows us to relate the correlation radius to the size of the particle-emitting source (see Sect. 2.1.3) and is commonly used in the BEC analyses, which makes it possible to compare the obtained results to those reported by other experiments.

A Bowler–Sinyukov formalism [230, 231] allows us to properly take into account the Coulomb interactions in an experimentally measurable correlation function. Firstly, it uses the $K(Q)$ term for Coulomb effects in extended sources which is given in its general form in Eq. (4.2). Secondly, it takes into account that not all of the selected SS pairs (*e.g.* pairs containing decay products of long-lived states) are affected by the Coulomb interactions and, therefore, no correction should be applied to them. A full parametrisation for the experimentally measurable correlation function, including the nonfemtoscopic background description and using the Bowler–Sinyukov formalism is given by

$$C_2(Q) = N\left[1 - \lambda + \lambda K(Q) \times C_{2,\mathrm{BEC}}(Q)\right] \times \Omega(Q) \ , \tag{4.7}$$

where λ denotes the intercept parameter (see Sect. 2.1.2), N is a normalisation factor and $\Omega(Q)$ is a general term for the nonfemtoscopic background contribution.

Contents of the bins in both the signal and the reference Q-variable histograms are Poisson-distributed (measuring the number of pairs corresponds to a counting experiment)

and hence a negative log-likelihood fit method is preferable for the BEC studies [95, 147]. In this approach, the following expression needs to be minimised in the fitting procedure

$$-2\ln L = 2\sum_i \left\{ A_i \ln\left[\frac{(1+C_i)\,A_i}{C_i\,(A_i+B_i+2)}\right] + (B_i+2)\ln\left[\frac{(1+C_i)\,(B_i+2)}{A_i+B_i+2}\right]\right\}\ , \quad (4.8)$$

where A_i and B_i are the bin contents of the signal and the reference Q histograms, and C_i corresponds to the fit function value at the bin centre. While the negative log-likelihood method is the statistically correct one for bins with a small number of counts, it is often that the high statistics available in the study lead to results that are consistent with a fit to the correlation function as defined in Eq. (4.1) performed with the least-squares approach.

It is worth noting at this point that the quality of fits in the BEC studies is not expected to be perfect. Due to the ad hoc descriptions of the unknown nonfemtoscopic background contribution as well as the compromise between the fit quality and interpretability of the measured correlation parameters (*e.g.* using the Lévy parametrisation with $\alpha_{\rm L}$ fixed to unity – see Sects. 2.1.3 and 2.1.4), the obtained χ^2/ndf values are often far from unity.

4.2 Data and simulation samples

Data used in the studies described in this thesis was collected by the LHCb experiment during the Run1 period (see Sect. 3.1.1). The sample employed in the *pp* analysis was recorded in 2011 at $\sqrt{s} = 7\,\text{TeV}$ centre-of-mass energy and corresponds to an integrated luminosity of $1.0\,\text{fb}^{-1}$. The 2013 proton-lead dataset that is used in the current study was registered at $\sqrt{s_{\rm NN}} = 5.02\,\text{TeV}$ centre-of-mass energy per nucleon (with beam energies of 4 and 1.58 TeV, respectively). Two collision modes were used with the beam directions reversed, which allows for the study of pPb collisions both in the forward and the backward region. The recorded pPb sample corresponds to an integrated luminosity of $1.6\,\text{nb}^{-1}$. All mentioned samples share similar data-processing stages (see Sect. 3.1.4 and Sect. 3.1.5 for the data-flow and the LHCb software description). The data is processed employing the DIRAC Grid infrastructure [235] and the final analysis is performed using the ROOT toolkit [181]. The *pp* (pPb) data and simulation samples are described in detail in Sect. 4.2.1 (Sect. 4.2.2) and a technical specification of the processing stages is provided in Appendix A.

4.2.1 Data and simulation samples – *pp* collisions

The average number of visible interactions per bunch crossing (see Sect. 3.1.1) during the 2011 *pp* data-taking period amounted to 1.4 [170]. The recorded minimum-bias sample used in the current study (see below for details on the trigger configuration) corresponds to 4×10^7 events. A simulated sample containing 2×10^7 minimum-bias events was produced using the standard LHCb simulation tools (see Sec. 3.1.5). Apart from the default LHCb configuration using PYTHIA8.1 version, two alternative ones utilising PYTHIA6.4 with the PERUGIA0 tuning [236] and a different Monte Carlo generator HERWIG++ [237] were employed to produce additional datasets. Each of these samples contains 1×10^7 events and they are used for the study of systematic uncertainties and for stability cross-checks. The simulated samples are needed to evaluate the effects related to the detector acceptance

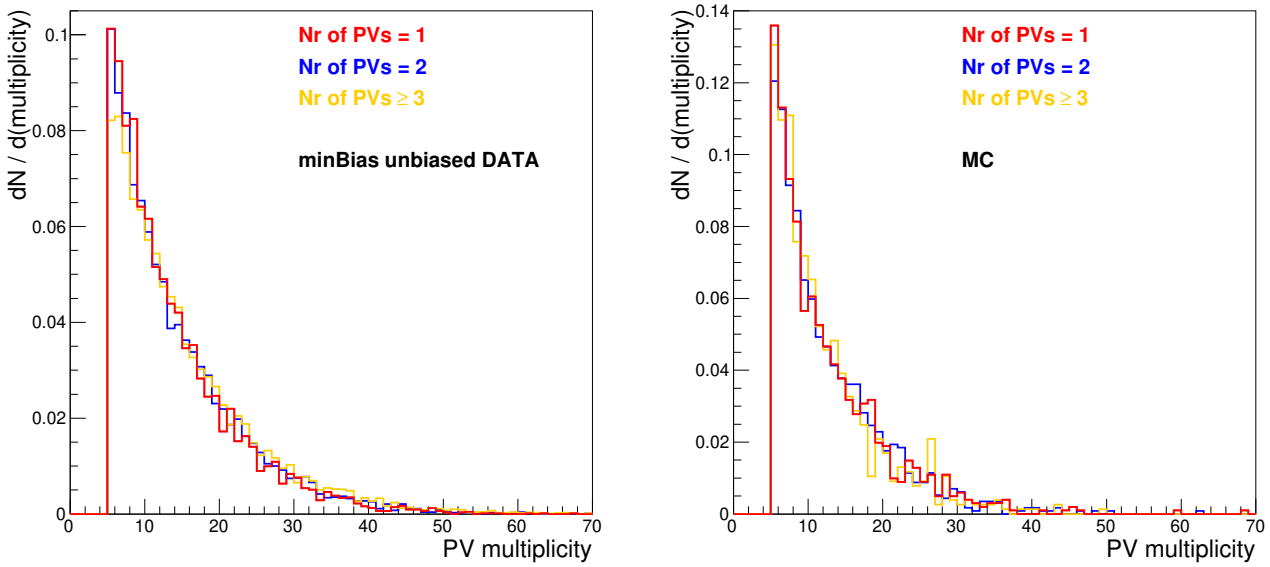

Figure 4.1: Multiplicity of the reconstructed VELO tracks assigned to a single PV in the events containing one (red), two (blue) and three or more (yellow) visible PVs (see Sect. 3.1.1), for the pp (left) data and (right) simulation. The distribution for the data is shown for the events selected with minimum-bias trigger after the unbiasing procedure.

and the imposed selection requirements. In all cases, modelling of the Bose–Einstein correlations is not activated in the simulation.

The described data sample contains a mixture of minimum-bias and no-bias events, and both can consist of multiple primary vertices. The no-bias ones correspond to events selected by the trigger completely randomly, while at least one reconstructed VELO track is required by the minimum-bias trigger [238]. The no-bias events are used directly in the analysis and a dedicated unbiasing procedure is applied to the minimum-bias ones. For the latter, one of the primary vertices with associated tracks that activated the minimum-bias trigger is randomly removed, thus providing a set of unbiased primary vertices. In the simulation, all events correspond in principle to the no-bias case, so no special treatment is applied. As shown in Fig. 4.1, the shape of the VELO-track multiplicity distribution for the data after the unbiasing procedure described above matches well to the one for the simulated no-bias events. The charged-particle multiplicity distributions for the pp data are not perfectly reproduced in the LHCb simulation [239]. Therefore, no ideal agreement in the absolute values of the reconstructed multiplicities between the data and the simulation is expected.

The samples are divided into three *event-activity classes* to study the dependence of the Bose–Einstein correlations on charged-particle multiplicity. The activity classes are defined based on the reconstructed VELO-track multiplicity distribution (see Fig. 4.2). This can be related to the overall charged-particle multiplicity in a primary vertex. The low activity class contains 48% of primary vertices with the lowest multiplicities, the medium one consists of 37% of PVs with higher multiplicities and the high activity class corresponds to the remaining 15% of PVs with the highest multiplicities.

The VELO-track multiplicity N_{VELO}, which is an LHCb-specific value that depends on

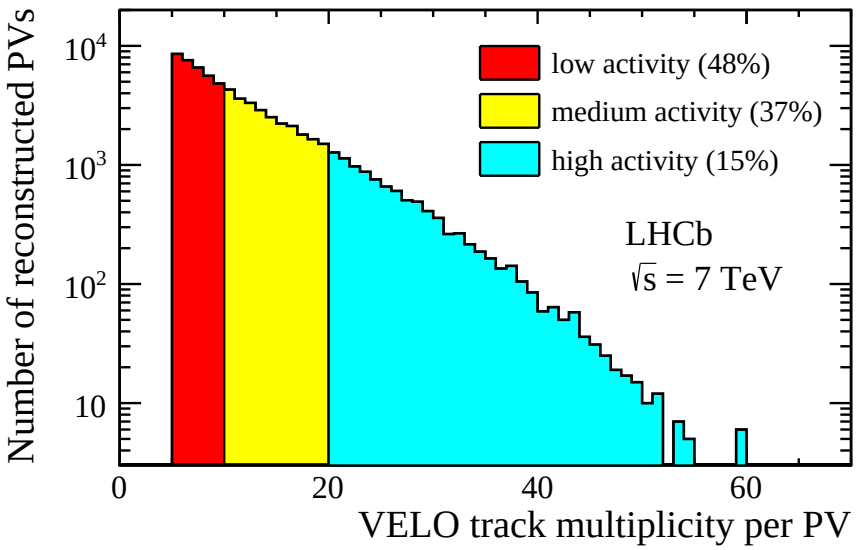

Figure 4.2: Distribution of the VELO-track multiplicity per PV in the 2011 no-bias pp data sample. Different colours indicate three activity classes defined as fractions of the full distribution. The minimum number of VELO tracks to accept a reconstructed PV is five [17].

Table 4.1: Activity classes in the pp data with the corresponding ranges of the VELO track and the charged-particle multiplicity.

activity class	low	medium	high
N_{VELO}	$\leqslant 10$	11–20	21–60
N_{ch}	$\leqslant (18 \pm 0.15)$	$(19 \pm 0.15) - (35 \pm 0.39)$	$(36 \pm 0.40) - (96 \pm 1.18)$

the reconstruction efficiency, is unfolded to the true charged-particle multiplicity N_{ch} using a Bayesian approach [240] within the ROOT unfolding framework (ROOUNFOLD) [241]. A simulation with PYTHIA8.1 using the standard LHCb configuration (see Sect. 3.1.5) is employed to produce an unfolding matrix illustrating the relation between the generated charged-particle multiplicity within the LHCb acceptance and the number of the reconstructed VELO tracks. This information is used to determine the corrected charged-particle multiplicities for the data. The relation between the reconstructed and unfolded multiplicity bin ranges is shown in Tab. 4.1. It agrees well with the dependencies estimated in the LHCb analysis of the charged-particle multiplicity [239].

4.2.2 Data and simulation samples – pPb collisions

The current study employs pPb collisions recorded by the LHCb experiment in 2013 [242, 243]. During this data-taking period, the instantaneous luminosity was kept at a level of five orders of magnitude lower ($\sim 5 \times 10^{27}\,\text{cm}^{-2}\,\text{s}^{-1}$) than what it was typically in the pp collisions, which lead to a very low value of the average number of visible interactions per bunch crossing (see Sect. 3.1.1). Only a small fraction of collisions in this dataset corresponds to multiple interactions and, furthermore, a dedicated selection requirement is applied to accept only the events with a single primary vertex (see Sect. 4.3.1). As already

mentioned, the data was collected in two beam modes: pPb, with the proton beam going along the LHCb spectrometer (towards the positive values on the z-axis in the LHCb coordinate system), and the Pbp one, with opposite directions of the beams. The recorded pPb collisions are asymmetric, which leads to a shift of ~ 0.465 between pseudorapidity values in the LHCb laboratory frame (η) and the nucleon-nucleon centre-of-mass system (η_{CMS})[1]. Thus, the two beam modes allow for the study of the pPb system in the forward ($1.5 < \eta_{\text{CMS}} < 4.5$) and the backward ($-5.5 < \eta_{\text{CMS}} < -2.5$) direction, additionally with an offset in the absolute pseudorapidity ranges. The analysis is performed in common VELO-track multiplicity bins (see below for details) to allow for a direct comparison between the two configurations and to study the potential dependence of the Bose–Einstein correlations on pseudorapidity (see Sects. 2.3.3 and 5). The full recorded pPb (Pbp) dataset corresponds to $\sim 2.0 \times 10^8$ (2.1×10^8) events.

Studies of the BEC effect require a sample of unbiased events. In the current analysis, it is acquired using a minimum-bias trigger that only requires minimal activity (at least one track) in the VELO detector. In the case of extremely low pileup (where most of the bunch crossings correspond to a single primary vertex), this is equivalent to accepting nearly all of the nonempty events. Hence, no specific feature is required for the events at the trigger level and the acquired sample is considered to be unbiased. Taking into account the stripping line and trigger requirements as well as the initial selection described in Sect. 4.3.1, the data samples available in the current study correspond to $\sim 6.3 \times 10^7$ (5.7×10^7) events for the pPb (Pbp) beam mode.

Simulated samples corresponding to the pPb data-taking conditions are produced using the standard LHCb simulation with the EPOS generator employed to model proton-lead collisions (see Sect. 3.1.5). Each of the pPb and the Pbp datasets contains $\sim 1.0 \times 10^7$ unbiased events with the number of interactions per bunch crossing fixed to unity. Additional samples are produced locally with a configuration exactly matching the central LHCb simulation to double the available statistics in both beam modes. The combined datasets correspond to $\sim 1.2 \times 10^7$ events after the initial selection described in Sect. 4.3.1 for each of the pPb and the Pbp cases. As in the pp analysis, the BEC effect is not present in the simulation. The simulated samples are used only for auxiliary studies (such as the optimisation of the particle identification described in Sect. 4.3.2), while the main results are obtained with a purely data-driven approach.

Similar to the pp study, the data and the simulation samples are divided into bins of the VELO-track multiplicity, which is used as a measure of the total charged-particle multiplicity in a primary vertex. The division is optimised to obtain high number of bins (to maximise the final number of data points for the correlation parameters) with statistics sufficient to perform the measurement. This procedure is based on the N_{VELO} distribution for the signal pairs (see Sect. 4.1.1) in the data, which is shown in Fig. 4.3 and allows for the selection of bins with similar signal yields for the final analysis. The integrated distribution presented in Fig. 4.3 (right) is particularly useful in optimising this choice. The study is performed in common bins for both the pPb and the Pbp modes, which enables a straightforward comparison between the two samples. Due to different shapes of the N_{VELO} distribution in the two cases, the bin ranges need to provide a compromise to ensure sufficient statistics for each beam configuration. The chosen bin boundaries

[1]Unless stated otherwise, references to pseudorapidity in this thesis correspond to the values in the LHCb laboratory frame. The use of a nucleon-nucleon centre-of-mass system is clearly indicated. For the symmetric pp collisions, pseudorapidity values in the two reference frames are equal ($\eta = \eta_{\text{CMS}}$).

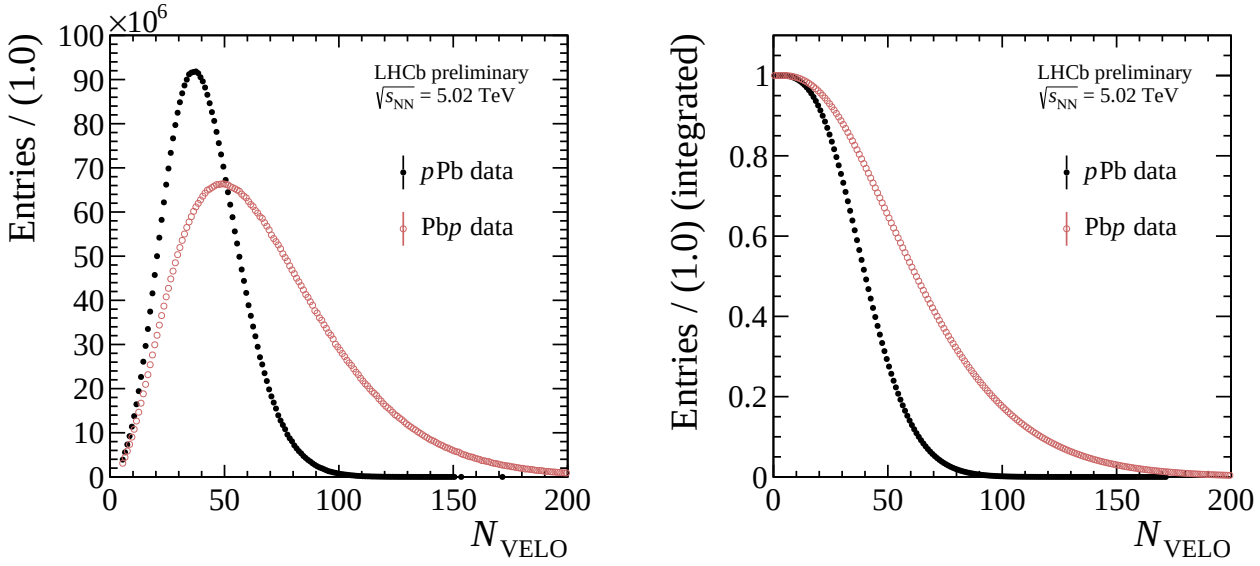

Figure 4.3: Distributions of the selected signal pairs originating from primary vertices with the given VELO-track multiplicity in the pPb (black dots) and the Pbp (red circles) data. (left) Regular distributions are shown together with (right) integrated ones, where the points illustrate the fraction of pairs corresponding to multiplicities higher than in the given bin.

are presented in Tab. 4.2, together with approximate fractions of the respective sample corresponding to the given bin, in order to illustrate the statistics available in each case.

4.3 Selection

A set of selection requirements is applied to both the data and the simulation samples. The main purpose of this procedure is to retain good-quality track pairs that are properly assigned to the given primary vertex. Variables used in the study and initial selection criteria are described in Sect. 4.3.1. The following Sect. 4.3.2 provides details on the particle identification, which is optimised to ensure a high-purity sample of pion candidates. Additional requirements are applied to reduce effects related to the fake and the clone tracks (see Sect. 4.3.3). A summary of the final selection criteria is given in Sect. 4.3.4.

As already mentioned, the studies of pp and pPb collisions share a similar selection strategy. The selection procedure is optimised independently for each analysis. However, in both cases the optimisations lead to consistent conclusions on the selection criteria, which is detailed in the following sections. If not otherwise stated, the described requirements are applied in both studies. Differences between the original selection for the pp analysis and the pPb study are clearly indicated.

4.3.1 Preselection criteria

The study of the BEC effect is performed using tracks within the LHCb detector acceptance ($2.0 < \eta < 5.0$) with reconstructed segments in the VELO detector and in the T stations downstream of the LHCb magnet (long tracks – see Sect. 3.1.2). Only the PVs corresponding to visible interactions (see Sect. 3.1.1) are taken into account. To

Table 4.2: Ranges of the VELO-track multiplicity bins in the pPb and the Pbp datasets. Approximate fractions of the relevant data sample corresponding to each bin are also indicated.

		sample fraction [%]	
bin#	N_{VELO}	pPb	Pbp
1	[5–10)	< 2	< 2
2	[10–15)	2	2
3	[15–20)	4	2
4	[20–25)	7	3
5	[25–30)	10	4
6	[30–35)	13	5
7	[35–40)	14	6
8	[40–45)	10	5
9	[45–50)	10	6
10	[50–55)	8	6
11	[55–60)	7	7
12	[60–65)	5	6
13	[65–80)	6	15
14	[80–90)	–	7
15	[90–100)	–	7
16	[100–115)	–	6
17	[115–140)	–	7
18	[140–180)	–	4

reduce the potential contamination from muons, the tracks with associated signatures in the muon stations are not accepted.

A number of variables is used in the analysis. Some of them, such as the track momentum p, transverse momentum p_{T} and impact parameter IP, are defined in Sect. 3.1. Another important quantity is track χ^2, which illustrates the fit quality in the track reconstruction. Combined information from different subdetectors is employed to obtain variables – *e.g.* $ProbNN(\mathrm{pion})$ – that are used for the particle identification (see Sect. 3.1.3 for details).

Examples of distributions of the basic variables mentioned above are shown for the pp data in Figs. 4.4, 4.5, 4.6 and 4.7. For some of these parameters, initial requirements are introduced at an early stage of the analysis. This procedure is referred to as preselection and some of the prerequisites are further optimised in the analysis process, which is described in the following sections. The preselection criteria are summarised in Tab. 4.3 and indicated in the figures, where applicable. The requirements imposed on the total and the transverse track momentum, track χ^2 and $ProbNN(\mathrm{ghost})$ ensure that only the tracks with a high-quality reconstruction are selected. The prerequisite of $IP < 0.4\,\mathrm{mm}$ is applied to remove tracks that are not originating directly from the given PV. A relatively loose requirement is chosen to retain most of the particles associated with a certain primary vertex. It is optimised using the simulation by studying the IP distribution for pions that are properly (or incorrectly) assigned to the respective PV (see Fig. 4.8). The distribution for the misassigned tracks is observed to be flat and hence the requirement on IP is chosen

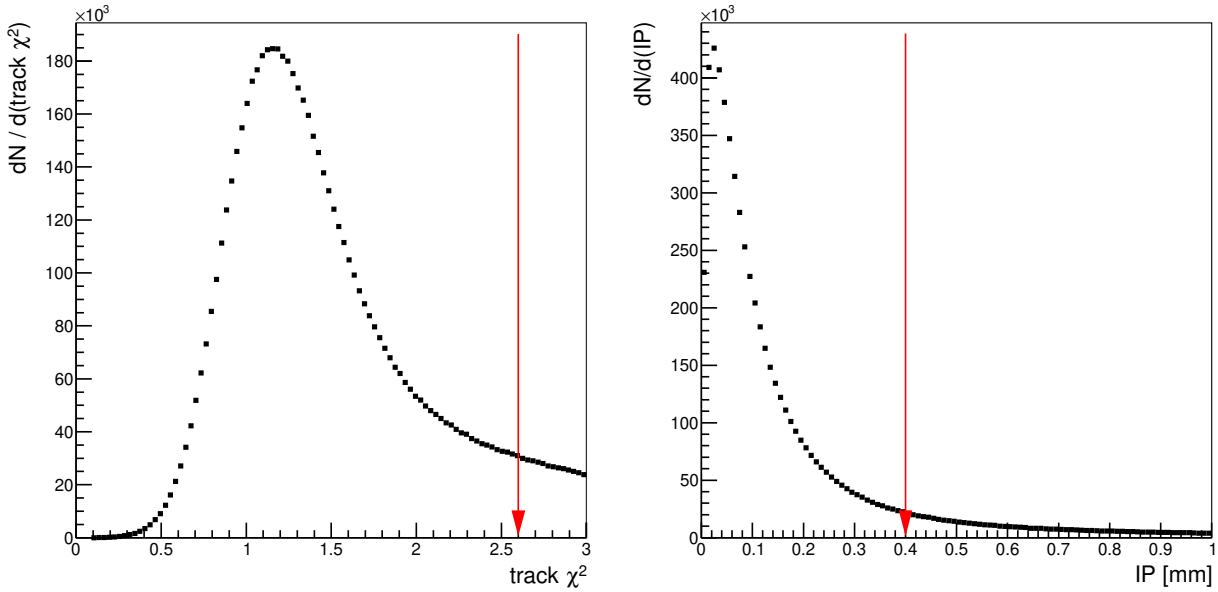

Figure 4.4: Distribution of the basic variables used in the study shown for the full (no selection) *pp* data sample: (left) reconstructed track χ^2 and (right) track impact parameter. The preselection requirements imposed on the given quantity are indicated by the red arrows.

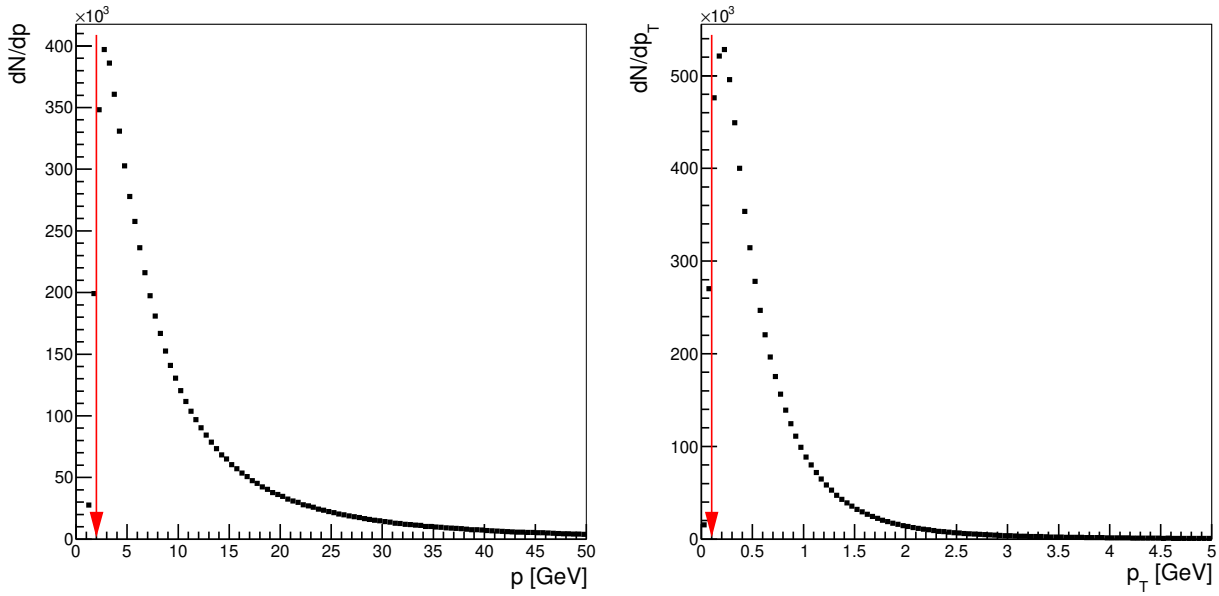

Figure 4.5: Distribution of the basic variables used in the study shown for the full (no selection) *pp* data sample: track (left) momentum and (right) transverse momentum. The preselection requirements imposed on the given quantity are indicated by the red arrows.

in a regime where fractions of the tracks that are properly or incorrectly assigned become comparable. A stability check is performed by varying this prerequisite in the range of 0.3–0.4 mm. It is observed that such variations have no significant impact on the shape of the correlation function for both the data and the simulation.

In the *p*Pb study, an additional requirement is introduced that only events with

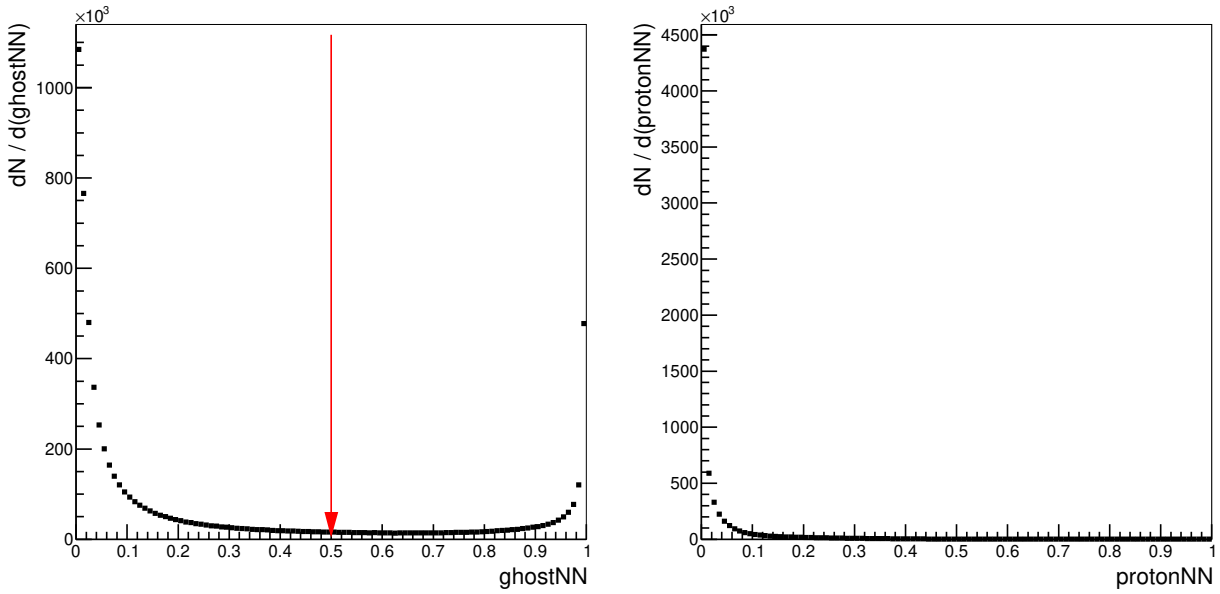

Figure 4.6: Distribution of the basic variables used in the study shown for the full (no selection) *pp* data sample: (left) $ProbNN$(ghost) and (right) $ProbNN$(proton). The preselection requirements imposed on the given quantity are indicated by the red arrows.

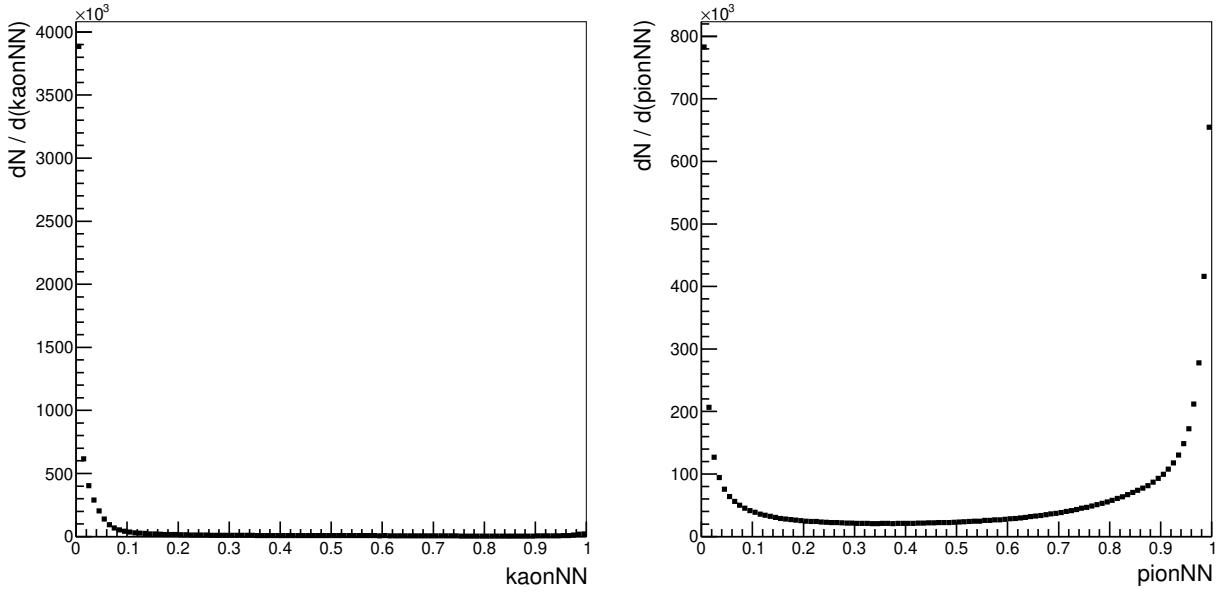

Figure 4.7: Distribution of the basic variables used in the study shown for the full (no selection) *pp* data sample: (left) $ProbNN$(kaon) and (right) $ProbNN$(pion).

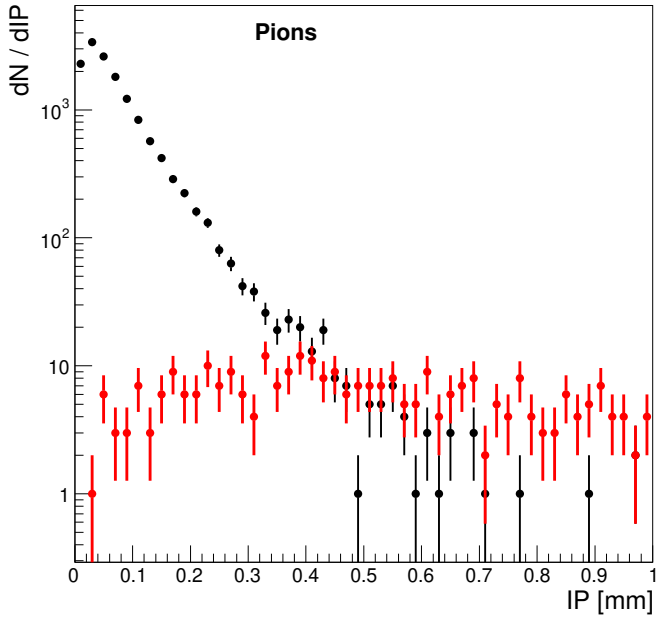

Figure 4.8: Distribution of the impact parameter for pions in the full (no selection) pp simulation sample (identification is based on the generator information). The black points correspond to the tracks properly assigned to the PV of their origin (which is verified using the simulation information), while the red points indicate the misassigned ones.

Table 4.3: Preselection criteria in the pp and the pPb analyses. See the description in text for more details on the imposed requirements.

selection	requirement
PV	visible PVs (both pp and pPb) single PV, $-160 < z_{\mathrm{PV}} < 60\,\mathrm{mm}$ (pPb only)
track type	long, charged, no muon signature
track η	2.0–5.0
track χ^2	< 2.6
track p	$> 2.0\,\mathrm{GeV}$
track p_{T}	$> 0.1\,\mathrm{GeV}$
track IP	$< 0.4\,\mathrm{mm}$
$ProbNN(\mathrm{ghost})$	< 0.5

a single primary vertex are accepted, which is related to the small fraction of collisions with multiple interactions in the pPb data (see Sect. 4.2.2). Furthermore, the primary vertex is required to be located within a limited distance from the centre of the VELO detector. Due to the limited VELO acceptance (see Fig. 3.5), some of the tracks can escape the subdetector, resulting in slightly different N_{VELO} distributions for PVs in various regions along the z-axis. This can lead to an effective bin migration and potentially affect the shape of the correlation function for the given VELO-track multiplicity. A requirement on the PV position along the z-axis (z_{PV}) is introduced to ensure uni-

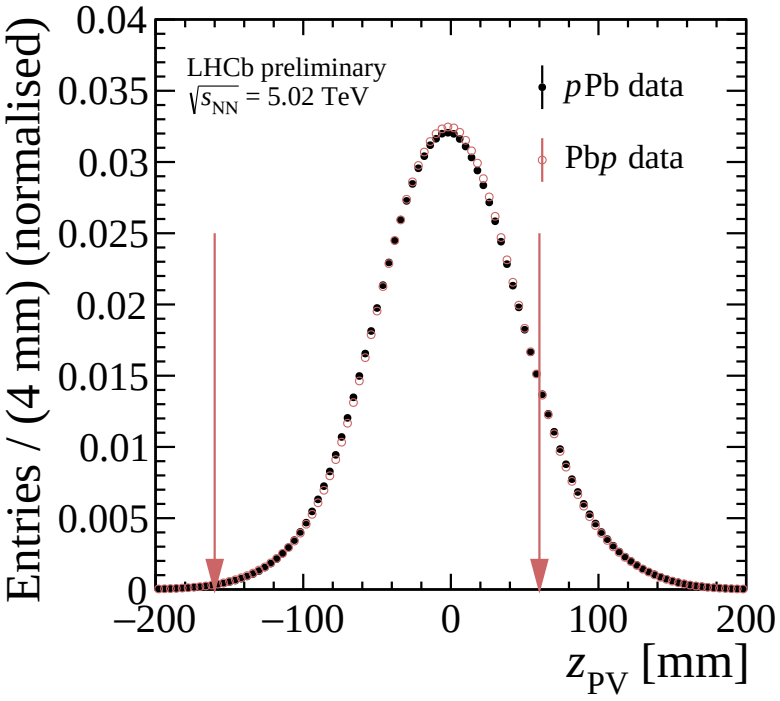

Figure 4.9: Distributions of the position along the LHCb z-axis for primary vertices associated with tracks in the pPb (black dots) and the Pbp (red circles) data after the preselection described in the text. The requirements imposed on the $z_{\rm PV}$ value are indicated by the red arrows. The distributions are normalised to unity.

formity of the $N_{\rm VELO}$ distribution in the overall sample. The optimal value for this selection criterion is determined by studying the $N_{\rm VELO}$ distributions across 10 mm-wide bins of the $z_{\rm PV}$ variable. First, a region of $z_{\rm PV}$ in which PVs have the highest average $N_{\rm VELO}$ values is found (which is equivalent to the PVs located around $z_{\rm PV} = -90$ mm). Then, the average $N_{\rm VELO}$ values at the centre of the VELO detector are determined, which is a region that gives the most significant contribution of tracks in both the pPb and the Pbp samples (see Fig. 4.9). The final requirement on the $z_{\rm PV}$ parameter imposes that the difference between the average $N_{\rm VELO}$ values in a given $z_{\rm PV}$ range and at the centre of VELO be no larger than the difference between the previously determined quantities. It corresponds to the prerequisite of $-160 < z_{\rm PV} < 60$ mm. The choice of this particular range is taken into account in the study of the systematic uncertainty (see Sect. 4.5.2.2).

4.3.2 Particle identification

The particle identification is optimised to provide a final sample of selected pions. It is important to ensure a sample with high purity, but a strict requirement on the $ProbNN$(pion) variable may also strongly affect the signal region of the correlation function, by suppressing low-momentum pions that contribute to the BEC effect. Two selection requirements in a *veto* configuration are imposed to reduce the fraction of misidentified kaons and protons in the final sample ($ProbNN(\text{kaon}) < 0.5$ and $ProbNN(\text{proton}) < 0.5$, respectively).

4.3.2.1 PID variables in the simulation

The PID variables in the LHCb simulation are not perfectly reproduced, so applying the same selection criteria on those parameters for both the simulation and the data could affect the samples differently. Due to that, the simulated quantities need to be corrected using PID calibration samples from the data. These samples contain high statistics of particles of a given type that are identified based on purely kinematic selections, without using the PID information [177]. The correction procedure is referred to as PID resampling and is performed using the PIDCalib package [244]. Standard algorithms for this operation are dedicated to correct only the tracks related to the given decay candidates. In the current study, correlations between all of the particles in the event are investigated. Therefore, each of the tracks needs to be taken into account in the resampling procedure and the available methods are adapted accordingly.

The *ProbNN* distributions for different particle types (π, K, p, e, μ) can be obtained from the calibration samples mentioned above. These quantities are dependent on the track momentum, pseudorapidity and the VELO-track multiplicity in the given PV, hence the relevant distributions in the calibration samples are provided in $(p, \eta, N_{\mathrm{VELO}})$-bins. In the simulation, the type of a generated particle is known, so the simulated PID variables can be replaced with random values using the related distribution from the calibration sample, for the given particle type and within the corresponding $(p, \eta, N_{\mathrm{VELO}})$-bin. Some of the bins in the calibration samples can be empty and the resampling cannot be performed. In such cases, the relevant distributions are integrated over N_{VELO}, so in the end, they are given only in the (p, η)-bins. It is found that the fraction of empty bins for pions, kaons and protons is below 0.1% in this procedure for the *pp* analysis.

The PID resampling is performed in both the *pp* and the *p*Pb studies using calibration samples dedicated for each collision type. The procedure is significantly more important in the *pp* case, where the simulation is used to construct the double-ratio function and to perform the final fits. An example of the distributions corrected using the strategy described in this section is shown in Fig. 4.10 for the kaon and proton identification in the *pp* analysis. The accuracy of the PID resampling method in this case is taken into account in the study of the systematic uncertainty (see Sect. 4.4.2.6).

4.3.2.2 Optimisation of the pion identification criteria

Purity of the final pion pair sample for different $ProbNN(\mathrm{pion})$ requirements is shown for the *pp* study in Tabs. 4.4, 4.5 and 4.6, and for two example N_{VELO} bins in the *p*Pb analysis in Tabs. 4.7 and 4.8. The fraction of pure kaon-kaon and proton-proton pairs (which could provide additional Bose–Einstein or Fermi–Dirac correlations, respectively, and distort the shape of the enhancement in the BEC-signal region for pions) is negligible in all cases. Mixed pairs comprise pairs of particles of different types (*e.g.* pion-kaon pairs) as well as pairs containing misreconstructed tracks (see Sect. 3.1.2). Pairs of the first kind do not affect the signal region by definition, since they do not exhibit Bose–Einstein or Fermi–Dirac correlations. The ones containing misreconstructed tracks could in principle impact the regime of the BEC enhancement. However, as indicated in Sect. 4.3.3, this kind of contribution is well-controlled experimentally using dedicated selection requirements.

The effect of different criteria for the pion identification on the $C_2(Q)$ shape is shown in Figs. 4.11 and 4.12 for the *pp* data and simulation, respectively. The final requirement of $ProbNN(\mathrm{pion}) > 0.65$ is chosen, being a value for which the enhancement due to the BEC

Table 4.4: Contributions of various pair types to the selected signal pairs in the *pp* simulation (using identification based on the generator information) with different requirements on the pion identification for the low-activity event class. The values are measured in the BEC-signal region of $Q < 1.0\,\mathrm{GeV}$.

	pair type [%]			
$ProbNN(\mathrm{pion})$	pions	kaons	protons	mixed
> 0.50	96.15	0.006	0.002	3.84
> 0.55	96.47	0.005	0.002	3.52
> 0.60	96.82	0.004	0.001	3.17
> 0.65	97.12	0.004	< 0.001	2.87
> 0.70	97.44	0.003	< 0.001	2.55
> 0.75	97.77	0.002	< 0.001	2.23
> 0.80	98.06	0.001	< 0.001	1.94

Table 4.5: Contributions of various pair types to the selected signal pairs in the *pp* simulation (using identification based on the generator information) with different requirements on the pion identification for the medium-activity event class. The values are measured in the BEC-signal region of $Q < 1.0\,\mathrm{GeV}$.

	pair type [%]			
$ProbNN(\mathrm{pion})$	pions	kaons	protons	mixed
> 0.50	95.72	0.008	0.003	4.27
> 0.55	96.09	0.006	0.002	3.90
> 0.60	96.47	0.005	0.002	3.52
> 0.65	96.82	0.004	0.001	3.18
> 0.70	97.18	0.003	< 0.001	2.81
> 0.75	97.57	0.002	< 0.001	2.43
> 0.80	97.93	0.001	< 0.001	2.07

Table 4.6: Contributions of various pair types to the selected signal pairs in the *pp* simulation (using identification based on the generator information) with different requirements on the pion identification for the high-activity event class. The values are measured in the BEC-signal region of $Q < 1.0\,\mathrm{GeV}$.

	pair type [%]			
$ProbNN(\mathrm{pion})$	pions	kaons	protons	mixed
> 0.50	94.55	0.013	0.005	5.43
> 0.55	95.04	0.010	0.004	4.94
> 0.60	95.53	0.008	0.003	4.46
> 0.65	95.98	0.006	0.002	4.01
> 0.70	96.46	0.004	0.001	3.54
> 0.75	96.98	0.003	< 0.001	3.02
> 0.80	97.50	0.002	< 0.001	2.50

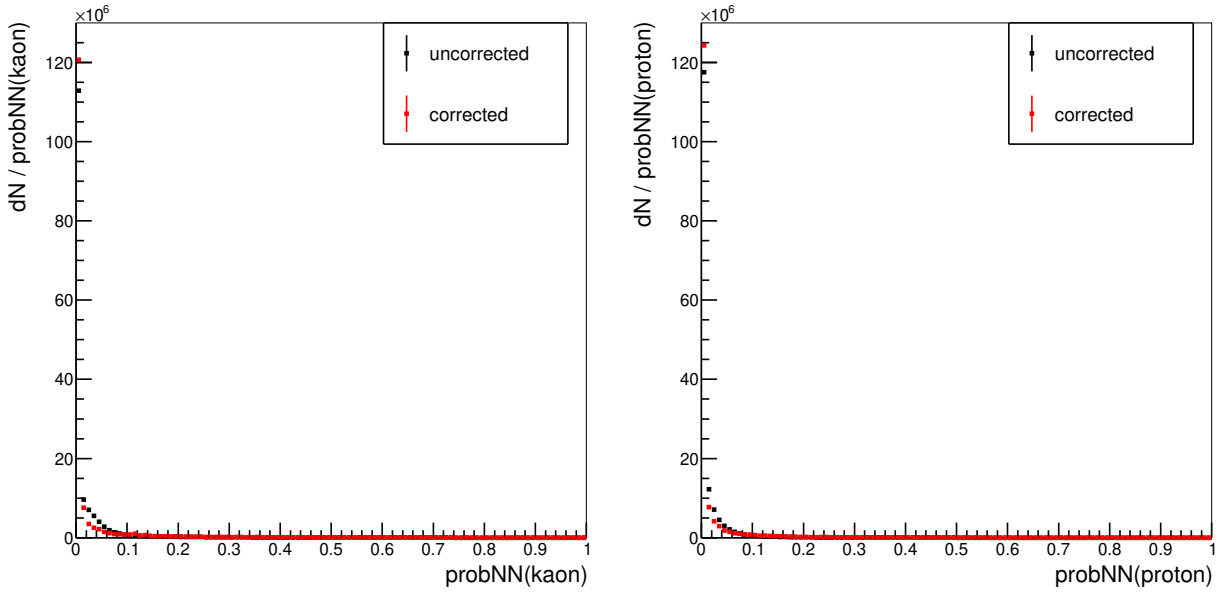

Figure 4.10: Example results of the PID variables resampling in the *pp* simulation for the identification of: (left) kaons and (right) protons. The black (red) points correspond to the uncorrected (corrected) distributions.

Table 4.7: Contributions of various pair types to the selected signal pairs in the simulation (using identification based on the generator information) with different requirements on the pion identification for a moderate-multiplicity region ($35 \leqslant N_{\mathrm{VELO}} < 40$) of the pPb sample. The values are measured in the BEC-signal region of $Q < 1.0\,\mathrm{GeV}$.

	pair type [%]			
$ProbNN(\mathrm{pion})$	pions	kaons	protons	mixed
> 0.50	91.13	0.001	0.001	8.86
> 0.55	91.50	0.001	0.001	8.49
> 0.60	91.83	< 0.001	< 0.001	8.16
> 0.65	92.11	< 0.001	< 0.001	7.88
> 0.70	92.42	< 0.001	< 0.001	7.56
> 0.75	92.81	< 0.001	< 0.001	7.17
> 0.80	93.20	< 0.001	< 0.001	6.79

signal begins to saturate (for the data), while the sample purity is still high ($\gtrsim 96\%$ pion pairs in all three activity bins). The choice of this selection requirement has no significant impact on the correlation function shape for the simulation (apart from the Q values below 0.05 GeV, which is a region not taken into account in the final analysis – see the following sections). Details of this optimisation strategy are investigated in terms of the studies of the systematic uncertainty (see Sect. 4.4.2.7).

A similar procedure is performed in the pPb study, leading to the same conclusion on the optimal requirement for pion identification ($ProbNN(\mathrm{pion}) > 0.65$). It is observed that for this prerequisite the enhancement due to the BEC signal begins to saturate in the correlation function for the data (see Figs. 4.13 and 4.14). The correlation function

Table 4.8: Contributions of various pair types to the selected signal pairs in the simulation (using identification based on the generator information) with different requirements on the pion identification for a high-multiplicity region ($100 \leqslant N_{\mathrm{VELO}} < 115$) of the Pb$p$ sample. The values are measured in the BEC-signal region of $Q < 1.0\,\mathrm{GeV}$.

	pair type [%]			
$ProbNN(\mathrm{pion})$	pions	kaons	protons	mixed
> 0.50	87.42	0.010	0.007	12.56
> 0.55	88.17	0.009	0.006	11.82
> 0.60	88.98	0.007	0.005	11.01
> 0.65	89.79	0.005	0.004	10.20
> 0.70	90.76	0.004	0.003	9.23
> 0.75	91.80	0.003	0.002	8.20
> 0.80	93.11	0.002	0.001	6.89

for the simulation is not taken into account in this case due to a purely data-driven method of obtaining the correlation parameters. Similar to the pp study, the impact of this requirement choice on the correlation parameters is investigated as an input to systematic uncertainty (see Sect. 4.5.2.2).

Although the purity of the final signal pairs is lower than in the pp case (one of the reasons is that the fraction of misreconstructed tracks is increasing with event multiplicity), it is still relatively high (close to 90%) and the competing increasing contribution corresponds to the mixed pairs, which do not contain the BEC or the FDC effects. More details on controlling the misreconstructed tracks that could affect the signal region are given in Sect. 4.3.3.

4.3.3 Study of the clone and the fake tracks

The presence of misreconstructed tracks in the final pion sample can affect the shape of the correlation function. Dedicated methods in the LHCb software (see Sect. 3.1.5) allow us to control the contribution from such tracks. The clone tracks are recognised by an algorithm based on the calculation of the Kullback–Leibler distance [245] between two tracks. In the standard LHCb reconstruction algorithms, tracks with values of Kullback–Leibler distance < 5000 are removed from further processing. Modifying this requirement is employed to evaluate the contribution of clone tracks to the systematic uncertainty (see Sect. 4.4.2.5). As mentioned in Sect. 3.1.2, the fraction of fake tracks is significantly reduced using a dedicated neural network classifier [175]. This contribution is further limited in the current study by applying relevant requirements on the PID variables. An optimal prerequisite of $ProbNN(\mathrm{ghost}) < 0.25$ is chosen due to a relatively low contribution of fake tracks in the selected pion samples (below 1% in most cases – see Tabs. 4.12 and 4.21). Also, it is observed that for this value the signal region of the correlation function is not significantly affected (as shown in Fig. 4.6 (left), most tracks have low values of the $ProbNN(\mathrm{ghost})$ variable and an overly strict selection may lead to the suppression of actual pions that contribute to the BEC signal). The impact of the choice of this particular prerequisite on the correlation parameters is studied as a contribution to the systematic uncertainty (see Sect. 4.4.2.4).

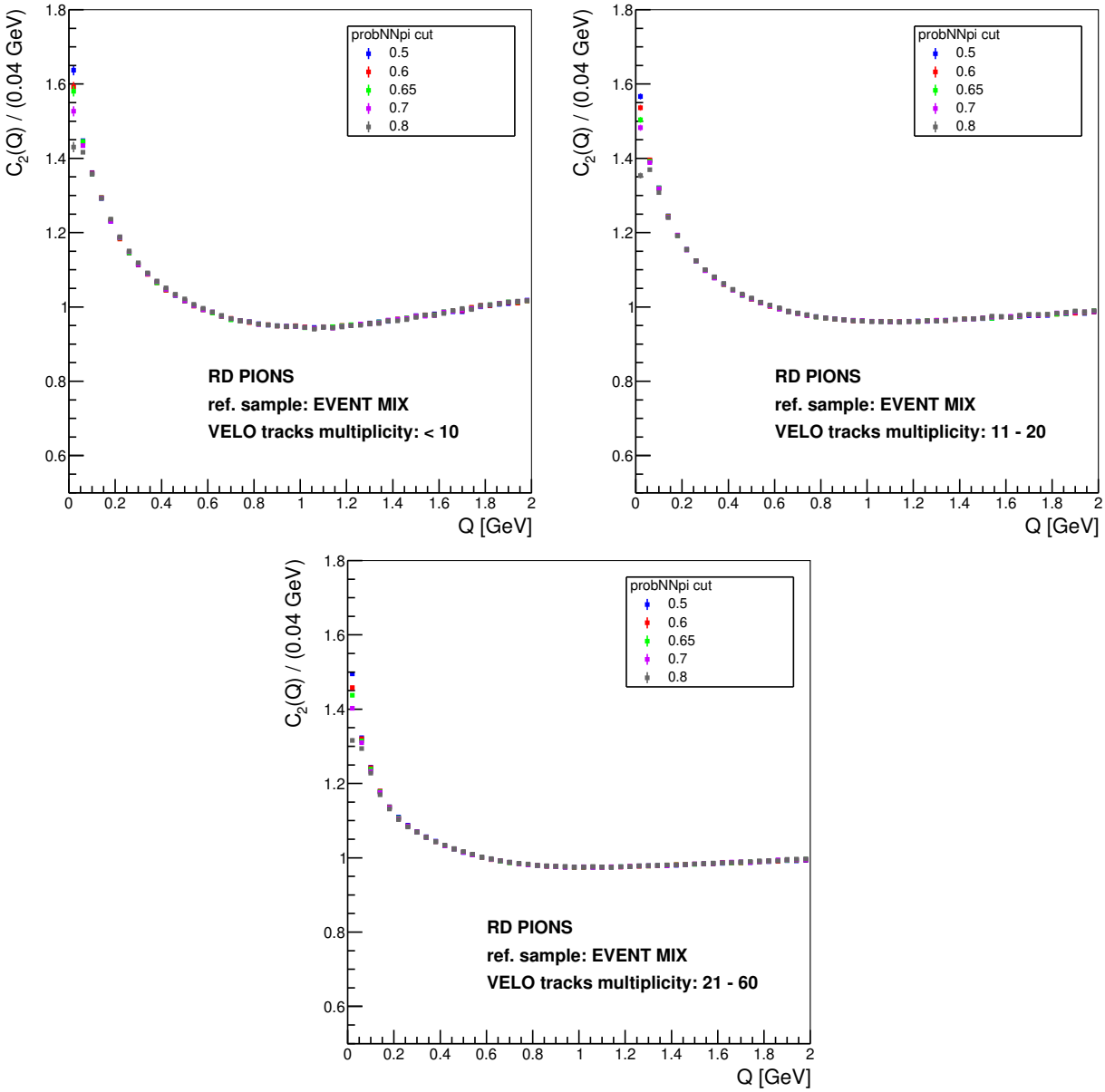

Figure 4.11: Correlation functions for the selected signal pairs (prior to the exclusion of the $Q < 0.05\,\text{GeV}$ region – see Sect. 4.3.3) in the pp data for the three event-activity classes. Different requirements on the pion identification are applied.

Despite the employment of the methods mentioned above, some of the misreconstructed tracks can still be present. An additional requirement is applied in the current study. If there are multiple tracks with all associated hits within the same segments of the VELO detector, then only one of these tracks (the one with the best reconstruction quality) is preserved. This significantly reduces the fractions of both the clone and the fake tracks.

The separation in momentum between two particles degrades for the very-low Q region, and this effect is not well reproduced in the simulation (which leads to increasing differences between the data and the simulation with the Q values approaching zero). Studies performed using the pp simulation suggest that without the dedicated requirements described in this section, the fractions of pion pairs containing a fake or a clone track in the region of $Q < 0.05\,\text{GeV}$ correspond to $\sim 25\%$ and 8–15%, respectively (the contributions increase

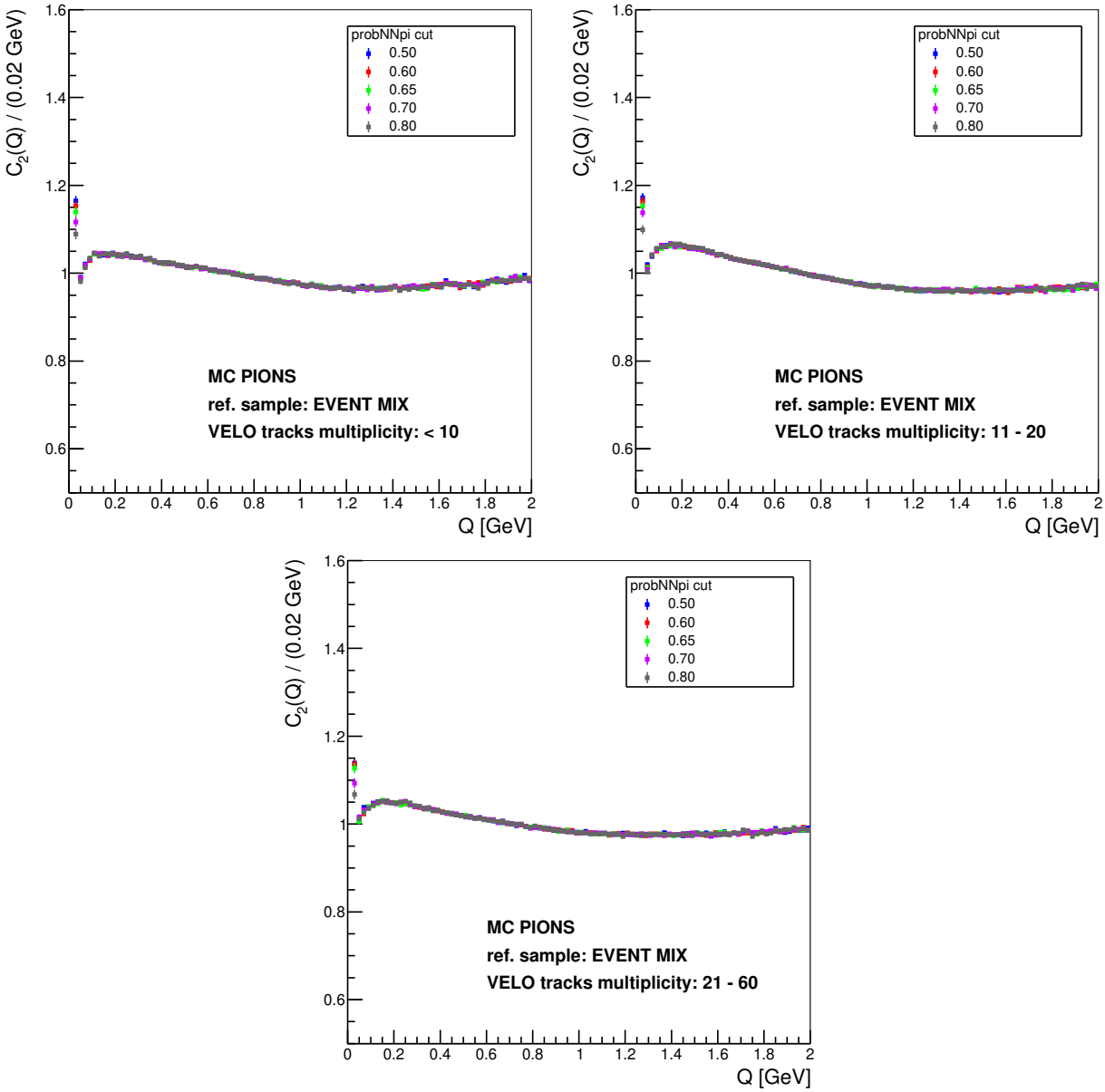

Figure 4.12: Correlation functions for the selected signal pairs (prior to the exclusion of the $Q < 0.05\,\mathrm{GeV}$ region – see Sect. 4.3.3) in the pp simulation for the three event-activity classes. Different requirements on the pion identification are applied.

with the event activity). Due to the significant influence of these misreconstructed tracks and the discrepancy in the track reconstruction between the data and the simulation, the $Q < 0.05\,\mathrm{GeV}$ region is excluded in further analysis.

Controlling the clone tracks is especially important due to their similar trajectories and the same electric charge. Pairs of the clone tracks can affect the correlation function shape in the low-Q region, where the BEC signal is expected. For this purpose, slopes of the track momenta[2] are studied. The clone tracks usually share a very similar trajectory, hence the difference in the relevant slopes in a particle pair – Δt_x and Δt_y – tends to be small. An example of Δt_x and Δt_y distributions for the pp data sample is shown in

[2]Slopes of the track momenta t_x (t_y) are defined as a ratio of the x (y) and z momentum components.

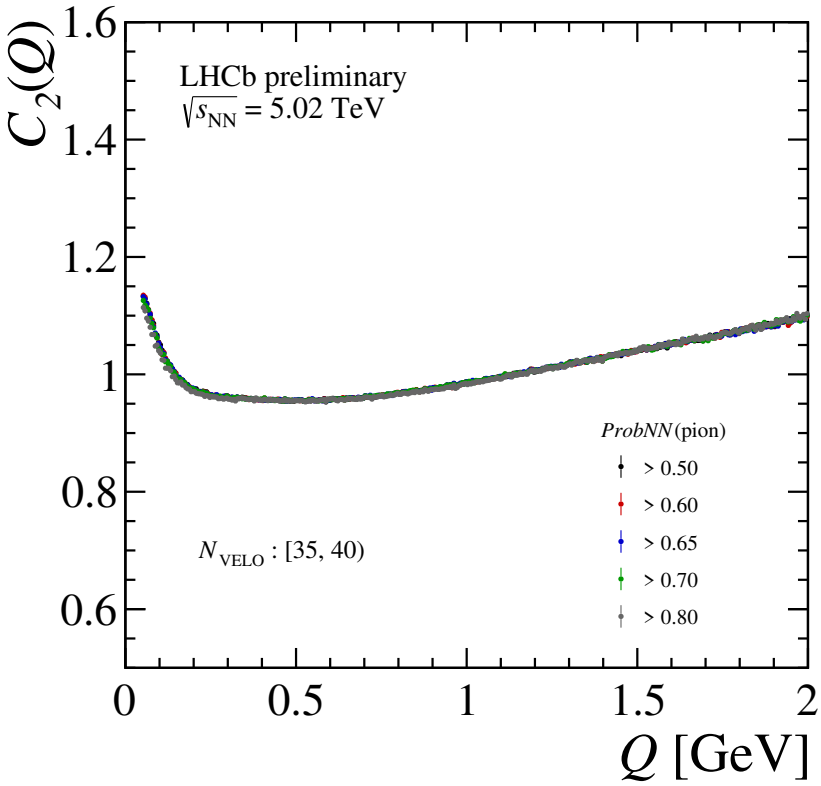

Figure 4.13: Correlation functions for the selected signal pairs in a moderate-multiplicity region ($35 \leqslant N_{\mathrm{VELO}} < 40$) of the pPb dataset. Different requirements on the pion identification are applied.

Fig. 4.15. It is visible that the final selection requirements (without the prerequisites on the Δt_x and the Δt_y values – see Sect. 4.3.4) already significantly reduce the number of pairs containing a clone track. However, to further limit this contribution in the current study, if both $|\Delta t_x|$ and $|\Delta t_y|$ values are smaller than 0.3 mrad, then such a pair is not accepted for the final analysis (the chosen value for this criterion is based on the shape of the structures in the presented distributions after applying the final selection requirements without the limits on Δt_x and Δt_y – see the red points in Fig. 4.15). With the selection procedure described in this section, the effect of the clone tracks on the pp analysis results is negligible (see Sect. 4.4.2.5).

As mentioned above, the fractions of signal pairs containing misreconstructed tracks tend to increase with the event multiplicity. It is reflected in the higher amounts of mixed pairs in the pPb study (see Sect. 4.3.2.2), where the available multiplicities are higher than in the pp case. For this reason, the control of the effects related to the clone tracks is further scrutinised in the pPb study. To illustrate the effect from the presence and removal of clone (but potentially also fake) tracks, the Δt_x and the Δt_y distributions for the pPb simulation are presented in Fig. 4.16. In both cases the points are shown after the preselection only, but the black dots correspond to a situation in which both the fake and the clone tracks are removed using the simulation information. For comparison, similar distributions for the data are shown in Fig. 4.17 – in this case, for the preselection and the final selection requirements. The final selection includes the strategies for

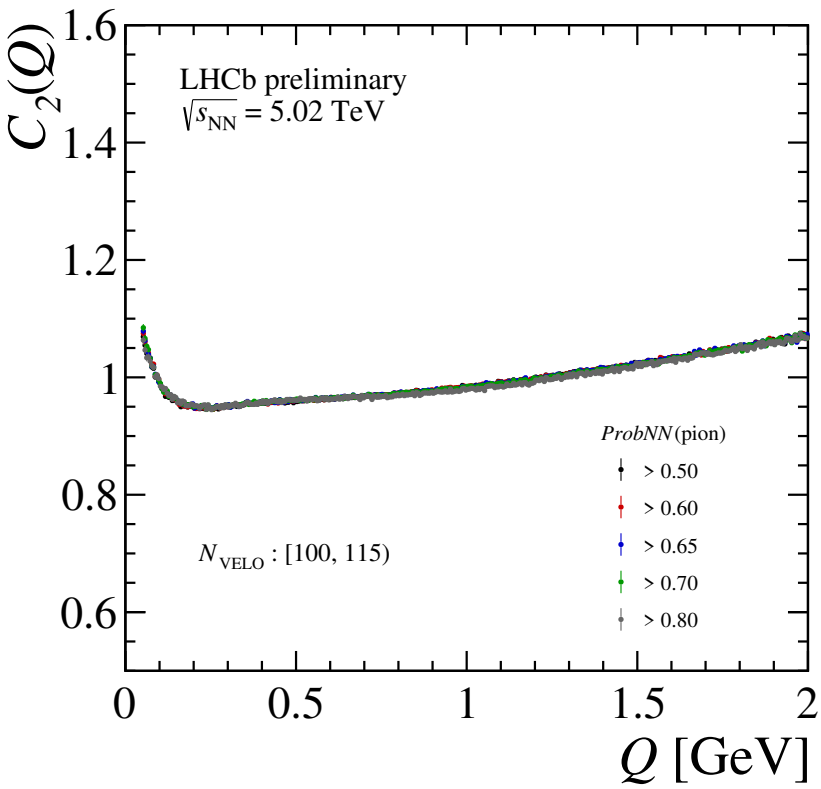

Figure 4.14: Correlation functions for the selected signal pairs in a high-multiplicity region ($100 \leqslant N_{\mathrm{VELO}} < 115$) of the Pb$p$ dataset. Different requirements on the pion identification are applied.

mitigating the contribution from misreconstructed tracks that are described in this section. The presented figures indicate that a set of tools mentioned above effectively removes the pairs with misreconstructed tracks from the BEC-signal region in the data sample, giving similar results as in the study performed using the simulation. As in the pp case, the clone and the fake tracks contributions are very well controlled in the pPb data.

4.3.4 Final selection requirements

A summary on the final selection criteria applied to the pp and the pPb datasets is given in Tab. 4.9. A detailed description of the optimisation of the requirements is provided in the previous sections. In particular, some of the initial prerequisites stated in Sect. 4.3.1 are more stringent in the final selection.

4.4 Study of the BEC effect in pp collisions

The analysis of the Bose–Einstein correlations for pairs of same-sign charged pions in pp collisions recorded by the LHCb experiment is described in this section. Details on the analysis methods, data sets and selection requirements are given in the previous Sects. 4.1, 4.2, 4.3. The main analysis procedure is presented in Sect. 4.4.1. It is followed by a description of the study of systematic uncertainties in Sect. 4.4.2. A summary of

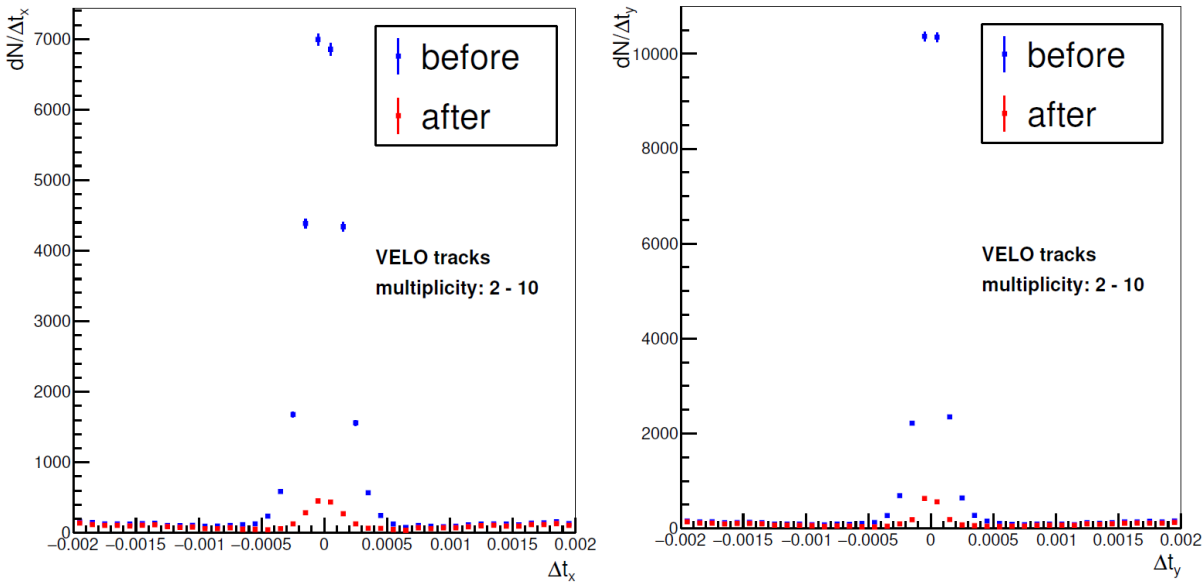

Figure 4.15: Example distribution of (left) Δt_x and (right) Δt_y for the signal pairs in the low-activity bin of the pp data (the values on the x-axis are given in radians). The blue (red) points are obtained using the preselection (the final selection without the criteria on Δt_x and Δt_y) requirements. The Δt_x (Δt_y) distribution is shown with an additional criterion that the Δt_y (Δt_x) value is lower than 0.5 mrad and the Q range is restricted to < 0.1 GeV (in order to illustrate the contribution of misreconstructed tracks in a region where it is most prominent).

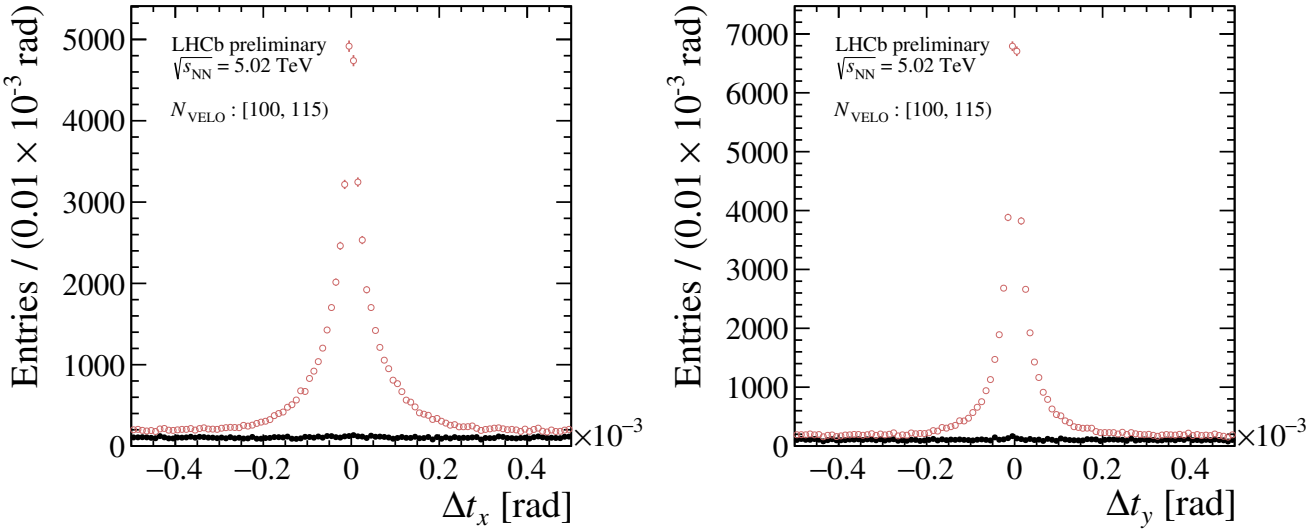

Figure 4.16: Example distribution of (left) Δt_x and (right) Δt_y for the signal pairs in a high-multiplicity region ($100 \leqslant N_{\text{VELO}} < 115$) of the Pb$p$ simulation. The red circles correspond to the usage of only the preselection requirements, while the black dots are obtained by further removing the pairs with misreconstructed tracks using the simulation information. The Δt_x (Δt_y) distribution is shown with an additional criterion that the Δt_y (Δt_x) value is lower than 0.5 mrad and the Q range is restricted to < 0.1 GeV (in order to illustrate the contribution of misreconstructed tracks in a region where it is most prominent).

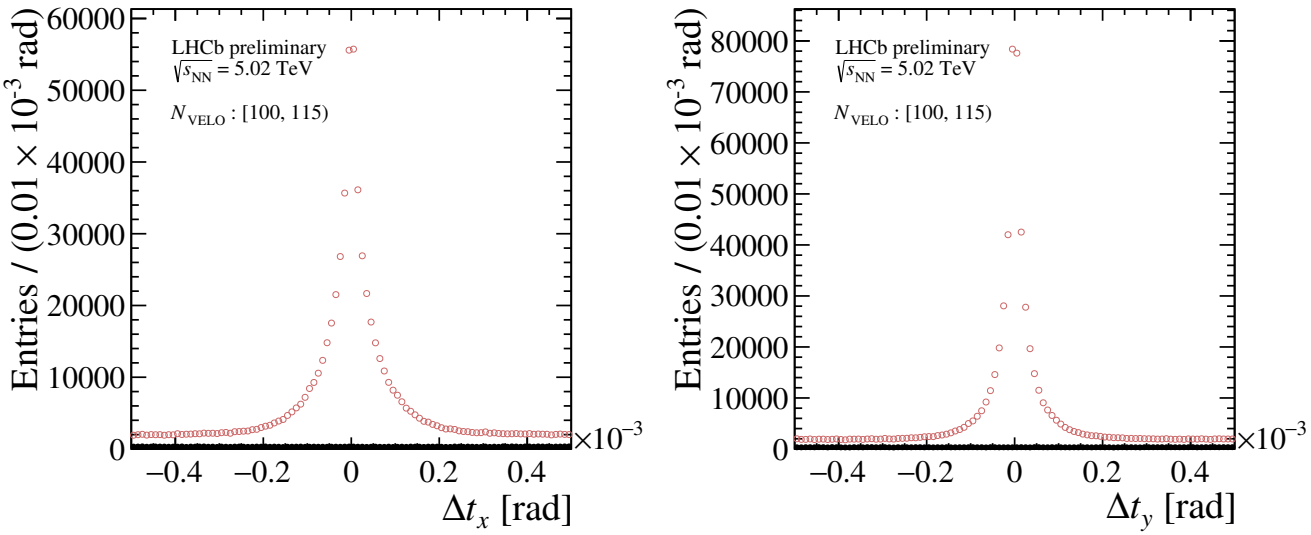

Figure 4.17: Example distribution of (left) Δt_x and (right) Δt_y for the signal pairs in a high-multiplicity region ($100 \leqslant N_{\mathrm{VELO}} < 115$) of the Pb$p$ data. The red circles correspond to the usage of only the preselection requirements, while the black dots are obtained after the final selection described in the text. The Δt_x (Δt_y) distribution is shown with an additional criterion that the Δt_y (Δt_x) value is lower than 0.5 mrad and the Q range is restricted to < 0.1 GeV (in order to illustrate the contribution of misreconstructed tracks in a region where it is most prominent).

Table 4.9: List of the final selection requirements in both the pp and the pPb studies. See the description in text for more details on the presented prerequisites.

selection	requirement
PV	visible PVs (both pp and pPb) single PV, $-160 < z_{\mathrm{PV}} < 60$ mm (pPb only)
track type	long, charged, no muon signature no shared VELO segments
track η	2.0–5.0
track χ^2	< 2.0
track p	> 2.0 GeV
track p_{T}	> 0.1 GeV
track IP	< 0.4 mm
$ProbNN$(ghost)	< 0.25
$ProbNN$(kaon,proton)	< 0.50
$ProbNN$(pion)	> 0.65
pair $\lvert\Delta t_x\rvert, \lvert\Delta t_y\rvert$	$\lvert\Delta t_x\rvert > 0.3$ mrad OR $\lvert\Delta t_y\rvert > 0.3$ mrad
pair Q	> 0.05 GeV

the analysis results is given in Sect. 4.4.3.

4.4.1 Analysis procedure

A general description of the methods used in the BEC studies is given in Sect. 4.1. Correlation functions for the OS and the SS pairs are determined in the Q range of 0.05–2.00 GeV with a bin width of 0.005 GeV. The limitation in the low-Q region is related to the high contribution of misreconstructed tracks below this regime and the related selection requirement described in Sect. 4.3.3. As detailed below in Sect. 4.4.1.4, the upper boundary is chosen based on a good description of the long-range correlations up to this limit in the simulation. In the current analysis for *pp* collisions, the reference sample in the correlation function is constructed using the event-mixing method (see Sect. 4.1.1). Reference pairs are created using particles originating from PVs from different events. The primary vertices are required to have the same VELO-track multiplicity to better reflect the properties of the Q distribution for the signal pairs.

Due to the relatively small size of the pion sources in *pp* collisions, a simplified correction for the Coulomb final-state interactions is applied using the Gamov factor for point-like sources (see Sects. 4.1.2 and 4.4.1.2). The double-ratio technique (see Sect. 4.1.3.1) is used to reduce effects related to the nonfemtoscopic background and specifics of the particular methods employed in the analysis (the optimisation of the selection requirements, the construction of the reference sample, *etc.*). Using this approach, the general parametrisation of the correlation function given in Eq. (4.7) can be simplified to

$$C_2(Q) = N\left(1 + \lambda e^{-|RQ|}\right) \times (1 + \delta Q) \ , \tag{4.9}$$

which corresponds to the Lévy-type correlation function with the index of stability α_{L} fixed to unity. The linear term with the δ parameter is introduced to account for the residual nonfemtoscopic background effects (mainly the long-range correlations, as the cluster contribution is reduced to a negligible level in this case – see Sect. 4.4.1.3). The correlation parameters are extracted by fitting the function given in Eq. (4.9) to the double-ratio distributions for the same-sign charged pion pairs.

4.4.1.1 Correlation functions for the data and the simulation

The correlation functions for the selected SS pion pairs in both the data and the simulation samples are shown in Figs. 4.18 and 4.19, respectively (no correction for the Coulomb interactions in the data is applied at this stage).

An enhancement due to the BEC signal is clearly visible in the correlation functions for the data in the low-Q region ($Q < 0.6$ GeV). Other structures are evident as well, such as a slight increase at higher Q values ($1.0 < Q < 2.0$ GeV), which is related to the contribution from the long-range correlations, and is visible in both the data and the simulation. For the simulation, no BEC signal is present as expected, however, an enhancement due to the cluster contribution can be seen in the low-Q region ($Q < 1.0$ GeV). This phenomenon is present in the data as well, although it is not visible there as clearly, due to the overlap with the BEC signal. Both nonfemtoscopic background contributions are significantly reduced in the final double-ratio distributions, which is controlled by studying the OS pairs (see Sect. 4.4.1.3 for details).

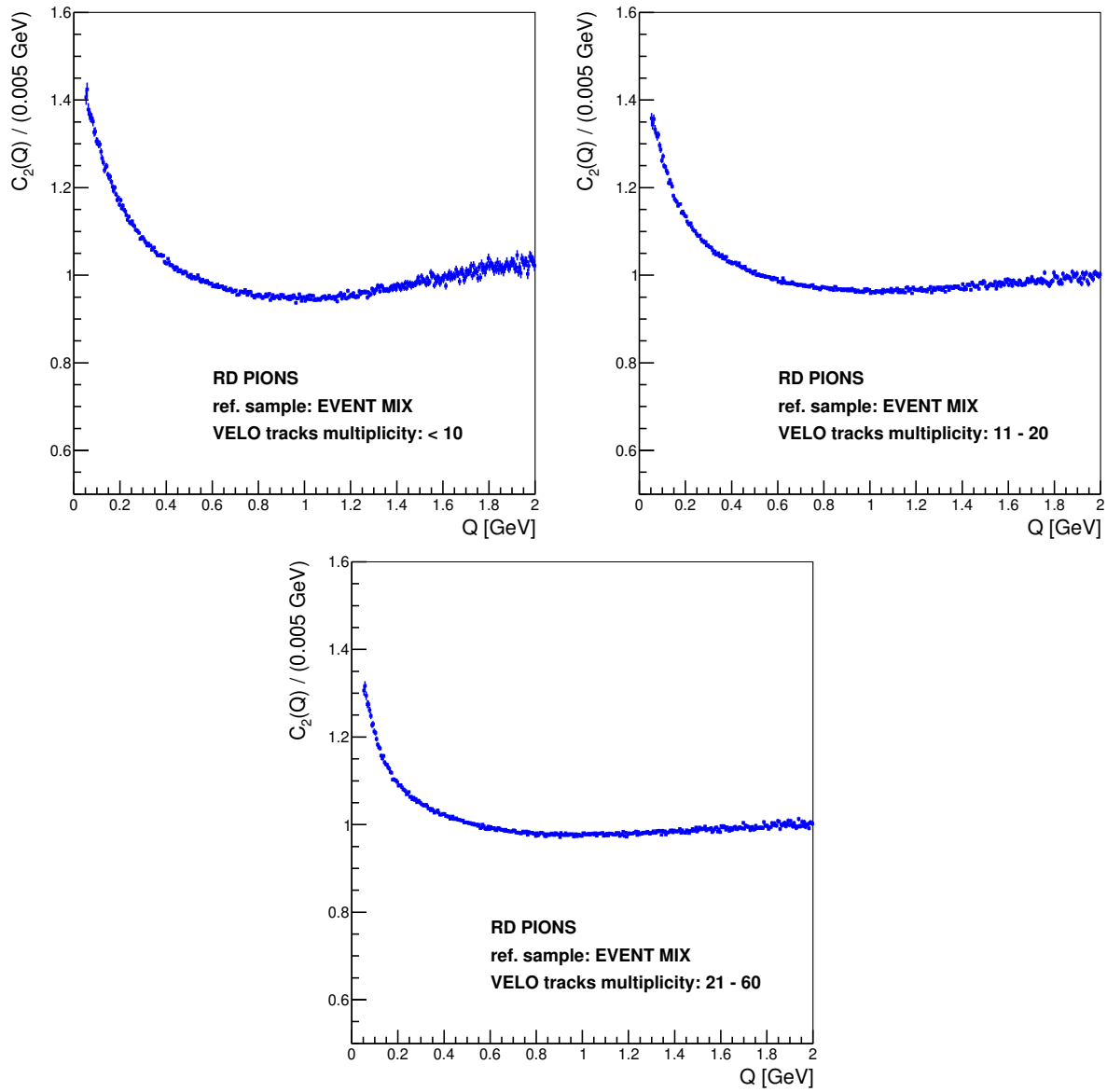

Figure 4.18: Correlation functions for the SS pairs in the three event-activity classes of the *pp* data sample (no correction for Coulomb interactions is applied at this stage – see Sect. 4.4.1.2).

4.4.1.2 Correction for the Coulomb interactions

The experimentally measured correlation functions are corrected for the Coulomb interactions by dividing them by the Gamov factor (see Eq. (4.4)) at the centre of the given bin of the Q variable. This procedure is applied to the correlation function for both the SS and the OS pion pairs in the data (the Coulomb interactions are not modelled in the simulation, so no correction is needed in this case)[3].

An example of the result of this correction is given in Fig. 4.20, where the double-ratio distribution for the opposite-sign charged pion pairs in the high-activity event class is shown. It is visible that the correction for Coulomb interactions affects only

[3]All of the correlation function and double-ratio distributions presented in the following sections on the *pp* study are shown – unless stated otherwise – with the correction for the Coulomb interactions applied.

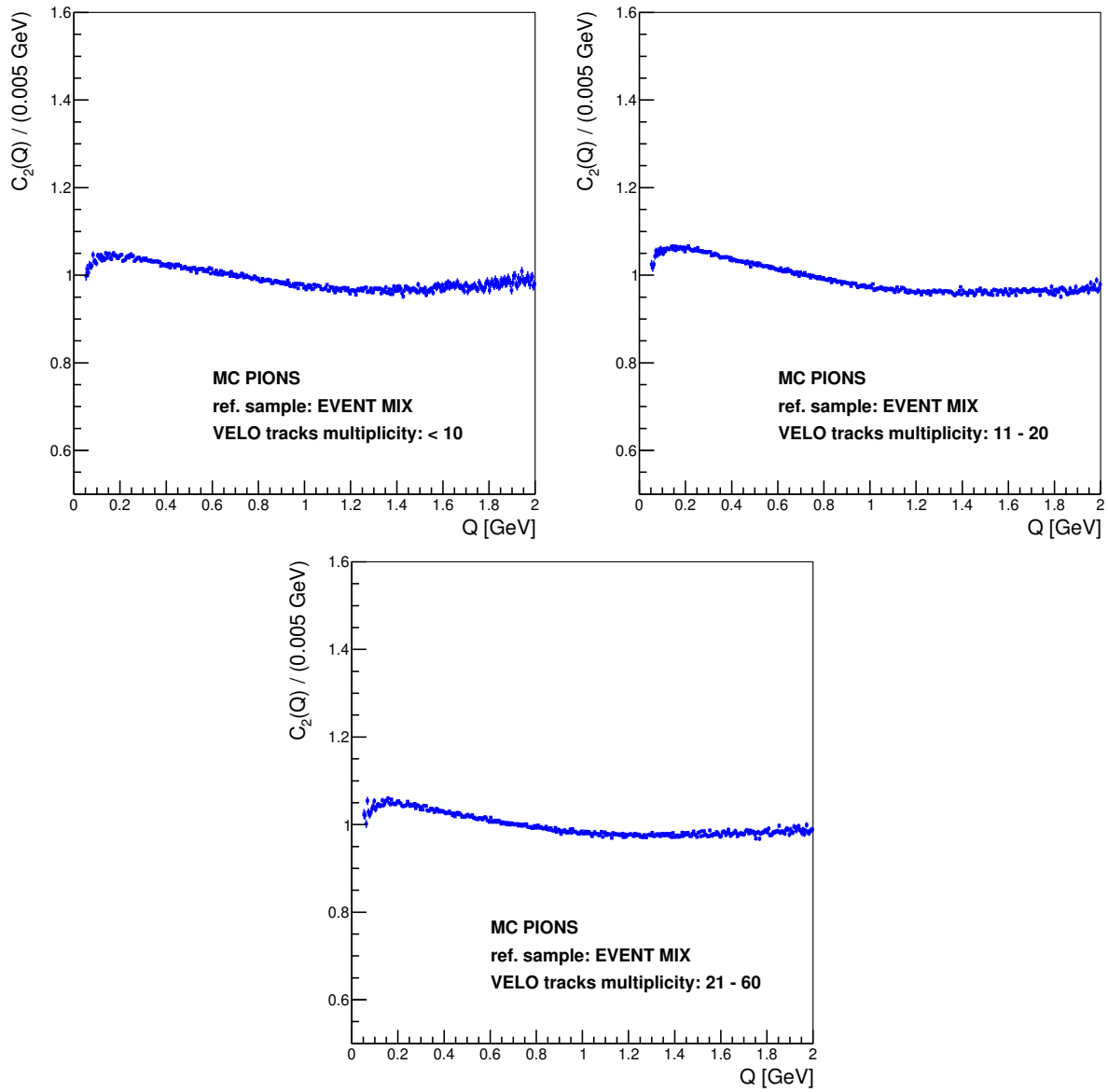

Figure 4.19: Correlation functions for the SS pairs in the three event-activity classes of the *pp* simulation sample.

the region of a very-low Q. After the correction, the double-ratio distribution in this regime ($Q < 0.2\,\mathrm{GeV}$) becomes nearly flat, which is an expected behaviour when the Coulomb interactions are properly corrected for (see Sect. 4.4.1.3 for a more detailed discussion of the OS double-ratio distributions). The Gamov factor provides only an approximated description of the Coulomb interactions, however, the observed effect suggests that this approach is sufficient in the case of pion pairs from *pp* collisions (due to the relatively small size of the particle-emitting source – see Sect. 4.1.2 for details). A potential impact of the inaccuracy of this method on the final results is taken into account in the systematic uncertainty studies described in Sect. 4.4.2.8.

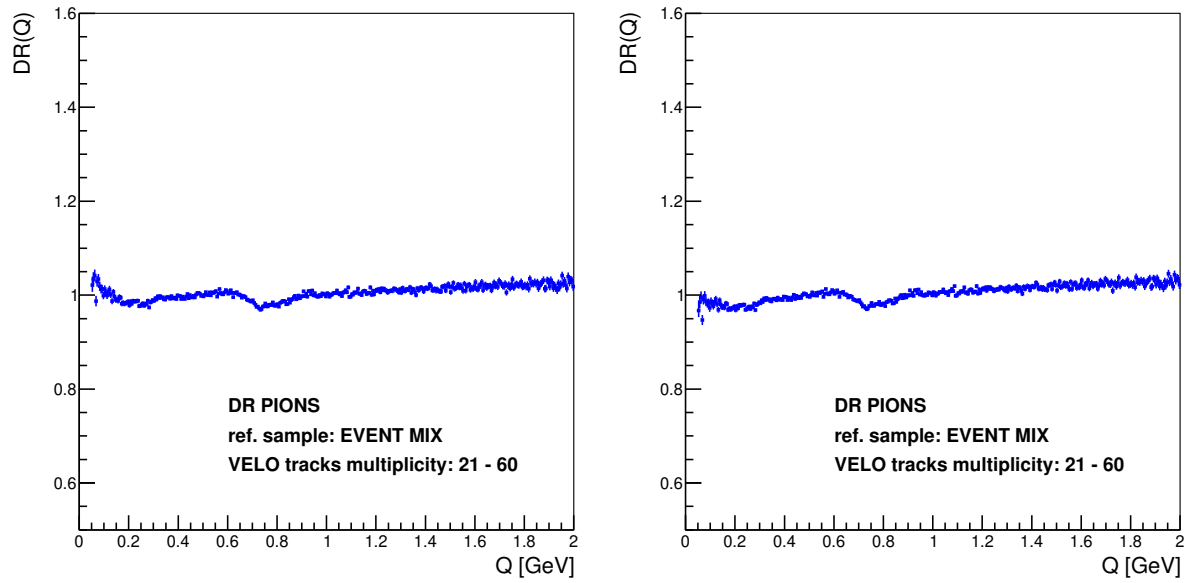

Figure 4.20: Double-ratio distribution for the OS pion pairs in the high-activity event class of the *pp* dataset (left) before and (right) after the correction for the Coulomb interactions using the Gamov factor.

4.4.1.3 OS pion pair sample as a control tool

The double-ratio distributions for the opposite-sign charged pion pairs (after the correction for the Coulomb interactions) are used as a natural tool to control the effects related to the nonfemtoscopic background (see Sect. 4.1.3.1). The constructed distributions (see Fig. 4.21) are approximately flat and close to unity (apart from the remaining structures emerging from resonance decays, which are not well reproduced in the simulation). The flat distribution for the OS pairs suggests that the nonfemtoscopic background effects seen in the correlation function for the data are properly modelled in the simulation and thus corrected for in the final double-ratio functions for both the same-sign and the opposite-sign charged pions. The potential discrepancy in the simulation of the two pair types by different Monte Carlo generators is accounted for in the systematic uncertainty studies (see Sect. 4.4.2.1).

4.4.1.4 Determination of the BEC parameters

Fits to the double-ratio distributions for the SS pion pairs using the parametrisation from Eq. (4.9) are performed in the full Q range (0.05–2.00 GeV) of the constructed correlation functions (see Sect. 4.4.1). As already mentioned, the lower range is limited due to the presence of misreconstructed tracks in the relevant region. In terms of the upper boundary, it is visible in the double-ratio functions for the SS pairs in Fig. 4.22 (but also for the OS ones in Fig. 4.21) that the obtained distributions are stable (nearly flat and close to unity) in the high-Q region. This indicates that the background related to the long-range correlations is well simulated in this range and the upper limit of the fit is chosen to be the maximum value of the Q variable in the constructed correlation functions. The agreement between the data and the simulation is observed to be worse

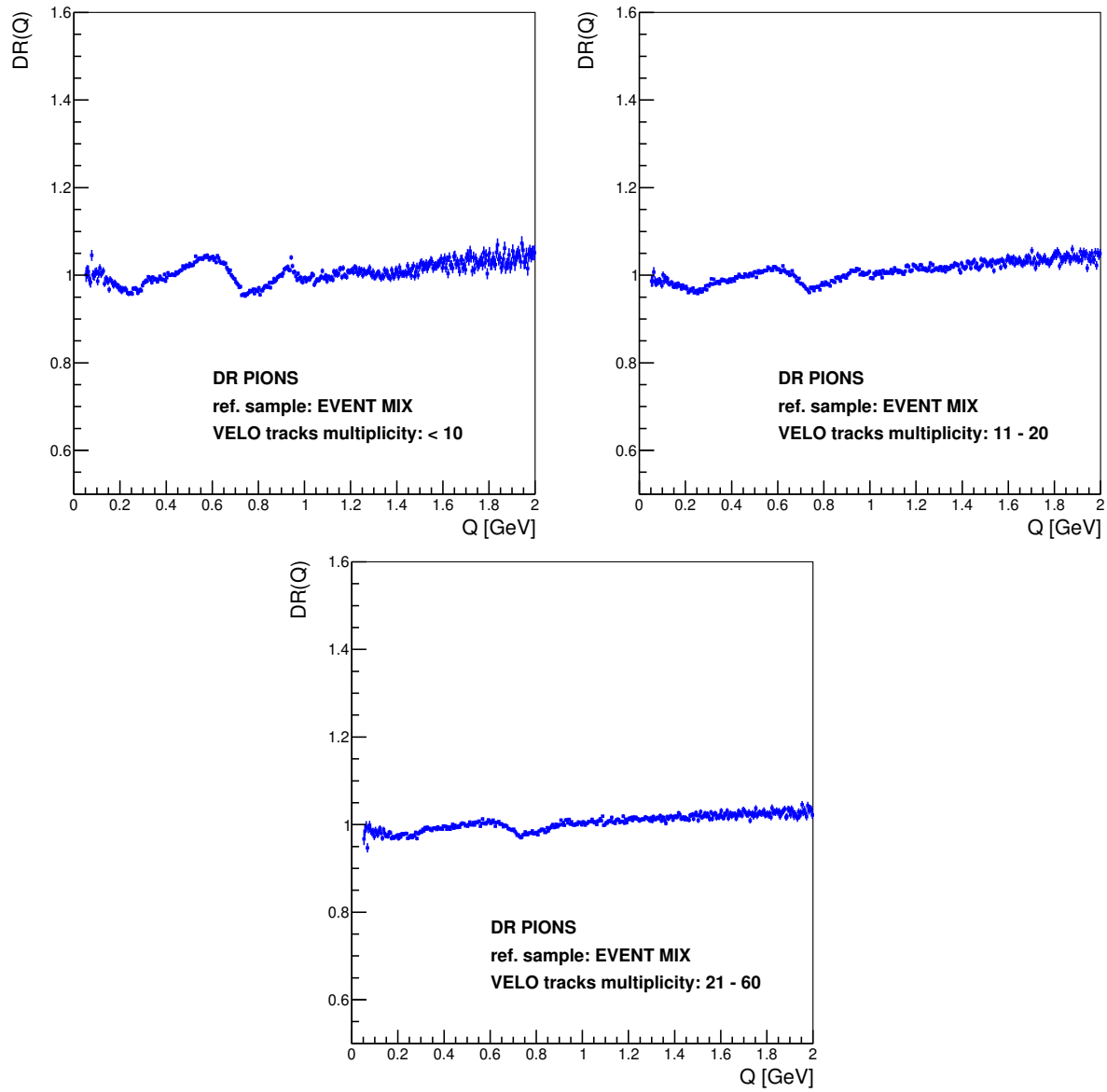

Figure 4.21: Double-ratio distributions for the selected OS pairs in the three event-activity classes of the *pp* dataset.

above $Q = 2\,\text{GeV}$. The impact of the selected fit boundaries on the final results is studied as a contribution to the systematic uncertainty in Sect. 4.4.2.9.

Results of the fits described above are shown in Fig. 4.22 and summarised in Tab. 4.10. An enhancement due to the BEC signal is clearly visible in the low-Q region for all of the activity classes. Only a slight increase is observed for the high Q-values, which is a residual effect related to the long-range correlations. This contribution is significantly reduced by employing the double-ratio method (compare to the correlation functions for the data in Fig. 4.18), which is also reflected in the values of the δ parameter that are close to zero. The χ^2/ndf values of the fits are relatively high (up to 1.6), which is a known effect in the BEC analyses related to the various assumptions in the signal parametrisation (see Sect. 2.1.3), together with unknown background effects, as there is no specific model for its proper description (see Sect. 4.1.4 for details). Respective pull

Table 4.10: Results of the fits to the double-ratio distributions for the signal pairs in the three activity classes of the *pp* dataset, using the parametrisation from Eq. (4.9). Only the statistical uncertainties of the fit parameters are shown.

activity class	R [fm]	λ	δ [GeV^{-1}]	N	χ^2/ndf
low	1.01 ± 0.01	0.72 ± 0.01	0.09 ± (<0.01)	0.89 ± (<0.01)	591/386
medium	1.48 ± 0.02	0.63 ± 0.01	0.05 ± (<0.01)	0.94 ± (<0.01)	623/386
high	1.80 ± 0.03	0.57 ± 0.01	0.03 ± (<0.01)	0.97 ± (<0.01)	621/386

distributions[4] correspond to values below or around three in the overall range of the fits. A summary of the measured correlation parameters R and λ is given in Sect. 4.4.3.

4.4.2 Systematic uncertainty

The correlation function distributions can be influenced by several different effects, thus impacting the final fit results. It includes various acceptance effects, imperfections of the chosen analysis method (*e.g.* due to the reference-sample construction and the selection requirements) or specifics of the fitting procedure. Many of these contributions, if properly modelled in the simulation, can be significantly reduced (often to a negligible level) employing the double-ratio method. However, in return, the double-ratio function is constructed using the simulation and it depends on the chosen Monte Carlo generator, which can be one of the major contributions to the overall systematic uncertainty. Also, some of the effects are not described very well (or not at all, like the Coulomb interactions) in the simulation and they require a specific treatment to be accounted for (*e.g.* as in the case of the clone and the fake tracks – see Sect. 4.3.3).

Various sources of the systematic uncertainty are described below. A general strategy is to perform the fitting procedure with relevant changes in different aspects of the analysis (*e.g.* introducing a difference in the selection requirements). The systematic uncertainty contribution from the given source is then calculated as an absolute difference between the central values of this new specific fit result and the central result (shown in Tab. 4.10). A summary of all of the nonnegligible sources of the systematic uncertainty is given in Tab. 4.11. Additional studies are performed to investigate a potential impact of beam-gas interactions as well as of residual acceptance effects in the double-ratio distributions. These contributions are found to be negligible.

4.4.2.1 Monte Carlo generator tuning

To study the effect of choosing a particular Monte Carlo generator, the double-ratio distributions are reconstructed employing the simulation sample produced using Pythia6.4 with the Perugia0 tuning (see Sect. 4.2.1 for details). This particular tuning is chosen for the systematic uncertainty studies due to its reasonable modelling of the long-range correlations. The results of the fits to the double-ratio distributions using Pythia6.4 with the Perugia0 tuning for the simulation are shown in Fig. 4.23. The choice of the Monte

[4]The values in a pull distribution are determined as a difference between the data point and the fit value in the given bin, divided by the statistical uncertainty of the data point. The pull values of ±3 are indicated in the pull distributions presented in this thesis for reference.

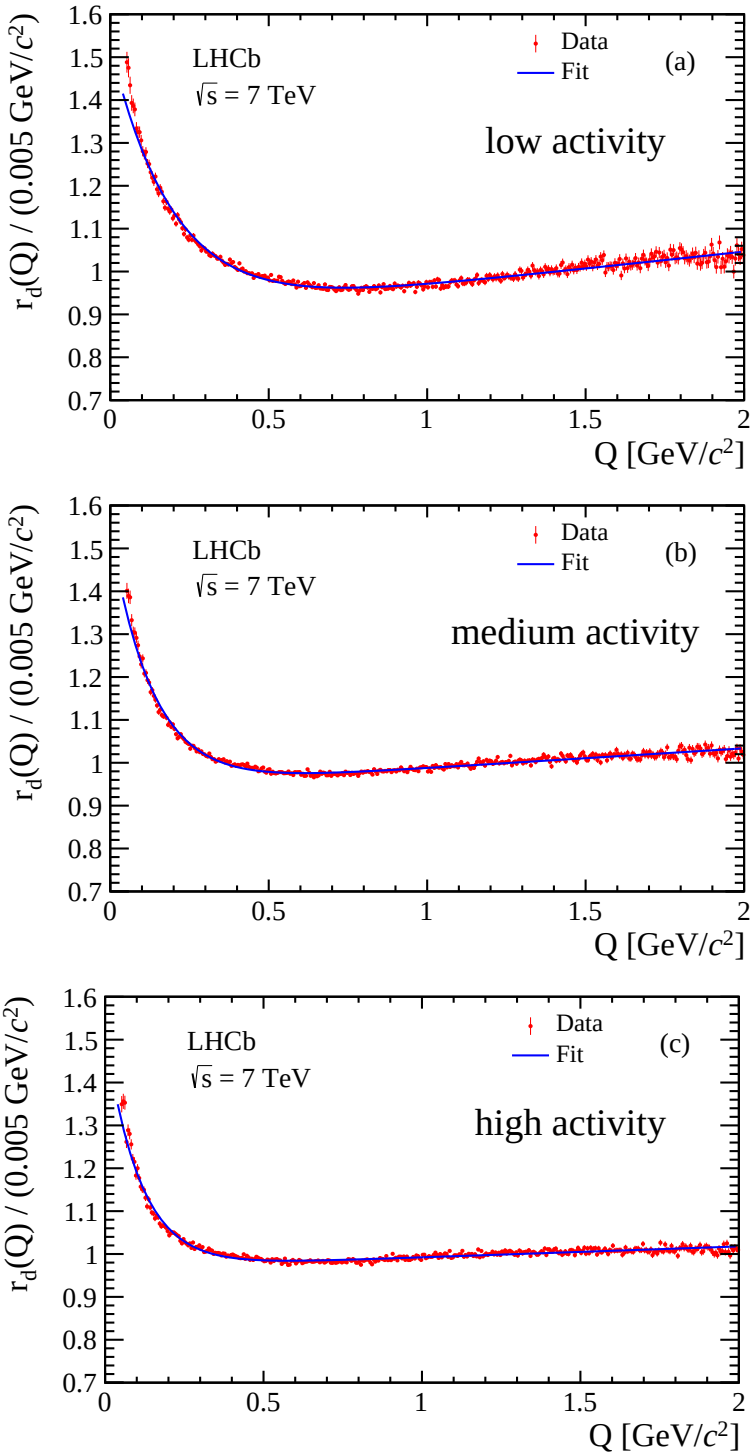

Figure 4.22: Results of the fits to the double-ratio distributions for the signal pairs in the three event-activity classes of the *pp* dataset: (a) low, (b) medium and (c) high. The blue solid lines illustrate the fit results using the parametrisation from Eq. (4.9). Only statistical uncertainties are shown for the data points [17].

Table 4.11: Individual contributions to the systematic uncertainty on the R and the λ parameters for the three activity classes in the *pp* study. Negligible contributions are not listed. The total uncertainty is a quadratic sum of the individual inputs [17].

	low activity		**medium activity**		**high activity**	
contribution	$\Delta R\,[\%]$	$\Delta\lambda\,[\%]$	$\Delta R\,[\%]$	$\Delta\lambda\,[\%]$	$\Delta R\,[\%]$	$\Delta\lambda\,[\%]$
generator tunings	6.6	4.3	8.9	3.5	6.5	1.5
PV multiplicity	5.9	5.8	6.1	4.5	3.9	4.3
PV reconstruction	1.8	0.1	1.4	1.2	0.1	<0.1
fake tracks	0.4	1.1	1.7	3.9	1.1	0.8
PID calibration	1.3	0.3	0.8	0.6	2.7	0.9
requirement on pion PID	2.9	1.8	1.6	0.1	1.3	0.1
fit range at low Q	1.2	1.0	1.2	1.5	1.8	2.7
fit range at high Q	1.8	0.1	2.1	0.8	2.4	1.4
total	9.8	7.6	11.4	7.3	8.8	5.6

Carlo generator tuning is found to be the dominant contribution to the systematic uncertainty of this study.

The sample simulated with the HERWIG++ generator is also employed to construct the double-ratio distributions. In this case however, as opposed to the PYTHIA6.4 simulation with the PERUGIA0 tuning, the long-range correlations are far from being properly reproduced, resulting in an uncontrolled behaviour of the double-ratio distributions in the high-Q region (see Fig. 4.24). Therefore, this sample cannot be used in the determination of the systematic uncertainty related to the Monte Carlo generators. However, studying the double-ratio distributions constructed using the simulation with HERWIG++ allows us to recognise the importance of the proper modelling of the nonfemtoscopic background effects in the simulation when employing the double-ratio method.

4.4.2.2 Number of primary vertices in an event

It is found that the number of primary vertices in an event can affect the shape of the double-ratio distribution. This is related to the imperfections of the reference sample construction, which can preserve residual correlations between the primary vertices within a single event in the correlation function. The double-ratio distributions for different number of PVs in the event are shown in Fig. 4.25. This contribution to the systematic uncertainty is calculated as a difference between the central values of the fit results for the two extreme cases: (*i*) a single PV and (*ii*) three or four PVs in the event. This effect is found to be the second-largest one in the overall systematic uncertainty.

4.4.2.3 Primary vertex reconstruction

The BEC signal originates from pairs of same-sign charged pions assigned to the same PV. Therefore, any inefficiencies of the PV reconstruction (which is also related to the presence of multiple PVs in an event) can affect the shape of the double-ratio distributions. To study the impact of this inefficiency, the PV reconstruction using the standard LHCb algorithms [173] is performed with 10% of the initially reconstructed tracks ran-

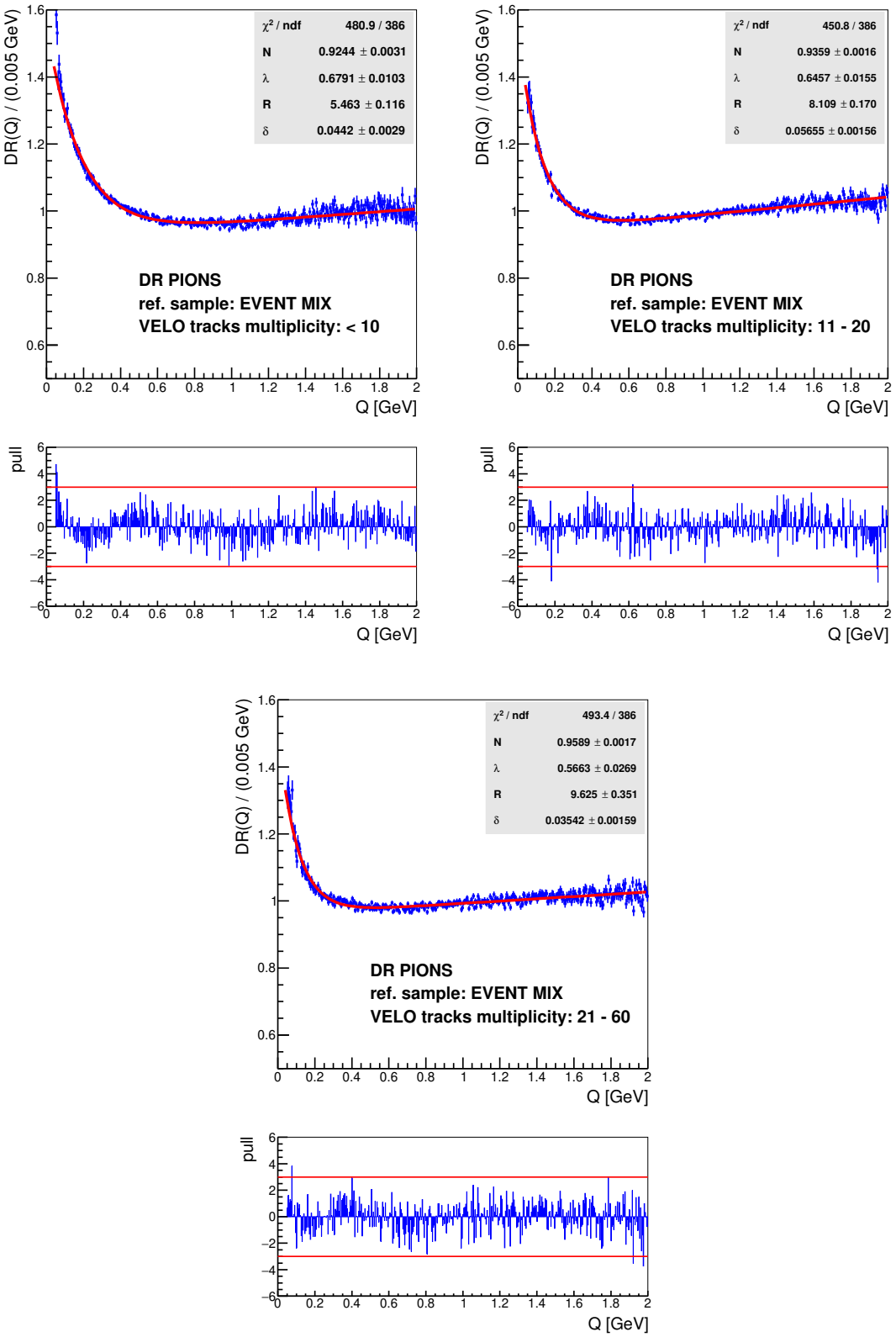

Figure 4.23: Results of the fits to the double-ratio distributions constructed employing the simulation sample produced using PYTHIA6.4 with the PERUGIA0 tuning for the signal pairs in the three event-activity classes of the *pp* dataset: (top left) low, (top right) medium and (bottom) high. The red solid lines illustrate the fit results using the parametrisation from Eq. (4.9). Pull distributions of the fits are given below the relevant figures. Only the statistical uncertainties are shown (both for the data points and for the fit results). The R values are given in GeV^{-1}.

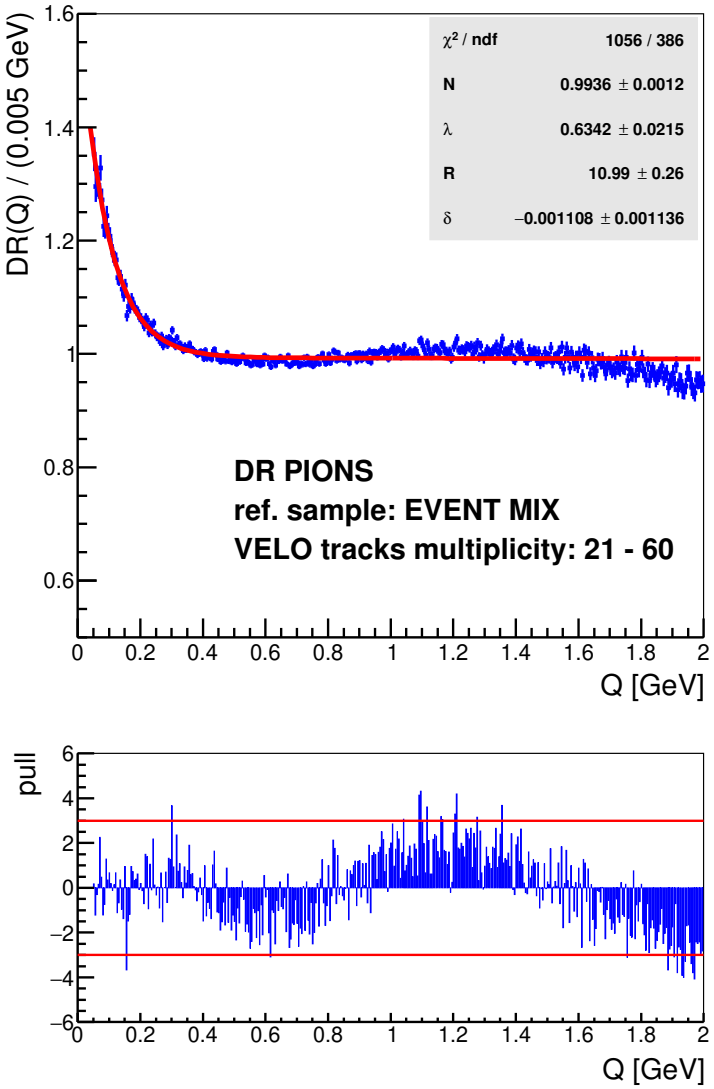

Figure 4.24: Example of a fit to the double-ratio distribution constructed employing the simulation sample produced using HERWIG++ for the signal pairs in the high-activity class of the *pp* dataset. The red solid line illustrates the fit result using the parametrisation from Eq. (4.9). A pull distribution of the fit is given below the main figure. Only the statistical uncertainties are shown (both for the data points and for the fit results). The R value is given in GeV^{-1}.

domly removed. Then, for such a sample, the fitting procedure is repeated to determine the systematic uncertainty.

4.4.2.4 Fake tracks

The presence of the fake tracks in the sample is strongly reduced by the dedicated selection requirements described in Sect. 4.3.3. Using the simulation, the fraction of the remaining fake tracks in the selected sample is found to be close to the value of 1% for all of the activity classes (see Tab. 4.12).

However, the fake tracks fraction from the data sample may not be perfectly re-

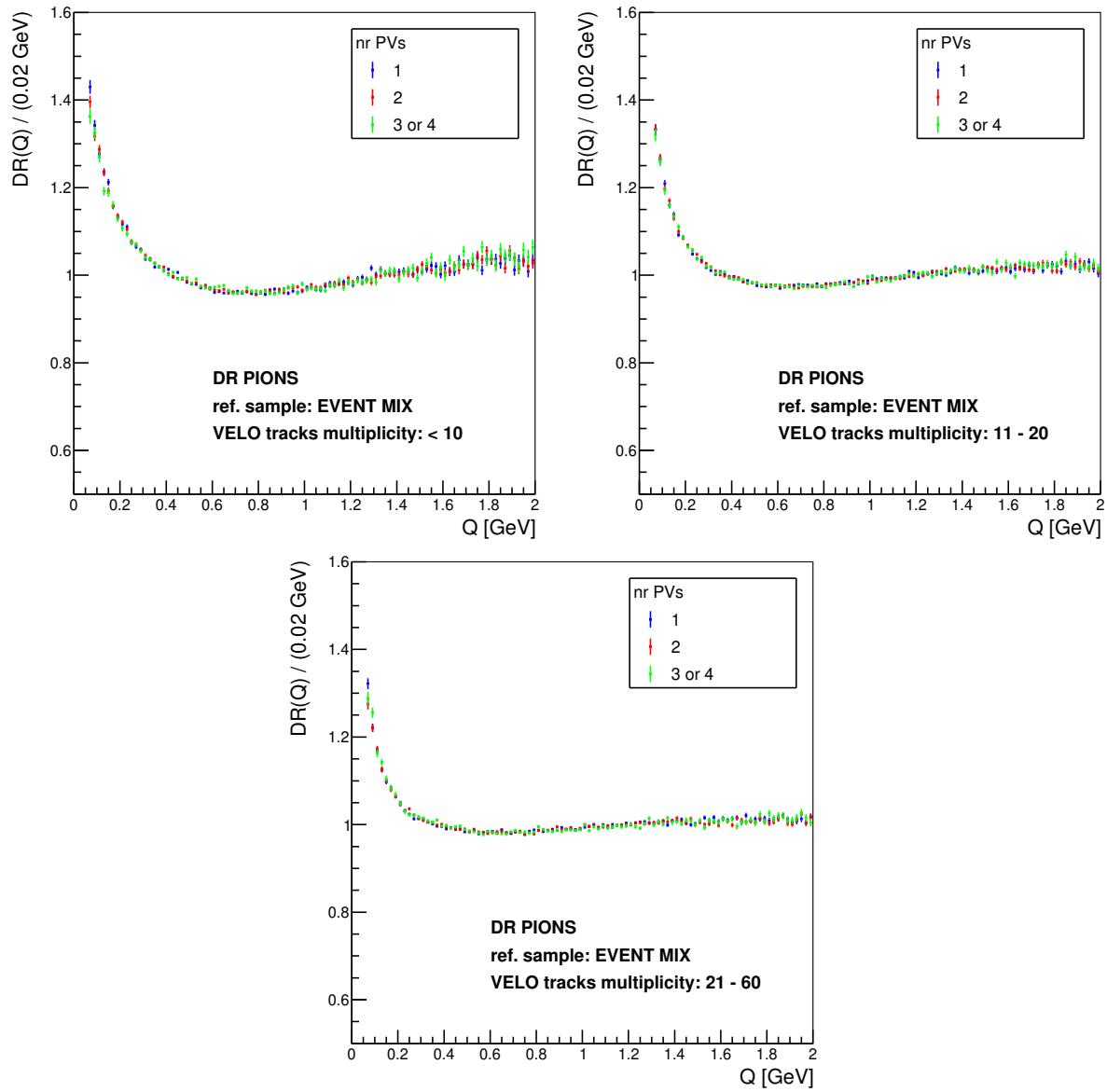

Figure 4.25: Double-ratio distributions for the signal pairs in the three event-activity classes of the *pp* dataset: (top left) low, (top right) medium and (bottom) high. The data points are obtained using the events with one (blue), two (red) and three or four (green) primary vertices.

produced in the simulation. Therefore, an additional study is performed with a very loose selection requirement on the fake tracks identification $ProbNN(\text{ghost}) < 0.50$ (instead of the $ProbNN(\text{ghost}) < 0.25$ value used in the final selection). The fake tracks fractions measured for these two requirements are shown in Tab. 4.12. Results of the fits to the double-ratio distributions constructed with the requirement of $ProbNN(\text{ghost}) < 0.50$ are used to determine the fake tracks contribution to the systematic uncertainty.

4.4.2.5 Clone tracks

Similar to the fake tracks, the clone tracks contribution is significantly suppressed and controlled using the methods reported in Sect. 4.3.3. The fraction of the selected signal

Table 4.12: Fractions of the fake tracks (using the simulation information) in the selected pion sample for the three activity classes in the *pp* dataset. The values are shown for two different selection requirements on the fake track identification.

	activity class		
$ProbNN(\text{ghost})$	low	medium	high
< 0.25	0.66%	0.88%	1.20%
< 0.50	0.75%	1.01%	1.37%

Table 4.13: Fractions of the pairs containing a clone track (using identification based on the simulation information) in the signal pairs for the three activity classes of the *pp* dataset.

activity class	low	medium	high
clone pairs fraction [%]	0.25	0.47	1.01

pairs containing a clone track is found using the simulation and is shown in Tab. 4.13. This contribution amounts to $\lesssim 1\%$ in all three activity classes.

The distribution of the Kullback–Leibler distance for tracks in the data sample and in the simulation is shown in Fig. 4.26. To study the clone tracks contribution to the systematic uncertainty, the analysis is performed with a more strict requirement on the Kullback–Leibler distance than in the standard LHCb algorithm for track reconstruction (see Sect. 4.3.3). The value of the requirement on the natural logarithm of the Kullback–Leibler distance is imposed to be higher than 12 in order to remove nearly all of track pairs containing a clone track, which is based on the Kullback–Leibler distance distribution shown in Fig. 4.26. In the original method, the value of $\log(5000) \approx 8.52$ is used. This contribution is found to be negligible.

4.4.2.6 PID variables calibration

Distributions of the variables used for the particle identification are not properly reproduced in the simulation. Thus, these distributions are resampled using the PIDCalib package. As described in Sect. 4.3.2.1, the corrected PID variable is assigned based on the distribution available in the PID calibration sample. This sample is divided into bins based on the particle momenta, pseudorapidity and the number of VELO tracks in an event. In this process, if for a given particle the corresponding bin in the calibration sample is empty, a distribution integrated over the number of VELO tracks in an event is used, to be able to perform the resampling. To study the systematic uncertainty related to this procedure, the analysis is repeated with the particles pointing to the empty phase-space bin removed from the simulation sample.

4.4.2.7 Optimisation of the pion identification

Although the impact of the selection requirements on the analysis results is limited due to the use of the double-ratio function, the purity of the final pion sample can affect the measurement. This effect is studied by using a less stringent requirement on the pion

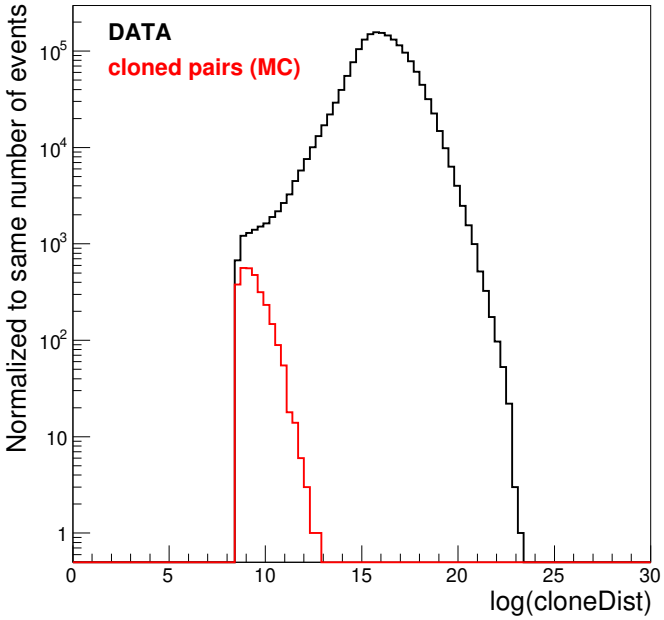

Figure 4.26: Natural logarithm of the Kullback–Leibler distance for all track pairs in the pp data (black) and only for pairs of clone tracks in the simulation (red).

identification ($ProbNN(\text{pion}) > 0.35$), which corresponds to increasing the fraction of the misidentified pions in the sample by 50%.

4.4.2.8 Correction for the Coulomb interactions

The simplified Gamov correction procedure applied to the correlation function for the data to correct for the Coulomb interactions is sufficient in the case of pion pairs originating from pp collisions (see Sects. 4.1.2 and 4.4.1.2). The impact of the potential inaccuracy of this method on the final results is studied by varying the applied correction factor by ±20%. It is found to be negligible.

4.4.2.9 Fit range and binning

In the present analysis, fits to the double-ratio distributions are performed within the Q-variable range of 0.05–2.00 GeV. The choice of these boundaries can affect the analysis results. To study this contribution to the systematic uncertainty, the fits are repeated with the lower limit changed by ±20% (which corresponds to 0.04 and 0.06 GeV), and the upper one by ±10% (which corresponds to 1.80 and 2.20 GeV). This check is performed with modifying only one of the boundaries at a time (separately for the lower and the upper one).

An additional investigation is carried out to estimate the impact of the bin width of the Q variable in the correlation functions on the final results. For this purpose, the analysis is repeated with the bin width changed from the nominal 0.005 GeV to 0.003 GeV. The contribution of this effect to the systematic uncertainty is found to be negligible.

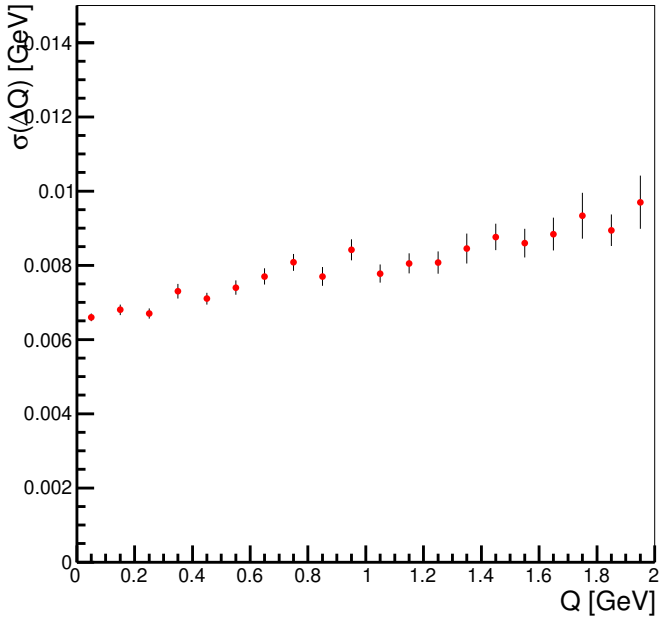

Figure 4.27: Resolution of the Q variable in the full range of Q used in the study.

4.4.2.10 Resolution of the Q variable

The resolution of the Q variable in the range of 0–2 GeV is shown in Fig. 4.27. It varies between 6–10 MeV within the considered Q-values. The systematic uncertainty related to the resolution of this variable is studied by smearing the Q values using a Gaussian function with a standard deviation corresponding to the determined resolution. This contribution is found to be negligible.

4.4.3 Summary

Results of the BEC analysis for same-sign charged pion pairs from pp collisions in the LHCb experiment are summarised in this section. The correlation parameters that are determined from the fits to the double-ratio distributions (see Sect. 4.4.1) in the three activity classes are given in Tab. 4.14, including the systematic uncertainties from the study described in Sect. 4.4.2. The ranges of the unfolded charged-particle multiplicity corresponding to the said activity classes (see Sect. 4.2.1) are also indicated. The systematic uncertainties dominate the statistical ones and are at the level of $\sim 10\%$ for the correlation radius (slightly less for the λ parameter) in all three activity classes. The major contributions to the systematic uncertainty are related to the choice of the Monte Carlo generator tuning and the effects originating from the presence of multiple primary vertices within individual events. The observed values of the correlation parameters as a function of the unfolded charged-particle multiplicity are shown in Fig. 4.28. This analysis has been published by the LHCb experiment in Ref. [17]. A discussion and interpretation of the presented results is given in Chap. 5.

Table 4.14: Results of the fits to the double-ratio distributions for the SS pion pairs using the parametrisation from Eq. (4.9) in the three event-activity classes of the *pp* dataset. Ranges of the unfolded charged-particle multiplicity corresponding to each of the classes are also indicated. The statistical and the systematic uncertainties are given separately, in this order [17].

activity class	N_{ch}	R [fm]	λ	δ [GeV^{-1}]
low	[8,18]	1.01 ± 0.01 ± 0.10	0.72 ± 0.01 ± 0.05	0.089 ± 0.002 ± 0.044
medium	[19,35]	1.48 ± 0.02 ± 0.17	0.63 ± 0.01 ± 0.05	0.049 ± 0.001 ± 0.009
high	[36,96]	1.80 ± 0.03 ± 0.16	0.57 ± 0.01 ± 0.03	0.026 ± 0.001 ± 0.010

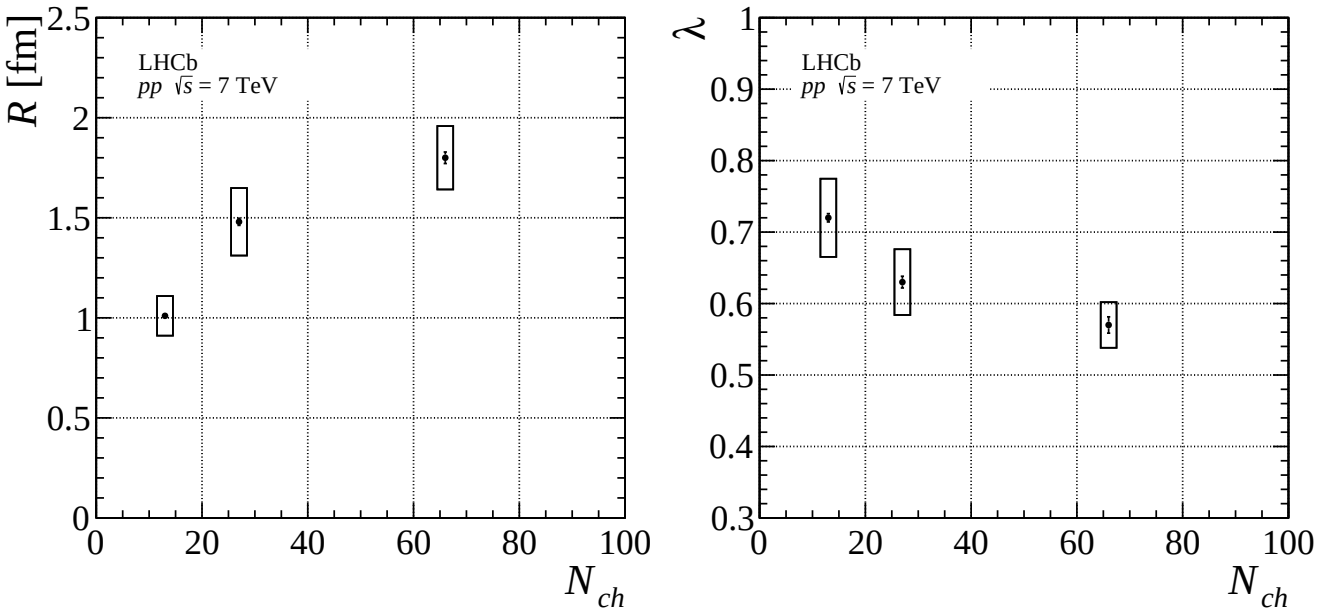

Figure 4.28: (left) Correlation radius R and (right) intercept parameter λ as a function of the unfolded charged-particle multiplicity in the *pp* study. Error bars indicate the statistical uncertainties, while boxes correspond to the systematic ones. The data points are positioned at the centres of the N_{ch} bins [17].

4.5 Study of the BEC effect in *p*Pb collisions

The analysis of the Bose–Einstein correlations for same-sign charged pions from *p*Pb collisions recorded by the LHCb experiment is described in this section (the presented results are preliminary). Details on the analysis method, data sets and selection requirements are given in Sects. 4.1, 4.2, 4.3. The main analysis procedure is presented in Sect. 4.5.1. It is followed by a description of the study of systematic uncertainties in Sect. 4.5.2. A summary of the analysis results is given in Sect. 4.5.3.

4.5.1 Analysis procedure

Correlation functions for both the SS and the OS pairs are constructed for the Q values between 0.05–2.00 GeV with a bin width of 0.005 GeV. This particular choice enhances consistency with the study performed for *pp* collisions (see more details in Sect. 4.4.1)

and allows for a direct comparison of the results between the two analyses. Clearly, the criterion of a stable double-ratio distribution in the high-Q regime does not apply here, but an extended range of the correlation function is important to obtain information on the nonfemtoscopic background far from the BEC-signal region. The reference sample for pPb correlation function is constructed using the event-mixing method. Particles in the reference pairs are required to originate from PVs with comparable VELO-track multiplicity (candidate particles for pairs are grouped into multiplicity bins of width of three) to ensure that the pair is created using a similar class of events. Furthermore, the two PVs associated with particles in the candidate pair must be located in a similar region along the z-axis in the LHCb coordinate system (a tolerance of 10 mm is imposed). This requirement provides consistent effects related to the detector acceptance for both particles in a pair.

Correlation function for the SS pairs is studied by means of Eq. (4.7) using the Bowler–Sinyukov formalism and the Lévy-type parametrisation for the BEC effect with index of stability fixed to unity, which leads to

$$C_2(Q) = N\left[1 - \lambda + \lambda K(Q) \times \left(1 + e^{-|RQ|}\right)\right] \times \Omega(Q) \ , \tag{4.10}$$

where the term $K(Q)$ representing the Coulomb interactions is given by Eq. (4.5). The effective radius in Eq. (4.5) is set to $R_{\text{eff}} = 2\,\text{fm}$ in the analysis, which is a value based on the expected order of magnitude of the measured correlation radii. The latter can be estimated *e.g.* by performing a simplified analysis using the Gamov factor (see Sect. 4.1.2) to correct for the Coulomb interactions. The effect of selecting this particular value of the R_{eff} parameter is taken into account in the study of systematic uncertainties in Sect. 4.5.2.3.

The analysis is performed based on the cluster subtraction method (see Sect. 4.1.3.2) to account for the nonfemtoscopic background effects (denoted by $\Omega(Q)$ in Eq. (4.10)). In this method, the nonfemtoscopic background description is found using the correlation functions for the OS pairs. The resulting contribution is then scaled and fixed in the final fits to the SS correlation functions. Details on this analysis procedure are provided in the following sections. All fits to the correlation functions are performed employing the negative log-likelihood method (see Sect. 4.1.4).

4.5.1.1 Correlation functions for OS and SS pairs

Examples of the correlation function for both the OS and the SS pion pairs in the pPb data are shown in Fig. 4.29. Various effects and distortions of the correlation function shape (see Sect. 4.1 for details) are visible. An increase due to the long-range correlations is prominent in the high-Q region ($1.0 < Q < 2.0\,\text{GeV}$) for both the OS and the SS pairs. Similarly, the cluster contribution is manifested in both cases for lower Q values ($Q < 1.0\,\text{GeV}$). However, for the SS pairs it is blended with the visible enhancement due to the BEC signal ($Q < 0.6\,\text{GeV}$). In the distribution for OS pairs, structures emerging due to two-body resonance decays to $\pi^+\pi^-$ are clearly visible.

The presence of structures related to resonance decays in the OS correlation functions degrades the quality of the fit from which the nonfemtoscopic background parameters are determined. Therefore, the affected regions are removed from the fit to the correlation function – see Tab. 4.15 for a list of visible resonances and the excluded Q ranges. The removed regions are optimised to provide a good quality of fits to the correlation functions for OS pairs and the choice of particular boundary values is taken into account

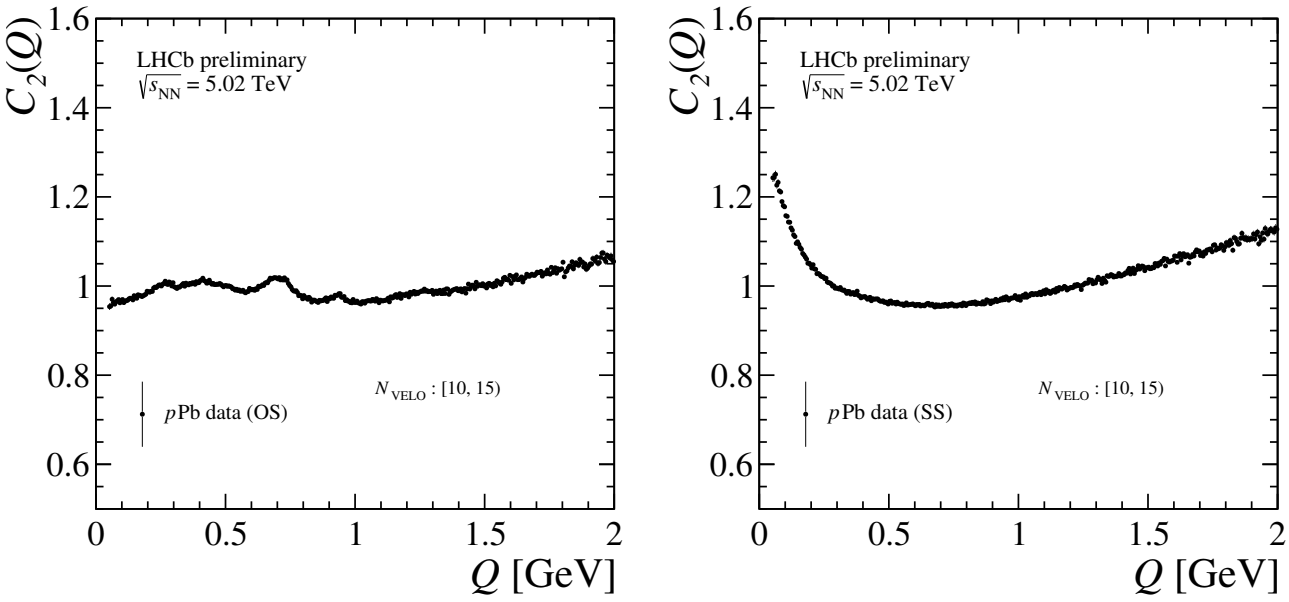

Figure 4.29: Example of a correlation function for (left) OS and (right) SS pairs in the pPb dataset. The presented distributions correspond to a low-multiplicity region of $10 \leqslant N_{\mathrm{VELO}} < 15$ (see Tab. 4.2).

Table 4.15: List of resonances decaying to $\pi^+\pi^-$ visible in the OS correlation functions and the Q regions removed from the latter in the pPb study. Masses of the resonances are given in MeV.

resonance	Q range [GeV]
$\rho^0(770)$	0.55–0.88
$K^0_{\mathrm{S}}(497)$	0.38–0.44
$f_0(980)$	0.91–0.97
$f_2(1270)$	1.21–1.27

in the study of the systematic uncertainty (see Sect. 4.5.2.1). The effect of the resonances removal is illustrated in Fig. 4.30 (left). It is worth noting that the impact of resonances is most prominent in the bins with low N_{VELO} values. The observed structures quickly diminish with increasing multiplicity due to a prevailing contribution from pairs of unrelated particles (see Fig. 4.30 (right)).

4.5.1.2 Description of the nonfemtoscopic background

As mentioned in Sect. 4.1.3.2, the OS correlation functions contain effects similar to the SS ones (apart from the BEC signal) and can be used to investigate the nonfemtoscopic background contribution. Since no BEC effect is present, Eq. (4.10) for the OS pairs is simplified to

$$C_2^{\mathrm{OS}}(Q) = N \times K(Q) \times \Omega(Q) \ . \tag{4.11}$$

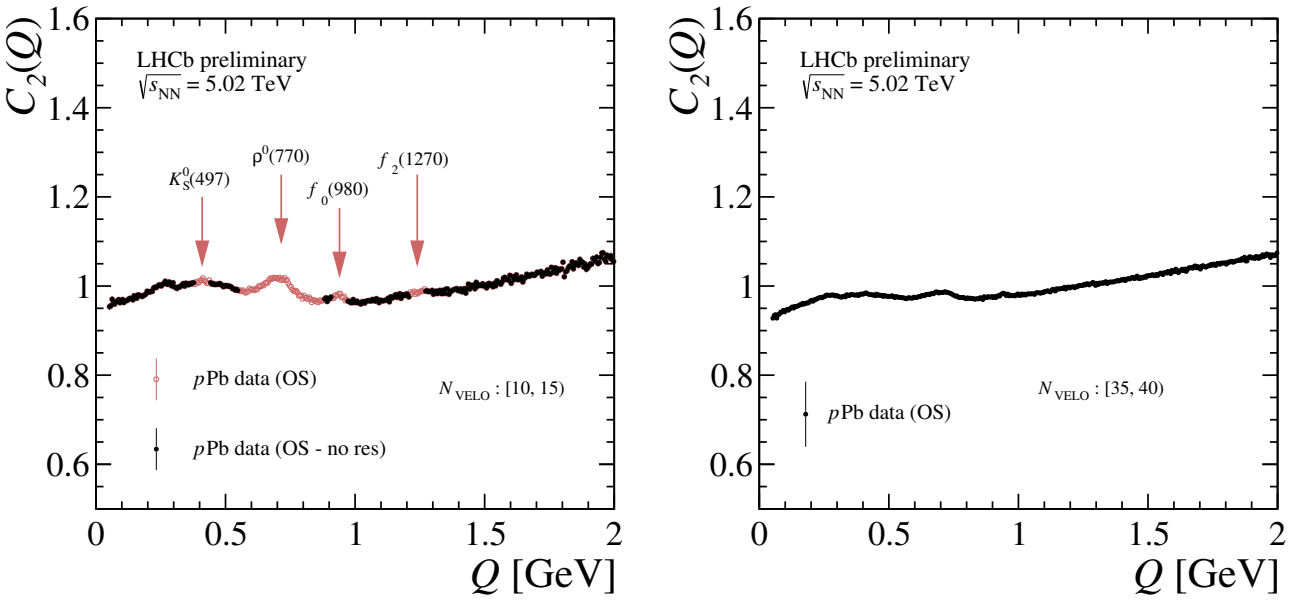

Figure 4.30: Example of a correlation function for the OS pairs in the pPb data illustrating (left) the removal of resonances and (right) the diminishing of the structures related to resonances with increasing multiplicity ($35 \leqslant N_{\mathrm{VELO}} < 40$ compared to $10 \leqslant N_{\mathrm{VELO}} < 15$ in the left figure). The left figure presents the OS correlation function before (red points) and after (black points) the Q ranges affected by resonances are removed. Regions affected by the given type of resonance are indicated by arrows.

A satisfactory description of the data is found using a simple Gaussian parametrisation for the cluster contribution [134], which gives

$$\Omega(Q) = (1 + \delta Q) \times \left[1 + z \frac{A_{\mathrm{bkg}}}{\sigma_{\mathrm{bkg}}\sqrt{2\pi}} \exp\left(-\frac{Q^2}{2\sigma^2_{\mathrm{bkg}}}\right)\right] , \tag{4.12}$$

where the linear term with δ corresponds to the long-range correlations (see Sect. 4.1.3) and the z parameter (fixed to unity in the OS fits) is a factor used for the background scaling between OS and SS pairs (see Sect. 4.5.1.3). The width σ_{bkg} and the amplitude A_{bkg} are multiplicity-dependent values that characterise the cluster contribution and are parametrised as

$$\sigma_{\mathrm{bkg}}\left(N_{\mathrm{VELO}}\right) = \sigma_0 + \sigma_1 \exp\left(-\frac{N_{\mathrm{VELO}}}{N_0}\right) , \tag{4.13}$$

$$A_{\mathrm{bkg}}\left(N_{\mathrm{VELO}}\right) = \frac{A_0}{\left(N_{\mathrm{VELO}}\right)^{n_A}} . \tag{4.14}$$

The fits to the correlation functions for OS pairs are performed simultaneously in all multiplicity bins available in the given sample (separately for the pPb and the Pbp dataset). In this procedure, the parameters from Eqs. (4.13) and (4.14) are common for all bins, while the N and the δ values are left free for each correlation function. A summed negative log-likelihood function (see Eq. (4.8)) is constructed for all of the N_{VELO} bins in the given dataset and minimised globally to obtain the best description of the data.

Table 4.16: Results of the global fits to the OS correlation functions using Eq. (4.12) to parametrise the cluster contribution in the pPb and the Pbp data. As described in the text, the N_0 parameter is fixed to the value of 15 for an optimal fit stability.

dataset	A_0 [GeV]	n_A	σ_0 [GeV]	σ_1 [GeV]	MinFcn/ndf
pPb	2.838 ± 0.109	0.8438 ± 0.0111	0.4799 ± 0.0018	0.1744 ± 0.0060	7944/3686
Pbp	1.107 ± 0.022	0.5036 ± 0.0049	0.5613 ± 0.0013	$\sim 10^{-11} \pm 10^{-3}$	9218/4424

The lower fit range for the OS pairs is limited with respect to the SS ones, due to a significant contribution of multibody resonance decays in the very low Q region [135,147]. The global fits are performed in the Q range of 0.25–2.00 GeV which gives an optimal quality of the fits. Also, it is found that the best stability of the global fits is obtained with a fixed value of the N_0 parameter from Eq. (4.13). It is set to 15 for both the pPb and the Pbp datasets, which is based on the fit results in the pPb case with that parameter left free. The impact of the choice of the lower Q boundary as well as the fixed N_0 value on the final correlation parameters is studied as a contribution to the systematic uncertainty (see Sect. 4.5.2.1).

Results of the global fits to the OS correlation functions for the pPb and the Pbp data are shown in Fig. 4.31 and summarised in Tab. 4.16. Quality of the fits[5] in both cases corresponds to MinFcn/ndf ~ 2, which is a satisfactory description in the BEC studies due to the unknown theoretical parametrisation of the nonfemtoscopic background (see Sect. 4.1.4). It is worth noting that for the Pbp sample, the best description of the cluster contribution in the data is found with the value of the σ_1 parameter consistent with zero.

4.5.1.3 Background scaling

As indicated in Sect. 4.1.3.2, the cluster contribution is expected to be larger for the OS pairs than for the SS ones due to the charge conservation in processes contributing to the cluster formation. For this reason, the amplitude of the cluster contribution for the SS correlation functions is multiplied by the scaling factor z (see Eq. (4.12)). To obtain a uniform background scaling across the N_{VELO} bins, this value is parametrised using a theoretically-motivated form based on the ratio of SS and OS pairs combinatorics [134]

$$z(N_{\mathrm{VELO}}) = \frac{aN_{\mathrm{VELO}} + b}{1 + aN_{\mathrm{VELO}} + b} \,, \tag{4.15}$$

where a and b are free parameters.

The parameters describing the background scaling are determined in the following way. First, individual fits to the SS correlation functions in all N_{VELO} bins are performed using Eq. (4.10) with the values characterising the cluster contribution fixed to the results of the OS fits specified in Tab. 4.16. In these fits, the z parameter is free. The measured z values are presented in Fig. 4.32, together with the results of fits performed using Eq. (4.15). The shifted z values that are used to study the contribution of the background scaling procedure to the systematic uncertainty (see Sect. 4.5.2.1) are also illustrated.

[5]Quantities denoted as MinFcn/ndf correspond to the final value of the function minimised in the fitting procedure (see Eq. (4.8)), divided by the number of degrees of freedom in the fit.

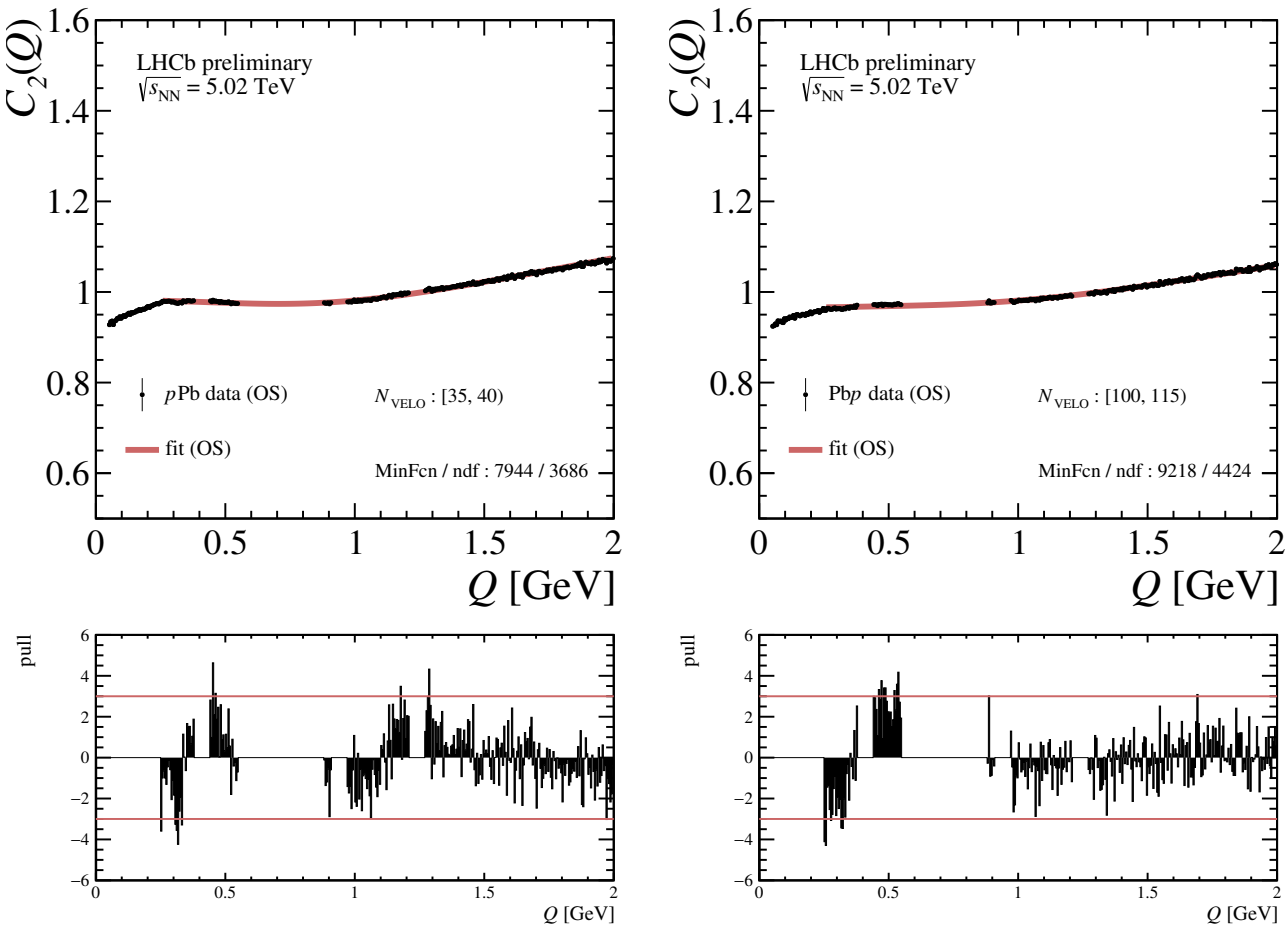

Figure 4.31: Examples of the OS correlation functions in individual N_{VELO} bins together with the global fit using Eq. (4.12) to parametrise the cluster contribution. The results are shown for (left) a moderate-multiplicity region ($35 \leqslant N_{\mathrm{VELO}} < 40$) of the pPb and (right) high-multiplicity regime ($100 \leqslant N_{\mathrm{VELO}} < 115$) of the Pb$p$ dataset. Pull distributions of the fits are given below the relevant figures. Only statistical uncertainties are shown for the data points.

Table 4.17: Results of fits using Eq. (4.15) to determine the parametrisation of the background scaling between the correlation functions for OS and SS pairs in the pPb study.

dataset	a	b	χ^2/ndf
pPb	0.0443 ± 0.0040	1.86 ± 0.12	112/13
Pbp	0.0749 ± 0.0069	3.12 ± 0.27	121/16

The a and the b parameters, which are determined independently for the pPb and the Pbp datasets, are presented in Tab. 4.17.

4.5.1.4 Determination of the BEC parameters

The correlation parameters are determined by performing fits to the SS correlation functions in each individual N_{VELO} bin using Eq. (4.10). In this procedure, the parameters

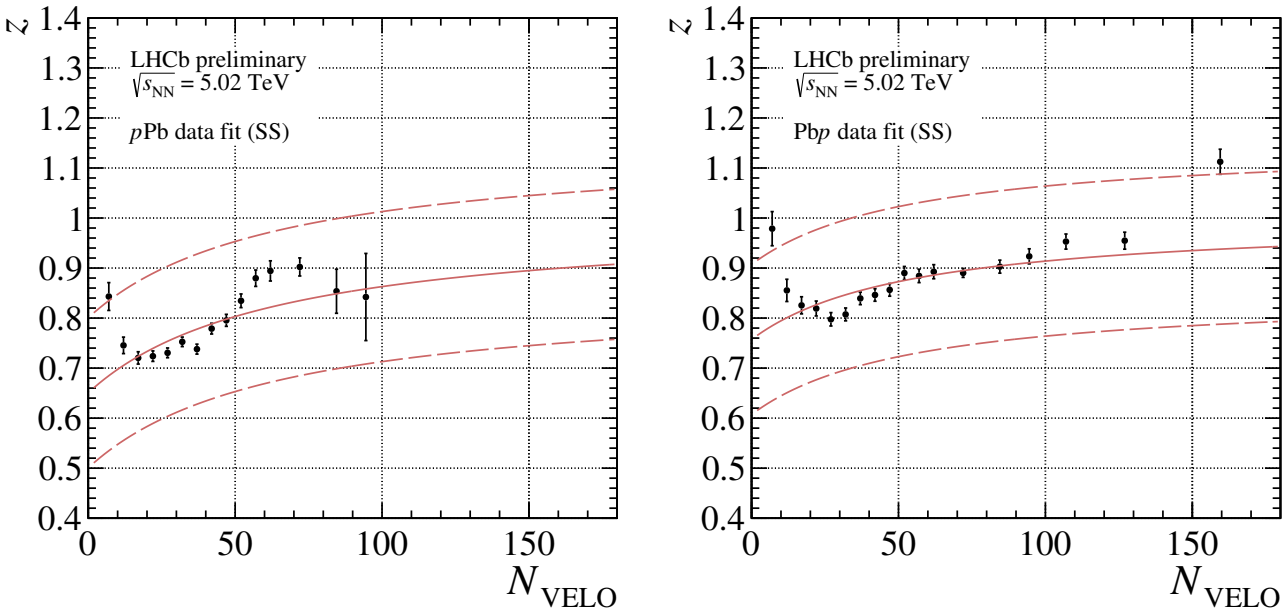

Figure 4.32: Results of the individual fits in N_{VELO} bins to the SS correlation functions employing Eq. (4.10) with a free z parameter for (left) the pPb and (right) the Pbp data sample. Parametrisations of the background scaling determined by fitting the data points using Eq. (4.15) are indicated by the red solid lines for each dataset. The dashed red lines illustrate shifted parametrisations used in the study of the systematic uncertainty related to the background scaling procedure. Error bars for the data points correspond to the statistical uncertainty of the fitted z values.

characterising the cluster contribution and the background scaling are fixed to the previously measured values, which are given in Tabs. 4.16 and 4.17, respectively. The fits are performed in the full range of the Q variable (0.05–2.00 GeV) in the constructed correlation functions (see more details in Sect. 4.5.1). The contribution from the choice of the fit range to the systematic uncertainty is investigated in Sect. 4.5.2.4.

Example results of the final fits to the SS correlation functions are presented in Fig. 4.33 and a summary of fits in all N_{VELO} bins is given in Tabs. 4.18 and 4.19. Quality of the individual fits corresponds to MinFcn/ndf ~ 4 at maximum (with values in the range of 1.1–4.3 in general), which is a reasonable result in this type of analysis, especially with an approach not employing the double-ratio method. As mentioned in Sects. 4.1.4 and 4.4.1.4, the fit quality in BEC studies is not expected to be ideal, due to various assumptions in the signal parametrisation (see Sect. 2.1.3) and unknown background effects. A summary of the measured R and λ parameters is given in Sect. 4.5.3.

4.5.2 Systematic uncertainty

The current section describes the study of various contributions to the systematic uncertainty of the measured correlation parameters. The final results can be affected *e.g.* by the method of the optimisation of the selection criteria, construction of the reference sample or parametrisation of the nonfemtoscopic background in the correlation function. The study is based on a similar investigation performed in the pp analysis (see Sect. 4.4.2).

Table 4.18: Results of the individual fits to the SS correlation functions using Eq. (4.10) for the pPb data. The statistical uncertainties of the fit parameters are indicated.

N_{VELO}	R [fm]	λ	δ [GeV^{-1}]	N	MinFcn/ndf
[5–10)	1.159 ± 0.010	0.860 ± 0.006	0.4043 ± 0.0027	0.6492 ± 0.0014	539/386
[10–15)	1.413 ± 0.010	0.635 ± 0.004	0.2503 ± 0.0010	0.7533 ± 0.0007	735/386
[15–20)	1.638 ± 0.011	0.562 ± 0.004	0.2071 ± 0.0006	0.7925 ± 0.0004	1095/386
[20–25)	1.790 ± 0.011	0.516 ± 0.004	0.1859 ± 0.0004	0.8131 ± 0.0003	1288/386
[25–30)	1.944 ± 0.012	0.476 ± 0.004	0.1669 ± 0.0003	0.8302 ± 0.0003	1645/386
[30–35)	2.088 ± 0.014	0.464 ± 0.004	0.1582 ± 0.0003	0.8392 ± 0.0002	1560/386
[35–40)	2.218 ± 0.016	0.452 ± 0.005	0.1522 ± 0.0003	0.8455 ± 0.0002	1650/386
[40–45)	2.364 ± 0.019	0.443 ± 0.005	0.1440 ± 0.0003	0.8529 ± 0.0002	1480/386
[45–50)	2.482 ± 0.023	0.435 ± 0.006	0.1404 ± 0.0003	0.8567 ± 0.0002	1374/386
[50–55)	2.575 ± 0.028	0.427 ± 0.008	0.1377 ± 0.0003	0.8595 ± 0.0003	1024/386
[55–60)	2.730 ± 0.036	0.443 ± 0.010	0.1332 ± 0.0004	0.8635 ± 0.0003	1006/386
[60–65)	2.799 ± 0.046	0.427 ± 0.012	0.1314 ± 0.0004	0.8652 ± 0.0003	732/386
[65–80)	2.972 ± 0.045	0.415 ± 0.011	0.1268 ± 0.0003	0.8696 ± 0.0003	805/386
[80–90)	3.462 ± 0.115	0.479 ± 0.033	0.1228 ± 0.0007	0.8733 ± 0.0006	461/386
[90–100)	3.535 ± 0.219	0.485 ± 0.062	0.1206 ± 0.0013	0.8752 ± 0.0011	435/386

Table 4.19: Results of the individual fits to the SS correlation functions using Eq. (4.10) for the Pbp data. The statistical uncertainties of the fit parameters are indicated.

N_{VELO}	R [fm]	λ	δ [GeV^{-1}]	N	MinFcn/ndf
[5–10)	1.227 ± 0.013	0.791 ± 0.007	0.3844 ± 0.0028	0.6723 ± 0.0015	528/386
[10–15)	1.469 ± 0.013	0.630 ± 0.005	0.2796 ± 0.0012	0.7408 ± 0.0007	688/386
[15–20)	1.658 ± 0.014	0.548 ± 0.005	0.2446 ± 0.0008	0.7691 ± 0.0005	987/386
[20–25)	1.801 ± 0.015	0.487 ± 0.005	0.2241 ± 0.0006	0.7862 ± 0.0004	1140/386
[25–30)	1.989 ± 0.017	0.467 ± 0.005	0.2049 ± 0.0005	0.8011 ± 0.0003	1389/386
[30–35)	2.130 ± 0.019	0.444 ± 0.005	0.1957 ± 0.0004	0.8091 ± 0.0003	1488/386
[35–40)	2.279 ± 0.021	0.433 ± 0.006	0.1881 ± 0.0004	0.8155 ± 0.0003	1592/386
[40–45)	2.380 ± 0.024	0.409 ± 0.006	0.1778 ± 0.0003	0.8232 ± 0.0003	1338/386
[45–50)	2.554 ± 0.027	0.415 ± 0.007	0.1718 ± 0.0003	0.8279 ± 0.0002	1252/386
[50–55)	2.725 ± 0.031	0.416 ± 0.008	0.1668 ± 0.0003	0.8320 ± 0.0002	1209/386
[55–60)	2.875 ± 0.035	0.420 ± 0.009	0.1597 ± 0.0003	0.8373 ± 0.0002	1057/386
[60–65)	2.972 ± 0.040	0.412 ± 0.010	0.1564 ± 0.0003	0.8398 ± 0.0002	959/386
[65–80)	3.322 ± 0.028	0.448 ± 0.007	0.1469 ± 0.0002	0.8471 ± 0.0001	1282/386
[80–90)	3.531 ± 0.043	0.449 ± 0.011	0.1362 ± 0.0002	0.8548 ± 0.0002	732/386
[90–100)	3.871 ± 0.052	0.513 ± 0.015	0.1273 ± 0.0003	0.8614 ± 0.0002	560/386
[100–115)	3.854 ± 0.049	0.513 ± 0.015	0.1160 ± 0.0003	0.8695 ± 0.0002	511/386
[115–140)	3.863 ± 0.049	0.555 ± 0.016	0.0971 ± 0.0003	0.8838 ± 0.0002	415/386
[140–180)	3.225 ± 0.053	0.487 ± 0.016	0.0711 ± 0.0003	0.9047 ± 0.0003	541/386

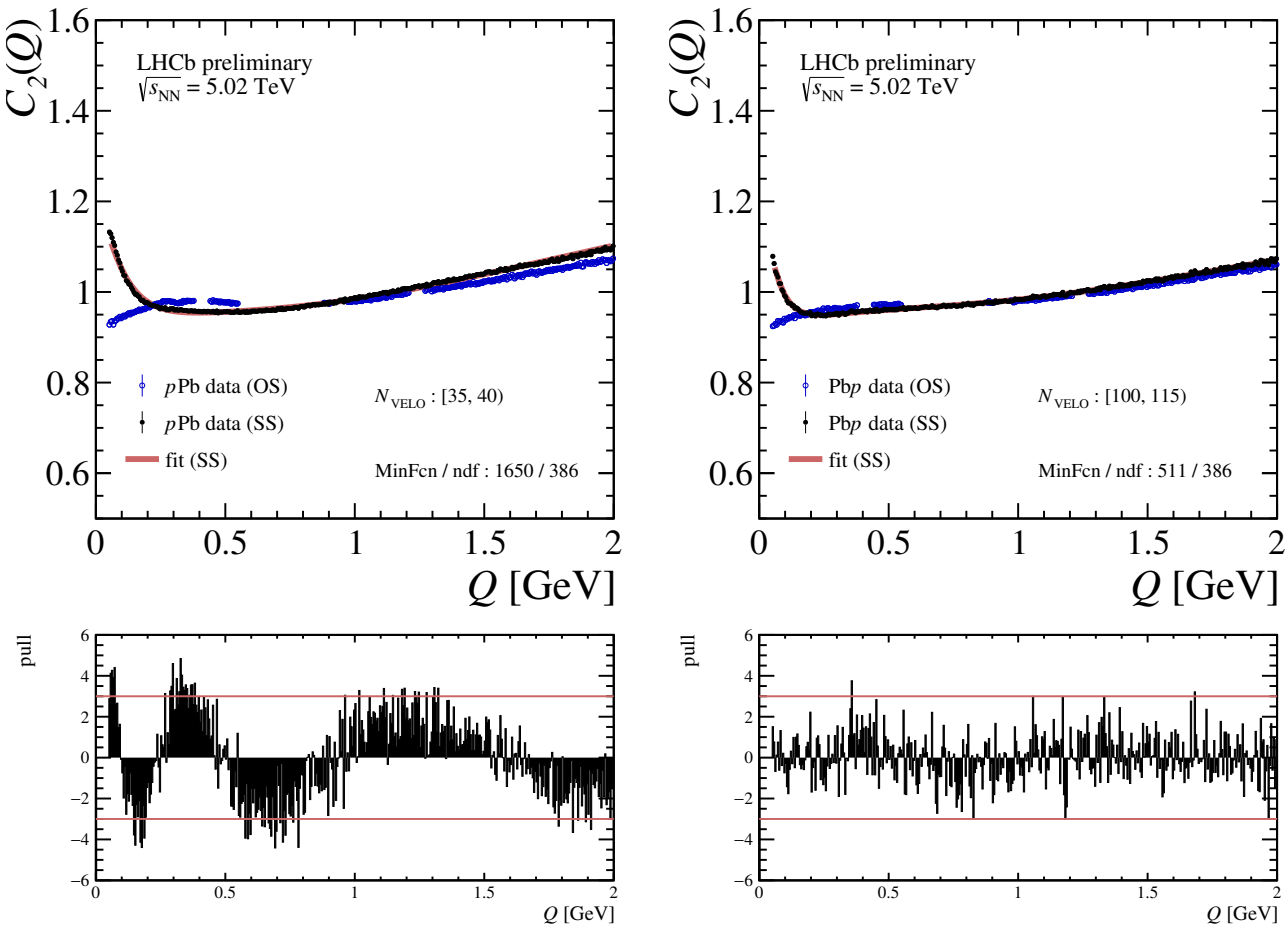

Figure 4.33: Examples of fits to the SS correlation functions in (left) a moderate-multiplicity region ($35 \leqslant N_{\mathrm{VELO}} < 40$) of the pPb and (right) a high-multiplicity regime ($100 \leqslant N_{\mathrm{VELO}} < 115$) of the Pb$p$ dataset. The black points correspond to the SS correlation functions, while the blue ones illustrate the OS correlation functions which are used to estimate the cluster contribution in the given N_{VELO} bin. The results of fits using Eq. (4.10) are indicated by the red solid lines. Pull distributions of the fits are given below the relevant figures. Only statistical uncertainties are shown for the data points.

However, many contributions specific to the data-driven approach are scrutinised in the current considerations (this refers mostly to the sources of uncertainty related to the nonfemtoscopic background description). It is worth noting that the importance of some contributions that are minor in the pp case may be more significant in the pPb study. It is related to differences in the analysis method, where employing the double-ratio approach can substantially reduce the impact of contributions that are properly modelled in the simulation (see Sect. 4.1.3.1). On the other hand, in the current analysis the systematic uncertainty related to the choice of the MC generator tuning, which is the dominant contribution in the pp study (see Sect. 4.4.2.1), is not present at all.

The general approach to determining the systematic uncertainty is to repeat the analysis procedure with appropriate modifications introduced to evaluate the contribution in question. Each input is assessed by comparing the refitted correlation parameters with the central results summarised in Tabs. 4.18 and 4.19. The determined contributions

are added in quadrature to receive the total systematic uncertainty. An improvement with respect to the method used in the *pp* study is introduced in calculating some of the contributions. In a case when the given input is related to a quantity that could be in principle arbitrarily chosen (such as a fit range or selection requirements), the systematic uncertainty is calculated as a root-mean-square (*RMS*) value of a set of results. This so-called RMS method allows us to represent well the spread of results obtained while varying the parameter in question and, furthermore, displays an improved stability with respect to potential fluctuations of the particular fits. If not otherwise stated, the individual contributions to the systematic uncertainty are calculated using the RMS method. In other situations, where the given input is related more to a robustness verification by changing a parameter value by a significant amount, the contribution is determined as an average or a maximum absolute difference between the modified and central results. Such cases are clearly indicated where applicable.

Studies of the individual contributions to the systematic uncertainty are described in the following sections. Some of the investigations performed in the *pp* analysis are not repeated in the current considerations, since, based on the previous results, their contributions to the systematic uncertainty are expected to be negligible or small. This refers to the common aspects for all analyses at LHCb, such as the efficiency of PV reconstruction or effects related to the resolution of the Q variable (see Sects. 4.4.2.3 and 4.4.2.10). A summary of individual contributions to the systematic uncertainty is provided in Tab. 4.20. This table presents ranges corresponding to the lowest and highest values of the determined inputs to the systematic uncertainty across most of the VELO-track multiplicity bins. Some bins display an outlying uncertainty of the correlation parameters, which is not representative of the other N_{VELO} regions. This refers to the two edge bins (corresponding to the tail in the N_{VELO} distribution at high values) in the pPb sample and a single edge bin in the Pbp data. In the former case, the two bins have significant statistical uncertainties on both the R and the λ parameters (up to 6% and 13%, respectively, while these uncertainties in the other N_{VELO} bins are limited to the ranges of 0.5–1.5% and 0.5–3.0%). This leads to potential fluctuations in individual fits and is reflected in anomalous values of the total systematic uncertainty – up to 18.0% (40.5%) for the R (λ) parameter in the bin with the highest multiplicity. For the edge bin in the Pbp sample, a contribution due to the PID optimisation procedure (see Sect. 4.5.2.2) is found to be significantly higher than in the other N_{VELO} regions. This sensitivity to the PID requirements may be related to the declining purity of the sample of selected pion pairs at higher multiplicities (see Sect. 4.3.2.2). The resulting total systematic uncertainty in this N_{VELO} region amounts to 30.0% and 19.5% for the R and the λ parameters, respectively. The total systematic uncertainty in all N_{VELO} bins is illustrated *e.g.* in Figs. 4.35 and 4.36. The anomalous uncertainties of the edge bins are indicated by large error bars – the edge bins are shown for completeness, however, their significance for a physics discussion is limited.

4.5.2.1 Parametrisation of the cluster contribution

Parametrisation of the nonfemtoscopic background in the correlation function is a key element in a data-driven BEC analysis. In the current study, this process can be divided into two steps. First, the parameters describing the cluster contribution (see Sect. 4.1.3.2) are established by investigating the OS pion pairs. This procedure can be impacted by

Table 4.20: Individual contributions to the systematic uncertainty on the R and the λ parameters in the pPb study. The listed ranges correspond to the lowest and highest values of the given input determined across most of the N_{VELO} bins in the pPb and the Pbp samples (see the description in text for details). Negligible contributions are not listed. The total uncertainty is a quadratic sum of the individual inputs.

	pPb dataset		**Pbp dataset**	
contribution	$\sigma_{\mathrm{syst}}(R)$ [%]	$\sigma_{\mathrm{syst}}(\lambda)$ [%]	$\sigma_{\mathrm{syst}}(R)$ [%]	$\sigma_{\mathrm{syst}}(\lambda)$ [%]
background scaling	4.5–9.0	3.5–11.0	4.5–6.5	3.0–9.5
background fit range	1.0–3.0	0.5–3.5	2.0–3.5	0.5–4.0
background fit – fixed N_0	0.5–3.0	0.5–3.0	< 0.5	< 0.5
background fit – resonances	0.5–4.0	0.5–4.0	1.5–3.0	0.5–3.5
PID optimisation	0.5–1.5	0.5–5.0	0.5–10.5	0.5–8.5
fake tracks	0.5–5.5	1.0–8.0	0.5–4.5	0.5–8.0
requirement on z_{PV}	0.5–1.5	0.5–3.0	0.5–2.0	0.5–3.5
Coulomb correction	0.5–1.5	1.0–2.5	0.5–2.0	0.5–3.0
SS fit range (min)	1.5–5.0	1.0–8.5	0.5–3.5	0.5–5.5
SS fit range (max)	0.5–1.0	0.5–2.0	0.5–2.0	0.5–3.0
reference sample	0.5–2.0	0.5–3.0	0.5–2.0	0.5–4.0
total	6.0–12.0	6.0–16.5	6.5–12.0	5.0–16.0

several effects related to the removal of structures originating from two-body resonance decays in the OS correlation function, the choice of the range of fit performed to determine the background parametrisation and fixing some of the fit parameters to improve the fit stability (see Sects. 4.5.1.1 and 4.5.1.2 for details). The second step refers to providing an appropriate scaling for the amplitude of the cluster contribution between the OS and the SS correlation functions, which is detailed in Sect. 4.5.1.3. An investigation on the impact of the mentioned effects on the final correlation parameters is discussed below.

The Q regions affected by structures related to two-body resonance decays are removed from the fits to the OS correlation function. The boundaries of the removed regions – see Tab. 4.15 – are optimised to provide a satisfactory fit quality (the emerging structures are not described at all by Eq. (4.12)). The impact of the particular choice of these limits is investigated by repeating the analysis with the widths of the defined regions increased and decreased by 20%.

Another important effect in the determination of the cluster contribution is related to the choice of the range of fits to the OS correlation functions. The lower boundary needs to be restricted due to a significant impact of multibody resonance decays in the low-Q region and is chosen as 0.25 GeV (see Sect. 4.5.1.2). The contribution of this effect to the systematic uncertainty is studied by varying its values within 10%. The analysis is repeated with the lower limit in the OS fit set to 0.225 and 0.275 GeV. Furthermore, it is found that an optimal fit stability is achieved with a fixed N_0 parameter. Its value is set to 15 based on the fit results for the pPb dataset obtained with this parameter left free. The impact of this technical measure on the final correlation parameters is investigated by modifying this value within $\sim 30\%$, *i.e.*, to 10 and 20.

The scaling of the cluster contribution amplitude between the OS and the SS pairs is an essential element of the CS method (see Sect. 4.1.3.2). The nominal parametrisations

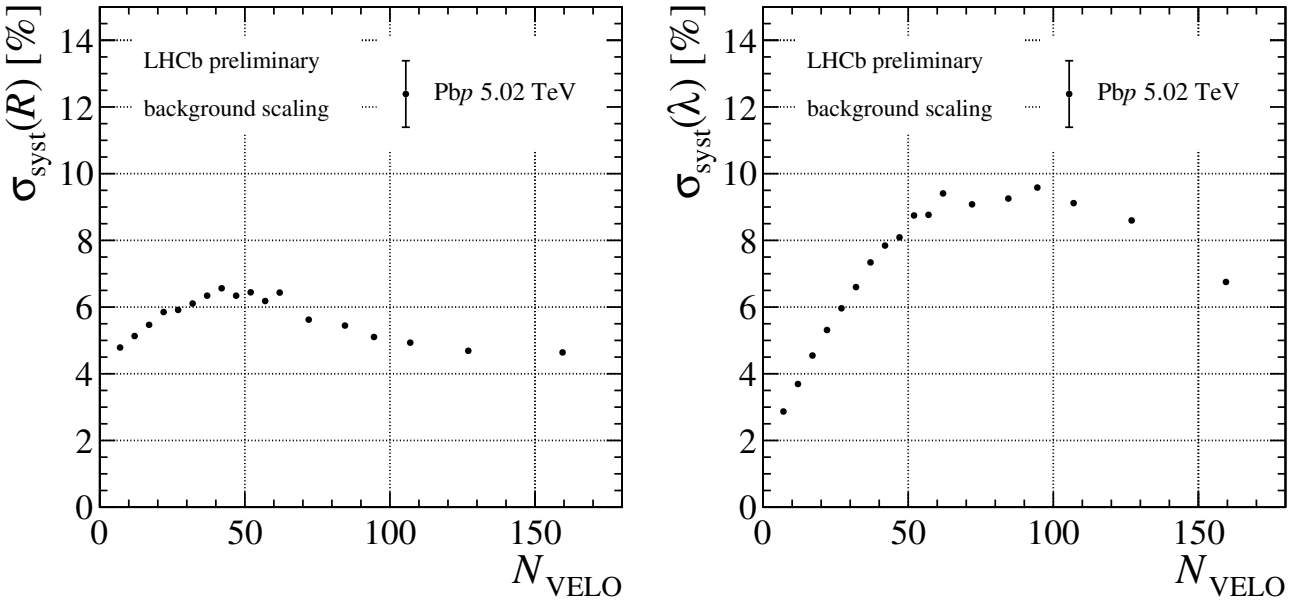

Figure 4.34: Systematic uncertainty due to the background scaling procedure for (left) the correlation radius and (right) the intercept parameter in a function of the VELO-track multiplicity in the Pbp dataset.

of the background scaling (see Eq. (4.15)) determined for the central results are shifted by an arbitrary value of ± 0.15 prior to their usage in the final SS fits to investigate the influence of this procedure on the measured correlation parameters. These values are chosen to comprise most of the individual results with the z parameter left free in the fits to the SS correlation function (see the corresponding bands around the nominal scaling parametrisations in Fig. 4.32). This contribution is calculated as an average absolute difference between the central and modified results for the final fits repeated with the higher and the lower z values. An example of the uncertainties related to this input determined in different N_{VELO} bins is shown in Fig. 4.34 for the Pbp dataset. The background scaling procedure is found to be the dominant contribution to the systematic uncertainty in the study, reaching up to 9% (11%) for the R (λ) parameter (see Tab. 4.20).

4.5.2.2 Selection criteria and the removal of misreconstructed tracks

Various selection requirements described in Sect. 4.3 are applied to the initial dataset to acquire a sample of well-identified prompt pions with high quality of reconstruction and assignment to a primary vertex. The optimisation of the selection criteria can influence the measured correlation parameters. The most important contributions to the systematic uncertainty in this procedure are studied in the current section.

One of the crucial prerequisites on the candidate tracks is related to the pion identification, which is done based on the $ProbNN(\text{pion})$ variable. The nominal requirement of $ProbNN(\text{pion}) > 0.65$, which is established in the optimisation process described in Sect. 4.3.2.2, is modified to a less strict $ProbNN(\text{pion}) > 0.30$ limit to assess how the correlation parameters are affected by the pion identification criteria. The chosen value represents a conservative approach, with the fraction of misidentified pions in

Table 4.21: Fractions of fake tracks in the selected pion sample and fractions of signal pairs (with Q values restricted to the $Q < 1.0\,\text{GeV}$ region) containing a fake track in the pPb and the Pbp datasets. The values are acquired using the simulation information and are shown for two different selection requirements on the fake track identification.

	single particle [%]		**particle pair [%]**	
$ProbNN(\text{ghost})$	pPb	Pbp	pPb	Pbp
< 0.25	0.51	0.43	1.06	0.81
< 0.50	0.57	0.48	1.19	0.91

the sample increasing by $\sim 50\%$ with respect to the final selection.

Misreconstructed tracks can also degrade the purity of the selected pions sample and affect the final results by their presence in the signal pairs (see the detailed discussion in Sect. 4.3.3). As illustrated in Fig. 4.17, the misreconstructed tracks (mostly the clone ones) that could directly contribute to the SS pairs in the BEC-signal region are well controlled in the data. Dedicated study is performed to evaluate the impact of additional fake tracks in the sample by modifying the selection requirement on such tracks from $ProbNN(\text{ghost}) < 0.25$ to $ProbNN(\text{ghost}) < 0.50$ (which corresponds to the maximum value available in the dataset after the preselection – Sect. 4.3.1). Fractions of fake tracks in the selected pion sample and fractions of signal pairs containing a fake track (values determined using the simulation) for these two criteria are given in Tab. 4.21. As the contribution of the clone tracks in the data sample is very well controlled (see Sect. 4.3.3), similar to the pp analysis where the uncertainty associated with the clone tracks is found to be negligible (see Sect. 4.4.2.5), the systematic uncertainty related to the clone tracks in the current study is expected to be negligible.

The contributions of the fake tracks and the pion selection criteria optimisation to the systematic uncertainty are calculated as the absolute difference between the results obtained with the modified selection requirements and the central ones. In a limited number of bins, these inputs constitute the most important contributions to the total systematic uncertainty (up to $\sim 10.5\%$ and 8.5% for the R and the λ parameters), together with the one related to the background scaling (see Sect. 4.5.2.1). These inputs are more significant than the relevant ones in the pp study, due to the fact that effects related to the optimisation of the selection criteria are considerably reduced in the pp case thanks to the use of the double-ratio method.

The last investigation on the systematic uncertainty related to the selection requirements takes into account the accepted position of primary vertices along the z-axis in the LHCb coordinate frame. This prerequisite is applied to ensure a uniform N_{VELO} distribution across the dataset to limit the multiplicity bin migration (see Sect. 4.3.1). The nominal range of the accepted z_{PV} values is extended or tightened by 20% to evaluate the systematic uncertainty related to this requirement. These variations correspond to repeating the analysis with limits on the PV position set to $-192 < z_{\text{PV}} < 72\,\text{mm}$ and $-128 < z_{\text{PV}} < 48\,\text{mm}$, respectively.

4.5.2.3 Correction for the Coulomb interactions

Final-state Coulomb interactions for both the SS and the OS pairs need to be taken into account in the fits to the correlation functions. In the current study, this is achieved using Eq. (4.5) with the effective radius set to $R_{\text{eff}} = 2\,\text{fm}$ which is based on the expected typical values of the measured correlation radii (see Sect. 4.5.1). The impact of this correction on the correlation parameters is investigated in two aspects. Firstly, the assumed R_{eff} values are halved and doubled (leading to 1 fm and 4 fm, respectively). This allows us to comprise all of the correlation radii actually measured in the study (see Fig. 4.35) and to take into account the expected differences in radii across the N_{VELO} bins. Secondly, the effective radius is not directly equivalent to the measured correlation radius. Although a proportionality between the two values can be expected (see Sect. 2.1), the exact relation may be not trivial. Therefore, for the purpose of this study, a simple proportionality of $R_{\text{eff}} = \epsilon R$ is assumed [147] and the ϵ values are varied between 0.5 and 2.0. This leads to the final values of R_{eff} corresponding to 0.5 fm and 8.0 fm, which are used to evaluate the systematic uncertainty related to the correction for Coulomb interactions. The described contribution is found to be relatively small, which indicates that the employed treatment of the Coulomb effects is highly robust with respect to the setting of the R_{eff} parameter.

4.5.2.4 Fit range and binning

The range and the binning used in the fits to the correlation functions for SS pairs can affect the final results. Details on the fitting procedure in the analysis are given in Sect. 4.5.1.4. The systematic uncertainty related to the SS fit range is investigated by varying its limits in both higher and lower values. The impact from the boundary in the low-Q region is evaluated by altering it within 20%, which corresponds to the values of 0.04 and 0.06 GeV. The relevant contribution is calculated using the RMS method (see Sect. 4.5.2). A similar procedure is implemented for the fit boundary at high Q values, where the modification at the level of 10% is applied, leading to the upper fit range being limited to 1.8 and extended to 2.2 GeV. Both contributions to the systematic uncertainty associated with the SS fit range are found to be relatively small (with the one of the lower limit being more significant than the one related to the upper boundary).

Another potential source of systematic uncertainty is studied by altering the binning of the Q variable in the correlation function. The bin width is doubled from the nominal 0.005 GeV to 0.010 GeV (see Sect. 4.5.1). The impact of this modification on the measured correlation parameters is negligible.

4.5.2.5 Reference sample construction

Construction of the reference sample is a basic aspect of the BEC analyses and details on this procedure in the current study are given in Sect. 4.5.1. As already stated, none of the available construction methods can provide a perfect reconstruction of the properties of the signal sample, and the measured correlation parameters can be affected by the particular implementation of this procedure. It is worth noting that reducing the effects related to the imperfections in the construction of the reference sample is one of the advantages of employing the double-ratio method (see Sect. 4.1.3.1). In the current study, the potential impact of the event-mixing implementation on the correlation parameters is assessed by varying the number of candidates available for mixing, which is

a parameter that can be tuned in the procedure. The nominal value of 10 is modified to 50 and 100, and the analysis is repeated using the updated settings. The final contribution to the systematic uncertainty is calculated using the RMS method (see Sect. 4.5.2) and found to be relatively small.

4.5.3 Summary

Results of the BEC study for same-sign charged pion pairs from pPb collisions in the LHCb experiment are described in this section. A data-driven analysis method is employed to account for effects related to the nonfemtoscopic background (see Sects. 4.5.1 and 4.1.3.2). The correlation parameters are determined in common bins of the VELO-track multiplicity for the two samples corresponding to different beam configurations (see Sect. 4.2.2). A summary of the obtained results, including the statistical and the systematic uncertainties is shown in Tab. 4.22. Similar to the pp study (see Sect. 4.4.3), the systematic uncertainties dominate the statistical ones. The systematic uncertainty is observed to reach up to $\sim 12.0\%$ (16.5%) for the measured R (λ) parameters in most of the N_{VELO} regions (except for the edge multiplicity bins – see Sect. 4.5.2). The most significant input to the systematic uncertainty is linked to the scaling of the cluster contribution amplitude between the OS and the SS pairs (see Sect. 4.5.2.1). Other important contributions originate from the selection requirements related to the particle identification and the fake tracks removal (see Sect. 4.5.2.2). Interestingly, the two latter inputs are significantly more prominent than in the pp study, where the double-ratio method essentially reduces the effects associated with the optimisation of the selection criteria. The correlation parameters measured in a function of the VELO-track multiplicity for both the pPb and the Pbp datasets are illustrated in Figs. 4.35 and 4.36. The publication procedure for the presented analysis is ongoing [18]. A detailed discussion on the BEC measurements performed in the LHCb experiment and described in this thesis (including as well the pp analysis) is provided in Chap. 5.

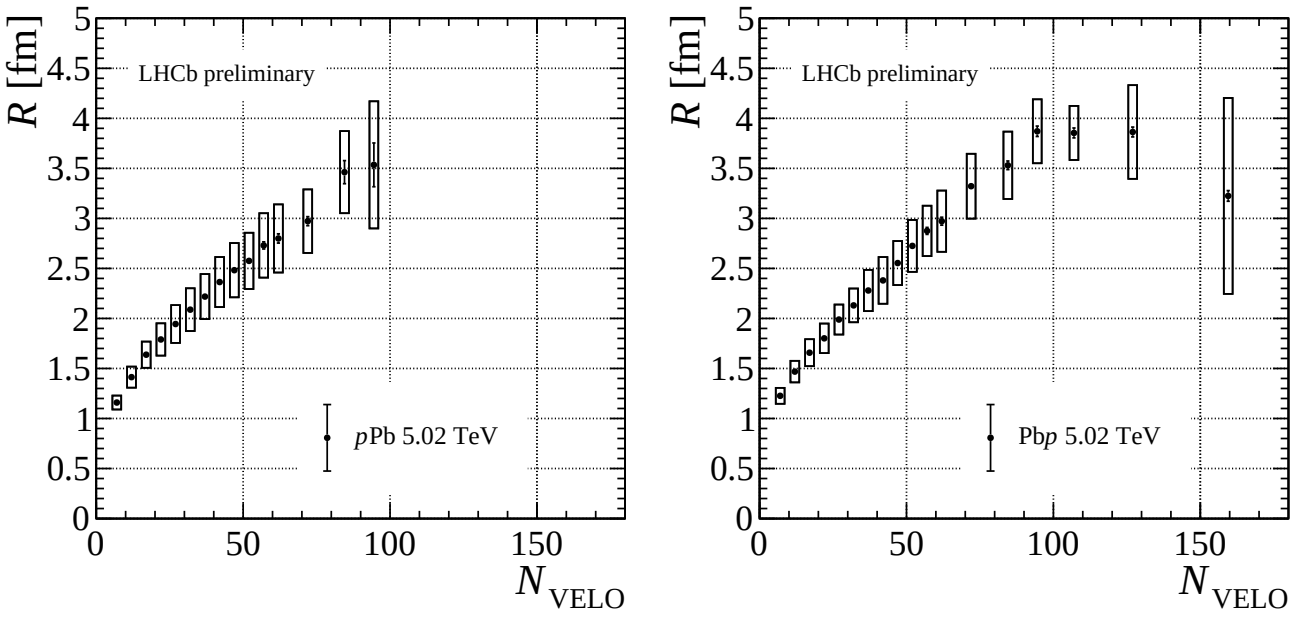

Figure 4.35: Correlation radius as a function of the VELO-track multiplicity measured in (left) the pPb and (right) the Pbp dataset. Error bars indicate the statistical uncertainties, while boxes illustrate the systematic ones. The data points are positioned at the centres of the VELO-track multiplicity bins.

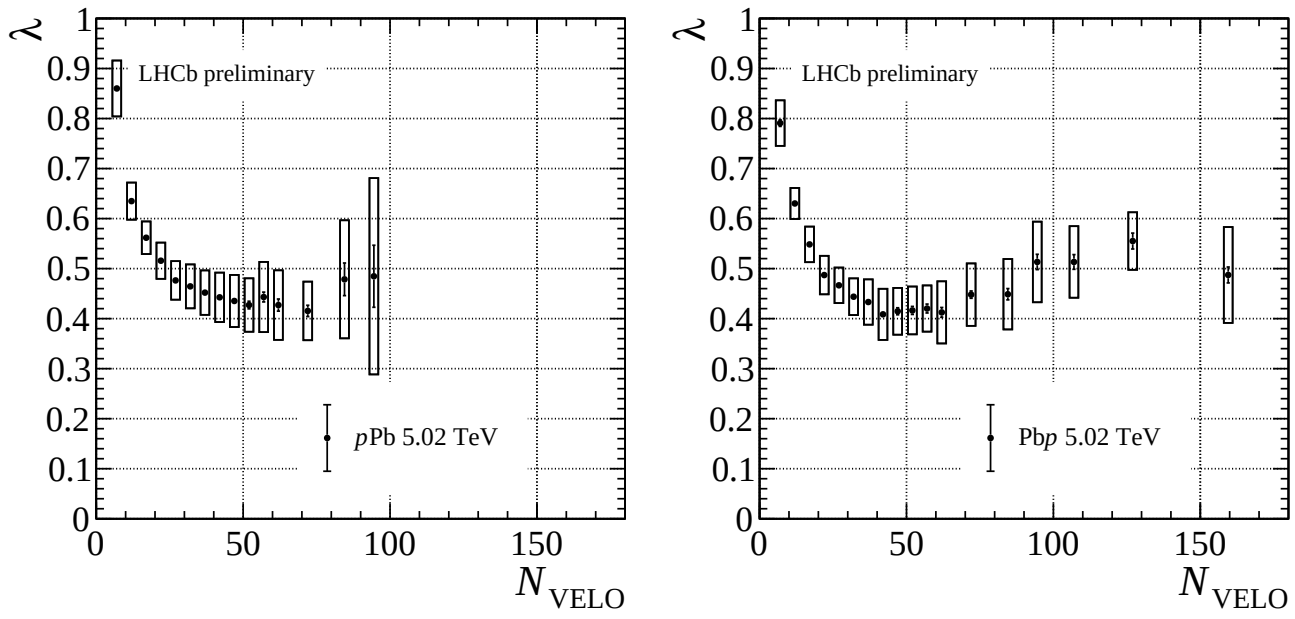

Figure 4.36: Intercept parameter as a function of the VELO-track multiplicity measured in (left) the pPb and (right) the Pbp dataset. Error bars indicate the statistical uncertainties, while boxes illustrate the systematic ones. The data points are positioned at the centres of the VELO-track multiplicity bins.

Table 4.22: Correlation parameters determined from the fits to the SS correlation functions using Eq. (4.10) in the N_{VELO} bins for the pPb and the Pbp datasets. Both the statistical and the systematic uncertainties are indicated, in this order.

	pPb dataset		**Pbp dataset**	
N_{VELO}	R [fm]	λ	R [fm]	λ
[5–10)	$1.159 \pm 0.010 \pm 0.070$	$0.860 \pm 0.006 \pm 0.056$	$1.227 \pm 0.013 \pm 0.080$	$0.791 \pm 0.007 \pm 0.045$
[10–15)	$1.413 \pm 0.010 \pm 0.105$	$0.635 \pm 0.004 \pm 0.037$	$1.469 \pm 0.013 \pm 0.108$	$0.630 \pm 0.005 \pm 0.031$
[15–20)	$1.638 \pm 0.011 \pm 0.131$	$0.562 \pm 0.004 \pm 0.033$	$1.658 \pm 0.014 \pm 0.135$	$0.548 \pm 0.005 \pm 0.036$
[20–25)	$1.790 \pm 0.011 \pm 0.161$	$0.516 \pm 0.004 \pm 0.036$	$1.801 \pm 0.015 \pm 0.148$	$0.487 \pm 0.005 \pm 0.038$
[25–30)	$1.944 \pm 0.012 \pm 0.189$	$0.476 \pm 0.004 \pm 0.039$	$1.989 \pm 0.017 \pm 0.150$	$0.467 \pm 0.005 \pm 0.036$
[30–35)	$2.088 \pm 0.014 \pm 0.214$	$0.464 \pm 0.004 \pm 0.044$	$2.130 \pm 0.019 \pm 0.169$	$0.444 \pm 0.005 \pm 0.037$
[35–40)	$2.218 \pm 0.016 \pm 0.225$	$0.452 \pm 0.005 \pm 0.044$	$2.279 \pm 0.021 \pm 0.206$	$0.433 \pm 0.006 \pm 0.045$
[40–45)	$2.364 \pm 0.019 \pm 0.250$	$0.443 \pm 0.005 \pm 0.049$	$2.380 \pm 0.024 \pm 0.233$	$0.409 \pm 0.006 \pm 0.051$
[45–50)	$2.482 \pm 0.023 \pm 0.271$	$0.435 \pm 0.006 \pm 0.052$	$2.554 \pm 0.027 \pm 0.220$	$0.415 \pm 0.007 \pm 0.047$
[50–55)	$2.575 \pm 0.028 \pm 0.281$	$0.427 \pm 0.008 \pm 0.053$	$2.725 \pm 0.031 \pm 0.259$	$0.416 \pm 0.008 \pm 0.048$
[55–60)	$2.730 \pm 0.036 \pm 0.322$	$0.443 \pm 0.010 \pm 0.070$	$2.875 \pm 0.035 \pm 0.252$	$0.420 \pm 0.009 \pm 0.046$
[60–65)	$2.799 \pm 0.046 \pm 0.341$	$0.427 \pm 0.012 \pm 0.070$	$2.972 \pm 0.040 \pm 0.306$	$0.412 \pm 0.010 \pm 0.062$
[65–80)	$2.972 \pm 0.045 \pm 0.318$	$0.415 \pm 0.011 \pm 0.059$	$3.322 \pm 0.028 \pm 0.324$	$0.448 \pm 0.007 \pm 0.062$
[80–90)	$3.462 \pm 0.115 \pm 0.410$	$0.479 \pm 0.033 \pm 0.118$	$3.531 \pm 0.043 \pm 0.337$	$0.449 \pm 0.011 \pm 0.070$
[90–100)	$3.535 \pm 0.219 \pm 0.635$	$0.485 \pm 0.062 \pm 0.196$	$3.871 \pm 0.052 \pm 0.320$	$0.513 \pm 0.015 \pm 0.081$
[100–115)	–	–	$3.854 \pm 0.049 \pm 0.270$	$0.513 \pm 0.015 \pm 0.072$
[115–140)	–	–	$3.863 \pm 0.049 \pm 0.468$	$0.555 \pm 0.016 \pm 0.057$
[140–180)	–	–	$3.225 \pm 0.053 \pm 0.979$	$0.487 \pm 0.016 \pm 0.096$

Chapter 5

Discussion of the analysis results

The current chapter presents a discussion of the results of the BEC analyses for same-sign charged pion pairs in pp [17] and pPb [18] collisions at LHCb reported in Chap. 4 (the results of the pPb study are preliminary). In the pPb case, two beam modes are used, which results in two samples corresponding to different regions of pseudorapidity in the pPb nucleon-nucleon centre-of-mass system (see Sect. 4.2.2). The R and the λ parameters (see Figs. 5.1 and 5.2) are measured in different bins of the reconstructed charged-particle multiplicity (N_{rec}), which is equivalent to the N_{VELO} values in the present studies. In the pp analysis, unfolded charged-particle multiplicity ranges corresponding to the reconstructed multiplicity bins are also provided.

The described results are complementary to measurements performed at the LHC energies in the central rapidity region [133, 134, 137, 144, 145, 147]. The observed behaviour of the correlation parameters is compatible with conclusions from the investigations done in the same collision systems by other experiments at the LHC (see also Sect. 2.3). In general, the correlation radius becomes larger with increasing event multiplicity, while the intercept parameter displays an opposite behaviour (some fluctuations from those patterns may be seen at higher multiplicity values). The trend for the λ parameter in the pPb study is less obvious, as it first becomes smaller with the rising multiplicity, while, in the higher multiplicity region ($N_{rec} > 50$), it tends towards a constant value. Similar behaviour of the λ parameter was observed in other investigations employing an analysis method analogous to the current one [134]. The determined R (λ) parameters vary within 1–4 fm (~ 0.40–0.85) in the presented studies and correspond to a region of up to ~ 180 in the reconstructed multiplicity values.

The measured correlation radii are observed to scale linearly with the cube root of the reconstructed charged-particle multiplicity. A simple fit illustrating this relationship is performed for each sample in the study (pp, pPb and Pbp) and the fit results are shown in Fig. 5.1 and Tab. 5.1. Only the statistical uncertainties of the measured R values are taken into account in the fitting procedure. As it can be seen in the figure, the data points in each sample are consistent, within the systematic uncertainties, with respective fit results. This kind of scaling was also reported by other experiments at the LHC for various collision systems [135, 145, 147] (see also Fig. 2.2). The slope of this relation in the pp study seems to be smaller than in the pPb case and the R values only start to diverge between those systems at higher multiplicities, being consistent in the low-multiplicity region. Similar behaviour is reported by the ALICE experiment [145]. However, the small number of data points in the pp study restrict stronger conclusions in this area.

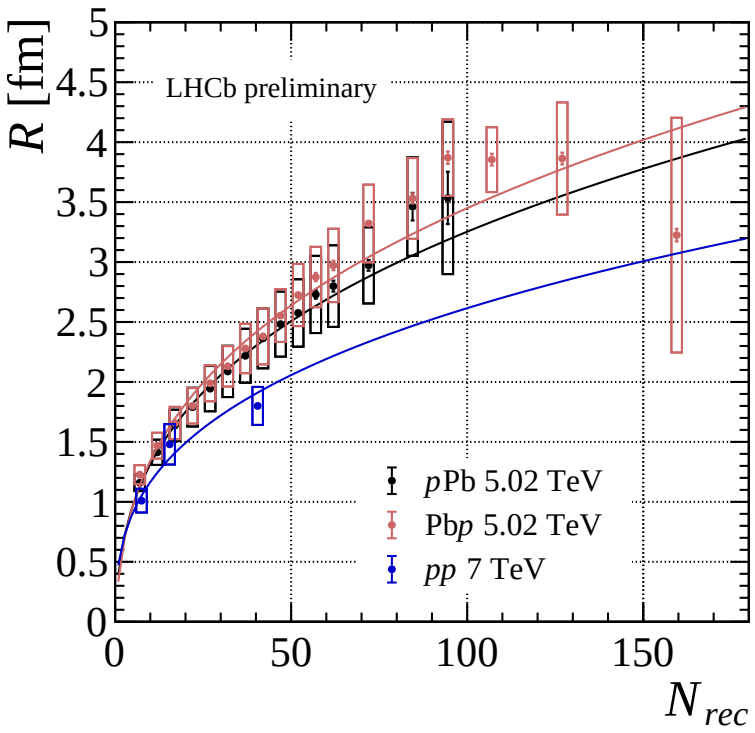

Figure 5.1: Correlation radius as a function of the reconstructed charged-particle multiplicity measured in the pp [17] and the pPb collision systems in the LHCb experiment. Error bars indicate the statistical uncertainties, while boxes illustrate the systematic ones. The data points are positioned at the centres of the multiplicity bins. Results of the fits to the observed radii linear in the cube root of the reconstructed multiplicity as described in the text are indicated by the solid lines.

Although the possibilities of direct comparisons to the predictions of the theoretical models on the correlation radii in the small systems are limited, some qualitative behaviour can be expected from those descriptions (see Sect. 2.3.1). The observed scaling of the R parameter is compatible with the models assuming a hydrodynamic expansion of the system [3, 152–154]. An alternative trend is expected from models based on the CGC framework, where the correlation radius tends to saturate at high multiplicity values [3,155,156]. While a potential saturation may be visible for the highest multiplicities available in the current pPb study ($N_{rec} > 100$), the systematic uncertainties in this region are significant and do not allow the clear identification of such behaviour. One of interesting perspectives for future BEC studies at LHCb would be to perform an analysis employing pp collisions at $\sqrt{s} = 13\,\text{TeV}$ centre-of-mass energy. It would allow the achievement of multiplicities higher than in the current pp study, enabling for more comprehensive comparisons to the pPb data in a broader multiplicity range as well as gaining more insight into the plausible saturation at high multiplicity values. As mentioned above, interpretations in these areas are currently limited. Potentially, the future analysis could also be extended into the k_{T} bins, which would provide more details for interpreting the evolution of the small systems in terms of the hydrodynamic expansion (see Sect. 2.3.2).

Another interesting observation can be made by comparing the correlation radii determined for the pPb and the Pbp samples (see Fig. 5.1), which correspond to the forward

Table 5.1: Results of the fits to the correlation radii as a function of the reconstructed charged-particle multiplicity presented in Fig. 5.1. The slope and the intercept parameters of the linear parametrisation are denoted as a and b, respectively. The fits take into account only the statistical uncertainties of the measured radii (hence the very high χ^2/ndf values, which are indicated here for completeness).

dataset	a	b	χ^2/ndf
pp	0.5830 ± 0.0206	-0.091 ± 0.048	68/1
pPb	0.7796 ± 0.0073	-0.365 ± 0.020	61/13
Pbp	0.8474 ± 0.0067	-0.484 ± 0.020	721/16

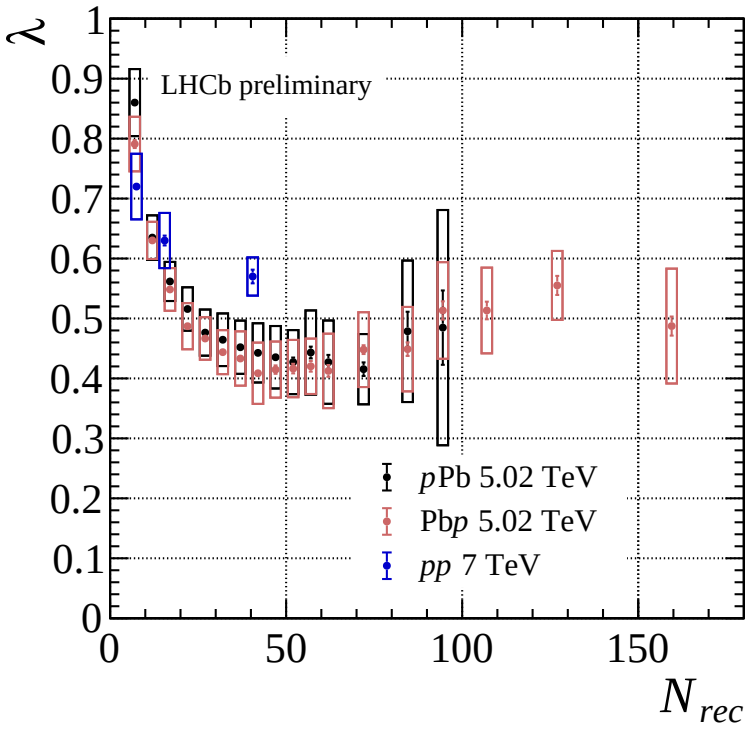

Figure 5.2: Intercept parameter as a function of the reconstructed charged-particle multiplicity measured in the pp [17] and the pPb collision systems in the LHCb experiment. Error bars indicate the statistical uncertainties, while boxes illustrate the systematic ones. The data points are positioned at the centres of the multiplicity bins.

and the backward region in the pPb collision system (see Sect. 4.2.2). Although the results in both samples agree well within the systematic uncertainties, it may be observed that the central values in the Pbp sample tend to radii which are systematically higher than in the pPb case, which becomes more prominent with increasing multiplicity. It could point to a behaviour similar to the one reported by the ATLAS experiment for pPb collisions [147], where a small (but nonnegligible) dependence of the R values on the pion pair rapidity was observed for high multiplicity values (see Fig. 2.6).

The potential sensitivity of the correlation radii to pseudorapidity can be further scrutinised by comparing the LHCb results in the forward region to the measurements

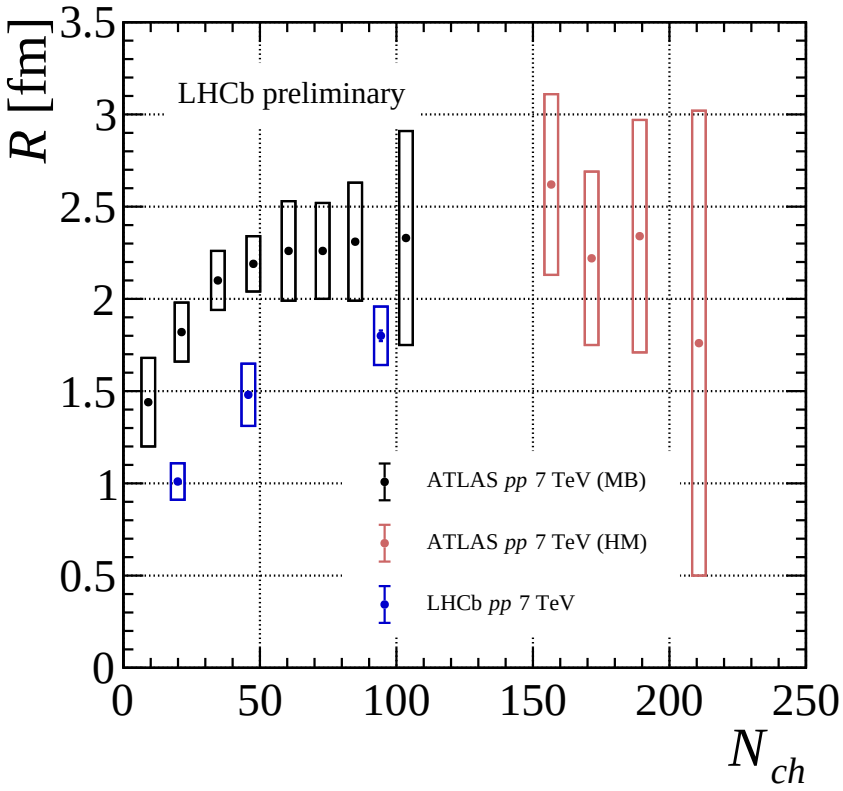

Figure 5.3: Correlation radii as a function of the unfolded charged-particle multiplicity, measured in the central pseudorapidity region (black and red dots) by the ATLAS experiment [144,246] and in the forward direction (blue dots) with the LHCb detector [17]. Values of the charged-particle multiplicities from the LHCb experiment are extrapolated to the ATLAS detector acceptance using the simulation (see description in the text for details). For the LHCb data points, error bars indicate the statistical uncertainties, while boxes illustrate the systematic ones. In the ATLAS case, boxes correspond to a quadratic sum of the statistical and the systematic uncertainties.

in the central rapidity range. A direct comparison of the correlation parameters determined in different experiments is not straightforward due to the nonidentical detector acceptances and the variations in the multiplicity determination methods. In the case of the ATLAS results for pp collisions at $\sqrt{s} = 7\,\text{TeV}$ [144] an unfolded charged-particle multiplicity is provided, so such a comparison seems feasible. A stand-alone Pythia8.1 simulation (without any LHCb-specific tuning) is used to find the relation between the unfolded charged-particle multiplicity in pp collisions at the LHCb ($2.0 < \eta < 5.0$, $p_{\text{T}} > 0.1\,\text{GeV}$) and ATLAS ($|\eta| < 2.5$, $p_{\text{T}} > 0.1\,\text{GeV}$) acceptances. This information is then employed to find the charged-particle multiplicity values in the ATLAS acceptance that correspond to the charged-particle multiplicity bins from the current pp study, and to superimpose the results, as shown in Fig. 5.3. The correlation radii measured at LHCb are systematically lower than those observed with the ATLAS detector in the corresponding charged-particle multiplicity regions.

Chapter 6

Summary and conclusions

The present thesis describes a study of the Bose–Einstein correlations for same-sign charged pions performed in the LHCb experiment with the pp and the pPb collisions at $\sqrt{s} = 7\,\text{TeV}$ centre-of-mass energy and $\sqrt{s_{\text{NN}}} = 5.02\,\text{TeV}$ centre-of-mass energy per nucleon, respectively (the results presented for the pPb system are preliminary). The correlation radii and the intercept parameters are determined in bins of the reconstructed charged-particle multiplicity covering values up to ~ 180. In the pp study, the unfolded charged-particle multiplicity ranges are also provided. Both analyses employ an event-mixing method for constructing the reference sample and assume a Lévy-type parametrisation of the correlation function with the index of stability fixed to unity. Two alternative strategies for the treatment of the nonfemtoscopic background effects are used. In the pp case, a double-ratio method is implemented to correct for such phenomena using the simulation. The pPb study employs a data-driven approach to parametrise the so-called cluster contribution based on the correlation functions constructed for opposite-sign charged pions.

The described measurements are the first of this type performed in the forward region at the LHC energies and are complementary to studies from other experiments at the LHC in the central rapidity region. In general, similar behaviour of the correlation parameters is observed. In particular, the measured correlation radii scale linearly with a cube root of the reconstructed charged-particle multiplicity, which is a tendency compatible with predictions of hydrodynamic models on the system evolution. The correlation radii determined for the pp and the pPb collisions seem to have comparable values in similar multiplicity ranges, but further conclusions in this area are restricted due to a limited number of multiplicity bins in the pp study and relatively large systematic uncertainties. The pPb system is investigated both in the forward and the backward direction due to asymmetric beams and hints for a potential sensitivity of the correlation radii to pseudorapidity are observed. This relationship is further explored by comparing the LHCb results for the pp collisions to a similar study in the central rapidity region. The correlation radii measured in the forward direction seem to be systematically smaller than the ones determined in the central rapidity region.

The presented studies provide an exclusive input for the development of phenomenological models describing the process of multiparticle production by supplying information on this phenomenon in a unique acceptance region of the LHCb detector. The measured trends, including the potential dependence of the correlation radii on pseudorapidity, can enable the addition of further constraints on the available models. It could contribute to the design of unified frameworks for modelling the pp, the pA and the AA collisions,

which is a perspective motivated by the observation of a significant number of similarities between these systems. One of the interesting prospects for future BEC studies at LHCb is the analysis using pp collisions at $\sqrt{s} = 13\,\text{TeV}$ centre-of-mass energy, which could also be potentially extended to include the k_T bins. A similar approach with added k_T bins could be taken to investigate the pPb system in more detail by employing collisions recorded in the LHCb experiment at $\sqrt{s_\text{NN}} = 8.16\,\text{TeV}$ centre-of-mass energy per nucleon. Other possibilities include multidimensional BEC analyses as well as performing measurements for the PbPb collisions.

This thesis also reports developments in the simulation of the LHCb RICH detectors for the Upgrade Ia period. In particular, modelling of specific sensor noise and accurate detector time-response has been implemented by the author. Additional methods have been designed to allow the assessment of the performance of various configurations of the RICH system using the simulation. The mentioned developments enable detailed modelling of the RICH detectors in Upgrade Ia, and they can also be potentially employed for research on the future upgrades. Studies conducted using the developed tools suggest that the modernised RICH detectors well satisfy their nominal performance requirements. Preliminary investigations also indicate that the employed strategies for mitigating the detector noise are very effective. It should allow for optimal detector operation in the Run3 period.

Appendix A

Details on the data and the simulation samples

Details on the data and the simulation samples used in the studies described in this thesis are given below. The official LHCb samples are managed centrally via an experiment *bookkeeping* system [235]. Auxiliary simulation samples were produced locally using the PLGrid Infrastructure [247].

A.1 Proton-proton data and simulation samples

Data-processing steps and storage details of the samples used in the *pp* study are described in this section. The data sample was obtained from a minimum-bias set of stripping lines (Reco14, Stripping20r1), while the simulation samples were produced using configuration Sim08 with Reco14.

The data sample:

- minimum-bias set of stripping lines (40M)

  40,104,766 minimum-bias data events in bookkeeping:

  LHCb/Collision11/Beam3500GeV-VeloClosed-MagDown/RealData/Reco14/90000000/MINIBIAS.DST

  LHCb/Collision11/Beam3500GeV-VeloClosed-MagUp/RealData/Reco14/90000000/MINIBIAS.DST

The simulation samples:

- Pythia8.1 MC2011 minimum-bias (event type = 30000000) (20M)

  20,057,476 MC2011 Sim08c Reco14a minimum-bias events in bookkeeping:

  MC/2011/Beam3500GeV-2011-MagDown-Nu2-Pythia8/Sim08c/Digi13/Trig0x40760037/Reco14a/Stripping20r1NoPrescalingFlagged/30000000/STREAMSDST

  MC/2011/Beam3500GeV-2011-MagUp-Nu2-Pythia8/Sim08c/Digi13/Trig0x40760037/Reco14a/Stripping20r1NoPrescalingFlagged/30000000/STREAMSDST

The samples were produced using Gauss v45r3, Boole v26r9 and Brunel v44r5. Packages Moore v14r8p1 and DaVinci v36r7 were used for the sample processing.

- (local) PYTHIA6.4 PERUGIA0 tuning MC2011 minimum-bias (event type = 30000320) (10M *MagDown*)

  10,000,000 MC2011 Sim08 Reco14 *MagDown* minimum-bias events simulated and stored locally using the PLGrid Infrastructure.

  The samples were produced with the same conditions as the PYTHIA8.1 MC2011 minimum-bias sample using Gauss v45r3, Boole v26r9 and Brunel v44r5.

- (local) HERWIG++ 2.7.1 MC2011 minimum-bias (event type = 30000000) (10M *MagDown*)

  10,000,000 MC2011 Sim08 Reco14 *MagDown* minimum-bias events simulated and stored locally using the PLGrid Infrastructure.

  The samples were produced with the same conditions as the PYTHIA8.1 MC2011 minimum-bias sample using Gauss v49r2, Boole v30r2 and Brunel v51r1.

A.2 Proton-lead data and simulation samples

Data-processing steps and storage details of the samples used in the pPb study are described in this section. The data samples were obtained from a *full* set of stripping lines (Reco14r1, Stripping20r3) with the following stripping lines to select minimum-bias events (together with a micro-bias trigger):

- StrippingIonProtonMagDownMinBiasLine
- StrippingIonProtonMagUpMinBiasLine
- StrippingProtonIonMagDownMinBiasLine
- StrippingProtonIonMagUpMinBiasLine.

The simulation samples were produced using configuration Digi14a (Sim09c, Reco14r1 and Stripping20r3p1). The stripping versions differ between the data and the simulation, since the minimum-bias lines were not included in the restripping campaign of Stripping20r3p1. The stripping for the data was done using DaVinci v32r2p10. EPOS MC2013 simulation samples were produced using Boole v30r2, Moore v20r4 and v14r12, Brunel v43r2p12 and DaVinci v32r2p18. Data analysis was performed using DaVinci v36r7.

The data samples:

- *full* set of stripping lines pPb (197M)

  197,258,115 events in bookkeeping:

  LHCb/Protonion13/Beam4000GeV-VeloClosed-MagDown/RealData/Reco14r1/Stripping20r3/90000000/ALL.DST

  LHCb/Protonion13/Beam4000GeV-VeloClosed-MagUp/RealData/Reco14r1/Stripping20r3/90000000/ALL.DST

- *full* set of stripping lines Pbp (211M)

  210,958,381 events in bookkeeping:

  LHCb/Ionproton13/Beam4000GeV-VeloClosed-MagDown/RealData/Reco14r1/Stripping20r3/90000000/ALL.DST

  LHCb/Ionproton13/Beam4000GeV-VeloClosed-MagUp/RealData/Reco14r1/Stripping20r3/90000000/ALL.DST

The simulation samples:

- EPOS MC2013 minimum-bias pPb (event type = 30000000) (10M)

  10,071,739 EPOS 2013 Sim09c Reco14r1 minimum-bias events in bookkeeping:

  MC/2013/pPb-Beam4000GeV-1580GeV-2013-MagDown-Fix1-Epos/Sim09c/L0Trig0x1710/Trig0x406a1710/Reco14r1/Stripping20r3p1/30000000/ALLSTREAMS.DST

  MC/2013/pPb-Beam4000GeV-1580GeV-2013-MagUp-Fix1-Epos/Sim09c/L0Trig0x1710/Trig0x406a1710/Reco14r1/Stripping20r3p1/30000000/ALLSTREAMS.DST

- EPOS MC2013 minimum-bias Pbp (event type = 30000000) (10M)

  10,167,211 EPOS 2013 Sim09c Reco14r1 minimum-bias events in bookkeeping:

  MC/2013/Pbp-Beam1580GeV-4000GeV-2013-MagDown-Fix1-Epos/Sim09c/L0Trig0x1710/Trig0x406a1710/Reco14r1/Stripping20r3p1/30000000/ALLSTREAMS.DST

  MC/2013/Pbp-Beam1580GeV-4000GeV-2013-MagUp-Fix1-Epos/Sim09c/L0Trig0x1710/Trig0x406a1710/Reco14r1/Stripping20r3p1/30000000/ALLSTREAMS.DST

- (local) EPOS MC2013 minimum-bias pPb and Pbp (event type = 30000000) (10M pPb + 10M Pbp)

  $2 \times$ 10,000,000 EPOS 2013 Sim09c Reco14r1 (*MagDown* and *MagUp*) minimum-bias events simulated and stored locally using the PLGrid Infrastructure.

  The samples were produced with the same conditions as the EPOS MC2013 minimum-bias sample using Gauss v49r8, Boole v30r2, Brunel v43r2p12 and DaVinci v32r2p18.

Bibliography

[1] E. V. Shuryak, *Quantum Chromodynamics and the Theory of Superdense Matter*, Phys. Rept. **61** (1980) 71.

[2] T. Csorgo, *Particle interferometry from 40 MeV to 40 TeV*, Acta Phys. Hung. A **15** (2002) 1, arXiv:hep-ph/0001233.

[3] B. Schenke and R. Venugopalan, *Eccentric protons? Sensitivity of flow to system size and shape in p+p, p+Pb and Pb+Pb collisions*, Phys. Rev. Lett. **113** (2014) 102301, arXiv:1405.3605.

[4] P. Romatschke, *Light-Heavy Ion Collisions: A window into pre-equilibrium QCD dynamics?*, Eur. Phys. J. C **75** (2015) 305, arXiv:1502.04745.

[5] P. Romatschke and U. Romatschke, *Relativistic Fluid Dynamics In and Out of Equilibrium*, Cambridge Monographs on Mathematical Physics, Cambridge University Press, 2019.

[6] R. Hanbury Brown and R. Q. Twiss, *A New type of interferometer for use in radio astronomy*, Phil. Mag. Ser. 7 **45** (1954) 663.

[7] R. Hanbury Brown and R. Q. Twiss, *Correlation between Photons in two Coherent Beams of Light*, Nature **177** (1956) 27.

[8] R. Hanbury Brown and R. Q. Twiss, *A Test of a new type of stellar interferometer on Sirius*, Nature **178** (1956) 1046.

[9] G. Goldhaber, S. Goldhaber, W.-Y. Lee, and A. Pais, *Influence of Bose-Einstein statistics on the anti-proton proton annihilation process*, Phys. Rev. **120** (1960) 300.

[10] G. Goldhaber, W. B. Fowler, S. Goldhaber, and T. F. Hoang, *Pion-pion correlations in antiproton annihilation events*, Phys. Rev. Lett. **3** (1959) 181.

[11] M. A. Lisa, S. Pratt, R. Soltz, and U. Wiedemann, *Femtoscopy in relativistic heavy ion collisions*, Ann. Rev. Nucl. Part. Sci. **55** (2005) 357, arXiv:nucl-ex/0505014.

[12] A. N. Makhlin and Y. M. Sinyukov, *Hydrodynamics of Hadron Matter Under Pion Interferometric Microscope*, Z. Phys. C **39** (1988) 69.

[13] Y. M. Sinyukov, *Spectra and correlations in locally equilibrium hadron and quark - gluon systems*, Nucl. Phys. A **566** (1994) 589C.

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

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

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

[17] LHCb collaboration, R. Aaij *et al.*, *Bose–Einstein correlations of same-sign charged pions in the forward region in pp collisions at* $\sqrt{s} = 7\,TeV$, JHEP **12** (2017) 025, arXiv:1709.01769.

[18] M. Kucharczyk and B. Malecki, *Study of the Bose–Einstein correlations of identical pions in p*Pb *collisions*, LHCb-ANA-2021-046, 2021.

[19] LHCb collaboration, B. Malecki, *Two-particle correlations in p-Pb collisions at the LHCb*, Acta Phys. Polon. B **48** (2017) 1067.

[20] LHCb collaboration, B. Malecki, *Bose-Einstein correlations and* $b\bar{b}$ *correlations in pp collisions with LHCb*, Nucl. Phys. A **982** (2019) 347, arXiv:1809.09898.

[21] LHCb collaboration, B. Malecki, *Two-particle correlations at the LHCb*, Acta Phys. Polon. B Proc. Suppl. **12** (2019) 223.

[22] LHCb collaboration, B. Malecki, *Bose–Einstein Correlations in pp and pPb Collisions at LHCb*, Universe **5** (2019) 95.

[23] *CERN laboratory*, https://home.cern/, accessed: 2021-02-26.

[24] M. Andreotti *et al.*, *Characterisation of signal-induced noise in Hamamatsu R11265 Multianode Photomultiplier Tubes*, arXiv:2110.00831, submitted to JINST.

[25] E. V. Shuryak, *Quark-Gluon Plasma and Hadronic Production of Leptons, Photons and Psions*, Phys. Lett. B **78** (1978) 150.

[26] PHENIX collaboration, K. Adcox *et al.*, *Formation of dense partonic matter in relativistic nucleus-nucleus collisions at RHIC: Experimental evaluation by the PHENIX collaboration*, Nucl. Phys. A **757** (2005) 184, arXiv:nucl-ex/0410003.

[27] BRAHMS collaboration, I. Arsene *et al.*, *Quark gluon plasma and color glass condensate at RHIC? The Perspective from the BRAHMS experiment*, Nucl. Phys. A **757** (2005) 1, arXiv:nucl-ex/0410020.

[28] PHOBOS collaboration, B. B. Back *et al.*, *The PHOBOS perspective on discoveries at RHIC*, Nucl. Phys. A **757** (2005) 28, arXiv:nucl-ex/0410022.

[29] STAR collaboration, J. Adams *et al.*, *Experimental and theoretical challenges in the search for the quark gluon plasma: The STAR Collaboration's critical assessment of the evidence from RHIC collisions*, Nucl. Phys. A **757** (2005) 102, arXiv:nucl-ex/0501009.

[30] ALICE collaboration, K. Aamodt *et al.*, *Elliptic flow of charged particles in Pb-Pb collisions at 2.76 TeV*, Phys. Rev. Lett. **105** (2010) 252302, arXiv:1011.3914.

[31] B. Muller, J. Schukraft, and B. Wyslouch, *First Results from Pb+Pb collisions at the LHC*, Ann. Rev. Nucl. Part. Sci. **62** (2012) 361, arXiv:1202.3233.

[32] K. Dusling, W. Li, and B. Schenke, *Novel collective phenomena in high-energy proton–proton and proton–nucleus collisions*, Int. J. Mod. Phys. E **25** (2016) 1630002, arXiv:1509.07939.

[33] J. L. Nagle and W. A. Zajc, *Small System Collectivity in Relativistic Hadronic and Nuclear Collisions*, Ann. Rev. Nucl. Part. Sci. **68** (2018) 211, arXiv:1801.03477.

[34] PHENIX collaboration, K. Adcox *et al.*, *Suppression of hadrons with large transverse momentum in central Au+Au collisions at* $\sqrt{s_{NN}}$ *= 130 GeV*, Phys. Rev. Lett. **88** (2002) 022301, arXiv:nucl-ex/0109003.

[35] CMS collaboration, V. Khachatryan *et al.*, *Charged-particle nuclear modification factors in PbPb and pPb collisions at* $\sqrt{s_{\mathrm{N\,N}}} = 5.02$ *TeV*, JHEP **04** (2017) 039, arXiv:1611.01664.

[36] U. Heinz and R. Snellings, *Collective flow and viscosity in relativistic heavy-ion collisions*, Ann. Rev. Nucl. Part. Sci. **63** (2013) 123, arXiv:1301.2826.

[37] PHENIX collaboration, A. Adare *et al.*, *Energy Loss and Flow of Heavy Quarks in Au+Au Collisions at* $\sqrt{s_{NN}}$ *= 200 GeV*, Phys. Rev. Lett. **98** (2007) 172301, arXiv:nucl-ex/0611018.

[38] P. Kovtun, D. T. Son, and A. O. Starinets, *Viscosity in strongly interacting quantum field theories from black hole physics*, Phys. Rev. Lett. **94** (2005) 111601, arXiv:hep-th/0405231.

[39] D. T. Son and A. O. Starinets, *Viscosity, Black Holes, and Quantum Field Theory*, Ann. Rev. Nucl. Part. Sci. **57** (2007) 95, arXiv:0704.0240.

[40] CMS collaboration, V. Khachatryan *et al.*, *Observation of Long-Range Near-Side Angular Correlations in Proton-Proton Collisions at the LHC*, JHEP **09** (2010) 091, arXiv:1009.4122.

[41] ATLAS collaboration, G. Aad *et al.*, *Observation of Long-Range Elliptic Azimuthal Anisotropies in* $\sqrt{s}$ *=13 and 2.76 TeV pp Collisions with the ATLAS Detector*, Phys. Rev. Lett. **116** (2016) 172301, arXiv:1509.04776.

[42] CMS collaboration, V. Khachatryan *et al.*, *Measurement of long-range near-side two-particle angular correlations in pp collisions at* $\sqrt{s}$ *=13 TeV*, Phys. Rev. Lett. **116** (2016) 172302, arXiv:1510.03068.

[43] CMS collaboration, V. Khachatryan *et al.*, *Evidence for collectivity in pp collisions at the LHC*, Phys. Lett. B **765** (2017) 193, arXiv:1606.06198.

[44] ALICE collaboration, B. Abelev *et al.*, *Long-range angular correlations on the near and away side in p-Pb collisions at* $\sqrt{s_{NN}} = 5.02$ *TeV*, Phys. Lett. B **719** (2013) 29, arXiv:1212.2001.

[45] ATLAS collaboration, G. Aad *et al.*, *Observation of Associated Near-Side and Away-Side Long-Range Correlations in $\sqrt{s_{NN}}$=5.02 TeV Proton-Lead Collisions with the ATLAS Detector*, Phys. Rev. Lett. **110** (2013) 182302, arXiv:1212.5198.

[46] CMS collaboration, S. Chatrchyan *et al.*, *Observation of Long-Range Near-Side Angular Correlations in Proton-Lead Collisions at the LHC*, Phys. Lett. B **718** (2013) 795, arXiv:1210.5482.

[47] LHCb collaboration, R. Aaij *et al.*, *Measurement of two-particle correlations in proton-ion collisions at $\sqrt{s_{NN}}$ =5 TeV*, Phys. Lett. **B762** (2016) 473, arXiv:1512.00439.

[48] PHENIX collaboration, A. Adare *et al.*, *Quadrupole Anisotropy in Dihadron Azimuthal Correlations in Central d+Au Collisions at $\sqrt{s_{NN}}$=200 GeV*, Phys. Rev. Lett. **111** (2013) 212301, arXiv:1303.1794.

[49] C. Loizides, *Experimental overview on small collision systems at the LHC*, Nucl. Phys. A **956** (2016) 200, arXiv:1602.09138.

[50] L. D. Landau, *On the multiparticle production in high-energy collisions*, Izv. Akad. Nauk Ser. Fiz. **17** (1953) 51.

[51] W. Florkowski, M. P. Heller, and M. Spalinski, *New theories of relativistic hydrodynamics in the LHC era*, Rept. Prog. Phys. **81** (2018) 046001, arXiv:1707.02282.

[52] P. Romatschke, *Do nuclear collisions create a locally equilibrated quark–gluon plasma?*, Eur. Phys. J. C **77** (2017) 21, arXiv:1609.02820.

[53] M. Spalinski, *Small systems and regulator dependence in relativistic hydrodynamics*, Phys. Rev. D **94** (2016) 085002, arXiv:1607.06381.

[54] J. L. Nagle *et al.*, *Minimal conditions for collectivity in e^+e^- and $p+p$ collisions*, Phys. Rev. C **97** (2018) 024909, arXiv:1707.02307.

[55] E. Iancu and R. Venugopalan, in *The Color glass condensate and high-energy scattering in QCD*, R. C. Hwa and X.-N. Wang, eds., 2003. arXiv:hep-ph/0303204.

[56] F. Gelis, E. Iancu, J. Jalilian-Marian, and R. Venugopalan, *The Color Glass Condensate*, Ann. Rev. Nucl. Part. Sci. **60** (2010) 463, arXiv:1002.0333.

[57] L. D. McLerran and R. Venugopalan, *Gluon distribution functions for very large nuclei at small transverse momentum*, Phys. Rev. D **49** (1994) 3352, arXiv:hep-ph/9311205.

[58] S. Pratt, E. Sangaline, P. Sorensen, and H. Wang, *Constraining the Eq. of State of Super-Hadronic Matter from Heavy-Ion Collisions*, Phys. Rev. Lett. **114** (2015) 202301, arXiv:1501.04042.

[59] S. A. Bass, J. E. Bernhard, and J. S. Moreland, *Determination of Quark-Gluon-Plasma Parameters from a Global Bayesian Analysis*, Nucl. Phys. A **967** (2017) 67, arXiv:1704.07671.

[60] J. E. Bernhard, J. S. Moreland, and S. A. Bass, *Bayesian estimation of the specific shear and bulk viscosity of quark–gluon plasma*, Nature Phys. **15** (2019) 1113.

[61] JETSCAPE collaboration, D. Everett *et al.*, *Multi-system Bayesian constraints on the transport coefficients of QCD matter*, Phys. Rev. C **103** (2021) 054904, arXiv:2011.01430.

[62] R. D. Weller and P. Romatschke, *One fluid to rule them all: viscous hydrodynamic description of event-by-event central p+p, p+Pb and Pb+Pb collisions at* $\sqrt{s} = 5.02$ *TeV*, Phys. Lett. B **774** (2017) 351, arXiv:1701.07145.

[63] J. L. Nagle *et al.*, *Exploiting Intrinsic Triangular Geometry in Relativistic He3+Au Collisions to Disentangle Medium Properties*, Phys. Rev. Lett. **113** (2014) 112301, arXiv:1312.4565.

[64] PHENIX collaboration, A. Adare *et al.*, *Measurement of long-range angular correlation and quadrupole anisotropy of pions and (anti)protons in central d+Au collisions at* $\sqrt{s_{_{NN}}}$*=200 GeV*, Phys. Rev. Lett. **114** (2015) 192301, arXiv:1404.7461.

[65] PHENIX collaboration, A. Adare *et al.*, *Measurements of elliptic and triangular flow in high-multiplicity* ^{3}*He+Au collisions at* $\sqrt{s_{_{NN}}} = 200$ *GeV*, Phys. Rev. Lett. **115** (2015) 142301, arXiv:1507.06273.

[66] PHENIX collaboration, C. Aidala *et al.*, *Measurement of long-range angular correlations and azimuthal anisotropies in high-multiplicity p+Au collisions at* $\sqrt{s_{_{NN}}} = 200$ *GeV*, Phys. Rev. C **95** (2017) 034910, arXiv:1609.02894.

[67] PHENIX collaboration, C. Aidala *et al.*, *Creation of quark–gluon plasma droplets with three distinct geometries*, Nature Phys. **15** (2019) 214, arXiv:1805.02973.

[68] B. Schenke, *Origins of collectivity in small systems*, Nucl. Phys. A **967** (2017) 105, arXiv:1704.03914.

[69] B. Schenke, C. Shen, and P. Tribedy, *Hybrid Color Glass Condensate and hydrodynamic description of the Relativistic Heavy Ion Collider small system scan*, Phys. Lett. B **803** (2020) 135322, arXiv:1908.06212.

[70] W. Kittel, *Bose-Einstein correlations in Z fragmentation and other reactions*, Acta Phys. Polon. B **32** (2001) 3927, arXiv:hep-ph/0110088.

[71] M. Gyulassy, S. K. Kauffmann, and L. W. Wilson, *Pion Interferometry of Nuclear Collisions. 1. Theory*, Phys. Rev. C **20** (1979) 2267.

[72] S. Pratt, T. Csorgo, and J. Zimanyi, *Detailed predictions for two pion correlations in ultrarelativistic heavy ion collisions*, Phys. Rev. C **42** (1990) 2646.

[73] S. Chapman and U. W. Heinz, *HBT correlators: Current formalism versus Wigner function formulation*, Phys. Lett. B **340** (1994) 250, arXiv:hep-ph/9407405.

[74] T. Csorgo, B. Lorstad, and J. Zimanyi, *Bose-Einstein correlations for systems with large halo*, Z. Phys. C **71** (1996) 491, arXiv:hep-ph/9411307.

[75] J. Bolz *et al.*, *Resonance decays and partial coherence in Bose-Einstein correlations*, Phys. Rev. D **47** (1993) 3860.

[76] T. Csorgo, S. Hegyi, and W. A. Zajc, *Bose–Einstein correlations for Lévy stable source distributions*, Eur. Phys. J. C **36** (2004) 67, arXiv:nucl-th/0310042.

[77] T. Csorgo, B. Lorstad, J. Schmid-Sorensen, and A. Ster, *Partial coherence in the core / halo picture of Bose-Einstein n particle correlations*, Eur. Phys. J. C **9** (1999) 275, arXiv:hep-ph/9812422.

[78] PHENIX collaboration, A. Adare *et al.*, *Lévy-stable two-pion Bose-Einstein correlations in* $\sqrt{s_{NN}} = 200$ *GeV Au+Au collisions*, Phys. Rev. C **97** (2018) 064911, arXiv:1709.05649.

[79] T. Csorgo, S. Hegyi, T. Novak, and W. A. Zajc, *Bose-Einstein or HBT correlations and the anomalous dimension of QCD*, Acta Phys. Polon. B **36** (2005) 329, arXiv:hep-ph/0412243.

[80] T. Csorgo, S. Hegyi, T. Novak, and W. A. Zajc, *Bose-Einstein or HBT correlation signature of a second order QCD phase transition*, AIP Conf. Proc. **828** (2006) 525, arXiv:nucl-th/0512060.

[81] M. Csanad, T. Csorgo, and M. Nagy, *Anomalous diffusion of pions at RHIC*, Braz. J. Phys. **37** (2007) 1002, arXiv:hep-ph/0702032.

[82] R. Metzler and J. Klafter, *The random walk's guide to anomalous diffusion: a fractional dynamics approach*, Physics Reports **339** (2000) 1.

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

[84] T. Sjostrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175.

[85] J. D. Bjorken, *Highly Relativistic Nucleus-Nucleus Collisions: The Central Rapidity Region*, Phys. Rev. D **27** (1983) 140.

[86] M. L. Miller, K. Reygers, S. J. Sanders, and P. Steinberg, *Glauber modeling in high energy nuclear collisions*, Ann. Rev. Nucl. Part. Sci. **57** (2007) 205, arXiv:nucl-ex/0701025.

[87] C. Loizides, *Glauber modeling of high-energy nuclear collisions at the subnucleon level*, Phys. Rev. C **94** (2016) 024914, arXiv:1603.07375.

[88] B. Schenke, P. Tribedy, and R. Venugopalan, *Fluctuating Glasma initial conditions and flow in heavy ion collisions*, Phys. Rev. Lett. **108** (2012) 252301, arXiv:1202.6646.

[89] C. Shen, *Observable signatures of initial state momentum anisotropies in nuclear collisions*, https://indico.cern.ch/event/981321/contributions/4133153 (with modifications to the captions), 2020.

[90] Y. Aoki *et al.*, *The Order of the quantum chromodynamics transition predicted by the standard model of particle physics*, Nature **443** (2006) 675, arXiv:hep-lat/0611014.

[91] Y. Aoki *et al.*, *The QCD transition temperature: results with physical masses in the continuum limit II.*, JHEP **06** (2009) 088, arXiv:0903.4155.

[92] F. Cooper and G. Frye, *Comment on the Single Particle Distribution in the Hydrodynamic and Statistical Thermodynamic Models of Multiparticle Production*, Phys. Rev. D **10** (1974) 186.

[93] S. A. Bass *et al.*, *Microscopic models for ultrarelativistic heavy ion collisions*, Prog. Part. Nucl. Phys. **41** (1998) 255, arXiv:nucl-th/9803035.

[94] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, *Parton Fragmentation and String Dynamics*, Phys. Rept. **97** (1983) 31.

[95] E802 collaboration, L. Ahle *et al.*, *System, centrality, and transverse mass dependence of two pion correlation radii in heavy ion collisions at 11.6-A-GeV and 14.6-A-GeV*, Phys. Rev. C **66** (2002) 054906, arXiv:nucl-ex/0204001.

[96] Axial Field Spectrometer collaboration, T. Akesson *et al.*, *Bose-einstein Correlations in $\alpha\alpha$, pp and $p\bar{p}$ Interactions*, Phys. Lett. B **129** (1983) 269.

[97] Axial Field Spectrometer collaboration, T. Akesson *et al.*, *Bose–Einstein correlations between kaons*, Phys. Lett. B **155** (1985) 128.

[98] Axial Field Spectrometer collaboration, T. Akesson *et al.*, *Evidence for a Directional Dependence of Bose-Einstein Correlations at the CERN Intersecting Storage Rings*, Phys. Lett. B **187** (1987) 420.

[99] UA1 collaboration, C. Albajar *et al.*, *Bose-Einstein Correlations in $\bar{p}p$ Interactions at $\sqrt{s}$ = 0.2 to 0.9 TeV*, Phys. Lett. B **226** (1989) 410, [Erratum: Phys.Lett.B 229, 439 (1989)].

[100] NA44 collaboration, H. Boggild *et al.*, *Directional dependence of the pion source in high-energy heavy ion collisions*, Phys. Lett. B **349** (1995) 386.

[101] CERES collaboration, D. Adamova *et al.*, *Beam energy and centrality dependence of two pion Bose-Einstein correlations at SPS energies*, Nucl. Phys. A **714** (2003) 124, arXiv:nucl-ex/0207005.

[102] ALEPH collaboration, D. Decamp *et al.*, *A Study of Bose-Einstein correlations in e+ e- annihilation at 91 GeV*, Z. Phys. C **54** (1992) 75.

[103] ALEPH collaboration, D. Buskulic *et al.*, *Production of K0 and Λ in hadronic Z decays*, Z. Phys. C **64** (1994) 361.

[104] ALEPH collaboration, R. Barate *et al.*, *Fermi-Dirac Correlations in Λ pairs in hadronic Z decays*, Phys. Lett. B **475** (2000) 395.

[105] ALEPH collaboration, A. Heister *et al.*, *Two-dimensional analysis of Bose-Einstein correlations in hadronic Z decays at LEP*, Eur. Phys. J. C **36** (2004) 147.

[106] DELPHI collaboration, P. Abreu *et al.*, *Bose-Einstein correlations in the hadronic decays of the Z0*, Phys. Lett. B **286** (1992) 201.

[107] DELPHI collaboration, P. Abreu *et al.*, *Kaon interference in the hadronic decays of the Z0*, Phys. Lett. B **379** (1996) 330.

[108] DELPHI collaboration, P. Abreu *et al.*, *Two-dimensional analysis of the Bose-Einstein correlations in e+ e- annihilation at the Z0 peak*, Phys. Lett. B **471** (2000) 460.

[109] L3 collaboration, M. Acciarri *et al.*, *Measurement of an elongation of the pion source in Z decays*, Phys. Lett. B **458** (1999) 517, `arXiv:hep-ex/9909009`.

[110] L3 collaboration, P. Achard *et al.*, *Bose-Einstein correlations of neutral and charged pions in hadronic Z decays*, Phys. Lett. B **524** (2002) 55, `arXiv:hep-ex/0109036`.

[111] L3 collaboration, P. Achard *et al.*, *Test of the τ-Model of Bose-Einstein Correlations and Reconstruction of the Source Function in Hadronic Z-boson Decay at LEP*, Eur. Phys. J. C **71** (2011) 1648, `arXiv:1105.4788`.

[112] OPAL collaboration, G. Abbiendi *et al.*, *Bose-Einstein correlations in K+- K+- pairs from Z0 decays into two hadronic jets*, Eur. Phys. J. C **21** (2001) 23, `arXiv:hep-ex/0001045`.

[113] OPAL collaboration, R. Akers *et al.*, *The Production of neutral kaons in Z0 decays and their Bose-Einstein correlations*, Z. Phys. C **67** (1995) 389.

[114] OPAL collaboration, G. Alexander *et al.*, *A First measurement of the $\Lambda\overline{\Lambda}$ and $\Lambda\Lambda$ ($\overline{\Lambda}\overline{\Lambda}$) spin compositions in hadronic Z0 decays*, Phys. Lett. B **384** (1996) 377.

[115] OPAL collaboration, G. Alexander *et al.*, *Multiplicity dependence of Bose-Einstein correlations in hadronic Z0 decays*, Z. Phys. C **72** (1996) 389.

[116] OPAL collaboration, G. Abbiendi *et al.*, *Transverse and longitudinal Bose Einstein correlations in hadronic Z0 decays*, Eur. Phys. J. C **16** (2000) 423, `arXiv:hep-ex/0002062`.

[117] OPAL collaboration, G. Abbiendi *et al.*, *Bose-Einstein correlations of π^0 pairs from hadronic Z0 decays*, Phys. Lett. B **559** (2003) 131, `arXiv:hep-ex/0302027`.

[118] STAR collaboration, C. Adler *et al.*, *Pion interferometry of $\sqrt{s_{NN}}$ = 130 GeV Au+Au collisions at RHIC*, Phys. Rev. Lett. **87** (2001) 082301, `arXiv:nucl-ex/0107008`.

[119] STAR collaboration, J. Adams *et al.*, *Three pion HBT correlations in relativistic heavy ion collisions from the STAR experiment*, Phys. Rev. Lett. **91** (2003) 262301, `arXiv:nucl-ex/0306028`.

[120] STAR collaboration, J. Adams *et al.*, *Azimuthally sensitive HBT in Au + Au collisions at $\sqrt{s_{NN}}$ = 200 GeV*, Phys. Rev. Lett. **93** (2004) 012301, `arXiv:nucl-ex/0312009`.

[121] STAR collaboration, J. Adams *et al.*, *Pion interferometry in Au+Au collisions at* $\sqrt{s_{NN}}$ *= 200 GeV*, Phys. Rev. C **71** (2005) 044906, arXiv:nucl-ex/0411036.

[122] STAR collaboration, B. I. Abelev *et al.*, *Pion Interferometry in Au+Au and Cu+Cu Collisions at RHIC*, Phys. Rev. C **80** (2009) 024905, arXiv:0903.1296.

[123] STAR collaboration, L. Adamczyk *et al.*, *Freeze-out dynamics via charged kaon femtoscopy in* $\sqrt{s_{NN}}$ *= 200 GeV central Au + Au collisions*, Phys. Rev. C **88** (2013) 034906, arXiv:1302.3168.

[124] STAR collaboration, L. Adamczyk *et al.*, *Beam-energy-dependent two-pion interferometry and the freeze-out eccentricity of pions measured in heavy ion collisions at the STAR detector*, Phys. Rev. C **92** (2015) 014904, arXiv:1403.4972.

[125] STAR collaboration, M. M. Aggarwal *et al.*, *Pion femtoscopy in* p^+p *collisions at* $\sqrt{s} = 200$ *GeV*, Phys. Rev. C **83** (2011) 064905, arXiv:1004.0925.

[126] PHENIX collaboration, S. S. Adler *et al.*, *Bose-Einstein correlations of charged pion pairs in Au + Au collisions at* $\sqrt{s_{NN}}$ *= 200 GeV*, Phys. Rev. Lett. **93** (2004) 152302, arXiv:nucl-ex/0401003.

[127] PHENIX collaboration, S. Afanasiev *et al.*, *Source breakup dynamics in Au+Au Collisions at* $\sqrt{s_{NN}}$ *= 200 GeV via three-dimensional two-pion source imaging*, Phys. Rev. Lett. **100** (2008) 232301, arXiv:0712.4372.

[128] PHENIX collaboration, A. Adare *et al.*, *Systematic study of charged-pion and kaon femtoscopy in Au + Au collisions at* $\sqrt{s_{_{NN}}}$*=200 GeV*, Phys. Rev. C **92** (2015) 034914, arXiv:1504.05168.

[129] PHOBOS collaboration, B. B. Back *et al.*, *Transverse momentum and rapidity dependence of HBT correlations in Au + Au collisions at* $\sqrt{s_{NN}}$ *= 62.4 GeV and 200 GeV*, Phys. Rev. C **73** (2006) 031901, arXiv:nucl-ex/0409001.

[130] T. Alexopoulos *et al.*, *A Study of source size in* $p\overline{p}$ *collisions at* $\sqrt{s}$ *= 1.8 TeV using pion interferometry*, Phys. Rev. D **48** (1993) 1931.

[131] H1 collaboration, C. Adloff *et al.*, *Bose-Einstein correlations in deep inelastic ep scattering at HERA*, Z. Phys. C **75** (1997) 437, arXiv:hep-ex/9705001.

[132] CMS collaboration, V. Khachatryan *et al.*, *First Measurement of Bose-Einstein Correlations in Proton-Proton Collisions at* $\sqrt{s}$ *=0.9 and 2.36 TeV at the LHC*, Phys. Rev. Lett. **105** (2010) 032001, arXiv:1005.3294.

[133] CMS collaboration, V. Khachatryan *et al.*, *Measurement of Bose-Einstein Correlations in pp Collisions at* $\sqrt{s} = 0.9$ *and 7 TeV*, JHEP **05** (2011) 029, arXiv:1101.3518.

[134] CMS collaboration, A. M. Sirunyan *et al.*, *Bose-Einstein correlations in pp,* pPb*, and PbPb collisions at* $\sqrt{s_{NN}} = 0.9 - 7$ *TeV*, Phys. Rev. C **97** (2018) 064912, arXiv:1712.07198.

[135] CMS collaboration, A. M. Sirunyan *et al.*, *Bose-Einstein correlations of charged hadrons in proton-proton collisions at* $\sqrt{s}$ *= 13 TeV*, JHEP **03** (2020) 014, arXiv:1910.08815.

[136] ALICE collaboration, K. Aamodt *et al.*, *Two-pion Bose-Einstein correlations in pp collisions at* $\sqrt{s} = 900$ *GeV*, Phys. Rev. D **82** (2010) 052001, arXiv:1007.0516.

[137] ALICE collaboration, K. Aamodt *et al.*, *Femtoscopy of pp collisions at* $\sqrt{s} = 0.9$ *and 7 TeV at the LHC with two-pion Bose-Einstein correlations*, Phys. Rev. D **84** (2011) 112004, arXiv:1101.3665.

[138] ALICE collaboration, B. Abelev *et al.*, $K^0_s - K^0_s$ *correlations in pp collisions at* $\sqrt{s} = 7$ *TeV from the LHC ALICE experiment*, Phys. Lett. B **717** (2012) 151, arXiv:1206.2056.

[139] ALICE collaboration, B. Abelev *et al.*, *Charged kaon femtoscopic correlations in pp collisions at* $\sqrt{s} = 7$ *TeV*, Phys. Rev. D **87** (2013) 052016, arXiv:1212.5958.

[140] ALICE collaboration, B. B. Abelev *et al.*, *Freeze-out radii extracted from three-pion cumulants in pp, p–Pb and Pb–Pb collisions at the LHC*, Phys. Lett. B **739** (2014) 139, arXiv:1404.1194.

[141] ALICE collaboration, J. Adam *et al.*, *Multipion Bose-Einstein correlations in pp,p -Pb, and Pb-Pb collisions at energies available at the CERN Large Hadron Collider*, Phys. Rev. C **93** (2016) 054908, arXiv:1512.08902.

[142] ALICE collaboration, S. Acharya *et al.*, *p-p, p-*Λ *and* Λ*-*Λ *correlations studied via femtoscopy in pp reactions at* $\sqrt{s} = 7$ *TeV*, Phys. Rev. C **99** (2019) 024001, arXiv:1805.12455.

[143] ALICE collaboration, S. Acharya *et al.*, *Study of the* Λ*-*Λ *interaction with femtoscopy correlations in pp and p-Pb collisions at the LHC*, Phys. Lett. B **797** (2019) 134822, arXiv:1905.07209.

[144] ATLAS collaboration, G. Aad *et al.*, *Two-particle Bose–Einstein correlations in pp collisions at* $\sqrt{\mathbf{s}} = 0.9$ *and 7 TeV measured with the ATLAS detector*, Eur. Phys. J. C **75** (2015) 466, arXiv:1502.07947.

[145] ALICE collaboration, J. Adam *et al.*, *Two-pion femtoscopy in p-Pb collisions at* $\sqrt{s_{\rm NN}} = 5.02$ *TeV*, Phys. Rev. C **91** (2015) 034906, arXiv:1502.00559.

[146] ALICE collaboration, S. Acharya *et al.*, *One-dimensional charged kaon femtoscopy in p-Pb collisions at* $\sqrt{s_{\rm NN}} = 5.02$ *TeV*, Phys. Rev. C **100** (2019) 024002, arXiv:1903.12310.

[147] ATLAS collaboration, M. Aaboud *et al.*, *Femtoscopy with identified charged pions in proton-lead collisions at* $\sqrt{s_{\rm NN}} = 5.02$ *TeV with ATLAS*, Phys. Rev. C **96** (2017) 064908, arXiv:1704.01621.

[148] ALICE collaboration, K. Aamodt *et al.*, *Two-pion Bose-Einstein correlations in central Pb-Pb collisions at* $\sqrt{s_{NN}} = 2.76$ *TeV*, Phys. Lett. B **696** (2011) 328, arXiv:1012.4035.

[149] ALICE collaboration, J. Adam *et al.*, *One-dimensional pion, kaon, and proton femtoscopy in Pb-Pb collisions at* $\sqrt{s_{\rm NN}}$ *=2.76 TeV*, Phys. Rev. C **92** (2015) 054908, arXiv:1506.07884.

[150] ALICE collaboration, S. Acharya *et al.*, *Kaon femtoscopy in Pb-Pb collisions at* $\sqrt{s_{\rm NN}}$ *= 2.76 TeV*, Phys. Rev. C **96** (2017) 064613, arXiv:1709.01731.

[151] Brookhaven-CERN-Copenhagen-Lund-Rutherford-Tel Aviv collaboration, H. Gordon *et al.*, *The Axial Field Spectrometer at the CERN ISR*, Nucl. Instrum. Meth. **196** (1982) 303.

[152] K. Werner *et al.*, *Evidence for hydrodynamic evolution in proton-proton scattering at 900 GeV*, Phys. Rev. C **83** (2011) 044915, arXiv:1010.0400.

[153] P. Bozek and W. Broniowski, *Size of the emission source and collectivity in ultra-relativistic p-Pb collisions*, Phys. Lett. B **720** (2013) 250, arXiv:1301.3314.

[154] V. M. Shapoval, P. Braun-Munzinger, I. A. Karpenko, and Y. M. Sinyukov, *Femtoscopic scales in* $p+p$ *and p+Pb collisions in view of the uncertainty principle*, Phys. Lett. B **725** (2013) 139, arXiv:1304.3815.

[155] A. Bzdak, B. Schenke, P. Tribedy, and R. Venugopalan, *Initial state geometry and the role of hydrodynamics in proton-proton, proton-nucleus and deuteron-nucleus collisions*, Phys. Rev. C **87** (2013) 064906, arXiv:1304.3403.

[156] L. McLerran, M. Praszalowicz, and B. Schenke, *Transverse Momentum of Protons, Pions and Kaons in High Multiplicity pp and pA Collisions: Evidence for the Color Glass Condensate?*, Nucl. Phys. A **916** (2013) 210, arXiv:1306.2350.

[157] I. G. Bearden *et al.*, *High-energy Pb + Pb collisions viewed by pion interferometry*, Phys. Rev. C **58** (1998) 1656.

[158] G. Alexander, I. Cohen, and E. Levin, *The dependence of the emission size on the hadron mass*, Phys. Lett. B **452** (1999) 159, arXiv:hep-ph/9901341.

[159] A. Bialas and K. Zalewski, *Mass dependence of the HBT radii observed in e+ e- annihilation*, Acta Phys. Polon. B **30** (1999) 359, arXiv:hep-ph/9901382.

[160] T. Csorgo and J. Zimanyi, *Pion interferometry for strongly correlated space-time and momentum space*, Nucl. Phys. A **517** (1990) 588.

[161] T. Csorgo and B. Lorstad, *Bose-Einstein correlations for three-dimensionally expanding, cylindrically symmetric, finite systems*, Phys. Rev. C **54** (1996) 1390, arXiv:hep-ph/9509213.

[162] *The LINAC2*, https://home.cern/science/accelerators/linear-accelerator-2, accessed: 2021-03-02.

[163] *The Proton Synchrotron Booster*, https://home.cern/science/accelerators/proton-synchrotron-booster, accessed: 2021-03-02.

[164] *The Proton Synchrotron*, https://home.cern/science/accelerators/proton-synchrotron, accessed: 2021-03-02.

[165] *The Super Proton Synchrotron*, https://home.cern/science/accelerators/super-proton-synchrotron, accessed: 2021-03-02.

[166] E. Mobs, *The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018*, OPEN-PHO-ACCEL-2018-005, 2018.

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

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

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

[170] LHCb collaboration, R. Aaij *et al.*, *Precision luminosity measurements at LHCb*, JINST **9** (2014) P12005, arXiv:1410.0149.

[171] LHCb collaboration, C. Elsasser, *$\bar{b}b$ production angle plots*, https://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb_ProductionAngles.html, accessed: 2020-10-27.

[172] R. Lindner, *LHCb layout*, LHCb-PHO-GENE-2008-002, 2008.

[173] M. Kucharczyk, P. Morawski, and M. Witek, *Primary Vertex Reconstruction at LHCb*, LHCb-PUB-2014-044, 2014.

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

[175] M. De Cian, S. Farry, P. Seyfert, and S. Stahl, *Fast neural-net based fake track rejection in the LHCb reconstruction*, LHCb-PUB-2017-011, 2017.

[176] LHCb RICH collaboration, A. Papanestis and C. D'Ambrosio, *Performance of the LHCb RICH detectors during the LHC Run II*, Nucl. Instrum. Meth. A **876** (2017) 221, arXiv:1703.08152.

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

[178] P. C. Bhat, *Multivariate analysis methods in particle physics*, Ann. Rev. Nucl. Part. Sci. **61** (2011) 281.

[179] H. Voss, A. Hoecker, J. Stelzer, and F. Tegenfeldt, *TMVA - Toolkit for Multivariate Data Analysis with ROOT*, PoS **ACAT** (2007) 040; A. Hoecker *et al.*, *TMVA 4 — Toolkit for Multivariate Data Analysis with ROOT. Users Guide.*, arXiv:physics/0703039.

[180] LHCb collaboration, *RTA and DPA dataflow diagrams for Run 1, Run 2, and the upgraded LHCb detector*, LHCB-FIGURE-2020-016, 2020.

[181] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. A **389** (1997) 81; I. Antcheva *et al.*, *ROOT: A C++ framework for petabyte data storage, statistical analysis and visualization*, Comput. Phys. Commun. **180** (2009) 2499, arXiv:1508.07749.

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

[183] LHCb collaboration, *LHCb Upgrade Software and Computing*, CERN-LHCC-2018-007, 2018.

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

[185] T. Pierog *et al.*, *EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider*, Phys. Rev. C **92** (2015) 034906, arXiv:1306.0121.

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

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

[188] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270; Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

[189] M. Clemencic *et al.*, *The LHCb simulation application, Gauss: Design, evolution and experience*, J. Phys. Conf. Ser. **331** (2011) 032023; LHCb collaboration, GAUSS *project*, https://lhcbdoc.web.cern.ch/lhcbdoc/gauss/, accessed: 2021-02-26.

[190] LHCb collaboration, BOOLE *project*, https://lhcbdoc.web.cern.ch/lhcbdoc/boole/, accessed: 2021-02-26.

[191] LHCb collaboration, MOORE *project*, https://lhcbdoc.web.cern.ch/lhcbdoc/moore/, accessed: 2021-02-26.

[192] LHCb collaboration, BRUNEL *project*, https://lhcbdoc.web.cern.ch/lhcbdoc/brunel/, accessed: 2021-02-26.

[193] LHCb collaboration, DAVINCI *project*, https://lhcbdoc.web.cern.ch/lhcbdoc/davinci/, accessed: 2021-02-26.

[194] G. Apollinari *et al.*, *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1*, CERN Yellow Reports: Monographs, CERN, Geneva, 2017.

[195] LHCb collaboration, *Physics case for an LHCb Upgrade II — Opportunities in flavour physics, and beyond, in the HL-LHC era*, arXiv:1808.08865.

[196] LHCb collaboration, *LHCb Tracker Upgrade Technical Design Report*, CERN-LHCC-2014-001, 2014.

[197] LHCb collaboration, *Letter of Intent for the LHCb Upgrade*, CERN-LHCC-2011-001, 2011.

[198] LHCb collaboration, *Framework TDR for the LHCb Upgrade: Technical Design Report*, CERN-LHCC-2012-007, 2012.

[199] LHCb collaboration, *LHCb Trigger and Online Technical Design Report*, CERN-LHCC-2014-016, 2014.

[200] LHCb collaboration, *LHCb VELO Upgrade Technical Design Report*, CERN-LHCC-2013-021, 2013.

[201] LHCb collaboration, *LHCb PID Upgrade Technical Design Report*, CERN-LHCC-2013-022, 2013.

[202] LHCb collaboration, *Computing Model of the Upgrade LHCb experiment*, CERN-LHCC-2018-014, 2018.

[203] M. Klusek-Gawenda, *Photon-Photon Scattering at the LHC in Collisions of Nuclei*, Acta Phys. Polon. B Proc. Suppl. **12** (2019) 333.

[204] LHCb collaboration, M. J. Charles and R. Forty, *TORCH: Time of Flight Identification with Cherenkov Radiation*, Nucl. Instrum. Meth. A **639** (2011) 173, arXiv:1009.3793.

[205] F. Keizer, *Sub-nanosecond Cherenkov photon detection for LHCb particle identification in high-occupancy conditions and semiconductor tracking for muon scattering tomography*, PhD thesis, Cambridge U., 2020, doi: 10.17863/CAM.45822.

[206] Y. Hu *et al.*, *Cherenkov Radiation Control via Self-accelerating Wave-packets*, Scientific Reports **7** (2017) 8695.

[207] J. Seguinot and T. Ypsilantis, *Photoionization and Cherenkov Ring Imaging*, Nucl. Instrum. Meth. **142** (1977) 377.

[208] LHCb collaboration, R. Forty, *RICH pattern recognition for LHCb*, Nucl. Instrum. Meth. A **433** (1999) 257.

[209] LHCb RICH collaboration, C. D'Ambrosio *et al.*, *The Future of RICH Detectors through the Light of the LHCb RICH*, Nucl. Instrum. Meth. A **876** (2017) 194, arXiv:1703.09927.

[210] M. Fiorini, *The upgrade of the LHCb RICH detectors*, Nucl. Instrum. Meth. A **952** (2020) 161688.

[211] T. Schneider, *The coatings of the RICH1 mirrors and quartz*, https://indico.cern.ch/event/1009732/contributions/4237181, 2021.

[212] L. Cadamuro *et al.*, *Characterization of the Hamamatsu R11265-103-M64 multianode photomultiplier tube*, JINST **9** (2014) P06021, arXiv:1403.3215.

[213] LHCb RICH collaboration, S. Gambetta, *First results from quality assurance testing of MaPMTs for the LHCb RICH upgrade*, Nucl. Instrum. Meth. A **876** (2017) 206.

[214] P. Carniti *et al.*, *CLARO-CMOS, a very low power ASIC for fast photon counting with pixellated photodetectors*, JINST **7** (2012) P11026, arXiv:1209.0409.

[215] M. Baszczyk *et al.*, *CLARO: an ASIC for high rate single photon counting with multi-anode photomultipliers*, JINST **12** (2017) P08019.

[216] T. Gys, *MaPMT QEs: update after corrections from filters and reference photodiode*, https://indico.cern.ch/event/1006764/contributions/4225797, 2021.

[217] LHCb RICH collaboration, *LHCb RICH upgrade – Production Readiness Review report for the Elementary Cell of the Photo-Detector*, https://lhcb-project-rich-upgrade.web.cern.ch/PRR_ofTheEC/Soi2ECka6KB1uORKd/EC-RR-v2r5.pdf, accessed: 2021-03-08.

[218] I. Slazyk, *Studies of the opto-electronic chain for the LHCb RICH upgrade*, PhD thesis, Institute of Nuclear Physics Polish Academy of Sciences, 2020, https://rifj.ifj.edu.pl/handle/item/339.

[219] LHCb RICH collaboration, M. P. Blago and F. Keizer, *High rate tests of the photon detection system for the LHCb RICH Upgrade*, Nucl. Instrum. Meth. A **876** (2017) 101, arXiv:1801.02623.

[220] M. McCann. Private communication, 2020.

[221] S. Easo *et al.*, *Simulation of LHCb RICH detectors using GEANT4*, IEEE Trans. Nucl. Sci. **52** (2005) 1665.

[222] G. Cavallero. Private communication, 2020.

[223] G. Cavallero and S. Gambetta. Private communication, 2020.

[224] F. Keizer. Private communication, 2020.

[225] F. Keizer, *Timing in Run 3 (and beyond)*, https://indico.cern.ch/event/951601/contributions/4005455, 2020.

[226] G. Baym, *The physics of Hanbury Brown–Twiss intensity interferometry: from stars to nuclear collisions*, Acta Phys. Polon. B **29** (1998) 1839, arXiv:nucl-th/9804026.

[227] G. Alexander, *Bose–Einstein and Fermi–Dirac interferometry in particle physics*, Rept. Prog. Phys. **66** (2003) 481, arXiv:hep-ph/0302130.

[228] M. Kucharczyk, *Determination of hadron emitter radii in Z0 decays to hadrons at LEP*, LAP Lambert Academic Publishing, 2011.

[229] D. Kincses, M. I. Nagy, and M. Csanad, *Coulomb and strong interactions in the final state of HBT correlations for Lévy type source functions*, Phys. Rev. C **102** (2020) 064912, arXiv:1912.01381.

[230] M. G. Bowler, *Coulomb corrections to Bose–Einstein correlations have been greatly exaggerated*, Phys. Lett. B **270** (1991) 69.

[231] Y. Sinyukov *et al.*, *Coulomb corrections for interferometry analysis of expanding hadron systems*, Phys. Lett. B **432** (1998) 248.

[232] S. Pratt, *Coherence and Coulomb effects on pion interferometry*, Phys. Rev. D **33** (1986) 72.

[233] M. Csanad, S. Lokos, and M. Nagy, *Expanded empirical formula for Coulomb final state interaction in the presence of Lévy sources*, Phys. Part. Nucl. **51** (2020) 238, arXiv:1910.02231.

[234] B. Kurgyis, *Coulomb interaction for Lévy sources*, 2020, arXiv:2007.10173.

[235] F. Stagni *et al.*, *Diracgrid/dirac: v6r20p15*, 2018. doi: 10.5281/zenodo.1451647; LHCb collaboration, *Lhcbdirac*, 2018. doi: 10.5281/zenodo.1451768.

[236] P. Z. Skands, *The Perugia tunes*, arXiv:0905.3418.

[237] M. Bahr *et al.*, *Herwig++ physics and manual*, Eur. Phys. J. C **58** (2008) 639, arXiv:0803.0883.

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

[239] LHCb collaboration, R. Aaij *et al.*, *Measurement of charged particle multiplicities and densities in pp collisions at* $\sqrt{s}$ *=7 TeV in the forward region*, Eur. Phys. J. **C74** (2014) 2888, arXiv:1402.4430.

[240] G. D'Agostini, *A Multidimensional unfolding method based on Bayes' theorem*, Nucl. Instrum. Meth. A **362** (1995) 487.

[241] T. Adye, *Unfolding algorithms and tests using RooUnfold*, arXiv:1105.1160.

[242] LHCb collaboration, R. Aaij *et al.*, *Study of J/*ψ *production and cold nuclear matter effects in pPb collisions at* $\sqrt{s_{NN}}$ *=5 TeV*, JHEP **02** (2014) 072, arXiv:1308.6729.

[243] LHCb collaboration, R. Aaij *et al.*, *Measurement of the nuclear modification factor and prompt charged particle production in p*Pb *and pp collisions at* $\sqrt{s_{\rm NN}} = 5\,\text{TeV}$, arXiv:2108.13115, submitted to Phys. Rev. Lett.

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

[245] S. Kullback and R. A. Leibler, *On Information and Sufficiency*, The Annals of Mathematical Statistics **22** (1951) 79; M. Needham, *Clone track identification using the Kullback–Liebler distance*, CERN-LHCB-2008-002, 2008.

[246] ATLAS collaboration, G. Aad *et al.*, *Two-particle Bose–Einstein correlations in pp collisions at* $\sqrt{\mathbf{s}}$ *= 0.9 and 7 TeV measured with the ATLAS detector. HEPData.*, 2015. doi: 10.17182/hepdata.70016.

[247] *The PLGrid Infrastructure*, https://www.plgrid.pl/en, accessed: 2021-03-02.