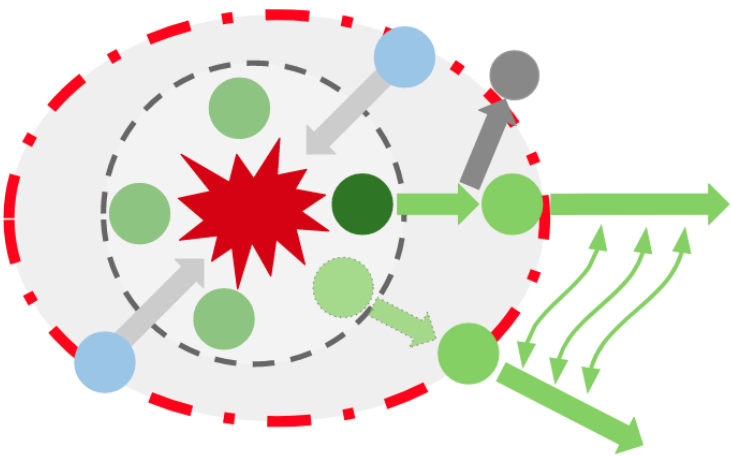
TUM School of Natural Sciences

TECHNISCHE UNIVERSITÄT MÜNCHEN

Fakultät für Physik

Master's Thesis

First Multiplicity and m_{T} Dependent Measurement of the Size of the Emission Source in pp Collisions at $\sqrt{s} = 13.6\,\mathrm{TeV}$ with ALICE



TECHNISCHE UNIVERSITÄT MÜNCHEN

TUM School of Natural Sciences

Fakultät für Physik

First Multiplicity and m_{T} Dependent Measurement of the Size of the Emission Source in pp Collisions at $\sqrt{s} = 13.6\,\mathrm{TeV}$ with ALICE

Erste Messung der Größe der Emissionsquelle in pp Kollisionen bei $\sqrt{s} = 13.6\,\mathrm{TeV}$ mit ALICE in Abhängigkeit der Multiplizität und der transversen Masse

Master's Thesis

Author: Georgios Mantzaridis
Supervisor: Prof. Dr. Laura Fabbietti
Advisor: Anton Riedel
Date: 28.12.2023

I confirm that the results presented in this master's thesis is my own work and I have documented all sources and materials used.

Ich versichere, dass ich diese Masterarbeit selbstständig verfasst und nur die angegebenen Quellen und Hilfsmittel verwendet habe.

Munich, 28.12.2023 Georgios Mantzaridis

Abstract

Femtoscopy at the LHC has developed into a powerful tool to measure the strong interaction between unstable hadrons like, for example, hyperons. It allows one to access systems that are not accessible using traditional experimental methods such as scattering or hypernuclei experiments. With the data collected with the ALICE detector during the Run 2 data-taking campaign, many multi-strange systems were measured along with the first femtoscopic measurement of charmed systems and three-body correlations. All these analyses are built on the measurement of the particle emitting source in high-multiplicity pp collisions and the observation that the emission is common for all hadron pairs and that it exhibits a common scaling with the average transverse mass (m_{T}) of the emitted particle pair. This finding allowed the constrain of the particle emission source, a necessary ingredient to use femtoscopy to study interactions.

The ALICE detector was recently upgraded to match the performance requirements for the Run 3 operation, which started in the summer of 2022. These upgrades will significantly increase the amount of collected data thanks to the implementation of software triggers to select rare events based on physics observables. This will bring previously inaccessible interaction channels within reach of measurement. However, the first step will be a precise measurement of the particle emission source at the new collision energy of $\sqrt{s} = 13.6\,\mathrm{TeV}$. The 500 billion minimum bias events collected in 2022 alone are the ideal dataset to measure the multiplicity-dependent m_{T} scaling for the first time in pp collisions at the LHC and thus extend the high-multiplicity measurement of Run 2. In this thesis, the upgrades made to the ALICE detector that facilitated this analysis are briefly discussed, followed by the description of the analysis procedure to measure the multiplicity-dependent m_{T} scaling of the hadron emission source using p–p and $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ correlations. The results agree with the expectation. However, the systematic uncertainties still play a dominant role in this analysis. The results are put into context with the previous Run 2 results, and the next steps and improvements to reduce the systematic uncertainties are discussed.

Contents

1 Introduction

1.1 From HBT correlations to Relativistic Heavy Ion Collisions

Femtoscopy has its origins in photon interferometry, where it was used to measure the size of light-emitting stellar objects by measuring the correlation of photon pairs emitted by them [1, 2]. Soon after, the method was adapted to measure pion–pion correlations to investigate their emission source in proton-antiproton collisions [3]. The correlations were assumed to arise from Bose-Einstein statistics and the properties of the annihilation and subsequent emission process. The final-state interactions were not explicitly considered. Subsequently, the formalism was further developed and extended to non-identical pairs, and the sensitivity to final state interactions was investigated and included in the formalism. In the heavy ion collision community, the method gained popularity for studying the size of the particle emitting source, i.e., the collision region from which hadrons emerge and are then measured in the detector [4].

In relativistic heavy ion collisions, this collision region is often referred to as the "fireball" [5] to emphasize the highly energetic and dynamical state of matter, the quark-gluon plasma (QGP), in which the quarks and gluons of the colliding nuclei are deconfined, i.e., they are not bound to their respective hadrons anymore. Experimentally, this state of matter can be created by increasing the density or the temperature of the baryonic matter. Both can be achieved in collider experiments, depending on the collision energies, the projectiles, and the targets used. The experimental efforts of the last decades are summarized in Fig. 1.1, which shows the QCD phase diagram in terms of the baryonic density on the x-axis and the temperature on the y-axis.

Various experiments and astrophysical objects, such as neutron stars, are placed at different points in the phase diagram, depending on the conditions of QCD that they can probe. In collider experiments, several observables, such as anisotropic flow, jet quenching, and strangeness enhancement, have been proposed to probe the properties of the QGP (see [5] for a comprehensive summary of the ALICE measurements). These observables probe the viscosity of the QGP in the sense that their measurements are sensitive to the collective behavior of the QGP. This collective behavior has been successfully interpreted within the framework of relativistic hydrodynamics. This led to the much-cited claim that the QGP behaves like an almost perfect fluid [6]. Indeed, a detailed analysis of these observables can help bring the properties of the QGP to light, and the comparisons with sophisticated models constrain the parameter space more and more [7].

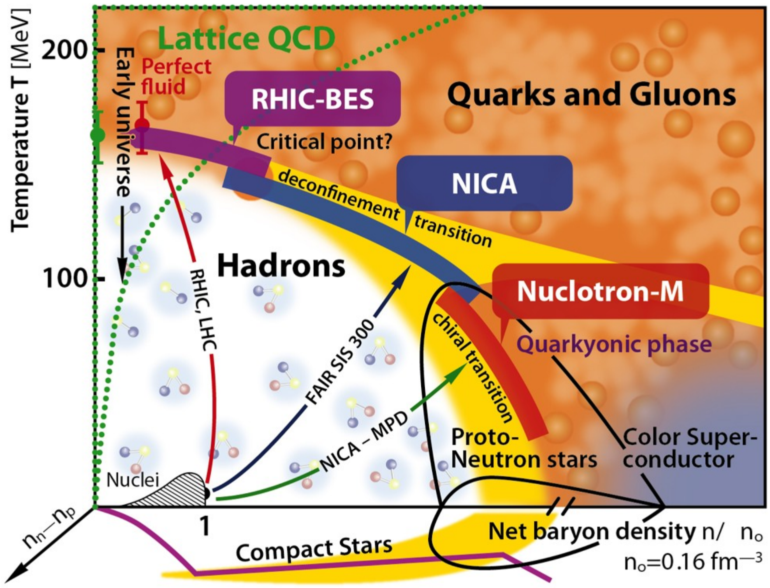


Figure 1.1: An overview of the QCD phasediagram. Figure taken from `https://nica.jinr.ru/physics.php`

Evolution of a Relativistic Heavy Ion collision

Taking all these observations together, the current understanding of relativistic heavy ion collisions is that the collision region evolves through several steps to produce final state particles that can be measured in the detector. A schematic overview is given in Fig. 1.2. A short pre-equilibrium phase is followed by the hydrodynamic evolution of the QGP. After the hadronization of the partons into hadrons and resonance states, the resulting hadron gas continues its evolution dominated by inelastic processes that can change the particle species. This happens until the chemical freeze-out temperature is reached. At that temperature, inelastic collisions cannot be sustained anymore and the particle's identities are fixed after this point. However, elastic collisions can still redistribute the momenta among the particles. Eventually, these collisions will also cease once the system reaches the kinetic freeze-out temperature. From there, the particles have a defined identity and momentum and are then measured directly with the detector.

Femtoscopy in relativistic heavy ion collisions measures the source size at a point in time between the chemical and kinetic freeze-out. In simple terms: Final state interactions can change the momentum distribution of the emitted particles, which means that the kinetic freeze-out is not complete. On the other hand, the identification and selection of final state particles for femtoscopic analysis implies fixed particle species, i.e., a complete chemical freeze-out.

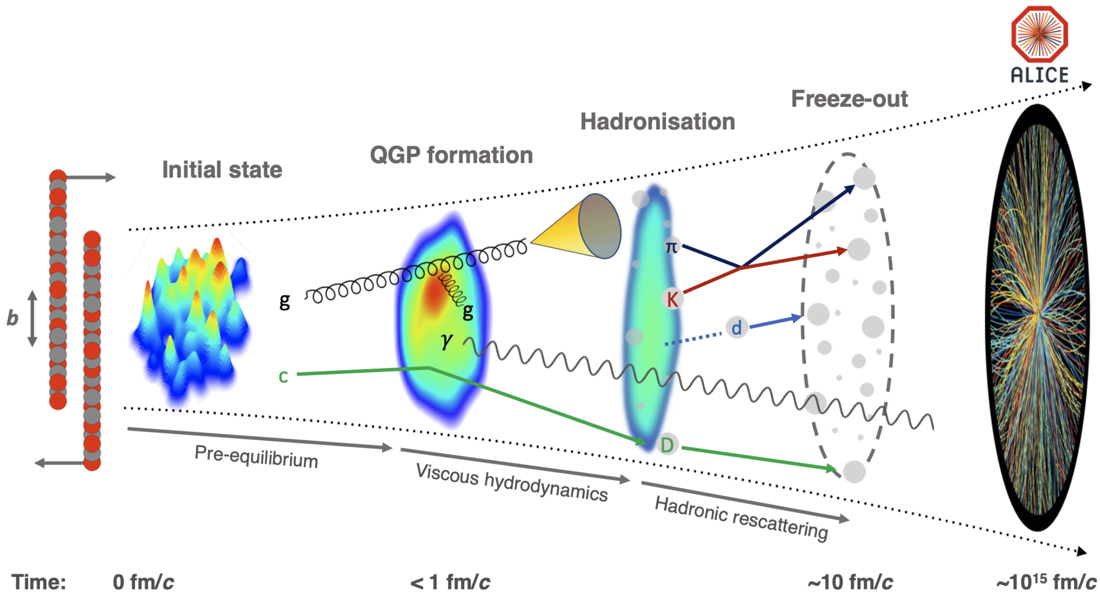


Figure 1.2: The evolution of a heavy-ion collision at LHC energies. Figure from [5]

1.2 From Relativistic heavy Ion Collisions to Final State interactions

The sensitivity to final state interactions was the driving factor for a revolution in femtoscopy, in which the paradigm is turned around: Rather than using particle pairs, for which the interaction is known or negligible, one can study simpler collision systems where the source can be constrained, and the unknown interactions between particles can be studied. Relativistic proton-proton (pp) collisions are a particularly promising system for this type of study due to two main advantages. First, such collisions are less complicated than heavy ion collisions. The mechanism of particle production is expected to be similar for all particles. If additional effects such as the collective evolution are absent in pp collisions, there is no reason for a difference in the particle emission because of the quark content. Second, based solely on the difference in spatial extension of the projectiles, one would expect a smaller source size in pp than in heavy ion collisions. As will be demonstrated in Chapter 2, the correlation function is a convolution of the source function and the two-particle wave function, which includes the interaction (see Equation 2.2). This illustrates the sensitivity to the effects of short-range potentials, as the strong interaction requires small source sizes.

Indeed, first femtoscopy measurements of proton-proton (p–p), proton-Lambda (p–Λ) and Lambda-Lambda (Λ–Λ) pairs with ALICE at the LHC, using the Run 1 data at $\sqrt{s} = 7\,\text{TeV}$ proved both of these claims to be true [8]. A subsequent analysis, in high-multiplicity triggered pp collisions in Run 2 at $\sqrt{s} = 13\,\text{TeV}$, went a step further and found evidence for a common hadron emission source, which scales with the transverse mass (m_{T}) of the particle pair. The transverse mass is defined by

$$m_{\mathrm{T}} = \sqrt{k_{\mathrm{T}}^2 + M^2}, \tag{1.1}$$

where M is the average mass of the particle pair [9]. The assumption of a common source was verified in that study using the p–p and p–Λ correlations, for which the strong interaction was modeled via state-of-the-art potentials from chiral effective field theory. In the meantime, the common m_{T} scaling has been observed for same charged pion pairs and K^+–p pairs as well [10].

If the common source assumption can be extended to all baryons, then this would allow the measurement of the interaction between all particles produced in a collision, provided that their lifetime is large enough to experience final state interactions (FSI). Strange hadrons, i.e., hyperons and strange mesons, fulfill these criteria. They are abundantly produced in pp collisions, and since they are decaying weakly, they have decay lengths of the order of centimeters. The interactions involving strange hadrons are an essential ingredient, for example, to constrain the equation of state of dense neutron stars [11]. However, with increasing strangeness content, it is increasingly difficult to access the interaction with common experimental methods such as scattering or hypernuclei experiments due to the small lifetime of these particles. Hyperon-hyperon interactions, for example, are nearly impossible with these methods.

Following the assumption of the common source, multi-strange channels like p–Ω^- [12] and p–Ξ^- [13] and hyperon-hyperon interactions like the Λ–Ξ^- [14] interaction were measured for the first time, while other systems, like the p–Λ or p–Σ^0 have been measured with unprecedented precision [15]. The measurements also extended to the baryon-meson sector, e.g., Λ–K^- [16] and p–φ [17]. A comprehensive review of the measurement of the strong interaction via femtoscopy can be found here [18]. Recently, the first attempts to constrain the interaction involving charmed hadrons [19] have been made and femtoscopy started expanding in the three-body sector [20]. However, these analyses reached the statistical limits of the available pp collision data collected by the ALICE detector, which were collected during the Run 2 period until the end of 2018.

1.3 The aim of this thesis

Following several significant upgrades, the ALICE detector is currently operating at the LHC in its latest iteration, the ALICE 2 detector. The data-taking strategy foresees the recording of all collisions, a complete reconstruction, and the application of software triggers to select events with rare observables, such as, for example, exotic pairs for femtoscopy measurement. This will significantly boost the precision of future measurements and make charmed and three-body systems accessible. The first step is the constraint of the source, as had been done for high-multiplicity pp collisions in Run 2 [9]. However, due to the data collection via software triggers, future analysis will not be bound to events with a well-defined multiplicity but rather to datasets with an arbitrary combination of different event multiplicities. Therefore, constraining the m_{T} scaling as a function of multiplicity is crucial for these future analyses. This work aims to study the m_{T} scaling of the source as a function of multiplicity for p–p and $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ pairs for the first time, paving the way for all future femtoscopy measurements with ALICE in Run 3. It is also the first femtoscopy analysis utilizing data from the upgraded ALICE detector and thus provides valuable insights into the data and reconstruction quality. The analysis will be done with the more than 500 Billion events collected at record a center of mass energy of 13.6 TeV by the ALICE detector in 2022 alone. This is the largest minimum bias dataset in the history of ALICE and surpasses the Run 2 minimum bias sample by about a factor of 500. This huge experimental success is made possible thanks to the combined efforts of the ALICE collaboration on the hardware and software side to build the detector and to develop an analysis framework capable of handling such data loads. The next chapter will briefly introduce the femtoscopy method. Afterward, the ALICE detector will be introduced in chapter 3, explaining the upgrades made for the Run 3 operation. This chapter ends with describing the new analysis framework and its design philosophy. In chapter 4 the event and particle selection will be explained, which leads to the sample of p–p and $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ pairs, from which the correlation function is constructed, as will be fleshed out in chapter 5, which will end with a description of the fitting procedure. Finally, the sources of systematic uncertainties will be discussed in chapter 6, followed by the discussion of the results in chapter 7 and a summary in chapter 8.

2 Femtoscopy

As discussed before, femtoscopy measures the correlation of particles in the momentum space, after the chemical freeze-out of the collision evolution. The main observable is the correlation function. The most general definition is given by

$$C\left(\vec{p}_1, \vec{p}_2\right) = \frac{P\left(\vec{p}_1, \vec{p}_2\right)}{P\left(\vec{p}_1\right) P\left(\vec{p}_2\right)} = \frac{E_1 E_2 dN / \left(d^3 p_1 d^3 p_2\right)}{\left(E_1 dN / d^3 p_1\right) \left(E_2 dN / d^3 p_2\right)}, \tag{2.1}$$

where $P\left(\vec{p}_1, \vec{p}_2\right)$ is the probability for finding a particle with momentum $\vec{p}_1$ when the second particle has been emitted with a momentum of $\vec{p}_2$, while $P\left(\vec{p}_1\right)$ and $P\left(\vec{p}_2\right)$ are the two independent probabilities for finding a single particle with the respective momentum. The interferometric origin in astronomy of femtoscopy can be understood when looking at (2.1) and by replacing $P\left(\vec{p}_i\right)$ in the denominator with the photon intensity $\langle I_i \rangle$ and the nominator with $\langle I_1 I_2 \rangle$. Finally, the correlation function can be calculated by the pair yield divided by the product of the two single particle invariant yields as shown in the right side of (2.1) [21].

2.1 The theoretical correlation function

The theoretical definition of the correlation function typically used within the ALICE collaboration reads

$$C(k^*) = \int S(\vec{r^*}) |\psi(k^*, \vec{r}^*)|^2 d^3 r, \tag{2.2}$$

where $S(\vec{r^*})$ is the source function and $\psi(k^*, \vec{r^*})$ is the pair wave function. It is defined in the rest frame of the particle pair's center of mass, which is denoted by the asterisk. Following that, $\vec{r^*}$ is the relative distance between the two particles and $\vec{k^*}$ the reduced relative momentum, defined as

$$\vec{k^*} = \frac{m_2 \vec{p}_1^* - m_1 \vec{p}_2^*}{m_1 + m_2}, \tag{2.3}$$

where $\vec{p}_i^*$ is the momentum of particle i in the particle rest frame.

The source function is a spatial probability density for the emission of a pair with the relative distance $\vec{r^*}$. The pair wave function depends the interaction potential of the particle pair. Both the source and the wave function can have a three-dimensional dependence. In this case, the coordinate system used for femtoscopy is the "out-side-long" frame, which is a longitudinal co-moving frame, where the "long" axis is placed in the direction of the beam axis, the "out" axis is placed along the total momentum of the pair and the "side" axis is perpendicular to both of them. It is related to the pair rest frame by a Lorentz Boost along the "out" direction by the combined pair momentum.

Equation 2.2 can be derived from (2.1) [4] under certain assumptions, including that the source function and the two-particle wave function can be factorized, i.e. the emission process and the interaction are independent. It also illustrates the interplay between the source function and the interaction: Due to the convolution, smaller sources will probe the two-particle wave functions at smaller relative distances and vice versa. As mentioned in the Introduction, historically, femtoscopy was used to study the size of the particle emitting source in heavy ion collisions by measuring pairs with a known or negligible interaction (i.e., pair wave function). This paradigm was turned around, and the m_{T} differential source size measurement of p–p and p–Λ correlations in high-multiplicity events confirmed the assumption of a universal hadron emission in ultra-relativistic high-multiplicity pp collisions [9]. Following this assumption, the emission source consists of a Gaussian emission profile given by

$$S_G\left(\vec{r^*}\right) = \left(4\pi r_{\mathrm{core}}^2\right)^{-3/2} \cdot \exp\left(-\frac{\vec{r^*}^2}{4r_{\mathrm{core}}^2}\right). \tag{2.4}$$

It depends only on the size parameter r_{core}. This Gaussian profile is common for all hadron pairs and scales with the average transverse mass of the pair (defined in Eq. 1.1). The common m_{T} scaling is broken because, additionally to the primordial yield of particles, there is a contribution from the decay of shortly lived resonances. Their decay lengths (typically less than 10fm) are small enough so that the daughters can experience final state interactions and thus be considered primary particles. The resonances themselves are assumed to not interact in such small time scales. This effectively extends the source function by an exponential tail. These resonances are hadron-specific. Equation 2.2 can still be used with the assumption of a Gaussian source function. In that case, the source size measurement will deliver the effective source size r_{eff}, which is larger than the core size due to the exponential tail. The contribution from shortly decaying resonances can be considered via simulations of the angular distributions, e.g., using the EPOS event generator [22]. The work presented here will measure the effective source size. One important thing to note is that the source function in (2.2) is the two-particle source function related to the particle pair under investigation. It is not equal to the single particle emission probability used, for example, in the (2.1). When measuring the source size in femtoscopy, one does not access the whole fireball but rather the so-called region of homogeneity [4], which is a region in the phase space, where the two particles have similar velocities. Naturally, both the fireball and the femtoscopic source are correlated, i.e., a larger fireball results in a larger region of homogeneity.

2.2 The experimental correlation function

The experimental correlation function can be obtained by the ratio of the k^* distribution of pairs in the same event (referred to as the same event distribution N_{same}) and an uncorrelated k^* distribution. For the uncorrelated k^* distribution, event mixing techniques can be used, where particles from different events are paired with each other to ensure the absence of any correlation other than the available phase space.

This distribution is therefore called the mixed event distribution N_{mixed}. The measured correlation function is given by

$$C(k^*) = \mathcal{N} \cdot \frac{N_{same}(k^*)}{N_{mixed}(k^*)}, \tag{2.5}$$

where $\mathcal{N}$ is the overall normalization parameter to ensure that the correlation function approaches unity for large k^* (typically $> 200\,\mathrm{MeV}/c$). For these relative momenta, one does not expect any correlation due to femtoscopic effects.

2.3 Comparing the experimental and theoretical correlation function

To compare or fit the experimental correlation function ((2.5)) with theoretical ((2.2)) correlation function, corrections for experimental effects have to be taken into account [8]. These fall into either of the two categories:

- non-femtoscopic background and
- contamination from feed down, miss-identified, and knock-out particles.

The first affects the correlation function outside of the femtoscopic region, i.e., for $k^* > 200\,\mathrm{MeV}/c$ in such a way that the correlation function departs from unity. The origin of these correlations is thought to be mainly coming from energy-momentum conservation or, in the case of mesons or baryon-antibaryon pairs, from mini-jets. They can be described with a polynomial which is multiplied with the theoretical correlation function. This so called baseline has no linear term to ensure a flat behavior at $k^* = 0$ (see Section 5.2.2)). Feed-down contributions carry residual correlations from their mother particle, that interacted with the other particle in the pair before the decay. Weak decays are a source of residual correlations due to the typically long lifetime of a weakly decaying particle. For example, the measured p–p correlation function contains residual contributions from p–Λ, where the Λ decayed into a proton, which is ultimately measured in the detector. The daughter proton will not interact anymore with other primordial particles, however, the residual correlation experienced by its mother particle needs to be taken into account. Miss-identified and knock-out particles (i.e., material particles originating from interactions of primary particles with the beam pipe or detector material) do not carry any correlation into the measurement and thus their presence decreases the signal strength. It is impossible to remove all non-genuine correlations from the sample, completely. Instead, their contributions can be corrected if their shape and relative contribution are known. The shape for feed-down contributions can be obtained by modeling the residual interaction and transforming it into the pair rest frame of the measured particles. Depending on the q values of this decay, the contribution can be washed out significantly. Correlations from miss-identified particles are modeled either via a flat contribution if they have no correlation or if it is weak enough to be neglected. If not, their correlation can be constrained in a data-driven way by measuring the correlation function with background particles (e.g., side-band corrections if the candidates are selected via their invariant mass). The relative contribution of the genuine, residual, and fake correlations can be estimated with the help of the λ parameters [8]. They are given by

$$\lambda_{ij} = \lambda_i \cdot \lambda_j = \mathcal{P}_i f_i \cdot \mathcal{P}_j f_j, \tag{2.6}$$

where f is the fraction of the fake candidates or of particles originating from the feed-down channel of interest and $\mathcal{P}$ is the purity of the respective particles. The purity can be extracted by consulting Monte Carlo simulations or in an entirely data-driven approach. The fractions need the input of Monte Carlo simulations as they require a simulation of the detector to properly account for, e.g., knock-out particles.

Knowing the non-genuine femtoscopic contributions, the femtoscopic part of the correlation function can be modeled with the help of the λ parameters by simply adding the correlations, weighted by their respective λ parameters:

$$C_{\text{femto}}\left(k^*\right) = 1 + \lambda_{\text{genuine}} \cdot \left(C_{\text{genuine}}\left(k^*\right) - 1\right) + \sum_{ij} \lambda_{ij} \left(C_{ij}\left(k^*\right) - 1\right), \tag{2.7}$$

where $C_{\text{genuine}}(k^*)$ is the theoretical correlation function from (2.5). This way one can add all contributions that are present in the measured correlation function to the model which can be used for fitting.

2.4 The CATS framework

The CATS framework [23] provides the utilities to compare theoretical with experimentally measured correlation functions. It implements the previously mentioned modeling steps, including the feed-down contributions, and provides tools to apply the momentum smearing due to the finite detector resolution (see Section **??**). At its core, it calculates the correlation function in (2.2) starting from pre-defined wavefunctions or interaction potentials. For the latter, it numerically solves the Schrödinger Equation given an interaction potential. These can be one of the pre-defined potentials for different YN interactions or provided as user input. The CATS framework provides full access to the λ, source, and interaction parameters in the modeled correlation function. This allows the fitting of the source radii or any other parameter regarding the model's interaction or decomposition.

3 Experimental Setup

3.1 The LHC at CERN

The Large Hadron Collider (LHC) is the largest and most powerful particle accelerator ever built. It is a circular particle accelerator located at the CERN facility in Geneva, around one hundred meters under the earth and with a circumference of about 27 km. In its main operating mode, it delivers pp collisions at ultra-relativistic energies of $\sqrt{s} = 13.6\,\text{TeV}$ to the four experiments (ATLAS, CMS, ALICE and LHCb) located at four of the interaction points along the ring. In addition to the pp operation, roughly one month per year is dedicated to special operation modes, where e.g. Pb–Pb beams are collided. The LHC is the final accelerator and storage ring after a series of pre-accelerators, as can be seen in Fig. 3.1. First, the linear accelerator Linac2 accelerates protons to 50 MeV, from where they are injected into the PS booster, which in turn accelerates them to 1.4 GeV. From there the protons are injected into the Proton Synchrotron and accelerated to 25 GeV and passed to the final pre-accelerator, the Super Proton Synchrotron (SPS), where they reach 450 GeV. From the SPS they are injected into the LHC, where they are accelerated to $\sqrt{s} = 13.6\,\text{TeV}$ stored in bunches with a separation of 25 ns, which corresponds to about 7.5 m, from each other [24].

The LHC is operated in periods of about three years, the so called Runs, which are followed by Long Shutdown (LS) periods of a similar timescale. The later are used for maintenance and major upgrades of the LHC and the detectors. The LHC started its operation in 2009 with the collision of two proton beams, whose energy was increasing up to $\sqrt{s} = 7\,\text{TeV}$. Following the LS1, the collision energy was increased to 13 TeV for the Run 2 period. At the time of writing, the LHC Run 3 period is running, for which the new world record collision energy of 13.6 TeV has been reached. While the collision energy determines the accessible physics processes, the number of collisions is represented by the luminosity, which is defined as

$$L = \frac{1}{\sigma}\frac{dN}{dt}, \tag{3.1}$$

where dN is the number of detected events in the time dt and σ is the cross section of the process of interest. The units are reported in number of events per area (in cm^2) and per second. The large statistical significance required for high-precision physics analysis profit from a high luminosity as it allows to record more collisions in the same amount of time. From a technical perspective, luminosity is controlled by the spacing between the proton bunches, the number of protons in each bunch, and the focusing of those bunches at the interaction points. From an experimental standpoint, high luminosities introduce a challenge, as the detectors and readout electronics must be designed to cope with the high interaction rates and to separate between multiple collisions per

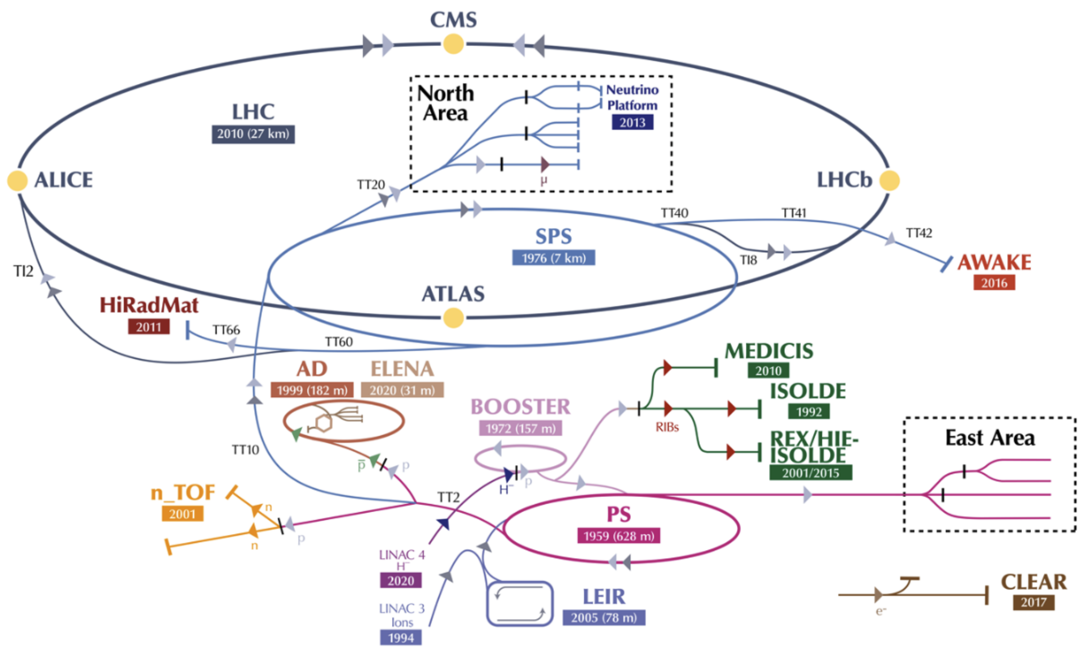


Figure 3.1: Overview of the CERN acceleration complex. The figure is taken from [24]

bunch crossing. The ALICE experiment, in particular, has physical limitations on its maximum luminosity imposed by the Time Projection Chamber (TPC), a gaseous detector that is utilized for particle identification and tracking and is the central detector component of the ALICE experiment. In fact, the planned high-luminosity Runs of the LHC will allow the accelerator to achieve a luminosity up to five times the current peak luminosity [25]. In such a scenario a gas detector will not be able to cope with the high interaction rates due to the physical limitations imposed by the drift velocities of charges in the gas. To be able to continue experimental operations during these Runs, the ALICE collaboration is preparing a silicon-only detector and large research and simulation efforts are invested into determining a way to match the PID performance of a gas-based PID detector. Currently, the luminosity at the ALICE experiment is lower than other LHC experiments. This is achieved by focusing the proton bunches to smaller angles such as to reduce the overlap between them. In Section 3.4, the specific detector issues and corresponding solutions will be discussed after an introduction to the ALICE experiment. Table 3.1 summarizes the Run and LS periods up until the Run 3 period, which provided the data to be analyzed in this work.

3.2 ALICE experiment

The ALICE experiment is located at the interaction point 2 of the LHC. A schematic overview of its current state - the ALICE 2 detector- is depicted in Fig. 3.2 along with an overview of its subdetectors. Since the beginning, the ALICE experiment was envisioned as a general-purpose heavy ion experiment, and the technical design and detector technologies were chosen accordingly to match the challenges and limitations

	time	$\sqrt{s_{pp}}$
Run 1	2009 - 2013	7 TeV
LS 1	2013 - 2015	
Run 2	2015 - 2018	13 TeV
LS2	2018-2022	
Run 3	since 2022	13.6 TeV

Table 3.1: Overview of the LHC Runs and the corresponding center of mass energies of the pp collisions.

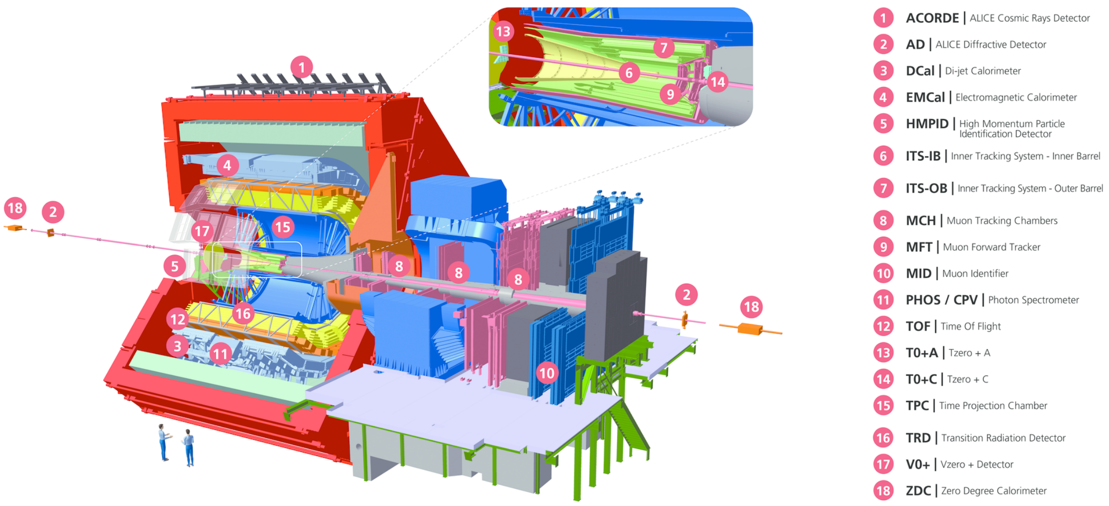


Figure 3.2: Schematic overview of the ALICE 2 detector after the upgrades during LS2. Figure taken from the ALICE Figure repository.

imposed by the physics program and the related observables. Above all, the detector was designed to perform at large final state multiplicities created by relativistic heavy ion collisions (up to the order of 30,000 final state particles in central Pb–Pb collisions [26]). To be versatile, it must cover a wide momentum range from low p_{T} of around 100 MeV/c necessary for correlation studies up to 100 GeV/c necessary for example, for jet physics. The low p_{T} capabilities of the detector necessitate a design with low material budget to reduce multiple scattering for low p_{T} particles, while the correlation studies rely on good PID capabilities. All these considerations are incorporated into the final ALICE design, which consists of a central barrel measuring hadrons, electrons, and photons and a forward muon spectromenter. Additional forward detectors estimate the multiplicity of a heavy ion collision and provide signals for triggering. The central barrel is surrounded by a solenoid magnet, which creates a magnetic field of 0.5 T, in which the particles are bent. Their curvature is used for momentum measurement. The central barrel itself consists of the Inner Tracking system (ITS), the Time Projection Chamber (TPC), the Transition Radiation Detector (TRD), the Ring Imaging Cherenkov Detector (HMPID), and the Time of Flight (TOF).

Upgrades leading to the ALICE 2 detector

During the Long Shutdown 2 (see Table 3.1), substantial upgrades to the ALICE detector were carried out, resulting in the ALICE 2 detector [27]. These upgrades aim to achieve two main objectives: An enhanced readout rate and an improved pointing resolution. The enhancement of the readout rate is necessary to cope with the large liminosity provided by the LHC and to collect larger data samples. Specifically, the readout is upgraded to a continuous readout for all sub-detectors except the EMCal, PHOS and HMPID, which were not upgraded and can be operated only in triggered mode. Additionally, all upgraded detectors support a triggered mode as well, which can be used, for example, in commissioning and calibration runs. In order to support a continuous readout, the readout electronics were replaced by the Common Readout Units (CRUs), which are the new interface between the detector front end and the further event processing pipeline [27]. In the case of the TPC, the entire readout technology was changed from multi-wired proportional chambers (MWPC) to GEM foil-based readout chambers (more on that in section Section 3.4). The improved pointing resolution is necessary, among others, to better distinguish collisions from the same bunch crossing through a better primary vertex resolution. In order to archive this, the detector was moved closer to the interaction point and the material budged of the ITS was reduced to minimize multiple scattering. For the same reason, the beam pipe around the collision point was replaced with a more lightweight beryllium-based pipe, whose radius was reduced to accommodate the smaller inner-most barrel of the ITS.

In the next sections, the main sub-detectors used for this work, namely the ITS, TPC, and TOF, will be discussed. A special emphasis will be given on the upgrades facilitating the aforementioned experimental improvements. For a comprehensive overview, the reader is directed to [28, 27].

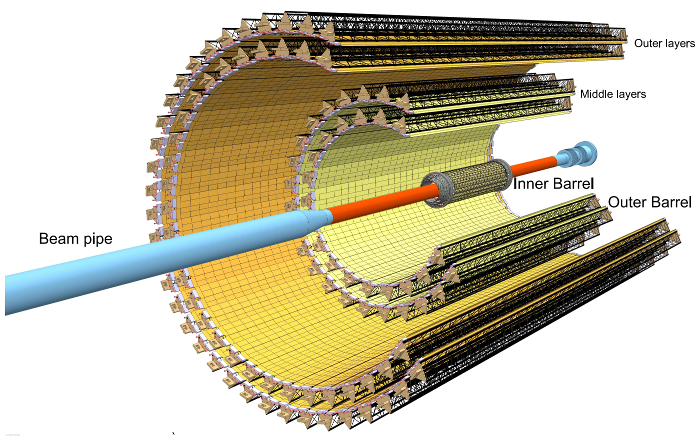


Figure 3.3: Schematic overview of the ITS2 detector. The three layer groups are shown together with the thinner beryllium beampipe. Figure taken from the ALICE Figure repository.

3.3 ITS

The Inner Tracking system is mainly used for the determination of the primary vertex and tracking of charged particles. The tracks can then be matched to the tracks detected in the TPC. The original ITS used in Run 1 and Run 2 was completely replaced by an upgraded version, the ITS2, with the primary aim of enhancing the precision of the collision vertex reconstruction along with that of the decay of heavy-flavored hadrons. Moreover, the detection of low p_T particles is improved. For that, one additional layer was added to the inner barrel of the ITS, which now consists of three layers. The outer barrel consists of four layers. A schematic view of the ITS2 is shown in Fig. 3.3. Its radial dimension, starting from the collision point extends from 22 mm to 395 mm, covering a rapidity of $|\eta| \leq 1.3$, which improved with respect to the previous iteration by 0.4 units of rapidity. The smaller inner radius is only possible due to the replacement of the beam pipe segment inside the ITS2 by a beryllium pipe with a smaller radius of 18 mm instead of 28 mm, as indicated in Fig. 3.3. Each layer of the ITS is based on silicon pixel detectors, which are realized by arrays of ALPIDE chips with a pixel pitch size of (27 µm × 29 µm). These upgrades reduced the spatial resolution (in $r \times \varphi$) from 10 µm × 100 µm in the first ITS to 5 µm × 5 µm and brought the material budget per layer down to $0.36\%X_0$ and $1.1\%X_0$ for the inner and the two outer layers, respectively [27]. The readout capabilities have been increased to 50 kHz for Pb–Pb runs.

ALPIDE chips

The ALPIDE chip is a CMOS Monolithic Active Pixel Sensor that has been developed specifically for the ALICE ITS upgrade. Derivatives of the chips are planned to be used in future upgrade projects like the ITS 3 and the outer tracker of the ALICE 3 detector [29]. The design required radiation hardness and efficient power consumption in order to minimize the material budget of the ITS. The ALPIDE chip meets these requirements due to the monolithic design, which fully incorporates the readout circuitry on the same

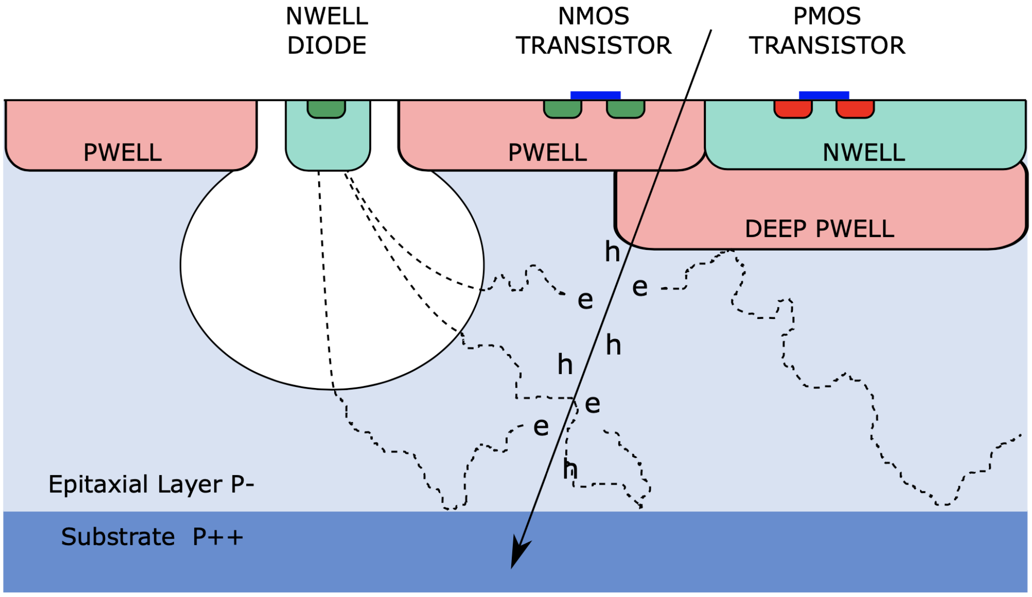


Figure 3.4: Cross section of an ALPIDE pixel. Figure taken from [27]

pixel. A cross-section of an individual pixel cell is shown in Fig. 3.4. Particular features are the small N-well diode for the readout, which maximizes the signal-over-noise ratio because of its small capacitance and the deep p-well shielding the CMOS circuitry from the epitaxial layer. This allows for a full-fledged CMOS circuitry that does not interfere with the charge collection at the anode [30]. Each chip measures 15 mm $\times$ 30 mm and consists of an array of 512×1024 pixels. Each pixel reports in a binary way whether it recorded a hit or not. This information is collected at the peripheral region, where the readout and interfacing functionalities for the entire chip are situated and shipped out of the detector [31].

3.4 TPC

The Time Projection Chamber is a gaseous detector and is used for charged particle identification (PID) and charged particle tracking. The advantage of using a gas-based detector as the main PID and tracking component lies in its ability to provide acurate PID and tracking even at large multiplicities. The drawback is the trade-off between the interaction rate and the space charge distortions from positively charged ions in the TPC drift volume, which change the drift paths of electrons and worsen the resolution of the detector. A correction during calibration and reconstruction of the data is necessary. The TPC's outer radius is 250 cm and with a length of 500 cm it covers the symmetric pseudo-rapidity range of $-0.9 < \eta < 0.9$ and has a full azimuth coverage. The PID is determined by measuring the energy loss, which can be related to the particle's charge and mass via the Bethe-Bloch formula. Additionally, charged particles ionize the TPC gas, and the free electrons drift to either one of the endplates of the TPC, where they are amplified and measured. As shown in Fig. 3.5, the TPC is divided into two

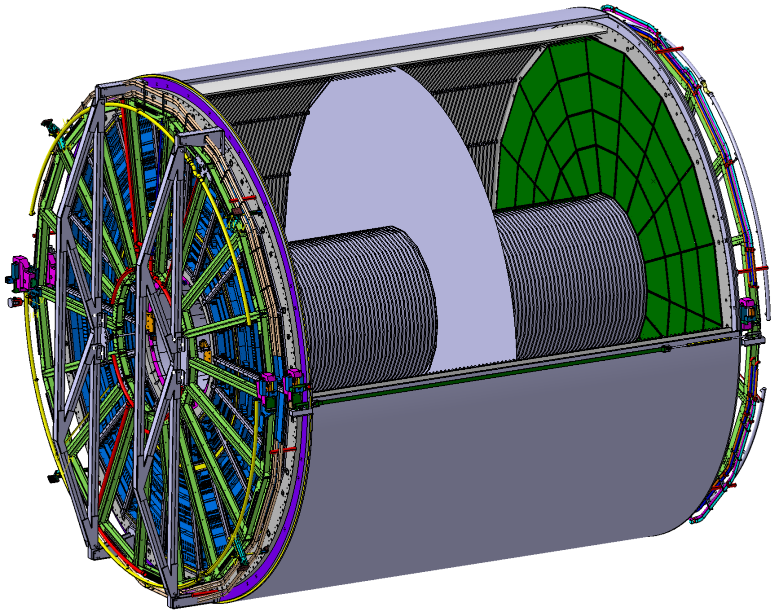


Figure 3.5: Schematic overview of the TPC. The trapezoidal segments with the readout chambers are shown as well as the dividing electrode in the middle. Figure taken from the ALICE Figure repository.

halves by a high voltage electrode, operating with 100 kV. Together with the endplates, they generate the electric field in which the free electrons drift. Both endplates are divided into 18 trapezoidal sectors, which contain the readout chambers. The outer part of the TPC cylinder provides electrical insulation to the rest of the detector via a CO_2 enclosing layer. The gas mixture used in the TPC is $\mathbf{Ne} - \mathbf{CO_2} - \mathbf{N_2}$ (90-10-5), which has mostly been used during the first two runs of the LHC as well. A **Ne** based gas in the TPC fulfills the requirements of a small material budget and a high ion mobility. Higher ion mobility means that the residual positive ions will travel faster to the electrode, where they can recombine with electrons, which reduces the contribution of space charge distortions and increases the resolution of the interaction vertex [32, 33].

Continous readout with GEM based readout chambers

These space charge distortions are enhanced by the contribution of positively charged ions drifting back from the amplification at the readout chambers into the TPC drift volume. In order to mitigate that, up until Run 2, the TPC utilized a gated grid in front of the multiwired proportional chambers (MWPCs), which were used for the readout [32]. The gated grid was connected to the triggering system of the detector and became transparent for about 100 µs, the maximum drift time of the electrons in the TPC when the triggering signal was received. After that, the grid hindered the passage of electrons and ions by applying an alternating voltage to it. The grid remained opaque for about 180 µs, the time needed by the gas mixture to drift from the anode wired to the gate. The entire process resulted in a dead time of the TPC of around 280 µs, capping out

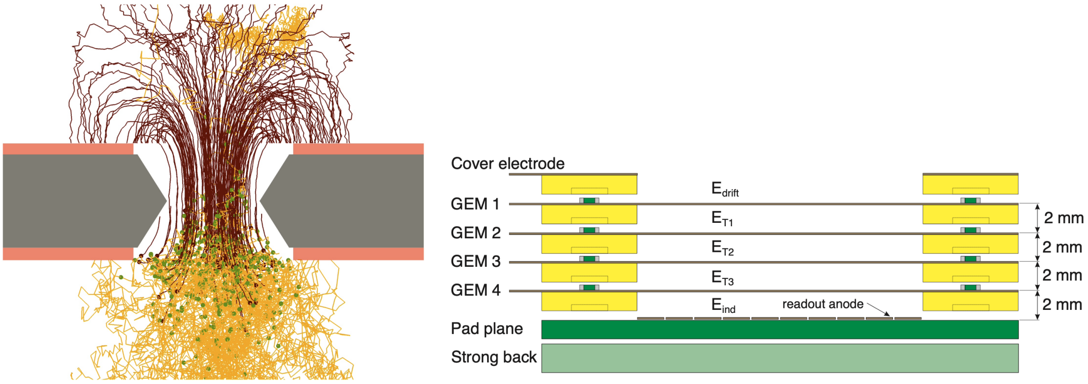


Figure 3.6: Left: Cross section of a hole in a GEM foil. The dark and light lines indicate the drift of ions and electrons, respectively, dots indicate places of ionization. It shows how incoming electrons enter the hole and multiply through ionization while most of the liberated ions are captured at the upper coated layer of the foil. Figure is taken from [34].
Right panel: The configuration of the four GEM foils in the readout detector in the TPC of ALICE. Details and R&D results can be found in [35]. Figure taken from [34].

the maximum possible readout rate at 3.5 kHz (even though the interaction rate during operation was at 300 Hz) [34]. This rate is more than one order of magnitude smaller than the aimed Pb–Pb interaction rate of 50 kHz aimed at in Run 3. With such rates, one expects tracks to be piled up from 5 events in the TPC simultaneously, on average. This necessitates a continuous readout of the detector, in which case the gating grid can not be used [34]. On the other hand, the ion backflow without the active gating would cause distortions in the TPC drift volume that are too large. An alternative technology for the readout is necessary, which allows for a continuous readout and mitigation of the ion backflow. The targeted value for the ion backflow (defined as the ratio of the cathode to anode current) is 1 %, which needs to be achieved while maintaining a high electron collection efficiency in order not to compromise on the $\mathrm{d}E/\mathrm{d}x$ resolution of the TPC [34]. Gas Electron Multipliers (GEMs) fulfill these requirements and were adopted as charge amplifiers for the upgrade of the TPC. They consist of multiple layers of thin foils on which holes are arranged in a regular grid. Each foil consists of an insulating material on which a conductive surface is coated. The holes are implemented through photo-lithographic processing [36]. The electron amplification happens between the foils of the GEM detector. Incoming electrons are guided with electric fields to the holes, where they ionize the present gas. Through the potential difference applied in each layer, consecutive electron avalanches are created between each layer, leading to a multi-step amplification of the incoming electrons, which is readout at the readout pads at the bottom of the GEMs. The strength and shape of the electric fields at the holes effectively block ions from drifting back into the TPC drift volume. This is shown in the left panel of Fig. 3.6 with a simulation using the Garfield/ Magboltz [37] packages [34]. The configuration of GEMs used in ALICE TPC is shown in the right panel of Fig. 3.6. Each readout module consists of four layers of GEM foils. The distance to each other, the size of the holes, and the potential differences applied in each layer are optimized to reach the targeted levels of ion backflow and electron collection efficiency [35].

3.5 TOF

The Time of Flight detector is used for particle identification of hadrons at the momentum range between 0.5 GeV/c and 2.5 GeV/c based on their measured time of flight. For this, it needs to have an intrinsic resolution better than 90 ps. This is archived by an array of Multi-gap Resistive-Plate Chambers (MRPC), which are arranged in strips. Each strip is read out by 96 pads with an area of $2.5 \times 3.5\,\mathrm{cm}^2$. Together, 91 strips form a supermodule. Like the TPC, it covers a pseudo-rapidity range of $|\eta| < 0.9$ and the full azimuth angle. It has an inner radius of 370 cm and an outer radius of 399 cm and a modular structure consisting of 18 sectors with five detector modules each. Each module contains 15 and 19 strips of Multi-gap Resistive-Plate Chambers (MRPC) in the inner and outer positions, respectively. The working principle of these detectors is a high electric field and a gaseous volume between to resistive plates, which causes traversing particles to create an electron avalanche. The MRPCs have multiple of these gaps (2x5, since it is a double MRPC design). This design allows the fast readout and high time resolution necessary for the high multiplicities archived at the ALICE experiment. In combination with the timing information, a precise time of flight measurement is made possible. The upgrades of the TOF during LS2 concerned only the readout system in order to enable continuous readout.

3.6 Datataking strategy of ALICE 2

With these upgrades, the detector can handle the increased interaction rates by two orders of magnitude compared to Run 2 for Pb–Pb collisions. The data-taking strategy was adapted to take full advantage of the new capabilities. An overview of the read-out and reconstruction pipeline is schematically shown in Fig. 3.7. The CRUs collect the signals of the continuous read-out detectors and are assembled in so-called Heartbeat Frames (HBF), which correspond in length to one LHC orbit, i.e., ∼89.4 µs. From there, they are combined to Sub-Time Frames (STF) by the first level processors (FLPs), which also compress the data throughput from 3.5 TB/s to 900 GB/s, can perform first calibrations and are used, among others, for data quality control. An event processing node (EPN) collects the STFs and combines them to a Time Frame (TF). A TF corresponds usually to 128 LHC orbits and represents the smallest entity to which tracks can be associated. Thus, the TFs replace the notion of events and collisions, which was to group track together when the detector was read out in triggered mode. One EPN performs the online (synchronous) reconstruction of one TF and reduces the data throughput to 130 GB/s. The largest contribution to the data load comes from the TPC, and the reduction is archived through a full track reconstruction and clustering of the TPC hits and the removal of background hits. Necessary spacepoints are efficiently stored as relative coordinates. A first calibration of the space charge corrections is also performed online. The output from the EPNs is stored in the Disk

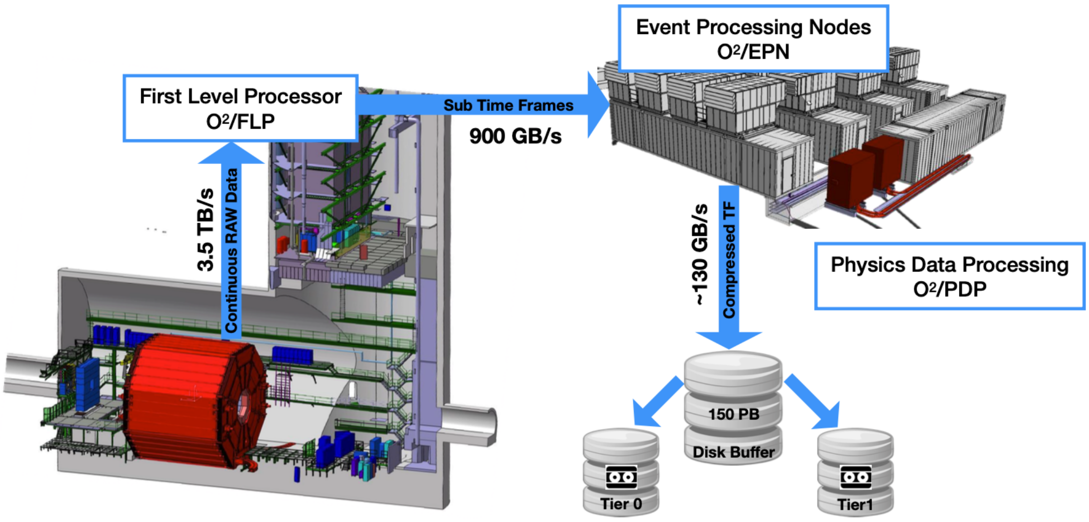


Figure 3.7: Overview of the data flow of ALICE in Run 3 during data taking and reconstruction. Figure taken from [27]

Buffer as a compressed TF (CTF), where the offline (asynchronous) reconstruction can be applied. For the asynchronous reconstruction, the full track information is used for the calibration. This includes an on-the-fly re-calibration of the TPC to take into account the contributions from secondary tracks, which originate from long-lived decays, e.g., weak decays, and have a displaced vertex with respect to the primary vertex. Once the offline reconstruction is complete, the dedicated software triggers can be applied, which select interesting events for physics analysis. Since they have access to the full event and track properties, the triggers can be based on complex physics observables, allowing the collection of dedicated datasets to access rare events precisely. This includes heavy flavor decays and three body femtoscopy triggers, which select three collimated particles with low relative momentum. Because of storage limitations, the reduction factor of all triggers is centrally decided and distributed among the physics working groups in ALICE. In order to free disc space for subsequent runs, the strategy foresees the deletion of all untriggered data except for a small subset which is kept as a minimum bias dataset. For Pb–Pb collisions as well as the pp reference runs at the same energy, no triggers are applied, and the full dataset is kept after the reconstruction. An exception to the strategy is done with the minimum bias dataset collected during 2022, which constitutes around 500 Billion pp collisions, the largest minimum bias sample ever collected. Due to the unprecedented number of events which allows for high precision multidimensional analysis it was decided to keep all of the reconstructed data. However, due to storage limitations, the raw data had to be deleted, making a re-calibration and reconstruction impossible on the full dataset. This situation will be discussed in more detail at the beginning of the next chapter. [1]

[1]This section is based on Chapter 5 of [27]

3.7 The O2 analysis Framework

The large amount of data that can be accumulated due to the upgrades and the offline reconstruction discussed in the previous sections requires an efficient analysis framework for analysis. The O^2 and O^2 Physics framework succeeds the previously used AliRoot and AliPhysics frameworks. O^2 stands for OnlineOffline, highlighting the fact that the framework implements the online readout and reconstruction tasks mentioned above as well as the offline analysis tasks. In the following, the basic concepts behind the offline component of the O^2 framework shall be summarized. A discussion of the specific tasks and software packages used for the analysis in this work can be found in Section 4.1. A general documentation of the framework can be found here [38], the source code in the GitHub repository here [39].

The data model is based on interconnected tables, where the column represents a physical property like, e.g., momentum in x-direction, reconstructed mass, etc., and a row represents an analysis object of the table, e.g. a track or a collision. The interconnectivity of the tables reflects the relations between the different objects. For example, multiple objects of the `Tracks` table can be linked to one object in the `Collisions` table. The framework allows for efficient filtering and partitioning of these tables, and a series of predefined macros and functions have very efficiently implemented common analysis steps such as event mixing and pairing of particles. The leading design principle is that the largest limitation of the analysis campaign will be the storage rather than the computing power. As a result, the offline reconstruction stores only the most important information in the Analysis Object Data, which are saved in the format of ROOT trees (AO2D.root files to be precise) and are used as the input data for each analysis. At that level, track candidates, for example, do not have an assigned momentum or distance to the primary vertex. This information is calculated on the fly during the analysis by dedicated helper tasks. These helper tasks create new tables in the same AO2D format in memory so that they can be consumed by subsequent tasks, which can either create new tables after an additional processing step or write the results to an `AnalysisResults.root` file, which contains all the histograms and QA plots needed for the analysis. This way, an analysis in the O^2 framework consists essentially of a series of chained tasks that create and consume tables while processing the data. The tables created by the helper tasks exist only during the run time of the analysis in memory and are deleted afterward. However, physics groups within ALICE are encouraged to develop their own data models that can be used for specialized analyses. These data models have their own producer tasks, which create tables similar to helper tasks. However, strict constraints on the reduction factor of their size with respect to the original input data were imposed. These tables can be saved as derived datasets and are available to all analyzers. The event and particle selection should be broad enough to accommodate different analyses, including systematic variations. This way, the steps common to a larger group of analyses, including the computationally expensive iteration over the whole dataset, need to be performed only once. Effectively, once the derived data are generated, this allows an analysis of a few hundred billion collisions over the course of an afternoon. The femtodream framework is an example for that as it is based around a common data model, as will be discussed in detail in Section 4.1.

4 Data Analysis

In this chapter, the analysis steps to obtain the same and mixed event distributions in Eq. 2.5 will be explained. This part of the analysis requires the largest computational effort, since it iterates through the entire dataset and calculate the k^* of each identified pair, which passes the selection criteria. The output of this analysis part is the nominator and denominator of Eq. 2.5 along with QA histograms to investigate the quality of the collected data including the purity. In a sense, this step represents the largest data compression, starting from an input of 3.1 PB to a few MB of output. This huge computational effort requires an efficient framework with large throughput such as the O2 framework [38], which is tailored for such purposes and has been discussed in the previous chapter. Specifically, the analysis is done by the tasks within the `FemtoDream` [40] package of the `O2Physics` [39] repository, which will be introduced in the next section.

4.1 The FemtoDream Framework

The development of the `FemtoDream` framework began within the group before this thesis. Most functionalities had already been available and validated using the pilot beam data from 2021. As part of the work presented in this thesis, the development was continued and the functionalities were extended. The core of `FemtoDream` is its own data model, which is designed in line with the specifications of the O^2 framework described in Section 3.7, i.e., for maximal data reduction and universal usability for all femtoscopy analysis. Two O^2 tables- one for particles and one for collisions- are created by the `FemtoDreamProducerTask`. It stores only events and particles that pass specific selection criteria, which will be specified in Sections 4.4 and 4.5. Only the minimal information necessary for femtoscopic analysis is stored for both collisions and particles. An overview of the stored variables can be found in Table 4.1.

Particles are saved in the `FDParticles` table. The kinematic variables p_{T}, η, and φ are stored for each candidate in order to compute their momentum and subsequently the k^* of the pairs. Additionally, a topological variable (`TempFitVar`) is needed for the estimation of the primary fraction of the candidate. The variable depends on the type of particle. In the case of tracks, the distance of closest approach to the primary vertex in the transverse plane (DCA_{xy}) is used. In the case of V0 candidates, which are neutrally charged particles reconstructed via their weak decay into two charged tracks, such as Λ ($\overline{\Lambda}$), the cosine of the pointing angle (CPA) is used. The pointing angle is defined as the angle between the momentum direction of the V0 candidate and the line connecting its decay vertex with the primary vertex. The particle type itself is stored in the integer variable `PartType`, which is used by the framework in an internal enumeration scheme in order to consistently associate the saved value in the

TempFitVar variable to the correct quantity. In the case of reconstructed particles such as V0s (but in principle extendable to more complicated decays), the invariant mass is stored in MLambda and MAntiLambda for the particle and antiparticle mass hypothesis, respectively. The hypothesis corresponds to the assumption made for the tracks upon reconstruction. The Cut variable stores the cuts that the particle has passed but whose values are not relevant for the computation of the observable in femtoscopy. These cuts primarily concern detector-specific variables, such as the number of TPC clusters or hits in the ITS, and topological variables, such as the DCA to the primary vertex in z direction. The cut is stored in the integer representation of a bitmask which is created internally by the framework, by ordering the variables with their respective cut options in a list and assigning a 1 if the cut is passed and a 0 if the cut is not passed by that particle. The PIDCut stores the PID agreement of a track with all by the framework supported track species in terms of the number of standard deviations ($n\sigma$). The currently supported species are Protons, Pions, Kaons, and Deuterons. Instead of saving the agreement with each hypothesis individually, different confidence intervals can be specified independently for the TPC and TOF PID. Similar to the Cut variable, they are ordered by the framework internally, and a cut bit mask stores for each particle if it lies within the confidence intervals of any of the PID hypotheses. With these restrictive measures, it is possible to reduce the disk usage of the FemtoDream datamodel to 20 Bytes per particle.

Each particle is linked to an object in the FDCollisions table. It stores the z position and the multiplicity of the event, which is necessary for the correct event mixing (more on that in Section 5.1.1). Two different values for the multiplicity are stored. The MultV0M contains the signal amplitude of the V0M detector, while MultNtr counts the number of tracks contributing to finding the primary vertex. Ideally, a calibrated multiplicity percentile based on the V0M detector should be used for a comparable multiplicity estimation with previous ALICE results from Run 2. As a further advantage, this reduces the bias from the multiplicity estimation as it relies on forward information, while the tracks are reconstructed mostly in mid-rapidity. At the moment of writing, the calibration was not available, and the variable MultNtr is used instead for multiplicity estimations in this work. The magnetic field is needed for the close pair rejection (see Section 5.1.1), and the Sphericity variable stores the sphericity of the event but is currently not in use.

The FemtoDreamProducerTask is used as a table producer task that should run once over the whole reconstructed data to create the derived datasets for femtoscopy in the form of tables, as explained above. The selection should be wide enough to include all candidates and systematic variations that will be used in the analysis. The analyzers can select a subset of the particles via the cut bits and perform the analysis directly on the derived datasets. This is done via pairing tasks; in particular, the PairTaskTrackTrack and the PairTaskTrackV0 are used in this work. They create the same and mixed event distributions needed in Eq. 2.5 and provide correlation histograms with the event properties for rudimentary QA. For further QA and inspections of the data, the producer task supports a debug mode, in which the FemtoDream data model is extended by continuous variables of all the cuts that are represented in the cut bit variables. However, these tasks cannot run on the whole dataset because of the

Name of the variable	datatype	Description of the Cut
FDParticles		
Pt	float	p_T of the particle
Eta	float	η of the particle
Phi	float	φ of the particle
PartType	integer	integer to
Cut	integer	cutbit for selection cuts
PIDCut	integer	cutbit for PID cuts
TempFitVar	float	variable for the template fits
Children	internal link	links a Λ ($\overline{\Lambda}$) candidate to its decay products
MLambda	float	invariant mass of Λ
MAntiLambda	float	invariant mass of $\overline{\Lambda}$
FDCollisions		
PosZ	float	z coordinate of the primary vertex
MultV0M	float	multiplicity based on the V0M signal
MultNtr	float	number of primary charged tracks with $\eta < 0.8$
Sphericity	float	event sphericity
MagField	float	strength of detector magnet

Table 4.1: Femto datamodel of the particle candidates and collisions

excessive consumption of resources needed (in particular memory). Instead, dedicated sampling datasets, which contain about 5% of all runs within a data period, can be used. The output of the `FemtoDreamProducerTask` in debug mode is further processed by dedicated debug tasks, which produce the necessary QA plots. Use cases include, for example, PID vs. p_T plots, which are used to calculate the purity of the proton candidates in Section 4.5.

4.2 Analysed datasets

The analyzed data in this work are the whole Minimum Bias (MB) datasets at nominal interaction rate that were collected by ALICE during the 2022 data taking. Together, they constitute about 500 billion events, the largest Minimum Bias dataset ever collected by the ALICE detector. The datasets are divided into five data-taking periods, which are summarized in Table 4.2, together with their original size, the size after the skimming (i.e. the size of the derived data using the FemtoDream datamodel). The large reduction factor of 200 highlights the efficiency of the FemtoDream framework and the strategy of using derived datasets, where the bulk of the analysis needs to be performed only once.

As described in the previous chapter, the ALICE experiment underwent major upgrades concerning the detector technology and the readout. The readout-technology, in particular, poses a challenge for the calibration of the data due to the necessary corrections of the space charge distortions in the TPC. Some further investigations were needed to fully understand the distortions and apply the corrections. This resulted in three additional iterations of the offline reconstruction, leading to the pass4 reconstruction, which will be used in this thesis. However, as will be shown in the following, not all problems could be solved with the pass4 reconstruction. For example, the PID separation between protons and Kaons drops drastically already at low momenta of $\approx 2\,\mathrm{GeV}/c$. As will be shown in the following chapters, it was possible to extract meaningful physics results from the data, which seem to be compatible with previous results. In fact, due to the overall good quality of the data and the great opportunity that 500 billion Minimum Bias events provide, the collaboration decided to keep the entire 2022 dataset for future analysis instead of deleting untriggered events. This decision is backed up by the work presented in this thesis. However, the limitations of the storage do not allow the storing of the raw data, i.e., the CTFs. Instead, the AO2Ds will be kept, which effectively freezes the status of the reconstruction and calibration as a recalibration and new reconstruction cannot be done without the raw data. Still, for a subset of the MB dataset, the so called "golden runs", the CTFs will be kept and possible new reconstructions can be performed with them. Investigating the differences between pass4 and a possible pass5 will be interesting when the latter becomes available.

dataset	original size	skimmed size	O2Physics tag of skimming
LHC22m_pass4	277.1 TB	1.31 TB	`O2Physics::daily-20231031-0100-1`
LHC22o_pass4	1800 PB	8.94 TB	`O2Physics::daily-20231031-0100-1`
LHC22p_pass4	162.2 TB	0.74 TB	`O2Physics::daily-20231031-0100-1`
LHC22r_pass4	410.3 TB	1.93 TB	`O2Physics::daily-20231031-0100-1`
LHC22t_pass4	361.1 TB	1.67 TB	`O2Physics::daily-20231031-0100-1`

Table 4.2: List of skimmed datasets with the corresponding derived datasets used for the analysis, and the ID and tag of the hyperloop train used for the skimming.

dataset	# events	# protons (antiprotons)
LHC22m	2.52 e10	2.51 e9 (2.19 e9)
LHC22o	1.72 e11	1.60 e10 (1.42 e10)
LHC22p	1.42 e10	1.37 e9 (1.17 e9)
LHC22r	3.72 e10	3.71 e9 (3.26 e9)
LHC22t	3.22 e10	2.97 e9 (2.63 e9)

Table 4.3: Summary of the number of events and proton (antiproton) candidates for each analised period.

4.3 Monte Carlo Dataset

At the time of writing, no MC dataset was available, which is anchored to the detector conditions of the measured events in this analysis. A general-purpose dataset, very limited in size, was available, which was intended for testing purposes and code validation before the start of Run 3. Therefore, it is not expected to reproduce the detector behavior accurately. The lack of Monte Carlo simulations made an estimation of the fractions impossible and necessitated a data-driven estimation of the proton purities.

4.4 Event Selection

The event selection is implemented in the `femtoDreamProducer` task, which was configured to select events using the *sel8*, which rejects physically uninteresting events such as collisions with residual gas in the beam pipe. Additionally, in order to minimize detector acceptance effects, events whose z-component of the primary vertex lied more then 10 cm away from the center of the detector, were rejected. The number of selected events are reported in Table 4.3, however, due to the structure of the `femtoDreamProducer` task, which simultaneously performs the candidate selection, the number of events reported here are the events with at least one proton or antiproton candidate.

4.5 Proton candidate selection

As a basis for the proton selection criteria, the cuts from previous analysis were used [8, 9]. Due to the discussed difficulties with the reconstruction and calibration, adjustments had to be made. The selection criteria are summarized in Table 4.4.

The DCA_{xy} and DCA_z are the distance of closest approach of the reconstructed particle track to the collision vertex in the xy plane and the z direction, respectively. They are obtained by extrapolating the track to the primary vertex. Cuts on these variables ensure the suppression of non-primary protons, that could originate, for example, from a weak decay of a Λ hyperon. The minimum number of TPC clusters maintains a good track quality, ensuring that there are enough hits in the TPC for a good momentum measurement and PID. The minimum cut on the number of crossed rows in the TPC suppresses the contribution from spiraling tracks in the magnetic field. In Run 2 there was an additional cut on the number of shared TPC clusters, where tracks with one or more shared clusters were removed from the analysis. This cut is removed for this analysis. It was found that this cut decreases the number of available protons in the sample too drastically while not improving the purity of the protons at the same time, as the systematic studies presented in the next section show. More thorough investigations revealed a bug in the reconstruction which was used to reconstruct the LHC22o dataset. This bug affects only the number of shared clusters in the TPC for the LHC22o period. With the removal of the cut the similar behavior of all periods is restored additional to the increased number of candidates. The opening of this cut will be justified in the following with a study of the proton purity and in Section 5.1.1 on the level of the correlation function. Because of storage limitations, it was decided that unmatched tracks, i.e., tracks in the TPC, that cannot be matched to hits in the ITS and have no associated weak decay, will be permanently deleted as they probably originate from material knock-out. In order to stay compatible with future iterations of the analysis, the tracks are required to be global tracks, i.e., to have at least one hit in the ITS. Finally, the pseudorapidity cut at $|\eta| < 0.8$, ensures that all tracks are within the full ITS, TPC and TOF acceptance.

The number of standard deviations $n\sigma$ between the measurement and the theoretical prediction of the energy loss and time-of-flight in the TPC and TOF, respectively, are used for the proton identification. Both detectors have different momentum ranges of optimal operation. To fully take advantage of both, proton candidates with a momentum smaller than 0.75 GeV/c were selected using exclusively the identification with the TPC. For proton candidates with a larger momentum, the TPC and TOF information are combined by calculating the geometric mean $n\sigma_{\mathrm{comb.}} = \sqrt{(n\sigma_{\mathrm{TPC}})^2 + (n\sigma_{\mathrm{TPC}})^2}$. The resulting PID distributions are shown in Fig. 4.1 for protons (upper row) and antiprotons (lower row) for the measurement in the TPC (left), the TOF (middle), and the combined PID (right). The distributions show no dominant contributions from contamination by other particle species at the selected momenta regions except for $p_{\mathrm{TPC}} > 2.0$ GeV/c, where the proton signal becomes less dominant compared to the background. This is mainly due to the contamination with Kaons because of the reduced separation capabilities of the TOF in the current state of the detector calibration. In Run 2, the separation capability between Kaons and protons used to drop significantly starting at p_{T} of around 4.0 GeV/c. The behavior of the purity as a function of p_{T} will be investigated in more detail in the next section. Here, the result is anticipated that the purity already drops below 80 % for proton candidates with p_{T} larger than 2.0 GeV/c, which motivates the maximum p_{T} cut at that value.

The resulting kinematic variables p_{T}, η and ϕ as well as the DCA$_{xy}$ after the described cuts are shown in Fig. 4.2. The proton candidates are shown in blue and the antiprotons in red. Their four-momentum can be fully reconstructed using these three kinematic variables. Both distributions for protons and antiprotons are compatible with each other. The structures in the ϕ distribution are due to the TPC sectors. and the matching inefficiencies between ITS and TPC. The η distribution shows structures related to the reconstruction of the tracks. The asymmetry in the middle is related to a dead TPC electrode while the reason for the other drops at $\eta \approx \pm 0.6$ is still under investigation. However, the fact that protons and antiprotons exhibit the same structure is a good sign.

Purity Estimation of the Proton Sample

In Run 2 analyses, the purity of tracks was usually determined with the help of Monte Carlo simulations. Since no anchored simulations were available at the time of writing, the purity estimation had to be done in a data-driven approach. For that, the PID selection in the TPC and TOF were widened in order to include a large portion of the background. The p_{T} dependent PID distribution, similar to Fig. 4.1, can then be subdivided into smaller p_{T} regions and projected on the $n\sigma$ axis. The distribution is fitted with a Gaussian function with an exponential tail on the right shoulder for the signal and an exponential plus a linear function to describe the background. The exponential tail was empirically found to be necessary to adequately describe the $n\sigma$ distribution in the TPC and TOF. The fit function is given by

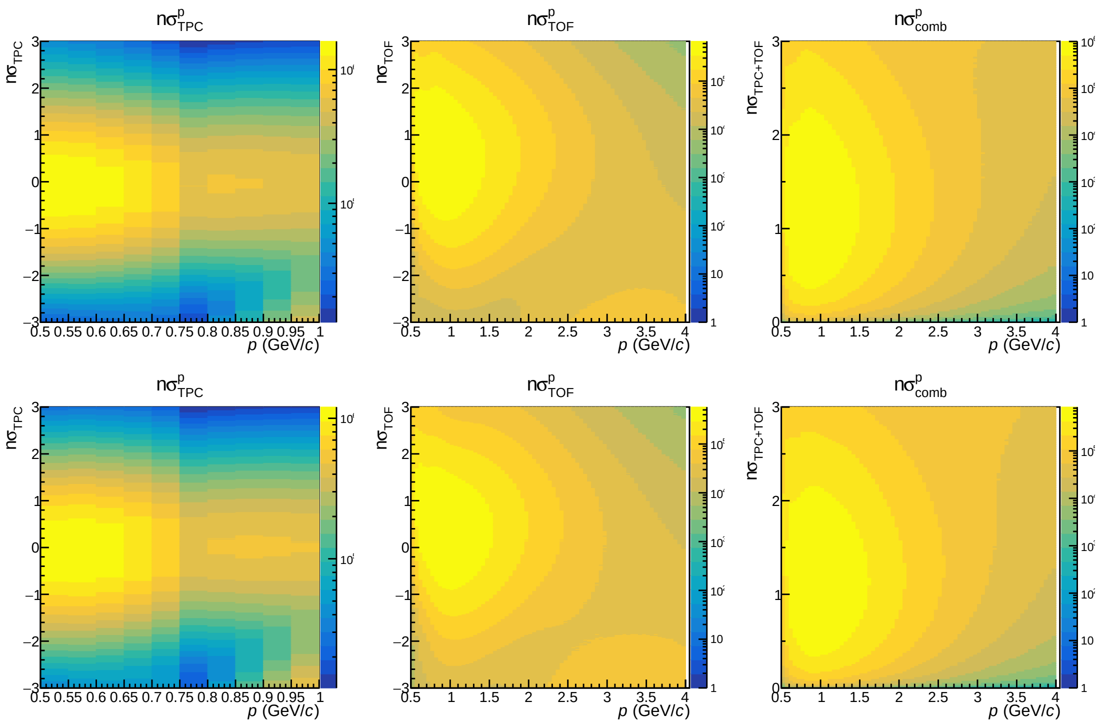


Figure 4.1: Particle identification for (anti)-protons. For $p < 0.75\,\mathrm{GeV}/c$ the specific energy loss information provided by the TPC in the form of its deviation from the theoretical expectation value is used (nσ) as shown on the left. For $p > 0.75\,\mathrm{GeV}/c$ also the time-of-flight shown on the right is used combined with the TPC information.

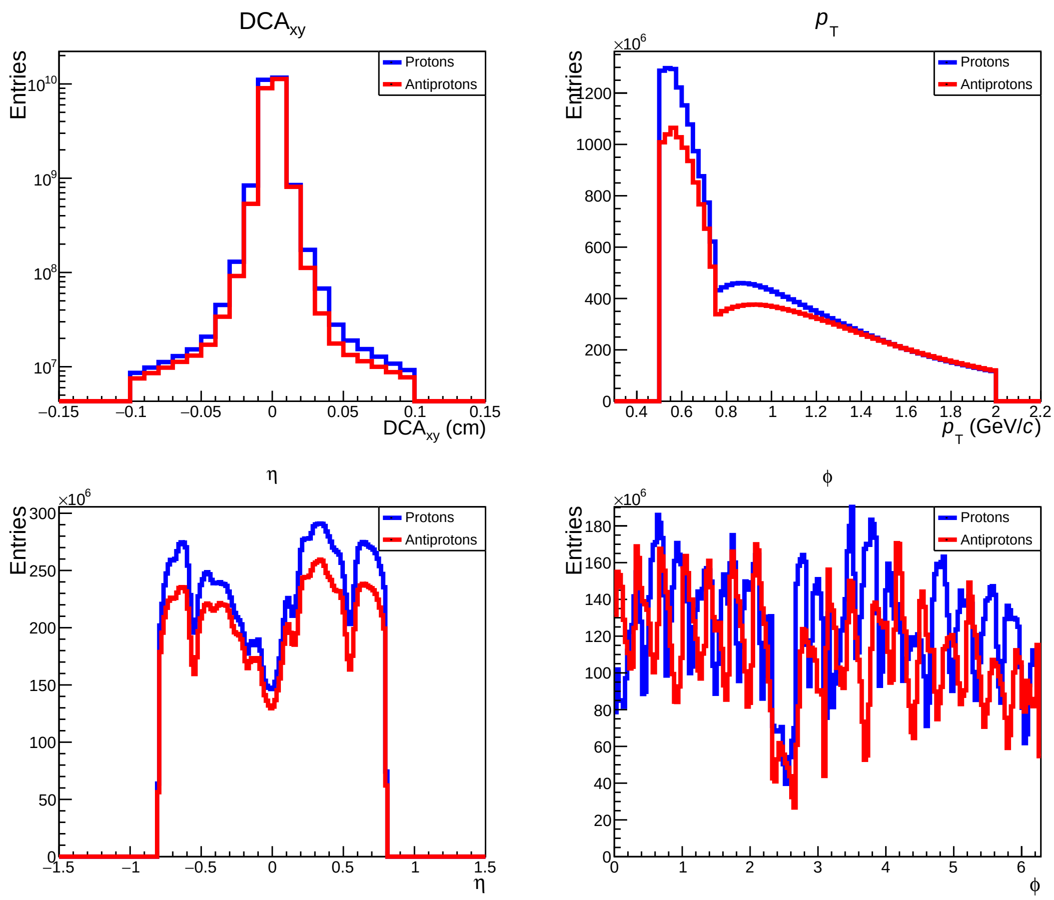


Figure 4.2: QA histograms show distributions of ϕ, η, p_{T}, and DCA$_{xy}$ for protons (in blue) and antiprotons (in red).

Selection criterion	Value
Pseudorapidity	$\lvert\eta\rvert$<0.8
Transverse momentum	0.5< p_T <2 GeV/c
TPC cluster	n_{TPC}>80
Crossed TPC rows	n_{crossed}>70
Tracks with shared TPC clusters	no Cut on the shared clusters applied
Distance of closest approach xy	$\lvert\mathrm{DCA}_{xy}\rvert$<0.1 cm
Distance of closest approach z	$\lvert\mathrm{DCA}_{z}\rvert$<0.2 cm
Particle identification	$\lvert n_{\sigma,\mathrm{TPC}}\rvert$ <3 for p <0.75 GeV/c $n_{\sigma,\mathrm{comb.}}$<3 for p>0.75 GeV/c

Table 4.4: Proton selection criteria.

$$f(x) = \begin{cases} [0] \cdot \mathrm{Gaus}(x, [1], [2]) & \text{for } x \leq [3] + [1] \\ [0] \cdot \mathrm{Gaus}([3] + [1], [1], [2]) \cdot \exp\left(-[3] \cdot \frac{x-[3]-[1]}{[2]2}\right) & \text{for } x > [3] + [1] \end{cases} \\ +[4] + [5] \cdot x + [6] \cdot \exp(-[7] \cdot x), \tag{4.1}$$

where the numbers in brackets indicate the fit parameters. The purity $\mathcal{P}$ can then be extracted as the ratio between the integral over the signal function (S) and the sum of the integral of the signal and background functions ($S + B$)

$$\mathcal{P} = \frac{S}{S+B}. \tag{4.2}$$

In order to investigate the influence of the cut on the shared clusters on the purity of the proton sample, the PID distributions were separately obtained for all five different run periods and for different maximal cuts on the shared clusters, namely no shared cluster and at most 80 shared clusters allowed. The resulting p_T dependent PID distributions were spitted up in the four different p_T ranges (0.5-0.6) GeV/c, (0.6-0.75) GeV/c, (0.75-1.25) GeV/c, and (1.25-1.75) GeV/c. For the first two ranges the PID distribution of the TPC was used, while the second two ranges were obtained from the TOF PID. These distributions were fitted with Eq. 4.1, and the purity was extracted according to Eq. 4.2. The results are shown in Fig. 4.3 and Fig. 4.5 for the range (0.5-0.6) GeV/c for protons and antiprotons, respectively, and in Fig. 4.4 and Fig. 4.6 for the range (0.75-1.25) GeV/c, respectively for protons and antiprotons. Each row represents another period while the and left and right panels correspond to the very strict cut (no shared clusters allowed) and open cut (80 shared clusters allowed at most), respectively. It should be noted that setting the cut on the number of shared clusters to 80 is almost equivalent to removing the cut, since the majority of the distribution lies below the value of 80 number of shared clusters. The values of the purity are reported in the bottom left part of the panels, together with the number of candidates (signal plus background). Figures 9.1 to 9.4 in the appendix show the same fits for the larger values of p_T.

Overall, the fit quality is not ideal in the background regions, as the functional form of Eq. 4.1 struggles to describe all structures in the PID distribution. These structures arise due to the imperfect calibration of the TPC and TOF. With this in mind, the fit describes the signal region very well, and the figures are plotted on a logarithmic scale, yielding better visibility at the cost of visually exaggerating the influence of the mismatch in the background shape. The differences between the fit function and the background shape are negligible compared to the difference of more than one order of magnitude between the signal and the background. The purities are compatible with each other across all five periods and for both cuts on the number of shared clusters. On the other hand, the increase of proton candidates is significant, most notably in the period LHC22o, as the number of candidates per events, reported in the figures with the fit results, suggest. Moreover, with the most open cut on the number of shared clusters, all five periods reach the same value of candidates per events, further supporting the hypothesis that the cut on the shared cluster is, at least in the current pass4 reconstruction, obsolete. This will be further crosschecked by comparing the correlation function in Section 5.1.1.

Using the same fitting strategy, the purity can be studied as a function of the transverse momentum in a finer binning. For that, the p_{T} vs $n\sigma$ distributions are subdivided in intervals with the width of 0.0625 GeV/c and the projection is fitted with (4.1), as before. This leads to an estimation of the proton and antiproton purity as a function of p_{T}, which is depicted in Fig. 4.7 in blue and red, respectively. It shows a flat behavior for transverse momenta up to about 1.4 GeV/c. Above these, the purity for protons and antiprotons declines steeply and drops below 80 % for transverse momenta slightly above 2.0 GeV/c. In order to maintain a high purity for a good signal extraction, the maximum p_{T} cut at 2.0 GeV/c was applied. This measure will not have a large impact on the low k^* region, where the signal of the correlation function is expected, since the low k^* pairs originate mainly from low p_{T} particles. However, this will limit significantly the available pairs with large m_{T} and thus decrease the statistical precision of the source size measurement in the largest m_{T} interval. More on this will be discussed in Section 5.1.1 and Section 7.

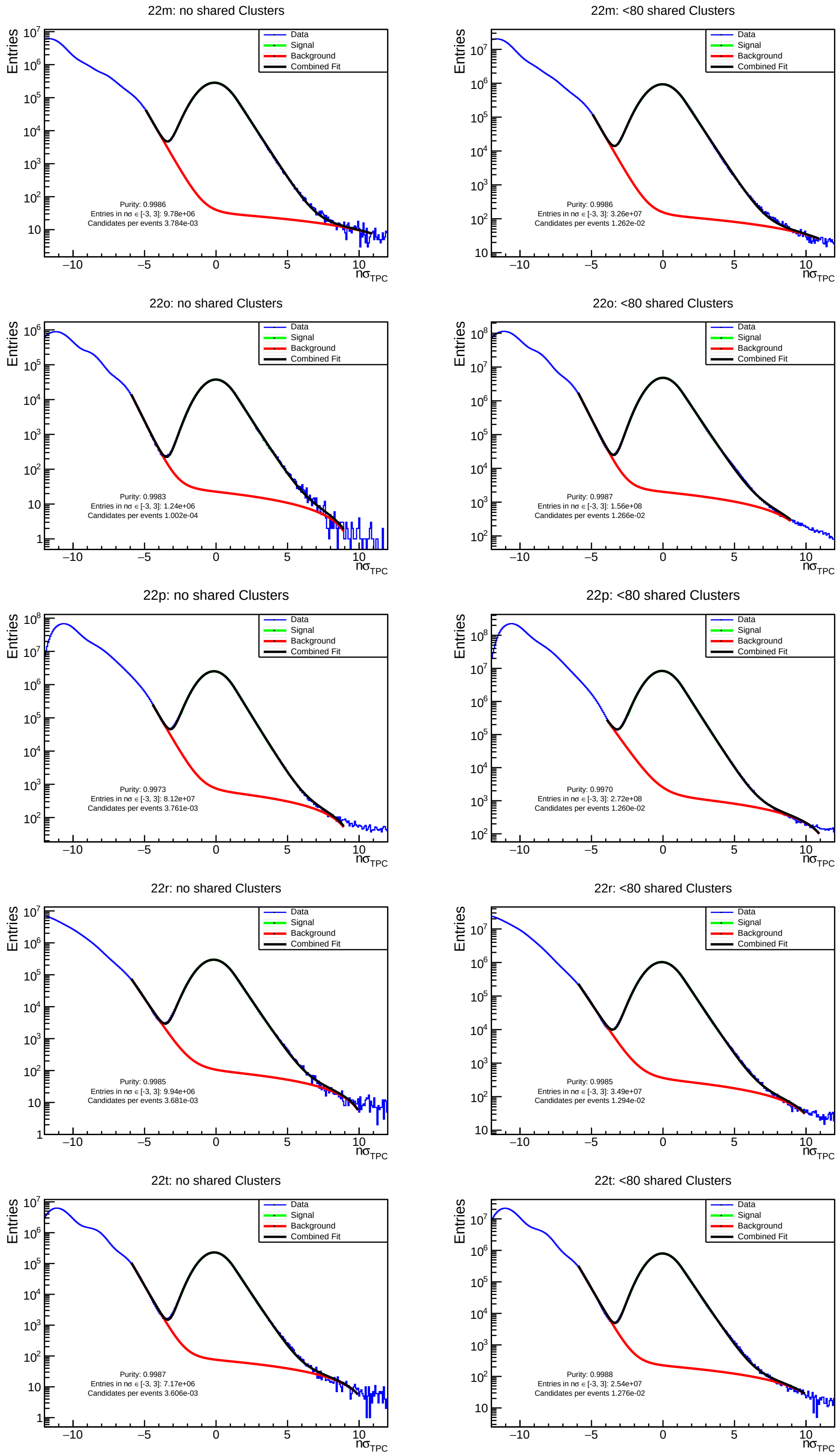


Figure 4.3: Fit of the TPC $n\sigma$ distribution for protons with $p_{\mathrm{T}} \in [0.5, 0.6)$ GeV/c for proton selection with no allowed shared clusters (left panels) and 80 allowed shared clusters (right panels). Each row corresponds to another datataking period of 2022.

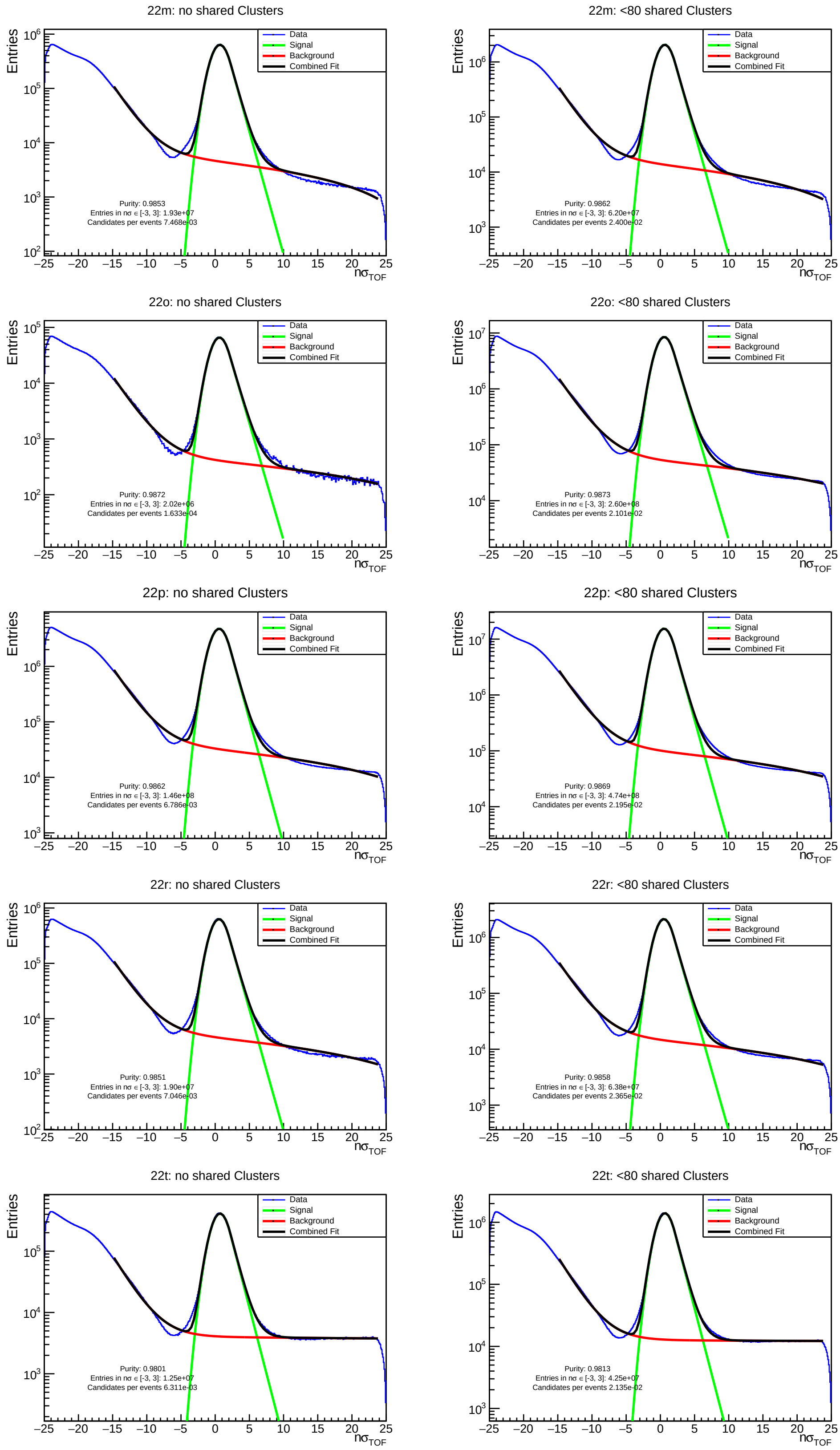


Figure 4.4: Fit of the TOF $n\sigma$ distribution for protons with $p_{\mathrm{T}} \in [0.75, 1.25)$ GeV/c for proton selection with no allowed shared clusters (left panels) and 80 allowed shared clusters (right panels). Each row corresponds to another datataking period of 2022.

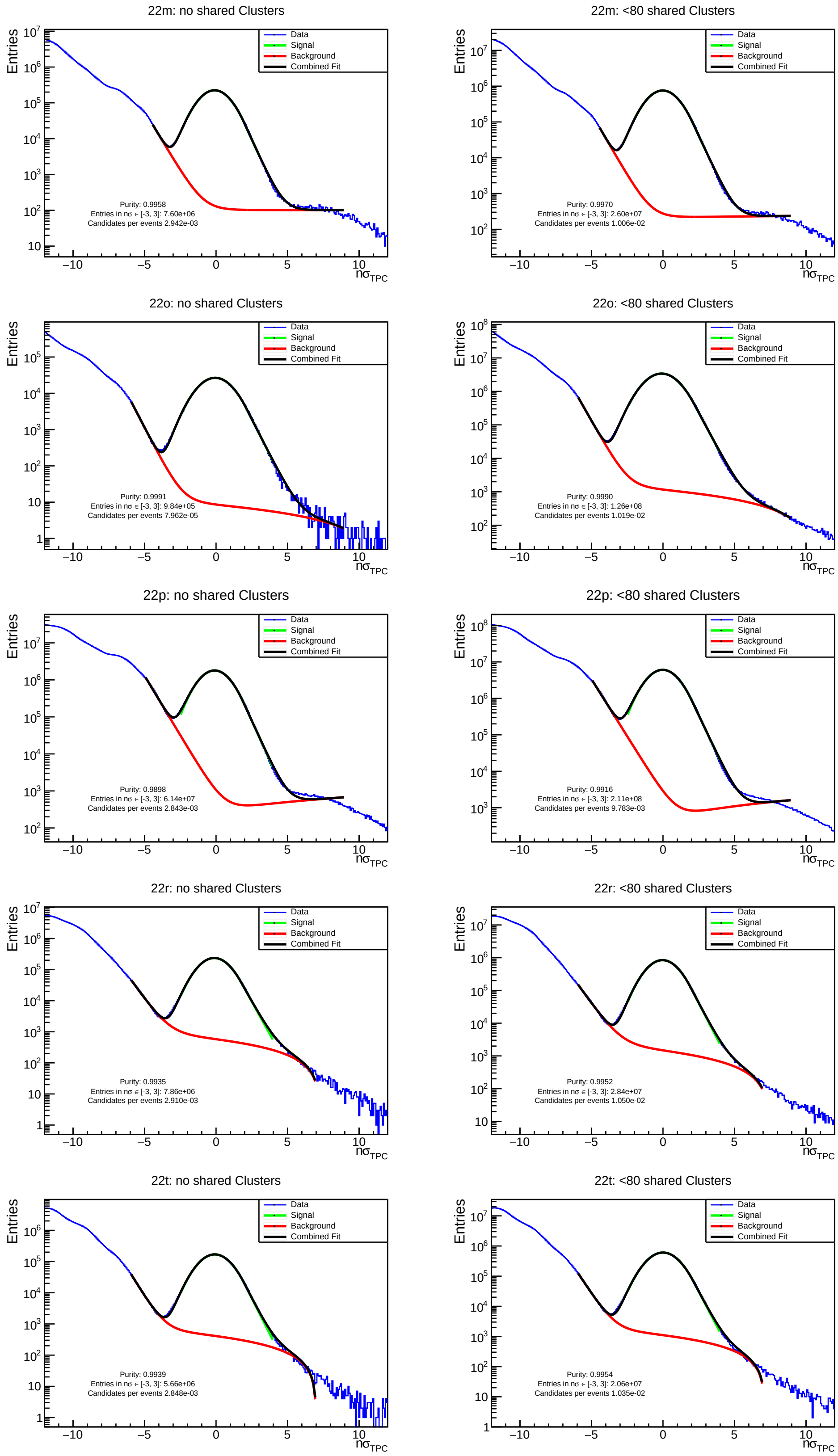


Figure 4.5: Fit of the TPC $n\sigma$ distribution for antiprotons with $p_{\mathrm{T}} \in [0.5, 0.6)$ GeV/c for proton selection with no allowed shared clusters (left panels) and 80 allowed shared clusters (right panels). Each row corresponds to another datataking period of 2022.

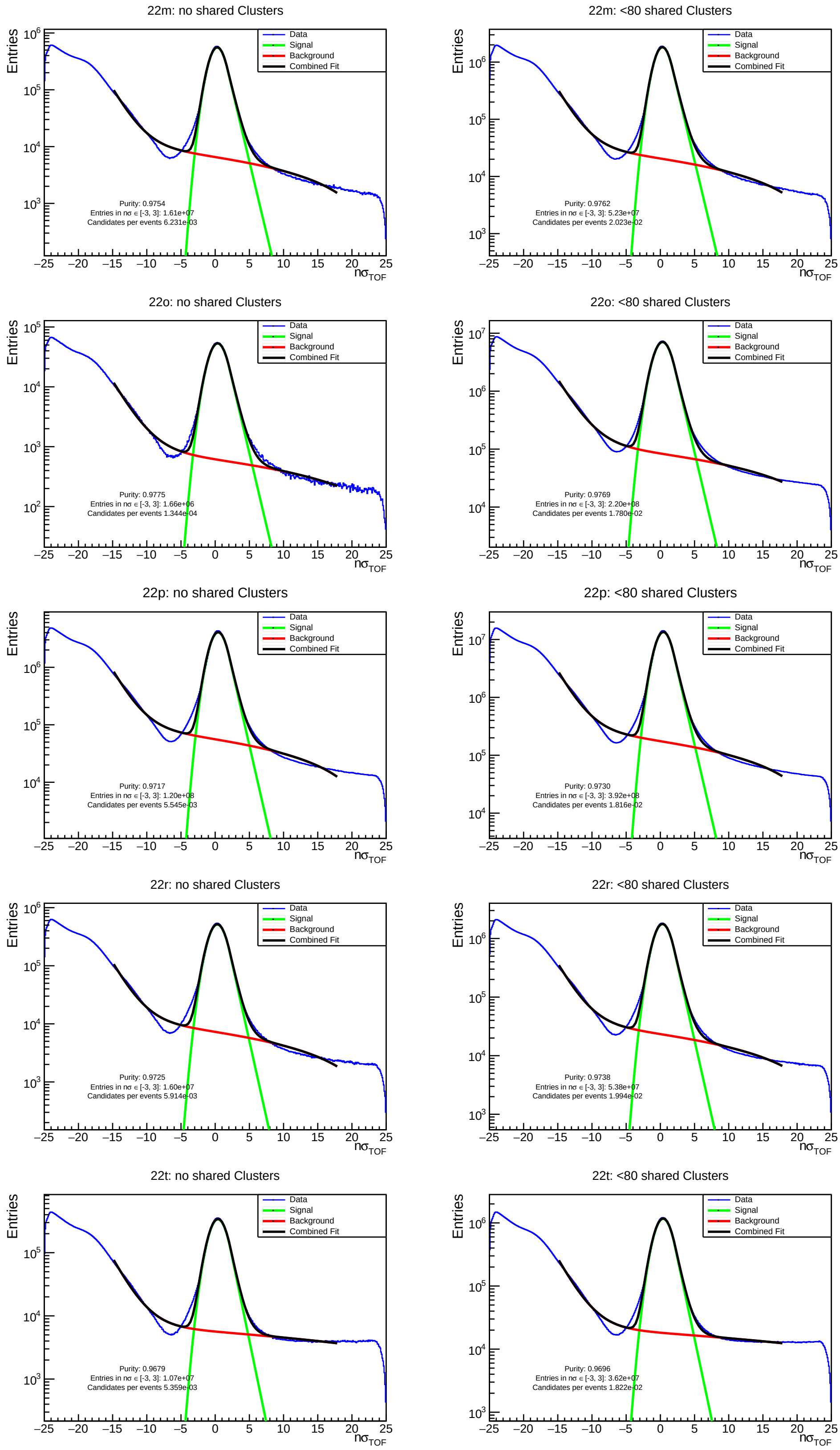


Figure 4.6: Fit of the TOF $n\sigma$ distribution for antiprotons with $p_{\mathrm{T}} \in [0.75, 1.25)$ GeV/c for proton selection with no allowed shared clusters (left panels) and 80 allowed shared clusters (right panels). Each row corresponds to another datataking period of 2022.

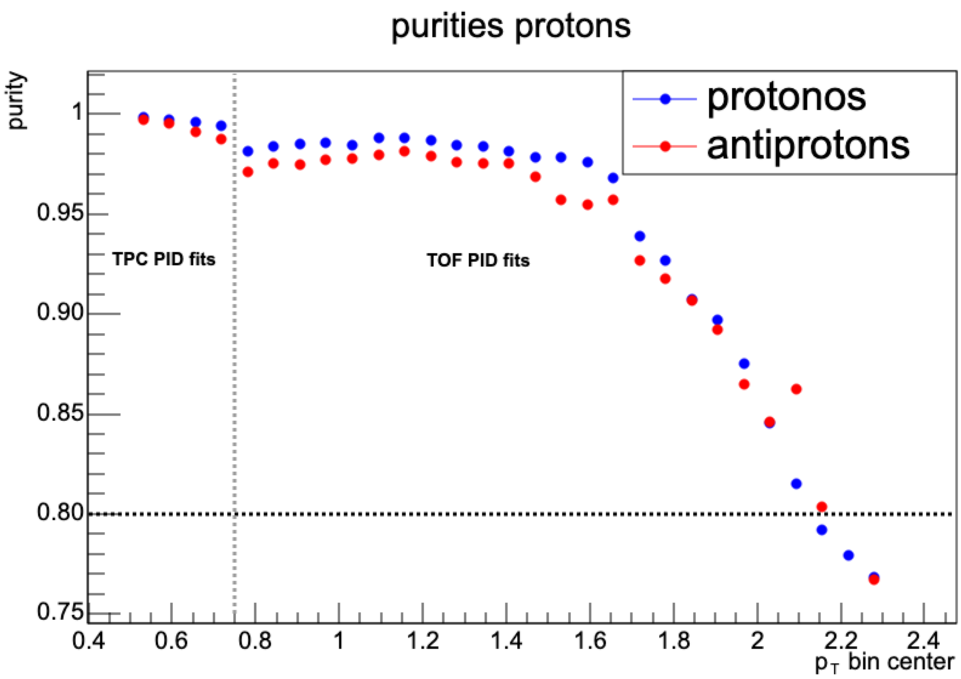


Figure 4.7: Proton and antiproton purities as a function of p_{T} in blue and red, respectively. The p_{T} interval in which the TPC and the TOF PID plots are used, is indicated and the limit of 80 % purity below which protons are rejected.

5 Femtoscopic Analysis

In this chapter, the selected proton candidates are used to construct the p–p correlation functions. In Section 5.1.1, the pairing and the event mixing are shown. This is followed by a detailed look at the experimental p–p correlation function. Section 5.2 discusses how the correlation functions are modeled from the theoretical side and how they are fitted to the data to measure the source size.

5.1 The experimental correlation function

5.1.1 Pairing and mixing

All candidates selected in the previous section are used to obtain the pairs for the same event distribution (SE). After a pair is constructed, one further check is applied, the so called close pair rejection. It removes auto-correlations caused by track splitting, a detector effect in which one physical track is reconstructed as two separate ones. A pair is rejected if the quadratic sum of the difference of angles η and φ of the two protons is smaller than 0.01 (i.e., $\sqrt{(\Delta\eta)^2 + (\Delta\varphi)^2} < 0.01$). Figure 5.1 shows the $\Delta\eta$ vs. $\Delta\varphi$ distribution before (right panel) and after (left panel) the close pair rejection. The dominant peak around $\Delta\eta = \Delta\varphi = 0$ in the left panel is clearly visible, showing the large contribution of auto-correlated pairs, that can be removed with the close pair rejection.

The k^* is calculated for all remaining pairs, and filled into a three-dimensional histogram, where the event multiplicity and the average m_T (as defined in Eq. 1.1) are the two other axis. From this three-dimensional histogram, it is possible to obtain the k^* distribution for different m_T and multiplicity intervals by slicing the histogram accordingly. As discussed in Section 4.1, the number of charged tracks in mid rapidity are used as a proxy for the event multiplicity. All p–p pairs are divided in six multiplicity classes, namely $[0,7), [7,11), [11,15), [15,20), [20,27)$, and $[27,200)$. The m_T intervals are chosen to be the same seven as in the previous analysis of the m_T scaling [9]. The number of pairs in the low k^* region ($k^* < 200\,\mathrm{MeV}/c$) for each multiplicity and m_T interval are summarized in Table 5.1.

The mixed distribution (ME) is obtained by pairing candidates from different events with each other. The underlying assumption is that efficiency and acceptance affect the same event and mixed event distribution in the same way. Due to the ratio in Eq. 2.5, both effects cancel in the calculation of the correlation function. Therefore, it is not necessary to apply efficiency and acceptance corrections. However, it requires mixing particles with similar event properties, i.e., a similar z coordinate of the primary vertex and a similar event multiplicity. The former ensures that both particles in a mixed

$N_{tr.}^{primary}$ / m_T [GeV/c]	pair	[0,7)	[7,11)	[11,15)	[15,20)	[20,27)	[27,200)
[1.02, 1.14)	p − p	7.9e+05	1.7e+06	2.4e+06	3.3e+06	4.1e+06	6.3e+06
	$\overline{p}-\overline{p}$	5.0e+05	1.1e+06	1.6e+06	2.3e+06	2.8e+06	4.4e+06
[1.14, 1.2)	p − p	2.8e+05	7.2e+05	1.1e+06	1.6e+06	2.0e+06	3.3e+06
	$\overline{p}-\overline{p}$	1.9e+05	5.0e+05	7.6e+05	1.1e+06	1.4e+06	2.3e+06
[1.2, 1.26)	p − p	1.5e+05	4.3e+05	6.9e+05	1.0e+06	1.4e+06	2.5e+06
	$\overline{p}-\overline{p}$	1.0e+05	2.9e+05	4.8e+05	7.3e+05	1.0e+06	1.7e+06
[1.26, 1.38)	p − p	1.0e+05	3.3e+05	5.7e+05	9.1e+05	1.3e+06	2.5e+06
	$\overline{p}-\overline{p}$	7.6e+04	2.4e+05	4.2e+05	6.7e+05	9.8e+05	1.9e+06
[1.38, 1.56)	p − p	5.5e+04	1.9e+05	3.7e+05	6.3e+05	9.7e+05	2.0e+06
	$\overline{p}-\overline{p}$	4.5e+04	1.6e+05	3.0e+05	5.3e+05	8.2e+05	1.7e+06
[1.56, 1.86)	p − p	2.1e+04	8.0e+04	1.6e+05	3.0e+05	4.9e+05	1.1e+06
	$\overline{p}-\overline{p}$	1.9e+04	7.4e+04	1.5e+05	2.8e+05	4.6e+05	1.1e+06
[1.86, 2.21)	p − p	3.3e+03	1.4e+04	3.0e+04	5.7e+04	9.9e+04	2.5e+05
	$\overline{p}-\overline{p}$	3.2e+03	1.4e+04	3.0e+04	5.8e+04	9.9e+04	2.5e+05

Table 5.1: Number of p–p and $\overline{p}$–$\overline{p}$ pairs with $k^* < 200$ MeVc for all 42 m_T and multiplicity intervals.

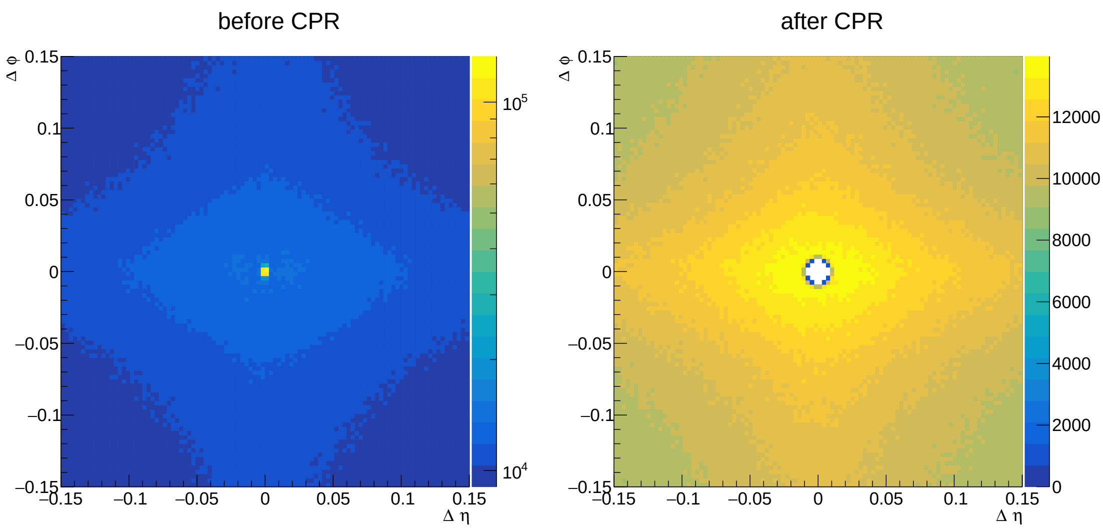


Figure 5.1: $\Delta\eta$ vs. $\Delta\varphi$ distributions before (left panel) and after (right panel) the close pair rejection.

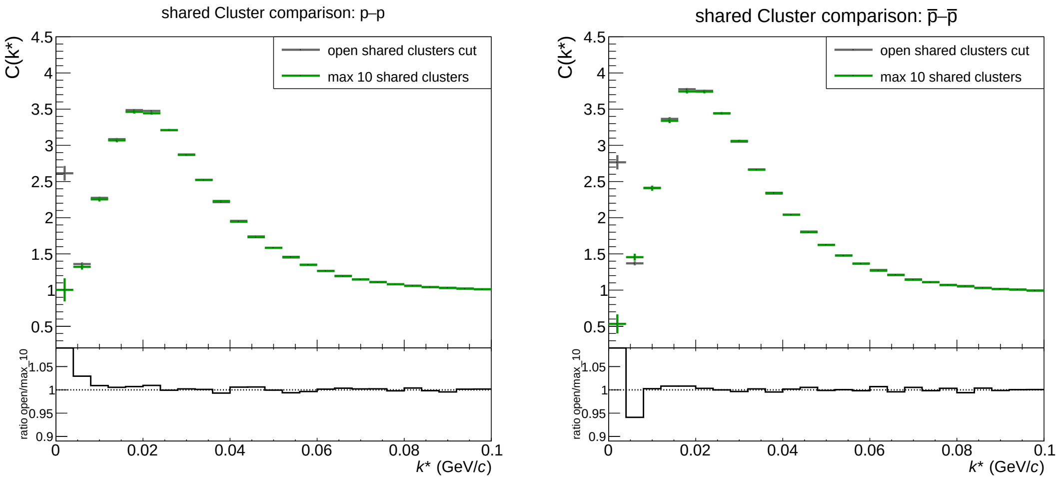


Figure 5.2: Comparison of the integrated correlation function in terms of the cut on the number of shared clusters in the TPC. The left panel shows the comparison for the p–p and the right panel for the $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ correlation function. The gray and green lines correspond to the open and closed cut, respectively.

event lie within the same detector acceptance. A similar event multiplicity is required because the shape of the p_{T} spectra, and therefore also the available phase space for k^*, depends on the multiplicity of the event. In order to mix only similar events with each other, the events are grouped according to their multiplicity and the z-position of the vertex in so called mixing boxes. They have a width of 2 cm (starting from -10 cm up to 10 cm) for the vertex position. For the multiplicity the boxes group events with a similar number of charged particles, in groups of four starting at 0 up to 100 and the last group from 100 to 200 charged particles.

In general, the obtained mixed event distribution has an unrealistic underlying multiplicity distribution because the mixing procedure does not preserve the statistical weights of the multiplicity, i.e., the number of pairs expected in each multiplicity interval. Due to the difference in phase-space depending on the event multiplicity, this can induce unphysical correlation signals in the correlation function. This can be corrected by re-weighting the mixed event distribution according to the yield of the same event distribution, interval by interval in the multiplicity, as described in [41]. The re-weight factor in each multiplicity interval is given by

$$\int_0^\infty N_{\mathrm{same},m}\left(k^*\right) dk^* = \omega_m \int_0^\infty N_{\mathrm{mixed},m}\left(k^*\right) dk^*. \tag{5.1}$$

With these weights, the reweighted mixed event distribution is given by

$$N'_{\mathrm{mixed}}\left(k^*\right) = \sum_m \omega_m N_{\mathrm{mixed},m}\left(k^*\right). \tag{5.2}$$

5.1.2 p–p ($\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$) correlation functions

The correlation functions from all five data-taking periods are combined by directly merging the same event and mixed event distributions. Concerning the influence of the cut on the number of shared clusters in the TPC 5.2, shows that the opening of this cut does not change the correlation function. The left and right panels shown the m_{T} and multiplicity integrated p–p and $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ correlation functions, respectively. The gray line corresponds to the correlation function obtained with an open cut on the shared clusters while the green corresponds to the cut at 10 maximally allowed shared clusters.[1]. Both of them are compatible with each other with the only significant difference in the first k^* bin. This deviation is due to the LHC22o period, which has a negligible role as long as the cut on the number of shared clusters is not removed due to the bug in the reconstruction mentioned earlier in Chapter 4. It was reported by the experts, that the LHC22o period suffers from resolution effects at low p_{T} which could explain the rise observed in Fig. 5.2. However, as will be shown in the following, there is no such effect significantly in the m_{T} and multiplicity differential correlation functions shown in Figures 5.3 and 9.5 to 9.20 (see next paragraph). The exact origin of this rise is under investigation. The compatibility of all periods in each m_{T} and multiplicity interval individually was confirmed before the merging.

The upper panels of Fig. 5.3 show the correlation functions for p–p and $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ pairs in blue and red, respectively, for the second largest m_{T} interval ($m_{\mathrm{T}} \in [1.56, 1.86)$) and all multiplicity intervals. The lighter and darker colors correspond to the multiplicity distribution before and after the reweighting. The lower panels show the ratio between these two, which is almost everywhere compatible with one. Thus, the reweighting does not seem to have an impact on the shape of the correlation function. This is expected since the multiplicity differential analysis already subdivides the multiplicity distribution into smaller intervals where the differences in the statistical weights are not too different. The effect of the reweighting on the mixed event and multiplicity distributions in this m_{T} interval is shown in Fig. 5.4 and Fig. 5.4, respectively. The lines follow the same color coding as the correlation functions in Fig. 5.3. The bottom three panels of Figure 5.5 nicely demonstrate how the fake multiplicity distribution due to the mixing process looks like and how it can be restored by the reweighting. Analogous figures for the other m_{T} intervals can be found in Figures 9.5 to 9.22 in the appendix.

Overall, the correlation functions of the p–p and $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ pairs follow the expected shape that has been observed in the previous analysis as well: The flat behavior at $k^* > 100\,\mathrm{MeV}/c$ is preceded by a rise at lower k^* due to the attractive strong interaction. Towards $k^* = 0$, the correlation function starts dropping due to the Pauli exclusion of the two protons. However, in the signal region, there is a discrepancy between the p–p and $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ correlation functions, which are expected to have identical behavior. The discrepancy is largest for the smallest m_{T} interval (see 9.5 in the appendix). The discrepancy is likely due to a difference in the material-induced protons contamination between the p–p and the $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ pairs. Material-induced protons are knock-out protons

[1] A previous comparison confirmed that the strictly closed cut and the cut at 10 shared clusters result both in essentially the same correlation function. Due to a new creation of a derived dataset in the meantime, which does not include the strictly closed cut, this comparison cannot be repeated. For the purposes of this comparisons the strictly closed and the cut at 10 clusters can be treated equivalently

from the interaction of highly energetic primary particles with the detector and beam pipe material. Antiprotons do not suffer from this contribution because they need to be produced in an inelastic scattering of primary particles with the detector material, which has a significantly lower cross-section. Material candidates do not introduce any correlation signal and contribute to a flat correlation function. The stronger the relative contribution of a flat correlation function, the more the strength of the signal is reduced. It is known that the material particles have a strong p_{T} dependence, with the largest contribution at low p_{T}. A close look at the p_{T} spectra in Fig. 4.2 shows that, indeed, the relative height of the first bin in the proton sample with respect to the peak height is larger than for the antiproton candidates. In order to confirm this assumption, the m_{T} and multiplicity dependent p–p and $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ correlation functions are obtained with a minimum p_{T} cut at $0.5\,\mathrm{GeV}/c$ and $0.6\,\mathrm{GeV}/c$ and shown in the left and right panel in Fig. 5.6, respectively. The lower panels show the ration between the $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ and p–p correlation functions with the dotted line indicating a ratio of one. The discrepancy between both correlation functions decreases in the case of the increased minimum p_{T} cut as one would expect if the protons have a considerably larger knock-out contribution at low p_{T} then the antiprotons.

5.2 Modeling the correlation function

The measured correlation function cannot be compared or fitted directly to theoretical predictions obtained from Eq. 2.2. Several experimental effects are present in addition to the genuine correlation and they need to be adequately modeled. First, the finite detector resolution has to be taken into account, as will be discussed in Section 5.2.1. Second, additionally to the genuine correlation of interest, other correlations to the genuine can be present. They can be classified in non-femtoscopic correlations (discussed in Section 5.2.2) and femtoscopic correlations (discussed in Section 5.2.3). Lastly, the modeling of the source will be discussed in Section 5.2.4.

5.2.1 Detector effects

The shape of the experimentally determined correlation function is affected by the finite momentum resolution of the detector. Instead of unfolding the data, one can smear the theoretical prediction with the detector resolution by transforming from the true k^* into the reconstructed momentum basis. This requires a momentum resolution matrix, which relates the true k^* on the abscissa with the reconstructed k^* on the ordinate. It can be obtained from the analysis of MC events, where for each reconstructed k^*, the true k^* on the level of the generator is known. This way, the smearing can be included in the fitting procedure. However, as discussed in Section 4.3, the available MC dataset does not implement all the reconstruction features of the detector and is limited in size. The obtained resolution matrix is shown in the left panel of Fig. 5.7. The amount of pairs in low k^* is insufficient to ensure a stable fit. To address both issues, the $k^*_{\mathrm{recon.}}$ distribution is fitted for each slice in k^*_{truth} with a Poisson distribution. From this distribution, one can sample more entries, thus artificially enhancing the number of pairs according to the underlying distribution. With this sampling method, it is possible

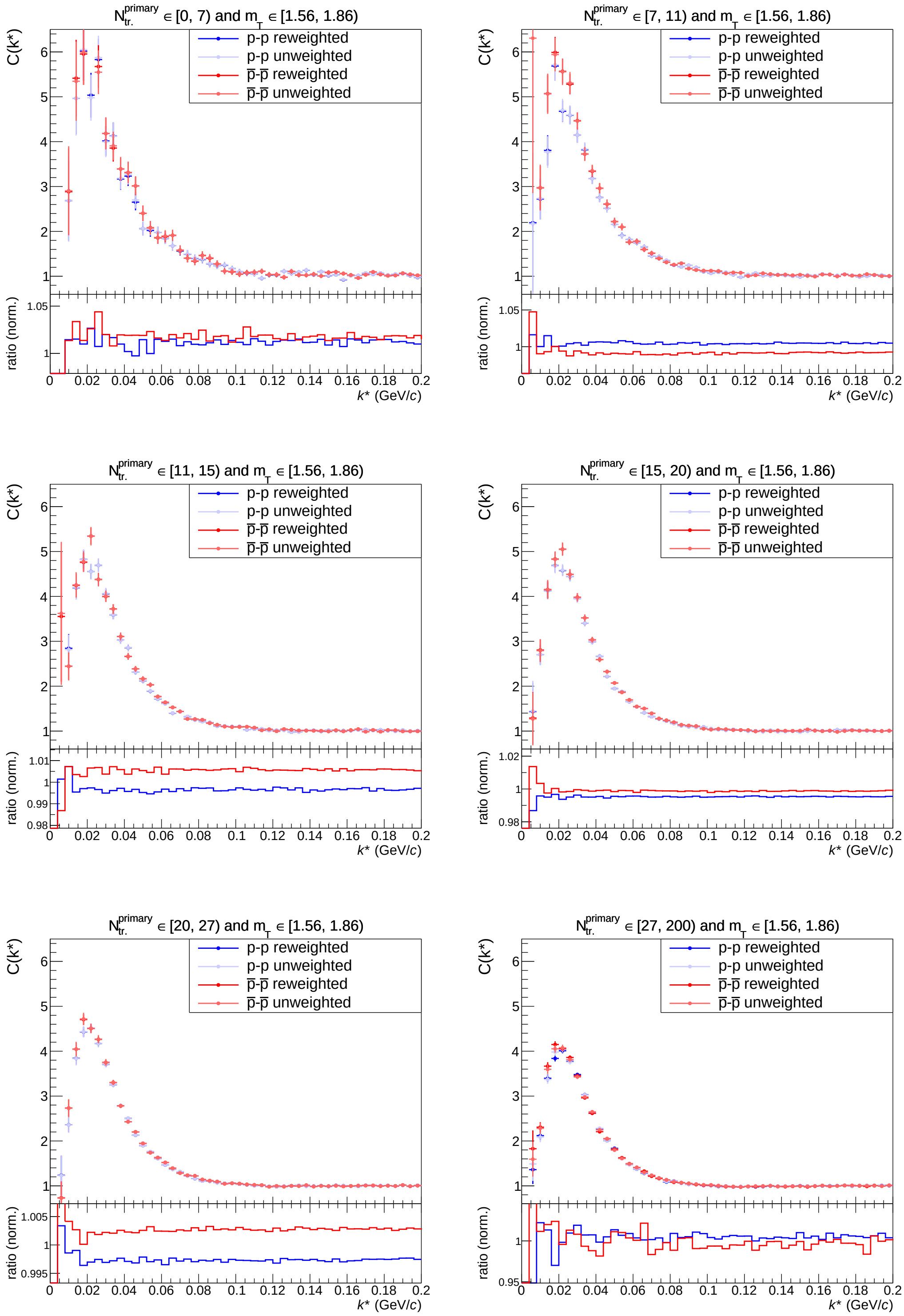


Figure 5.3: Upper panel: Correlation function for the second largest m_T interval ($m_T \in [1.56, 1.86)$) and all multiplicity intervals. The blue and red lines correspond to the p–p and $\overline{\text{p}}$–$\overline{\text{p}}$ correlation function, respectively, the lighter and darker colors to unweighted and reweighted correlation functions, respectively. Lower panel: Ratio between re-weighted and unweighted correlation functions.

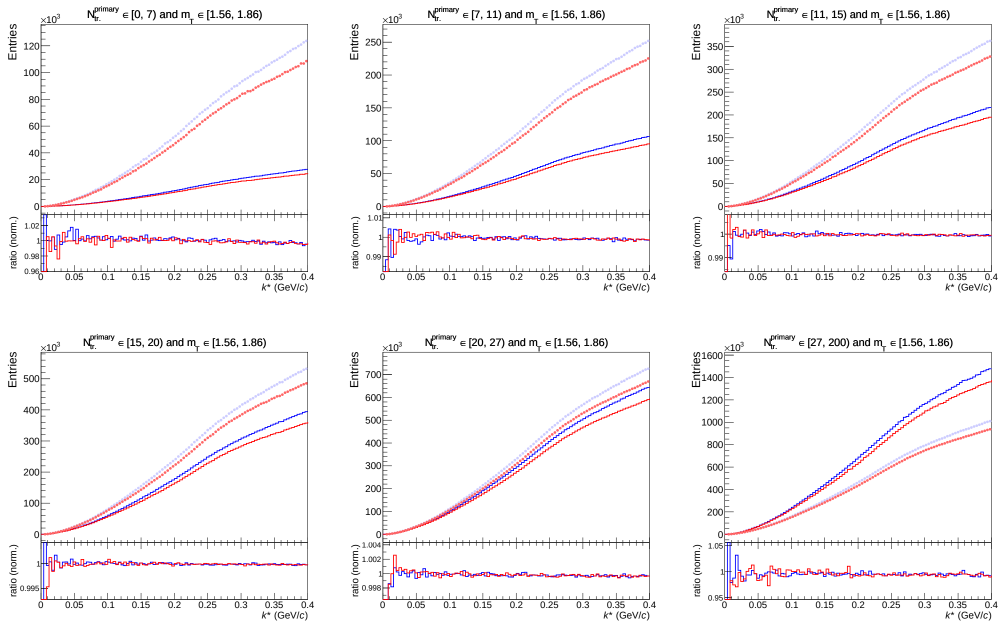


Figure 5.4: Reweighted and unweighted mixed event distributions (upper panel) and their ratio (lower panel) for p–p and $\overline{\text{p}}$–$\overline{\text{p}}$ in blue and red, respectively, for $m_{\text{T}} \in [1.02, 1.14)$ and all multiplicity intervals.

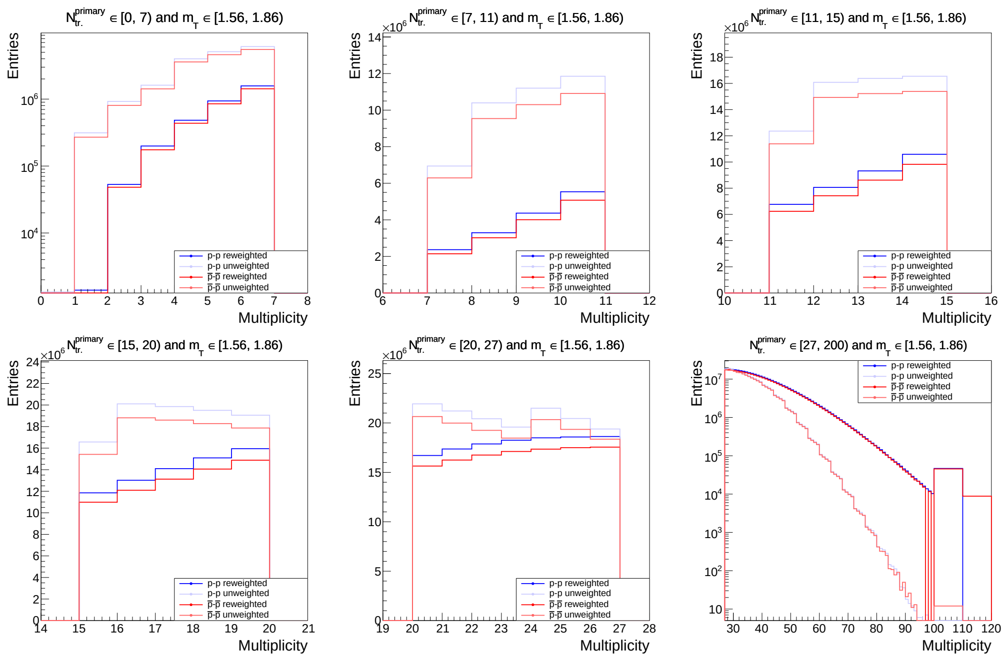


Figure 5.5: Reweighted and unweighted multiplicity distributions for p–p and $\overline{\text{p}}$–$\overline{\text{p}}$ in blue and red, respectively, for $m_{\text{T}} \in [1.02, 1.14)$ and all multiplicity intervals.

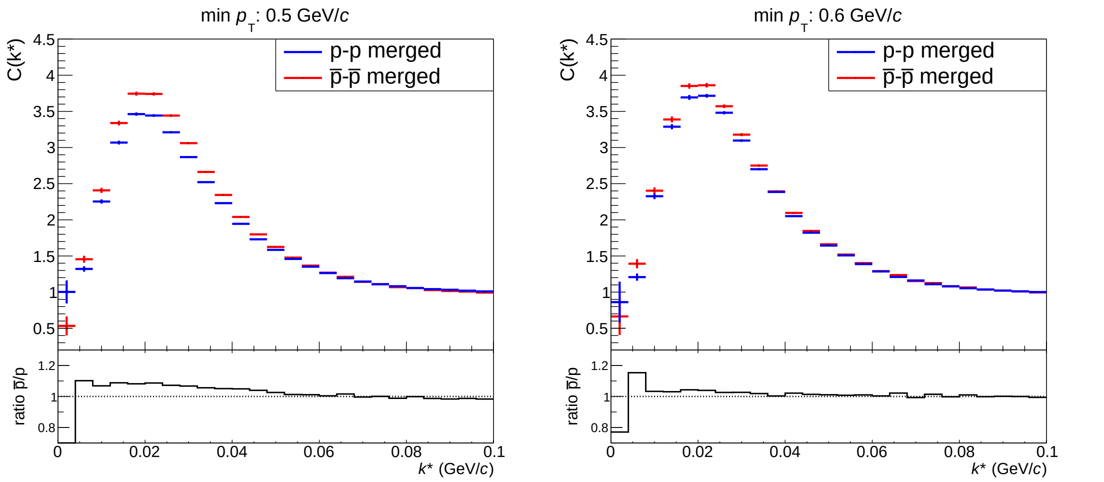


Figure 5.6: Comparison of the m_T and multiplicity integrated correlation functions using proton candidate selection with $p_T > 0.5\,\mathrm{GeV}/c$ (left panel) and $p_T > 0.6\,\mathrm{GeV}/c$ (right panel). The discrepancy between p–p and $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ becomes smaller with the increased cut, which is in line with the suspicion that the origin of this discrepancy lies in the presence of knock-out protons in the datasample.

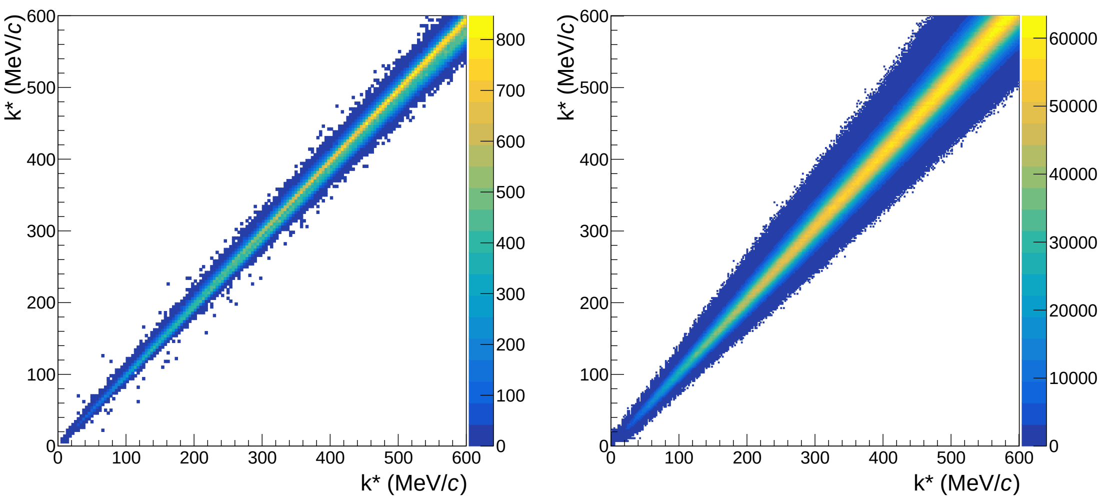


Figure 5.7: Left panel: Smearing matrix for p–p pairs obtained from unanchored, general purpose MC simulations for Run 3. Right panel: Smearing matrix obtained by sampling from the fitted slices obtained from the left panel. More details in the text.

to test a worse detector resolution by widening the distribution by any desirable factor while leaving the mean of the distribution unchanged. The right panel of Fig. 5.7 shows the smearing matrix obtained by enhancing the number of pairs from the sampling as well as increasing the resolution by a factor of two. This approach is ignorant of the effects of the momentum resolution introduced by the data reconstruction and seemingly reproduces the observed behavior of the detector. In order to account for the uncertainties introduced by this smearing procedure, different versions of the smearing matrix will be tested in the fitting procedure as systematic variations. This will be discussed in detail in Section 6.2.

5.2.2 Non-femtoscopic correlations

For sufficiently large relative momenta (k^*>200 MeV/c), the FSI among the particles fades out, and hence the correlation function should approach unity. However, the measured correlation functions often exhibit an enhancement for large k^* values above 500 MeV/c. Such non-femtoscopic effects are suspected to come from energy-momentum conservation and are more pronounced in small colliding systems, where the average particle multiplicity is low [4]. They are described via multiplication of a baseline to the modeled correlation function, which is fitted at large values of k^* to the data. The functional form of the baseline is given by

$$C(k^*)_{\mathrm{non-femto}} = 1 + b \cdot k^{*2} + c \cdot k^{*3} \tag{5.3}$$

which is a third-degree polynomial function with a missing linear term. The missing linear term ensures that the correlation function is flat at $k^* = 0\,\mathrm{MeV}/c$.

5.2.3 Modeling of the femtoscopic p–p correlation

Genuine p–p correlation function

The genuine p–p correlation function is modeled considering the Coulomb and the strong interaction and the antisymmetrization of the wave functions. For the strong interaction the state-of-the-art Argonne v_{18} [42] potential is used. It implements s and p waves and has been tested with high precision in scattering experiments before.

Feed down to p–p

Feed down contributions in the p–p system come from weak decays into protons. The sample of all measured p–p pairs is decomposed with the following contributions

$$\begin{aligned}\{\mathrm{p–p}\} = &\ \mathrm{p–p} + \mathrm{p_\Lambda–p} + \mathrm{p_\Lambda–p_\Lambda} + \mathrm{p_{\Sigma^+}–p} + \mathrm{p_{\Sigma^+}–p_{\Sigma^+}} \\ &+ \mathrm{p_\Lambda–p_{\Sigma^+}} + \tilde{\mathrm{p}}\mathrm{–p} + \tilde{\mathrm{p}}\mathrm{–p_\Lambda} + \tilde{\mathrm{p}}\mathrm{–p_{\Sigma^+}} + \tilde{\mathrm{p}}–\tilde{\mathrm{p}},\end{aligned} \tag{5.4}$$

where $\tilde{X}$ refers to misidentified particles of species X. The particle in the subscript denotes the mother particle, from which the proton originates, e.g., p_Λ and p_{Σ^+} are protons coming from the Λ and Σ^+, respectively. The relative weight of each is evaluated using the λ parameter formalism, as discussed in Section 2.3. The purities

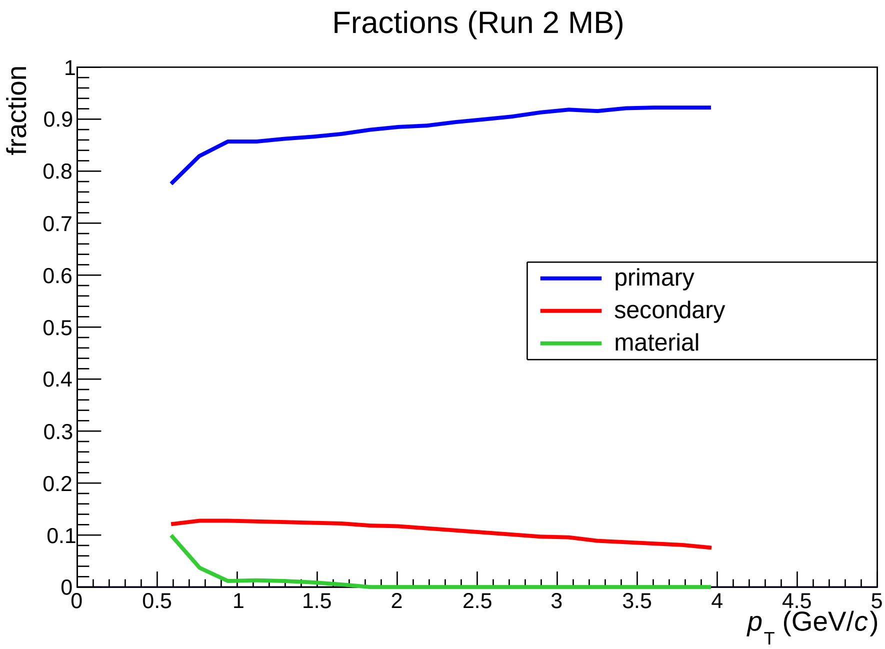


Figure 5.8: Fractions of primary, secondary and material contributions. The values are from the minimum bias analysis in Run 2 and were reported in [43]

have already been calculated in Section 4.5. The calculation of the fractions relies on template fits, for which the DCA_{xy} distribution of each contribution is obtained separately from MC simulations. They are used as templates to fit the experimentally measured DCA_{xy} distribution and the fractions given by the relative weight of each contribution in the fit. The DCA_{xy} selection has to be widened for the template fits, since the differences are mostly affecting the tails of the distributions and most of the sensitivity lies in that region. All other selections remain at the analysis cut.

Since there are no anchored MC productions available, this procedure can not be applied. Instead, the fractions obtained in the minimum bias analysis in Run 2 [43] are used as an estimation for the fractions in Run 3. The fractions for the feed-down contributions are not expected to change significantly between Run 2 and Run 3 since they are related to branching ratios and production yields, which should not be affected significantly by the change of collision energies from Run 2 to Run 3. However, the contribution from material protons is expected to be lower in Run 3 than Run 2 because of the reduced material budget in the detector, particularly in the ITS. In this work, it is assumed that the material contribution for antiprotons is vanishing and for the protons, the values of the Run 2 analysis will be used. This represents the two extreme cases of the expected material contribution that can be considered as a systematic variation to cover the uncertainty concerning the material contribution. The fractions from Run 2 are provided as a function of p_T and are shown in Fig. 5.8. They are used together with the p_T dependent purities to calculate the λ parameters following Eq. 2.6 and weighted with the p_T distributions for protons and antiprotons. The resulting λ parameters are reported in Table 5.2. The leading non-genuine contribution comes from the feed-down from Λ hyperons. The other contributions are summed together and assumed to be

	λ [%]	
Pair	p–p	$\overline{\text{p}}$–$\overline{\text{p}}$
genuine	66.27	72.70
p_{Λ}–p	13.90	15.21
p_{Σ^+}–p	5.96	6.52
p–$\text{p}_{\text{flat}_{\text{f}}\text{eed-down}}$	9.91	1.62
p–p_{fake}	3.96	3.96

Table 5.2: Weight parameters of the individual components of the p–p and $\overline{\text{p}}$–$\overline{\text{p}}$ correlation functions.

flat. The same is assumed for the contribution of miss-identified protons, which are summarized by the λ parameter for the p–p_{fake} contribution.

The p–Λ correlation function is modeled using the potential from the revised chiral effective field theory calculation to next-to-leading order (NLO) [44]. It has been used in the previous source size measurement in high multiplicity Run 2 data [9] and resulted in the best compatibility with the data in a dedicated study using femtoscopy [15]. The p–Λ and correlation functions has to be transformed into the p–p system and smeared for their decay kinematics. The matrices for that are obtained and have been used previously in [41]. Note that the momentum resolution of the p–p pairs is applied to the feed-down contributions as well because, after all, the protons are measured in the detector, independent of their primary or secondary origin.

5.2.4 Modeling of the Source

The source is modeled assuming a Gaussian profile, given by

$$S(r^*) = \frac{1}{(4\pi r_0^2)^{3/2}} \exp\left(-\frac{r^{*2}}{4r_0^2}\right), \tag{5.5}$$

which is isotropic in space. This reduces the parameters of the source to a single parameter, the source size r_0, which is related to the width of the distribution. Since the source is assumed to be spherically symmetric, it is often more convenient to integrate out the angular dependence, which leads to a multiplication of $S(r^*)$ by $4\pi r^{*2}$

$$S_{4\pi}(r^*) = \frac{2\sqrt{\pi} r^{*2}}{r_0^3} \exp\left(-\frac{r^{*2}}{4r_0^2}\right). \tag{5.6}$$

This essentially is the probability to emit two particles at a certain relative distance r with respect to each other.

5.3 Femtoscopic fit

The modeling of the femtoscopic correlations following the description above are implemented in the CATS framework. The only free parameter is the source size of the Gaussian source. All other parameters, in particular the λ parameters and any parameters related to the interaction, are fixed. With CATS the femtoscopic correlation function can be evaluated for an arbitrary k^* and be used as a function in the fitter. It is multiplied with a function of the form Eq. 5.3 representing the baseline, where all the parameters are left free. Thus, the fitting determines, along with the source size, the two baseline parameters and the normalization parameter. In order to capture the behavior of the non-femtoscopic correlations, the fitting range is extended up to a k^* of $400\,\mathrm{MeV}/c$, while the strong interaction is computed only up to $300\,\mathrm{MeV}/c$. All correlation functions are normalized to unity at $k^* \in [240, 340\,\mathrm{MeV}/c]$, because the correlation function exhibits a flat behavior for these values of k^*. However, the normalization is arbitrary and is selected at these values only for visual purposes. Any physical bias is reabsorbed by the free fit parameter N. The p–p and $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ correlation functions are fitted separately with their proper λ parameters.

6 Systematic uncertainties

6.1 Systematic uncertainties of the data

The systematic uncertainties of the candidate selection were evaluated by varying the cuts on the proton candidates according to the values reported in Table 6.1. These variations were combined randomly until 15 sets were found, which changed the integrated yield of pairs with $k^* < 300\,\mathrm{MeV}/c$ by less than 10%. The correlation function is obtained for each of these variations, resulting in a spread of values in each k^* bin. Since the systematic variations are evenly spread, one can assume a flat distribution and thus calculate the root-mean-square (RMS) in each bin by dividing the difference between the largest and the lowest deviation by the square root of twelve. This results in a bin-wise systematic uncertainty, which is shown in Figures 6.1 and 6.2 for the p–p and $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ correlation functions, respectively, for all multiplicity intervals in the second largest m_T interval (i.e. for $m_\mathrm{T} \in [1.56\,\mathrm{GeV}/c,\ 1.86\,\mathrm{GeV}/c)$) in the second highest multiplicity interval. The analysis of the whole dataset with all systematic variations is computationally expensive and exceeds the assigned computing budget. Therefore, the correlation functions were obtained from a sampled dataset, which provides about 5 % of all the data for analysis. With such a small dataset there is the danger of overestimating the systematic uncertainties if the variations are dominated by statistical fluctuations. Indeed, as will be shown in the next chapter, the systematic uncertainties of the data are larger than the statistical uncertainties which is contrary to the expectation and the experience with previous analysis. In the future, the limitations of the framework can be overcome by a more efficient handling of the derived dataset, which will allow to perform the systematic variations on a larger dataset and minimize the contribution from statistical fluctuations. This will be implemented in future improvements of this analysis. The variations presented in Figures 6.1 and 6.2 are computed in a broader binning of $16\,\mathrm{MeV}/c$ to enhance the statistical significance of each bin. The systematic uncertainties are fitted with an exponential function plus a constant term, which is shown as a red line in Fig. 6.1 and Fig. 6.2. This way, the systematic uncertainty is interpolated between the bins and can be evaluated at an arbitrary value of k^*. Figures 6.1 and 6.2 show that the systematic uncertainty decreases with larger k^*, consistently with an exponential decay. In some m_T and multiplicity intervals, around the position of the peak in the correlation function, the systematic uncertainty overshoots the exponential decay. This is likely due to the variation of the low p_T cut, which changes the contribution of material protons, as discussed in Section 5.1.2. Consistent with this interpretation is that this effect is generally more pronounced in the p–p than in the $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ correlation functions. The figures for the systematic uncertainties in the other m_T and multiplicity bins can be found in Figures 9.23 to 9.34 in the appendix.

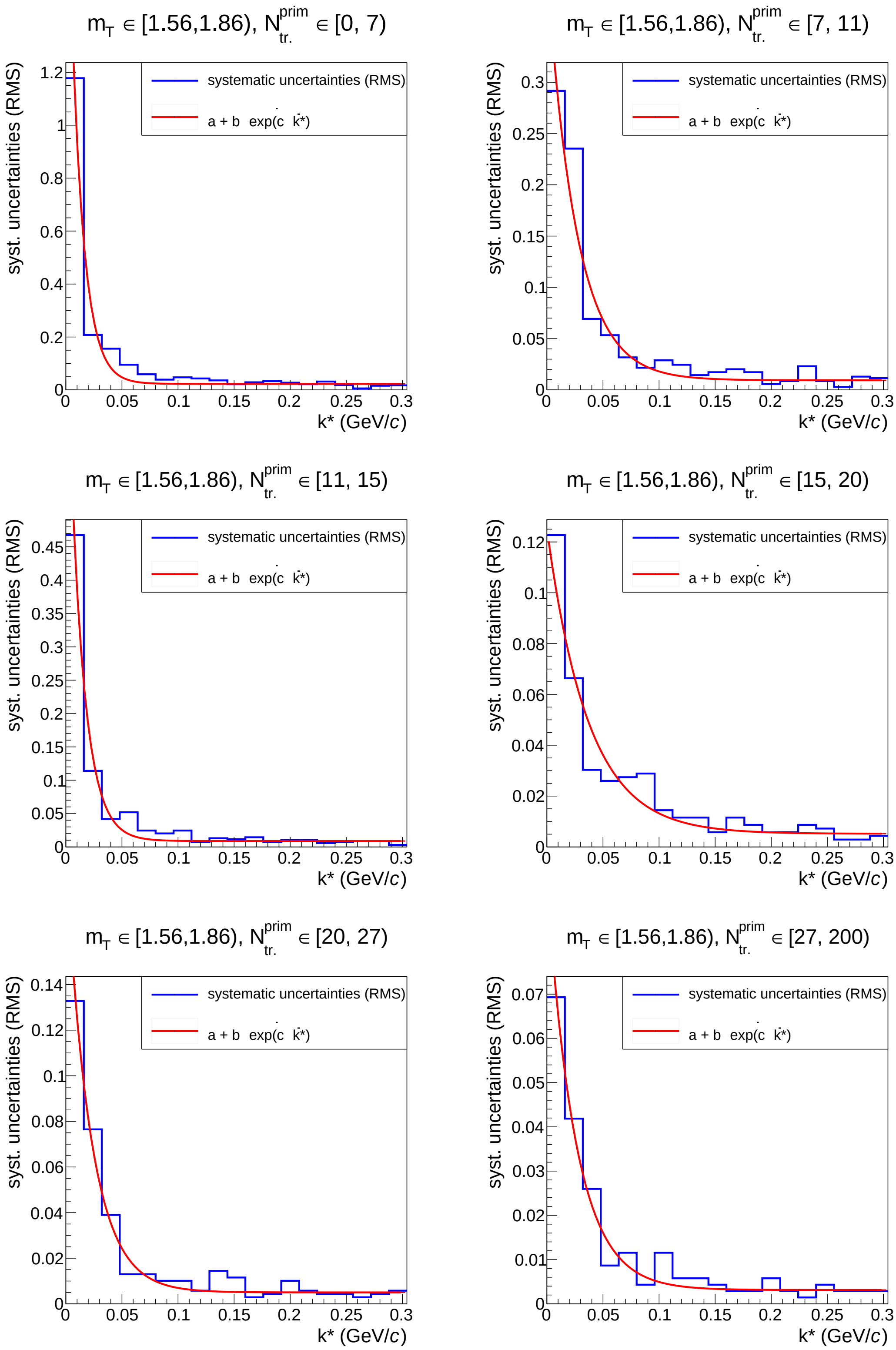


Figure 6.1: Systematic uncertainties of the p–p correlation functions for all multiplicity intervals in the second largest m_T interval ($m_T \in [1.56\,\mathrm{GeV}/c,\ 1.86\,\mathrm{GeV}/c)$)

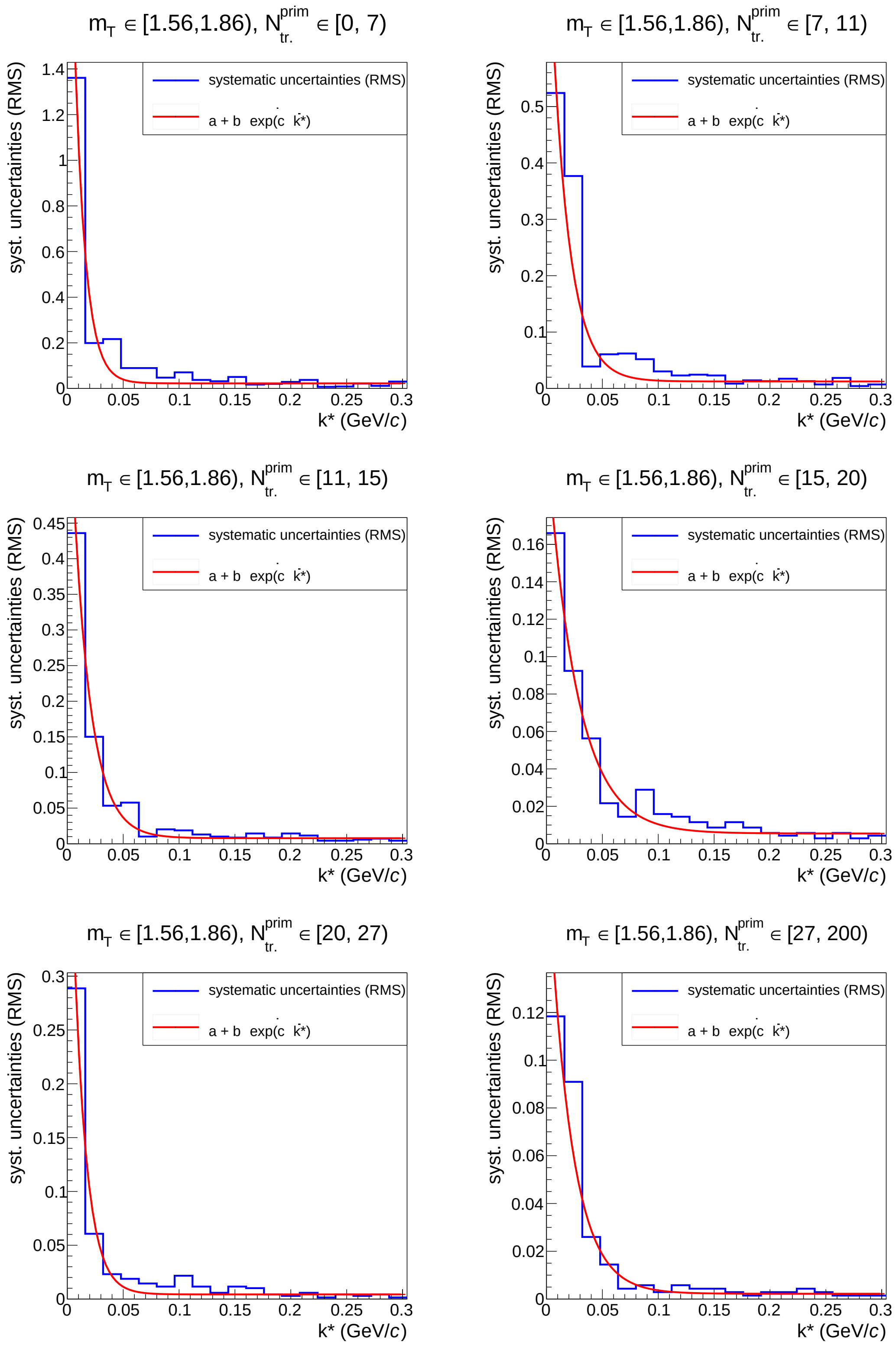


Figure 6.2: Systematic uncertainties of the $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ correlation functions for all multiplicity intervals in the second largest m_T interval ($m_\mathrm{T} \in [1.56\,\mathrm{GeV}/c,\ 1.86\,\mathrm{GeV}/c)$)

Variable	Default	Variation
Proton Cuts		
min p_{T} (GeV/c)	0.5	0.4, 0.6
max $\lvert\eta\rvert$	0.8	0.77, 0.83
max n_{σ}	3	2.5, 3.0
max Number of Clusters TPC	80	70, 90

Table 6.1: Variations of different selection criteria on the proton candidates

6.2 Systematic of the p–p fits

The uncertainties of the fitting are evaluated by performing the fit with different configurations. The variations are summarized along with the standard configuration in Table 6.2. The range in k^*, in which the strong interaction is considered for the computation of the correlation function is varied to 240 MeV/c and 360 MeV/c. The global range of the fit (i.e. the non-femtoscopic baseline) is increased from 400 MeV/c to 500 MeV/c and the baseline itself is also varied by fixing the third-degree term of the polynomial to zero. The λ parameters are varied by decreasing and increasing the primary contribution by 20 % and re-scaling the remaining contributions equally for compensation. Finally, in order to capture the for the systematic uncertainties the fit related to the momentum resolution, the fitting is done without a smearing matrix and using the smearing matrix in the right panel of Fig. 5.7 as a variation. No momentum smearing most probably underestimates the effect of the momentum resolution and the right smearing matrix of Fig. 5.7 is probably overestimating it, while the true effect should lie between these two cases. The last two variations, i.e., the variation of the λ parameters and the momentum resolution, represent the largest uncertainties due to the lack of MC simulations. They have been generously considered and are the main contributors to the systematic uncertainties of the fit. All these variations are combined in all possible permutations, resulting in a total of 108 variations of the fit for each m_{T} and multiplicity interval. All fit variations are superimposed and the values of the correlation function and the baseline are evaluated in the bin center of each k^* bin of the measured correlation function. Finally, the fit value and the systematic uncertainty are evaluated by placing the bin center at the mean of the distribution and setting the uncertainties to the difference between the maximum and minimum value, divided by $\sqrt{12}$.

Parameter	Default	Variation 1	Variation 2
Range of the strong interaction	300 MeV/c	240 MeV/c	360 MeV/c
Range of the fit	400 MeV/c	500 MeV/c	—
Region of Normalization	240-340	220-320	260-360
Baseline function	$1 + b \cdot k^{*2} + c \cdot k^{*3}$	$1 + b \cdot k^{*2}$	—
$\lambda_{\mathrm{p–p}}$	66.27 %	53.02 %	79.52 %
$\lambda_{\overline{\mathrm{p}}–\overline{\mathrm{p}}}$	72.70 %	58.15 %	87.23 %
Momentum smearing	no smearing	smearing using unanchored MC with worse resolution	—

Table 6.2: Systematic variations of the fit. Details are explained in the text.

7 Results

7.1 Fit Results and measured correlation functions

The measured correlation functions are shown for the second largest m_{T} interval and all multiplicity intervals in 7.1. The other m_{T} intervals can be found in Figures 9.35 to 9.40 in the appendix. The upper panels show the correlation functions and the fits together with the baselines for p–p and $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ pairs in blue and red, respectively. The inset shows the intermediate k^* range in a different scale to highlight the behavior of the depletion. The lower panels show the agreement of the measurement with the fit expressed in terms of the number of standard deviations ($n\sigma$). It is calculated by taking into account both the systematic and statistical uncertainties of the data via

$$\sigma = \frac{C(k^*)_{\mathrm{fit-edge}} - C(k^*)_{\mathrm{measured}}}{\sqrt{\sigma^2_{\mathrm{stat}} + \sigma^2_{\mathrm{syst}}}}. \tag{7.1}$$

The upper and lower edges of the bands correspond to the agreement of the data with the upper and lower edges of the bands of the fit, respectively. The fit captures the experimental data and reproduces the behavior of the depletion at intermediate k^* values. It shows clearly the attractive strong interaction in the enhancement for $k^* < 100\,\mathrm{MeV}/c$ and the effect of the Pauli Blocking towards $k^* = 0\,\mathrm{MeV}/c$. However, the uncertainties of the fitting procedure are very large, which is represented by the large width of the fit band. In most cases, the data points in the signal region sit on the upper edge of the fit band, as the systematic variations of the fit tend to have a less pronounced peak. The large uncertainties in the fit are related to the lack of MC simulations, which forced a generous assignment of uncertainties for the momentum resolution and the λ parameters, as discussed in the previous chapter. From experiences with previous analyses, it is expected that the uncertainty of 20 % in the primary λ parameter and the variation of the smearing matrix overestimate the true uncertainty. This will have to be improved in the future, as soon as reliable MC simulations are available. Further improvement is expected by extracting the λ parameters more differentially, e.g., as a function of m_{T}, instead of obtaining one integrated (and averaged) value for all multiplicity and m_{T} intervals. Both purity and material contributions are known to have a p_{T} dependence, which would translate into an m_{T} dependence of the λ parameters. Since the largest contribution of proton candidates comes from the low p_{T} region (see upper right panel of Fig. 4.2), it could be that the integrated λ parameter is not correctly representing the sample composition in larger m_{T} intervals.

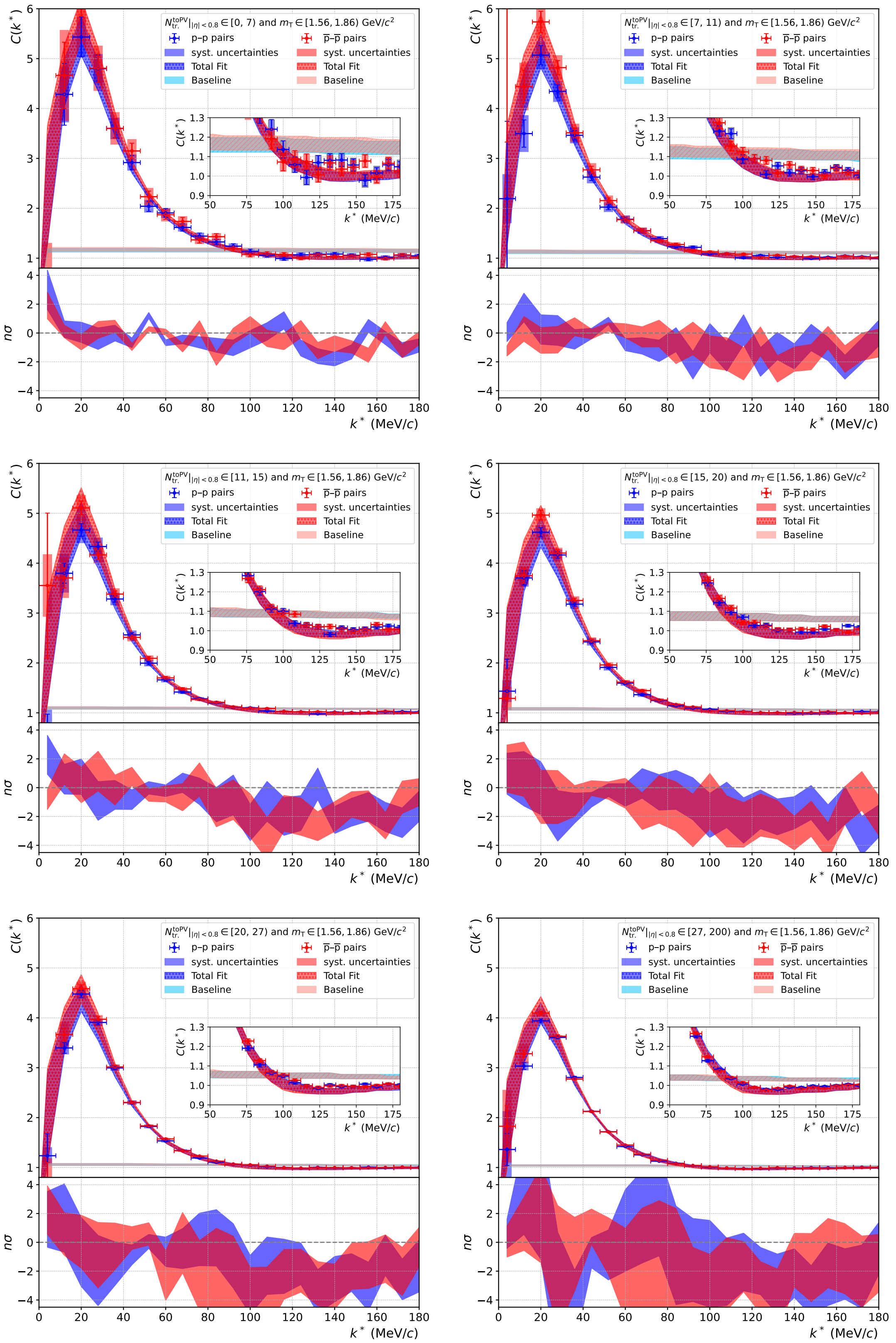


Figure 7.1: Correlation functions for the second largest m_T interval and all multiplicity intervals. The blue plots correspond to the p–p correlation function and the red plots to the $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ correlation function.

7.2 Effective source size pairs as a function of m_T and multiplicity

From the fits, one can extract the source parameters in each multiplicity and m_T interval. Figure 7.2 shows the extracted radii as a function of m_T for each multiplicity interval separately. The radii extracted from the correlation functions of p–p and $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ pairs are shown in blue and red, respectively. The markers are placed on the x-axis according to the mean value of the m_T distribution for pairs in this particular m_T and multiplicity interval. The uncertainties of the markers correspond to the uncertainties of the fit, while the shaded boxes correspond to the systematic uncertainties that are obtained from the fit variations. The value for the source size is extracted for each fit variation and the systematic uncertainties are calculated by the difference between the largest and smallest value divided by $\sqrt{12}$. The systematic uncertainties of the radii are larger than the uncertainties of the fit and due to the reasons discussed previously. The source sizes extracted from p–p and $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ pairs are compatible with each other within the systematic uncertainties, most of the time by less than half a standard deviation. In both cases they follow the expected m_T scaling, i.e. they decrease with increasing m_T and increase with increasing multiplicity. However, in all multiplicity intervals, the source sizes of the $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ pairs are systematically larger than the source sizes of the p–p pairs, except in the first m_T interval, where the situation is reversed. This could be an indication that the m_T dependence of the λ parameters can not be neglected. In general, an overestimation of the material contribution leads to a smaller source size because the fit needs to reproduce the strength of the correlation to compensate for the missing pairs in the sample. The compensation happens by reducing the source size, such as reducing the average distances between proton pairs and thus enhancing the effect of the strong interaction. If the assumption, that the $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ sample does not have a sizable material contribution, is correct, it could mean that at the moment the material contribution for all m_T intervals but the first one is overestimated while it is underestimated in the first m_T interval. It is known that the knock-out protons concentrate mostly around low p_T, which is in line with this interpretation.

Figure 7.3 shows the m_T scaling of the source for all multiplicity intervals superimposed. The radii from the p–p and $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ correlations have been combined and weighted by the number of pairs in each interval. The statistical and systematic uncertainties have been propagated and the difference between the radii from p–p and $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ pairs is considered as an additional uncertainty included in the shaded box. The black markers are the effective source radii extracted from the Run 2 analysis of the high-multiplicity pp dataset at 13 TeV [9]. The radii are larger than the radii for the largest multiplicity sample in the Run 3 analysis, although their uncertainties slightly overlap. This difference is expected because both datasets cover a different multiplicity region. The high-multiplicity Run 2 data corresponds to 0.17 % of all events with the largest multiplicity while the multiplicity class $N_{\mathrm{tr.}}^{\mathrm{primary}} \in [27, 200)$ in Run 2 corresponds roughly to the largest 17 % of all events with at least one proton in terms of multiplicity. A robust crosscheck for the correctness of this analysis would be the measurement of the m_T multiplicity class which is comparable to the high-multiplicity class in Run 2. This, however, requires a multiplicity calibration in terms of multiplicity percentiles, which was not available at the time of writing. Furthermore, the comparison with the Run 2 results shows the reduced sensitivity of the current analysis in large m_{T} due to the p_{T} cut at $2.0\,\mathrm{GeV}/c$ as well as the comparatively large systematic uncertainties in Run 3, which were discussed previously. With available Monte Carlo simulations and further improvements in the estimation of the λ parameters, the uncertainties in the Run 3 measurement will approach the size of the uncertainties of the Run 2 high-multiplicity measurement.

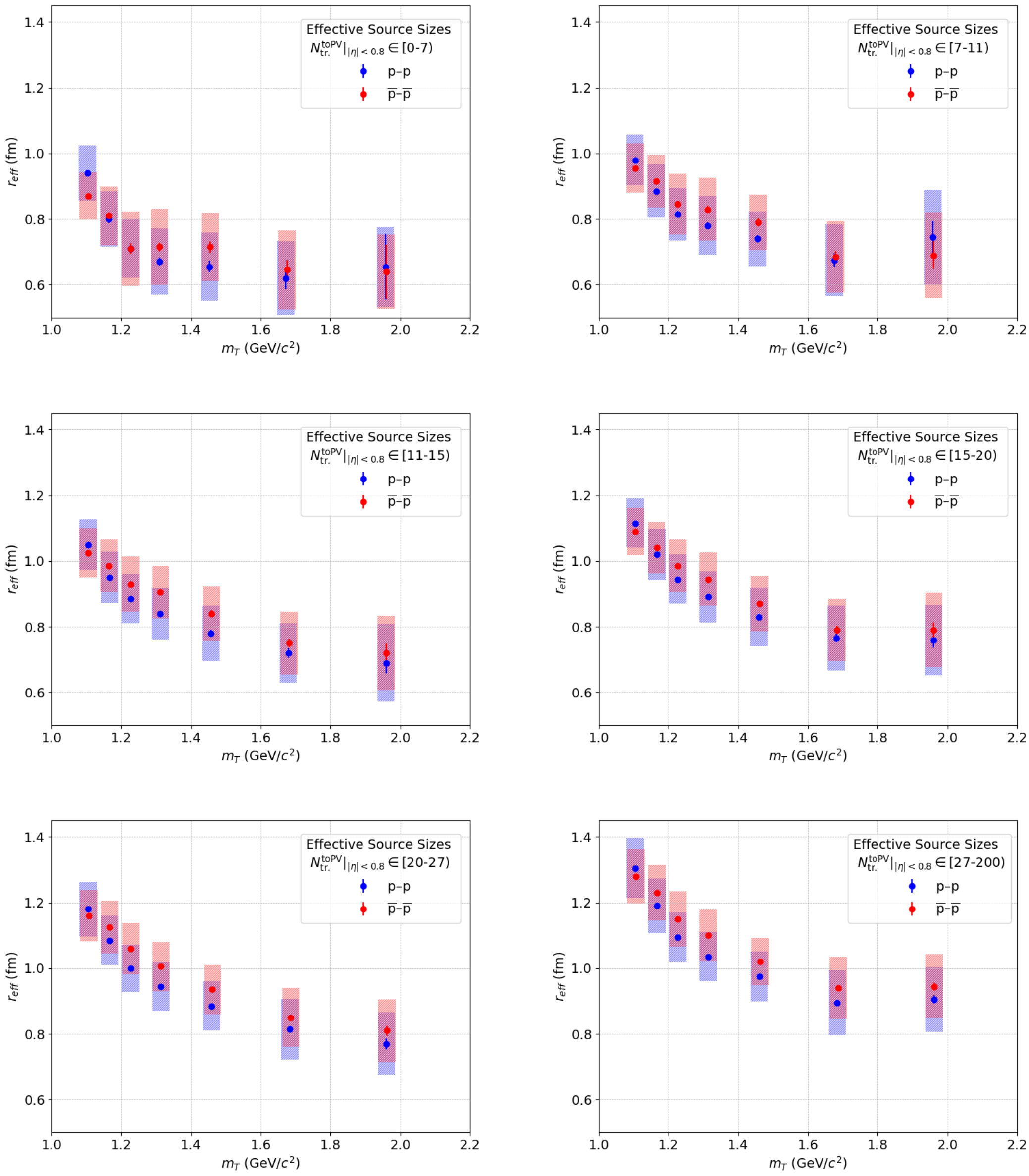


Figure 7.2: The Gaussian source radius as a function of m_{T} for all multiplicity intervals separately. The radii extracted from the p–p and $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ correlations are shown in blue and red, respectively. The markers represent the uncertainties of the fit with the standard configuration. The shaded boxes represent the systematic uncertainties obtained from all the fit variations

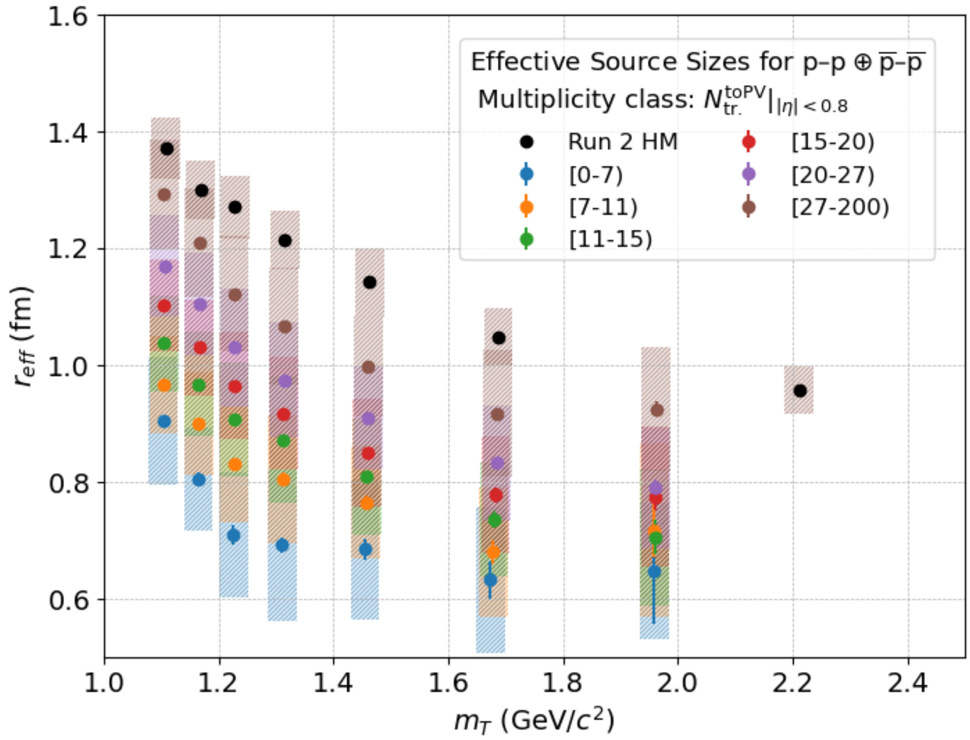


Figure 7.3: $m_{\rm T}$ scaling of the effective source size of p-p pairs in pp collisions at 13.6 TeV in various multiplicity classes.

8 Summary and Outlook

This thesis presents the first multiplicity differential analysis of the m_T scaling of the femtoscopic source in pp collisions at LHC energies. The expected scaling has been observed, namely a decrease of the radii with increasing m_T and an increase of the radii with increasing event multiplicity. For a precise measurement, a multi-differential analysis such as this one requires a lot of data, in the ballpark of a few hundred billion events. This has been possible thanks to the upgraded ALICE detector and the new data-taking paradigm, in which the detector is read out continuously and the reconstruction and triggering are performed offline on all recorded events. The necessary detector upgrades were discussed at the beginning of this work. This was followed by an introduction to the O^2 and `FemtoDream` analysis frameworks, that enable an efficient analysis of these enormous amounts of data.

Future improvements and next steps

The next steps of the analysis will have to be the reduction of the systematic uncertainties, the enhancement of the statistical significance of the largest m_T class, and the definition of the multiplicity classes in terms of calibrated multiplicity percentiles. Currently, systematic uncertainties are dominating the statistical uncertainties of the data due to the uncertainties regarding the composition of the sample with non-genuine contributions and the momentum resolution of the detector. Getting them under control will require Monte Carlo simulations with the correct modeling of the detectors and their conditions during data-taking. The results discussed in the previous sections show signs of a m_T dependence of these corrections. This can be addressed as well with the help of simulations. The statistical significance of the largest m_T interval is decreased compared to the other m_T intervals because of the p_T cut at $2.0\,\mathrm{GeV}/c$, which was introduced because of the observed drop in the purity for protons with larger transverse momenta. Ideally, improved data calibration and reconstruction will improve the separation capabilities of the TOF in this momentum region, which is the main reason for this drop in the purity, and extend the number of pairs, especially in the largest m_T interval. In any case, simulations could help to recover the lost candidates at least partially, by correcting for the contamination with adequately estimated λ parameters. Finally, a calibrated multiplicity percentile will provide a more solid definition of the multiplicity classes, which will also allow for a better comparison with the previous results. As a validation measurement, one could reproduce the multiplicity class of the high-multiplicity data of Run 2 [9] in Run 3 and directly compare them in the same multiplicity. Moreover, the multiplicity classes can be redefined in such a way, as to distribute the available p–p pairs equally among all multiplicity classes. Once done, the core radii can be extracted and the analysis can be repeated with p–Λ pairs to investigate the universality of the m_{T} scaling in more multiplicity intervals and complement the measurement in high multiplicity pp collisions [9].

Beyond constraining the source in pp collisions

Beyond the scope of this work, but certainly within reach given the improvements mentioned above, the precise data of the p–p correlation function can help to test the state-of-the-art interaction models to unprecedented precision. Recently, the evaluation of the Run 2 p–p correlation function using a novel source model, CECA [45], revealed a tension between the experimental data and the modeled correlation function for the largest m_{T} interval. In that study, as in the here presented work, the p–p interaction was modeled using the Argonne v_{18} [42] potential, which is anchored to the vast amount of available scattering data. With femtoscopy, one can very precisely access the interaction at low momenta, a region in which scattering experiments are the weakest due to the kinematic constraints of firing projectiles on targets and recovering the scattered particles. However, the tension could be related to the femtoscopic source as well. The largest pair m_{T} corresponds to the smallest source sizes and it is not known what the lowest value of the femtoscopic source is. Physically, the size of the proton should play a role in this. Precise data of the p–p correlation functions as a function of multiplicity could provide the necessary experimental input to explore these tensions and test both the state-of-the-art potentials and the femtoscopic source in extreme cases.

Finally, and to close the circle with the beginning of the introduction, these data could help approach the question of hadronization in pp collisions and understand the differences to relativistic nucleus-nucleus collisions. It was long believed that the QGP can not be produced in elementary pp collisions. At the very least, it is difficult to apply the traditional experimental methods used to understand the hydro-dynamical evolution of heavy ion collisions to pp collisions because of the lower multiplicities involved in the latter. Recent measurements in pp collisions found signatures that were interpreted as evidence for QGP formation in nucleus-nucleus collisions, like, for example, the enhancement of strangeness production in high-multiplicity pp collisions [46] or long-range azimuthal correlations [47]. Similarly, the m_{T} scaling is a well-known feature in relativistic heavy ion collisions and is understood to arise from collective effects [48]. Therefore, a precise measurement of the femtoscopic source sizes could provide valuable experimental data to understand the origin of m_{T} scaling and the differences with Pb–Pb collisions. Some attempts to interpret these results suspect high-multiplicity pp collisions to partially develop properties of nucleus-nucleus collisions with the possibility of forming small droplets of QGP [5, 49]. In this context, the different collision systems can be interpreted as points on a multiplicity scale, spanning four orders of magnitude, starting at pp collisions with low multiplicities over to high-multiplicity pp collisions, p–Pb, and finally, peripheral, semi-central, and central Pb–Pb collisions. In that sense, the multiplicity-dependent measurement of the m_{T} scaling complements the measurement in high-multiplicity events and helps fill the missing points on the multiplicity scale.

9 Appendix

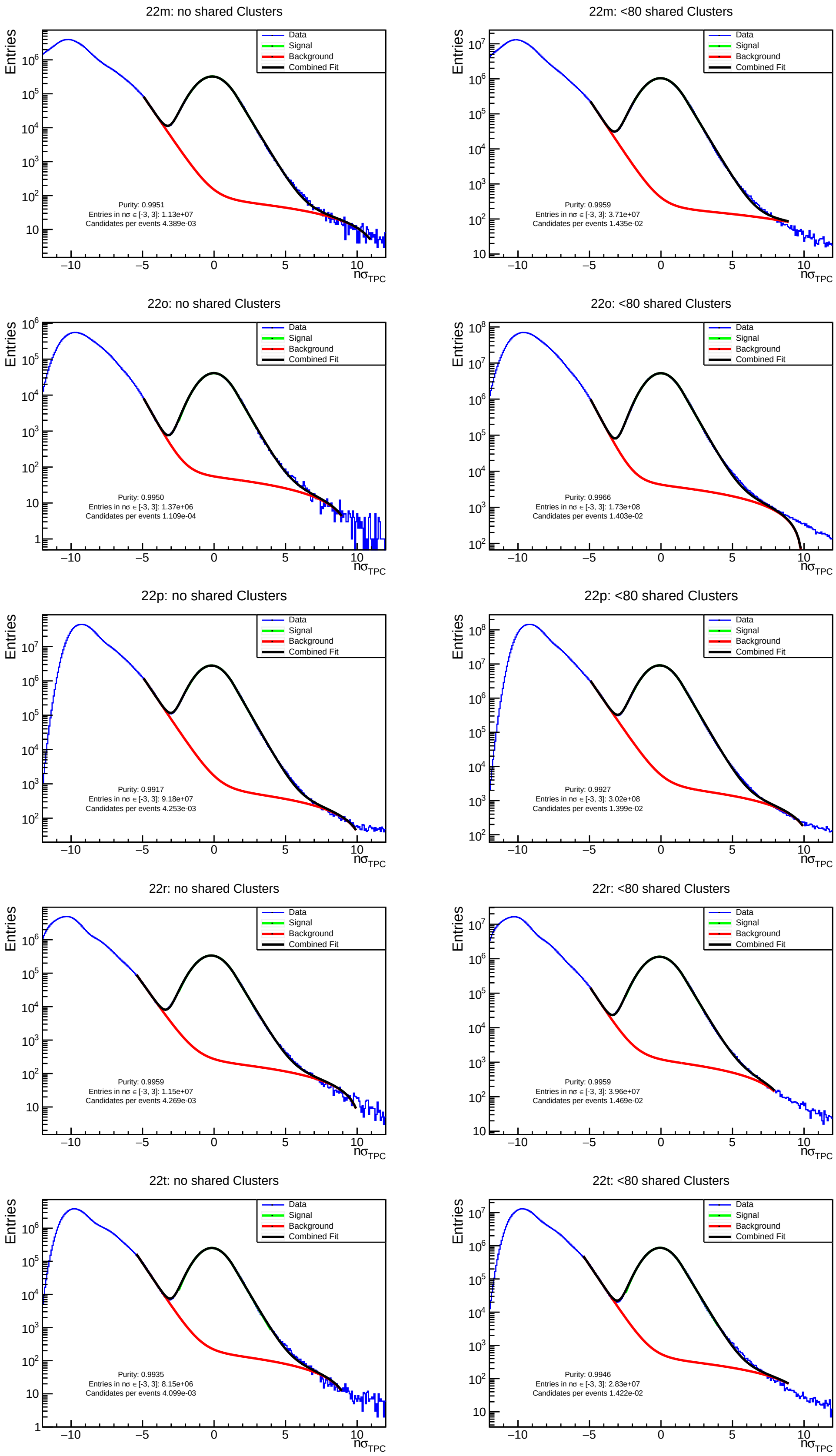


Figure 9.1: Fit of the TPC $n\sigma$ distribution for protons with $p_{\mathrm{T}} \in [0.6, 0.75)$ GeV/c for proton selection with no allowed shared clusters (left panels) and 80 allowed shared clusters (right panels). Each row corresponds to another datataking period of 2022.

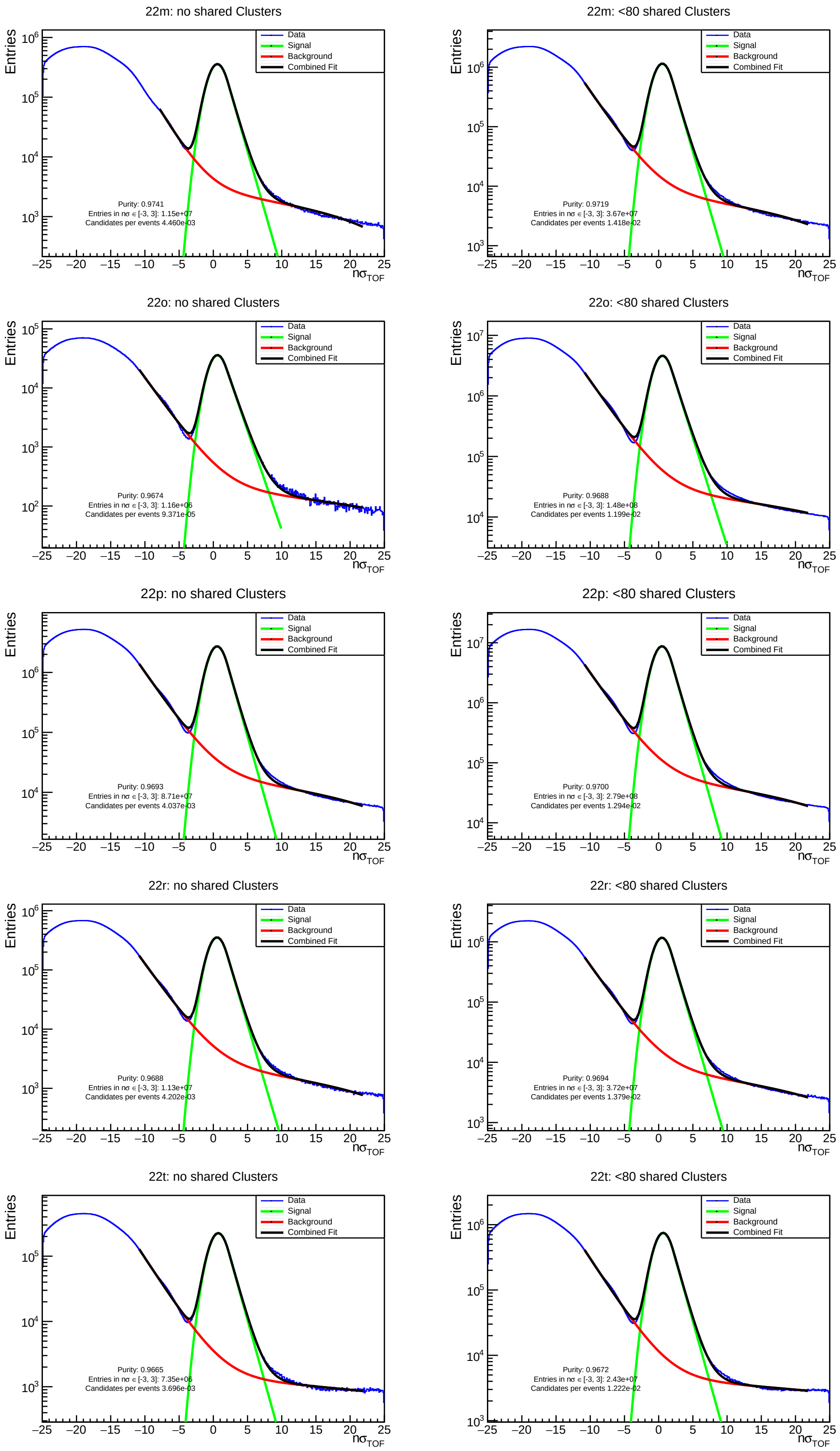


Figure 9.2: Fit of the TOF $n\sigma$ distribution for protons with $p_{\mathrm{T}} \in [1.25, 1.75)$ GeV/c for proton selection with no allowed shared clusters (left panels) and 80 allowed shared clusters (right panels). Each row corresponds to another datataking period of 2022.

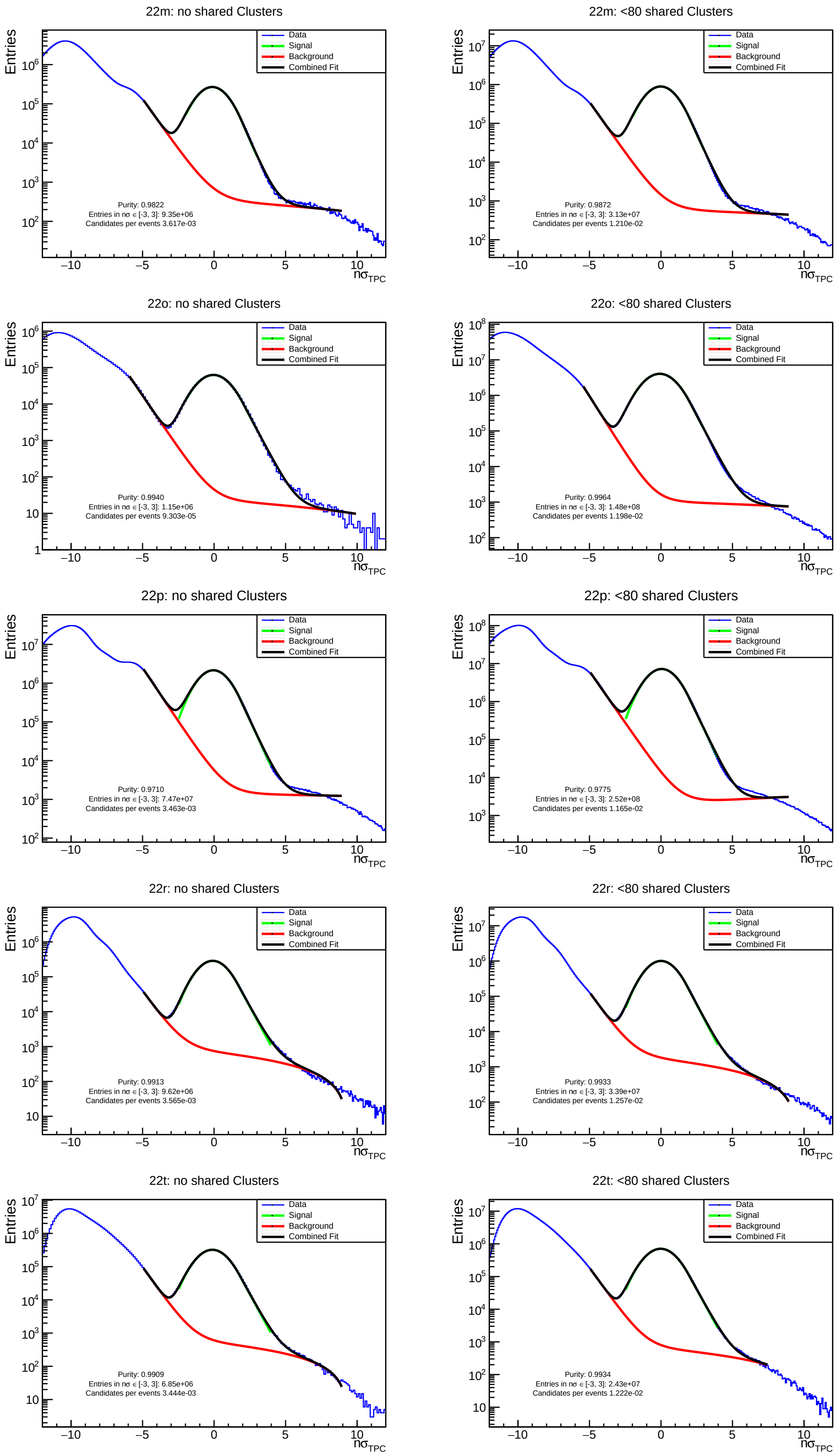


Figure 9.3: Fit of the TPC $n\sigma$ distribution for antiprotons with $p_{\mathrm{T}} \in [0.6, 0.75)$ GeV/c for proton selection with no allowed shared clusters (left panels) and 80 allowed shared clusters (right panels). Each row corresponds to another datataking period of 2022.

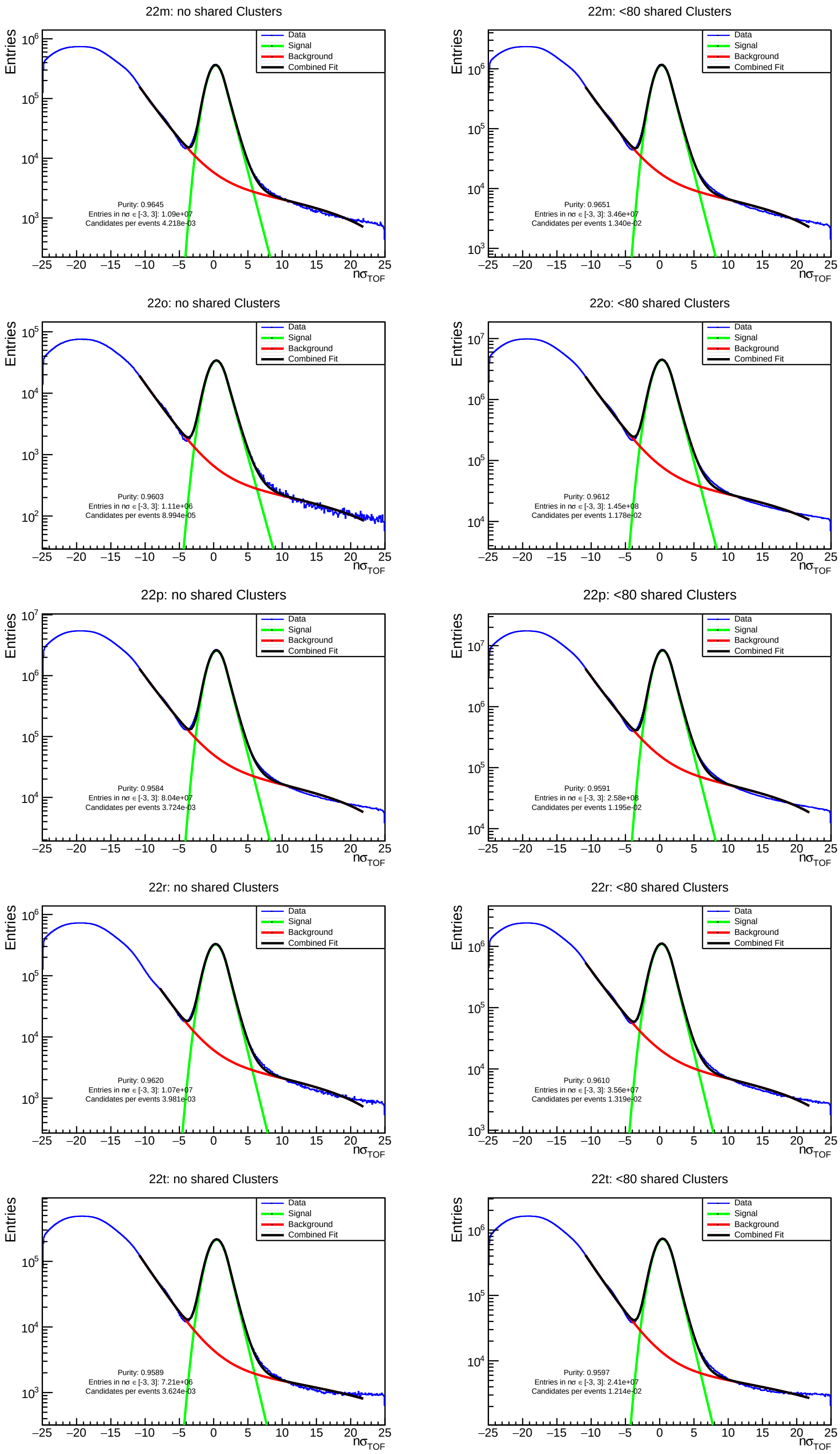


Figure 9.4: Fit of the TOF $n\sigma$ distribution for antiprotons with $p_{\mathrm{T}} \in [1.25, 1.75)$ GeV/c for proton selection with no allowed shared clusters (left panels) and 80 allowed shared clusters (right panels). Each row corresponds to another datataking period of 2022.

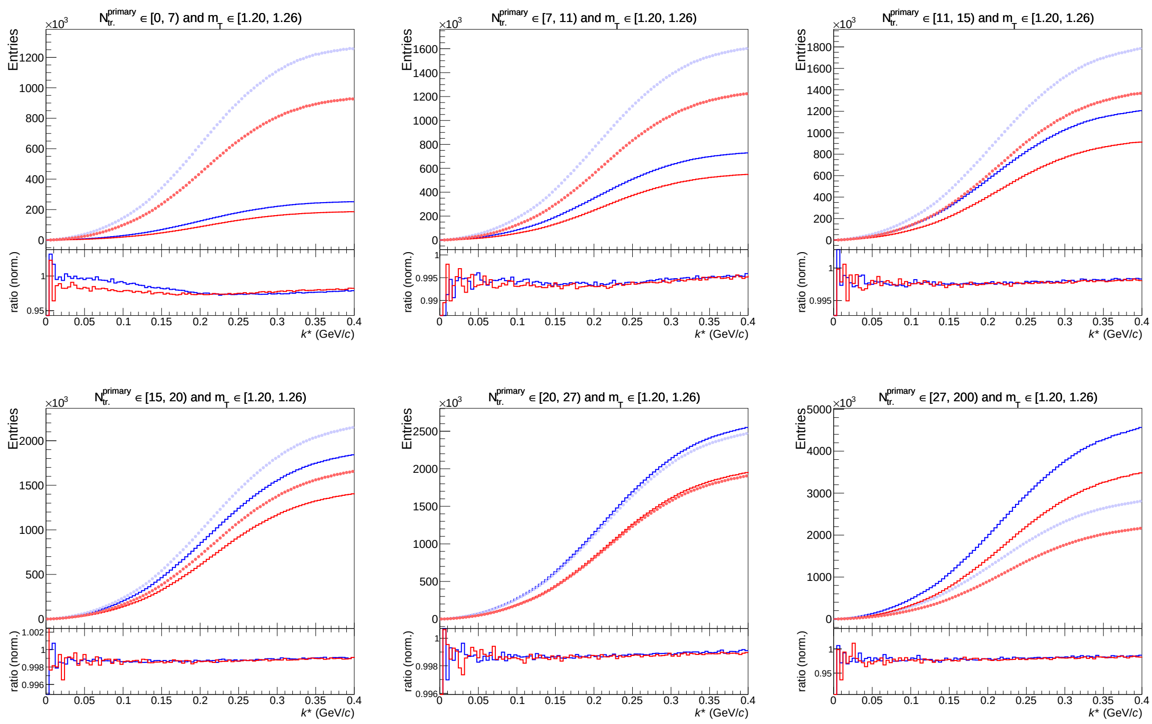


Figure 9.12: Reweighted and unweighted mixed event distributions (upper panel) and their ratio (lower panel) for p–p and $\overline{\text{p}}$–$\overline{\text{p}}$ in blue and red, respectively, for $m_{\mathrm{T}} \in [1.20, 1.26)$ and all multiplicity intervals. More detail in 5.1.2

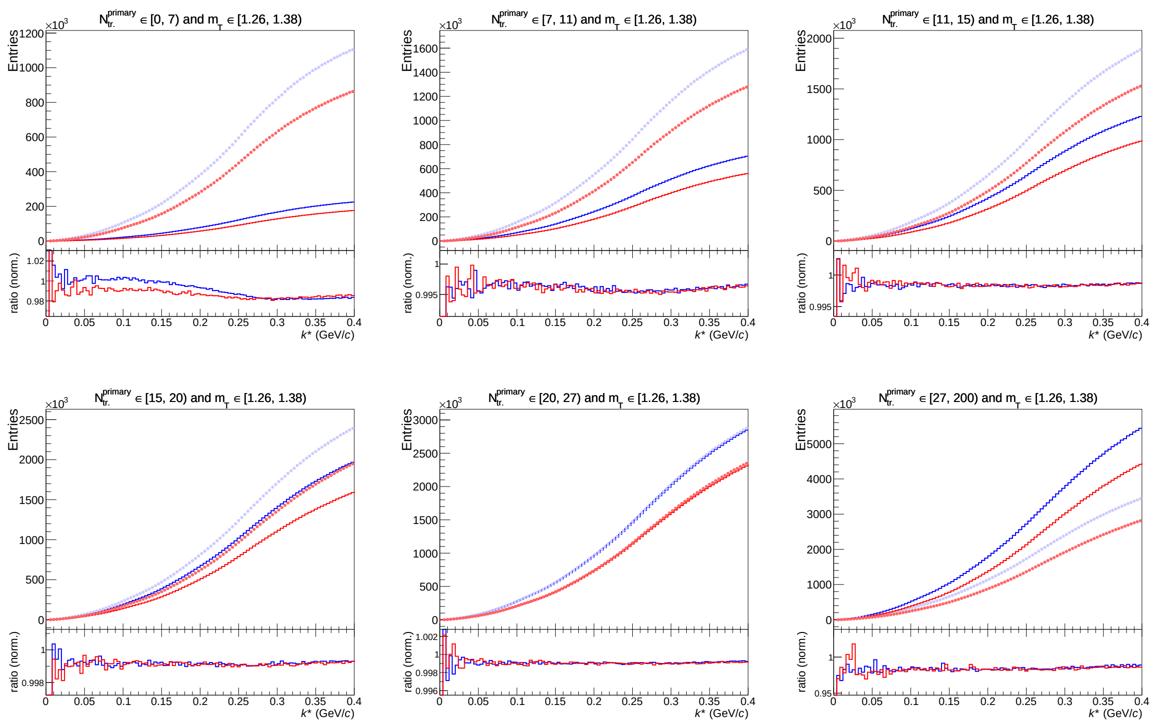


Figure 9.15: Reweighted and unweighted mixed event distributions (upper panel) and their ratio (lower panel) for p–p and $\overline{\text{p}}$–$\overline{\text{p}}$ in blue and red, respectively, for $m_{\mathrm{T}} \in [1.26, 1.38)$ and all multiplicity intervals. More detail in 5.1.2

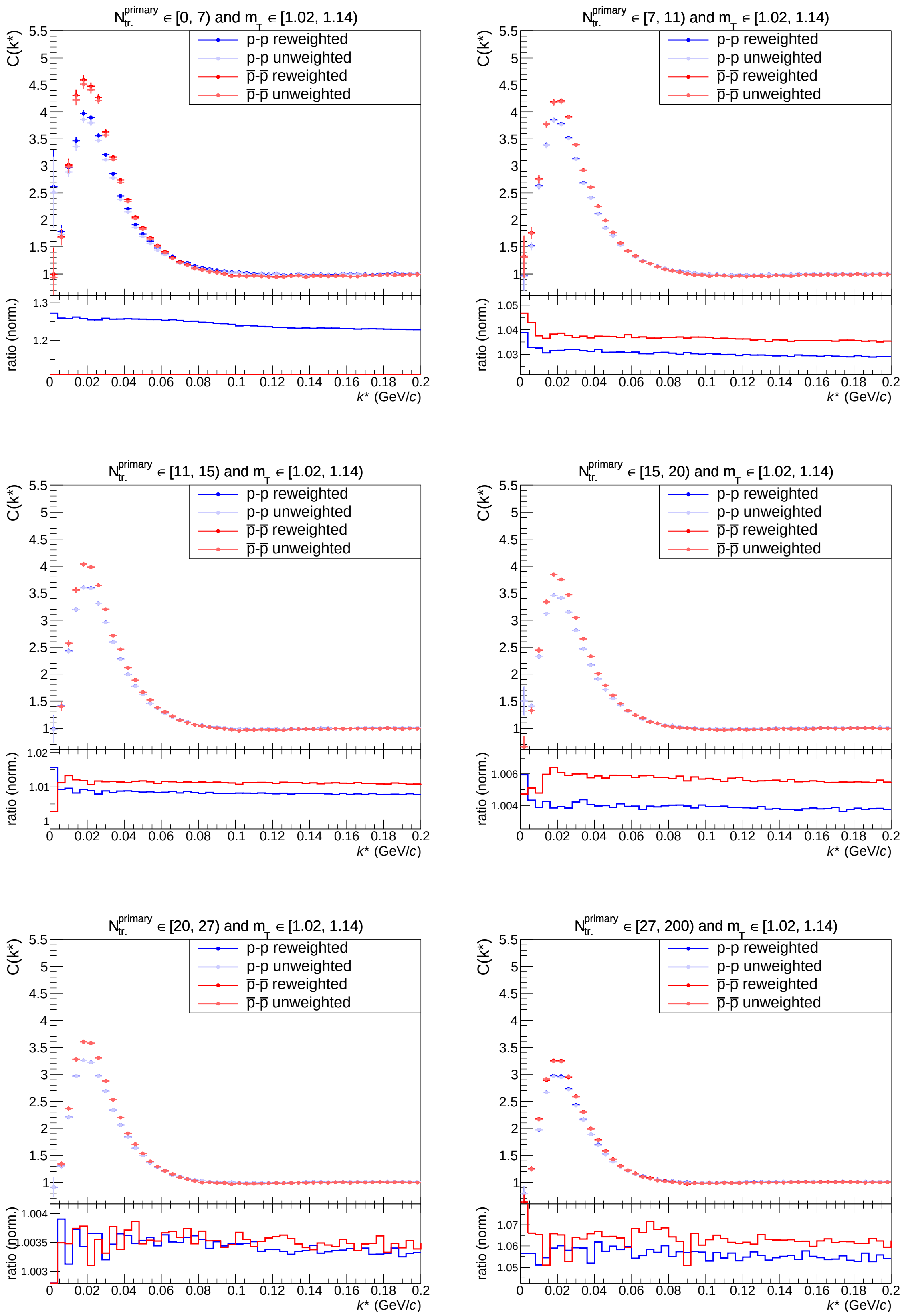


Figure 9.5: Reweighted and unweighted correlation functions (upper panel) and their ratio (lower panel) for p–p and $\overline{\text{p}}$–$\overline{\text{p}}$ in blue and red, respectively, for $m_{\text{T}} \in [1.02, 1.14)$ and all multiplicity intervals. More detail in 5.1.2

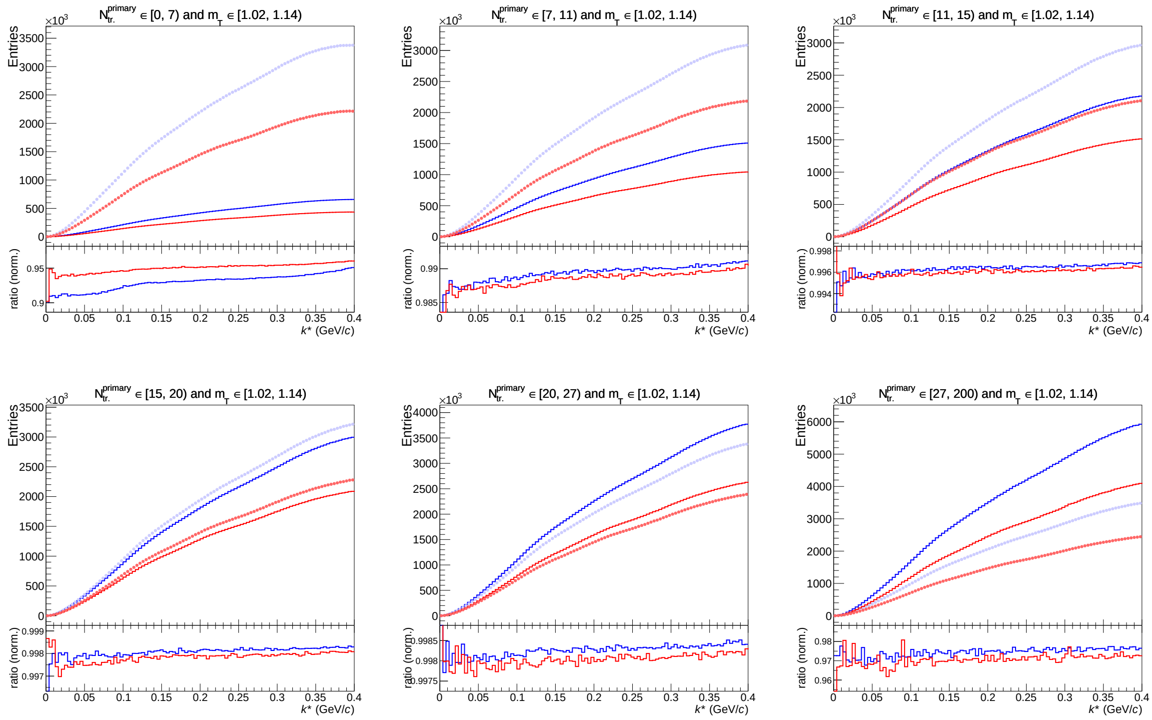


Figure 9.6: Reweighted and unweighted mixed event distributions (upper panel) and their ratio (lower panel) for p–p and $\overline{\text{p}}$–$\overline{\text{p}}$ in blue and red, respectively, for $m_T \in [1.02, 1.14)$ and all multiplicity intervals. More detail in 5.1.2

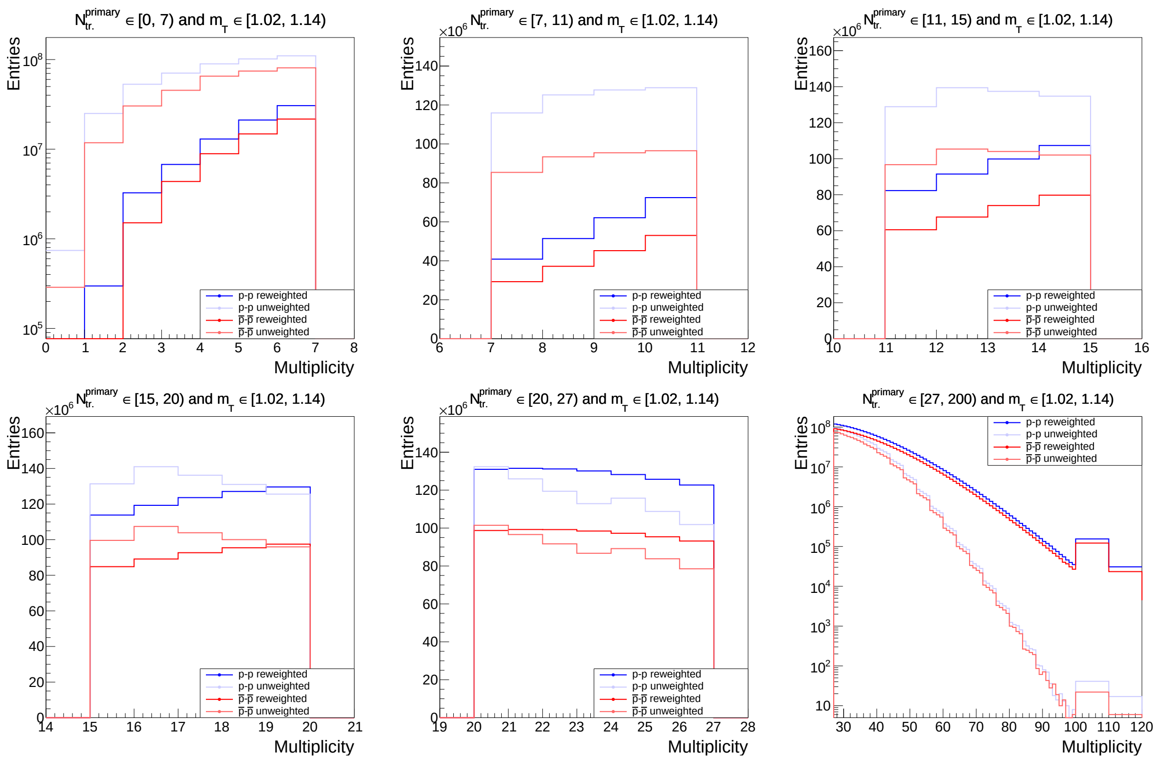


Figure 9.7: Reweighted and unweighted multiplicity distributions for p–p and $\overline{\text{p}}$–$\overline{\text{p}}$ in blue and red, respectively, for $m_T \in [1.02, 1.14)$ and all multiplicity intervals. More detail in 5.1.2

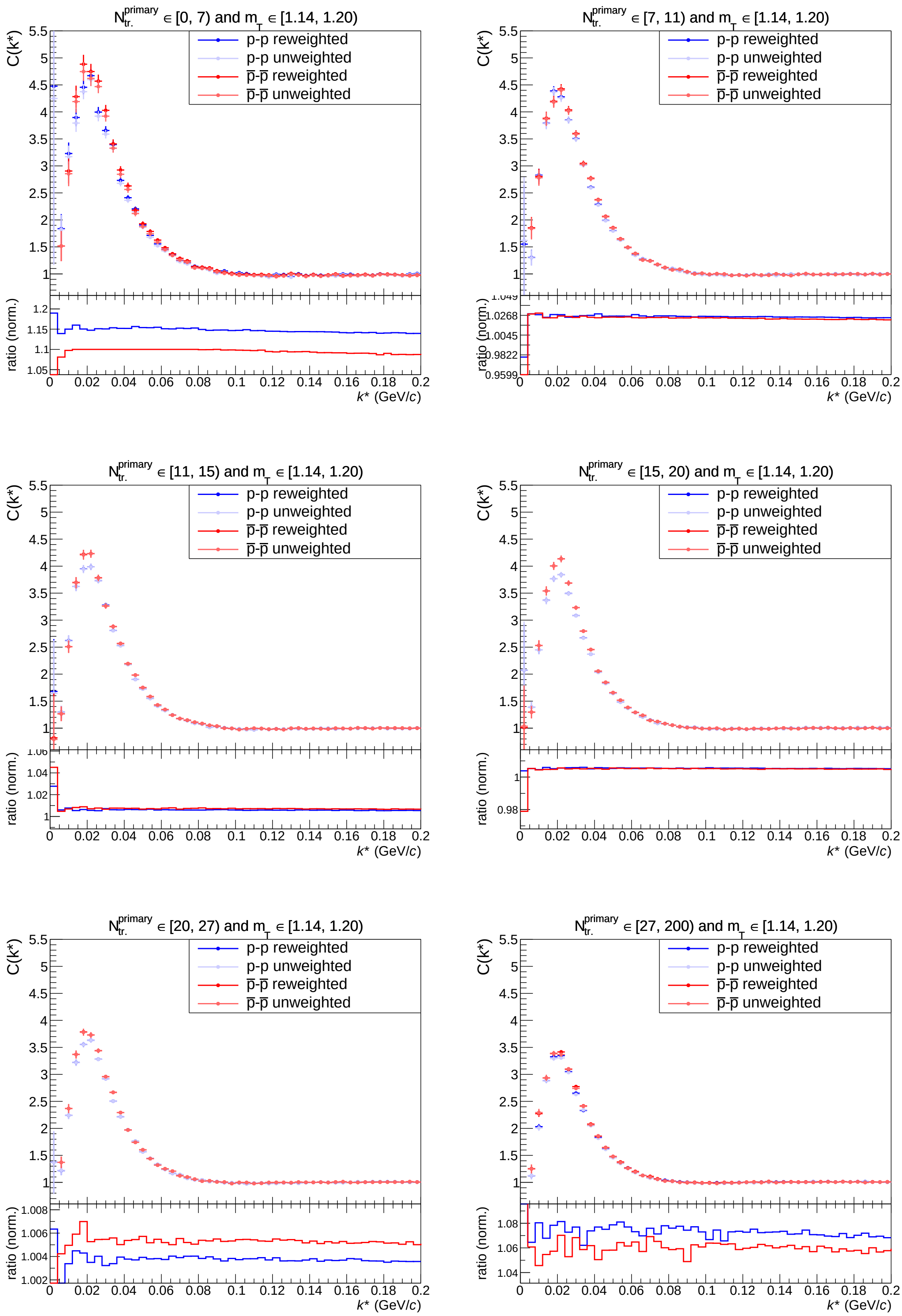


Figure 9.8: Reweighted and unweighted correlation functions (upper panel) and their ratio (lower panel) for p–p and $\overline{\text{p}}$–$\overline{\text{p}}$ in blue and red, respectively, for $m_{\text{T}} \in [1.14, 1.20)$ and all multiplicity intervals. More detail in 5.1.2

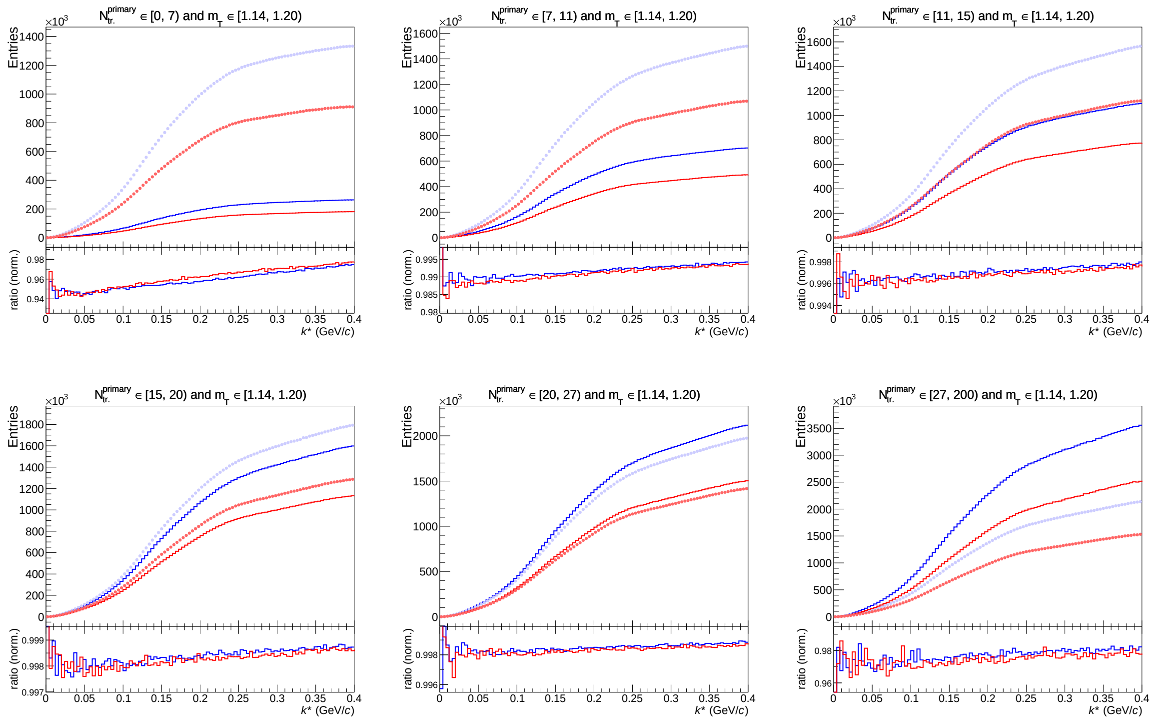


Figure 9.9: Reweighted and unweighted mixed event distributions (upper panel) and their ratio (lower panel) for p–p and $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ in blue and red, respectively, for $m_{\mathrm{T}} \in [1.14, 1.20)$ and all multiplicity intervals. More detail in 5.1.2

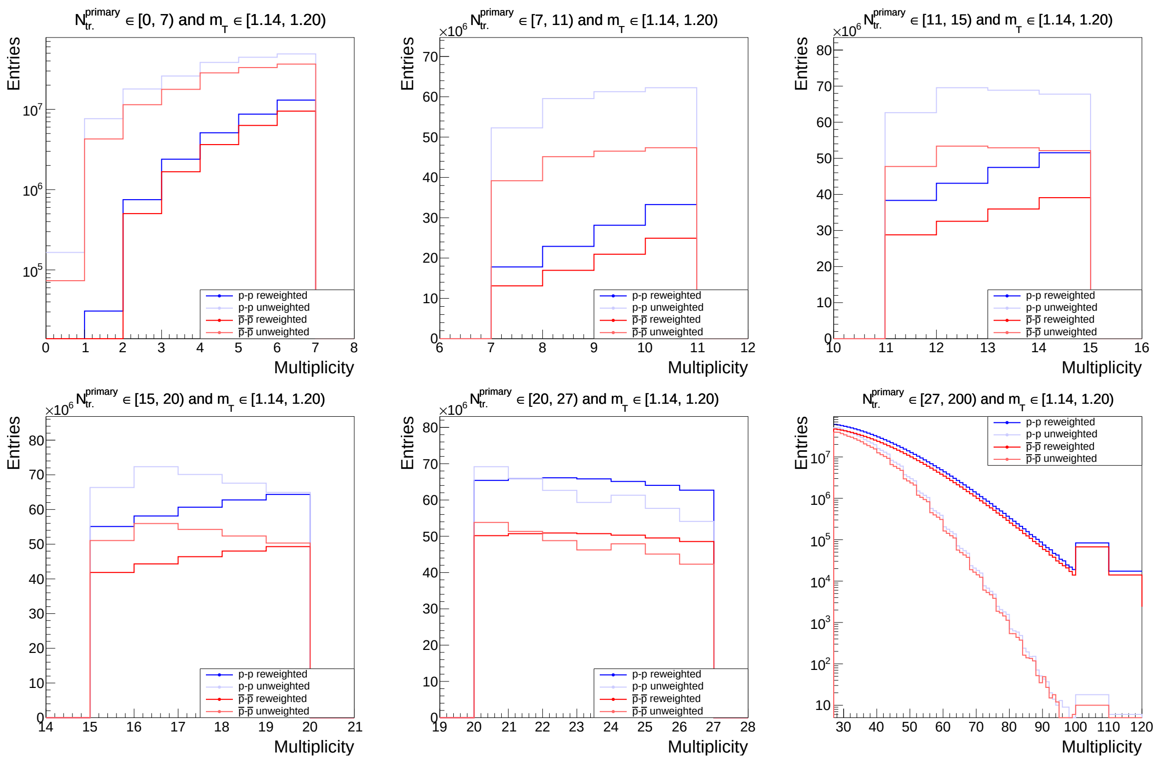


Figure 9.10: Reweighted and unweighted multiplicity distributions for p–p and $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ in blue and red, respectively, for $m_{\mathrm{T}} \in [1.14, 1.20)$ and all multiplicity intervals. More detail in 5.1.2

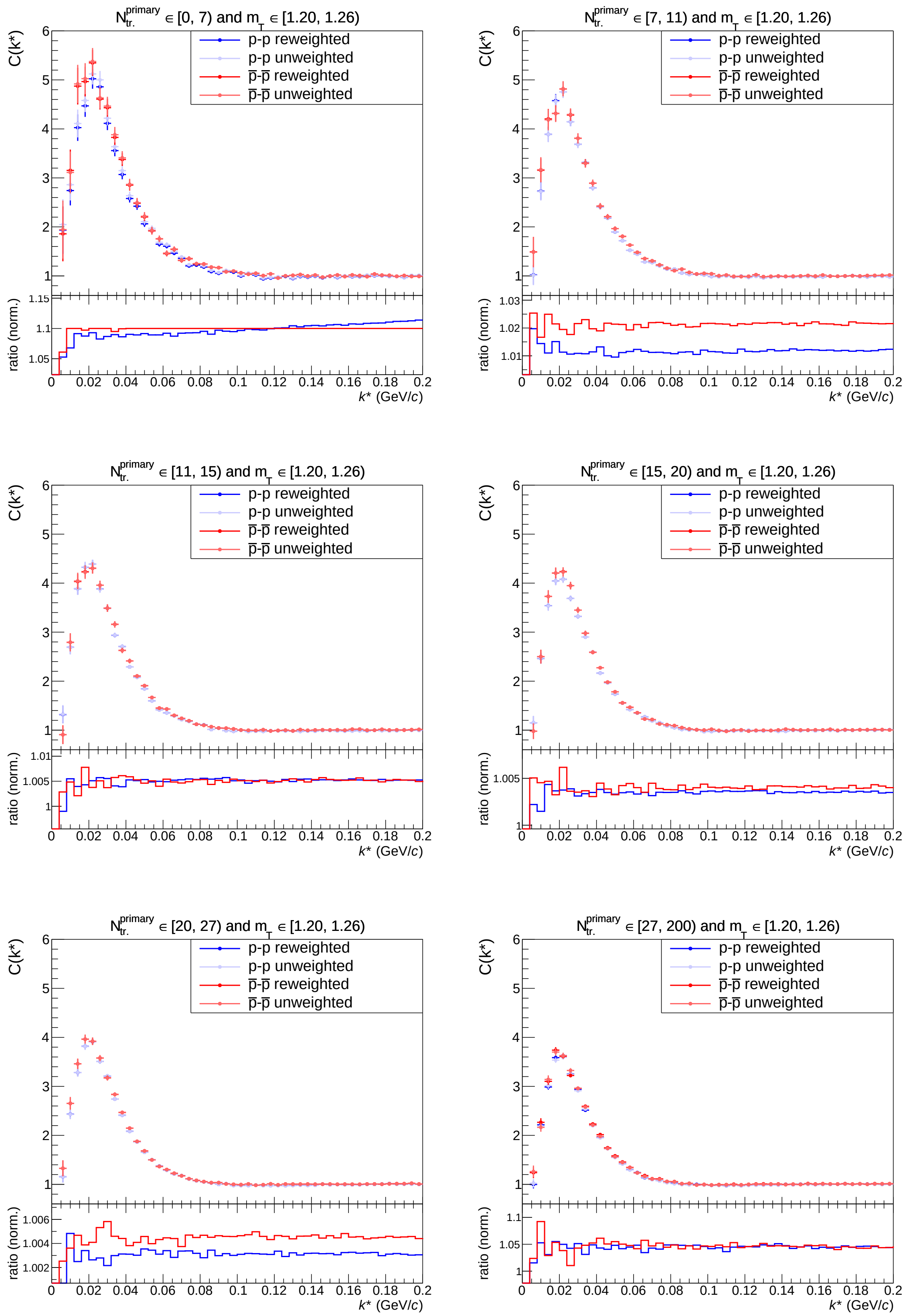


Figure 9.11: Reweighted and unweighted correlation functions (upper panel) and their ratio (lower panel) for p–p and $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ in blue and red, respectively, for $m_{\mathrm{T}} \in [1.20, 1.26)$ and all multiplicity intervals. More detail in 5.1.2

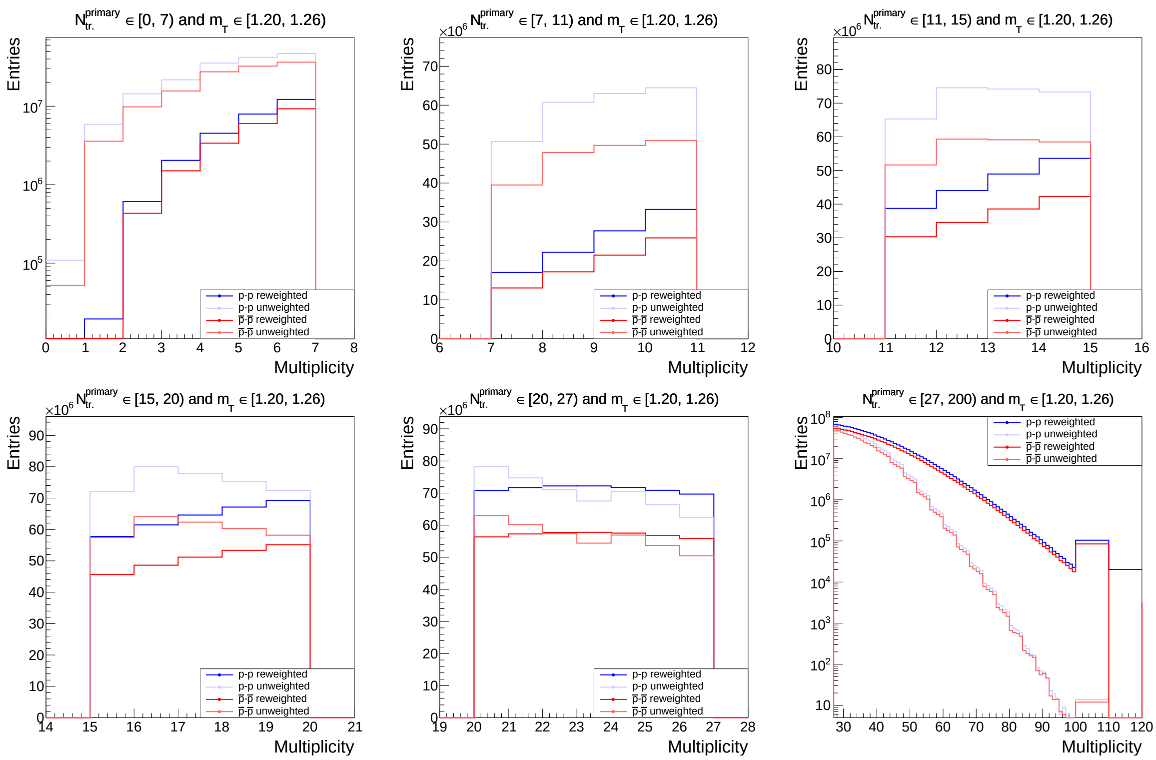


Figure 9.13: Reweighted and unweighted multiplicity distributions for p–p and $\overline{\text{p}}$–$\overline{\text{p}}$ in blue and red, respectively, for $m_{\text{T}} \in [1.20, 1.26)$ and all multiplicity intervals. More detail in 5.1.2

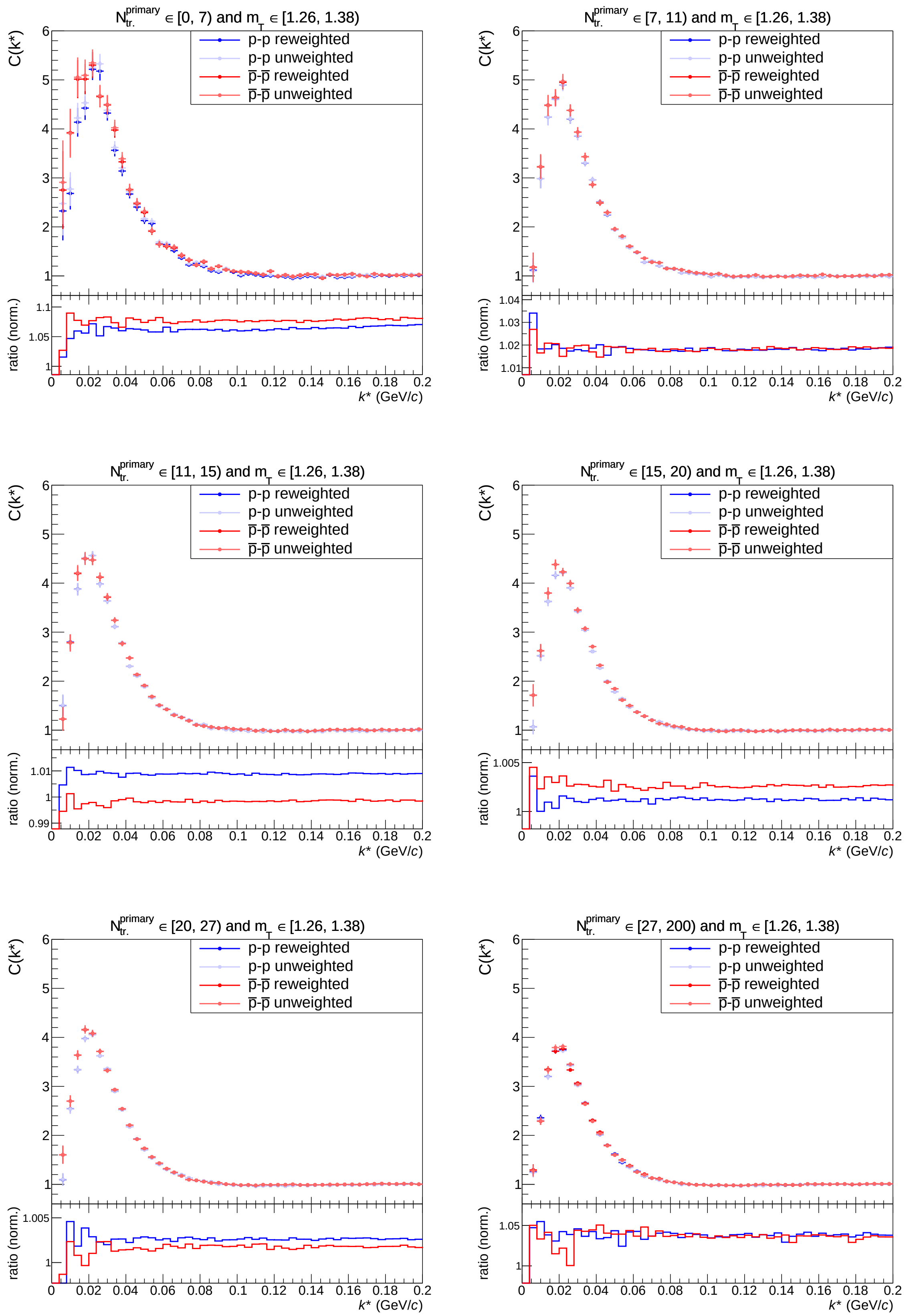


Figure 9.14: Reweighted and unweighted correlation functions (upper panel) and their ratio (lower panel) for p–p and $\overline{\text{p}}$–$\overline{\text{p}}$ in blue and red, respectively, for $m_{\text{T}} \in [1.26, 1.38)$ and all multiplicity intervals. More detail in 5.1.2

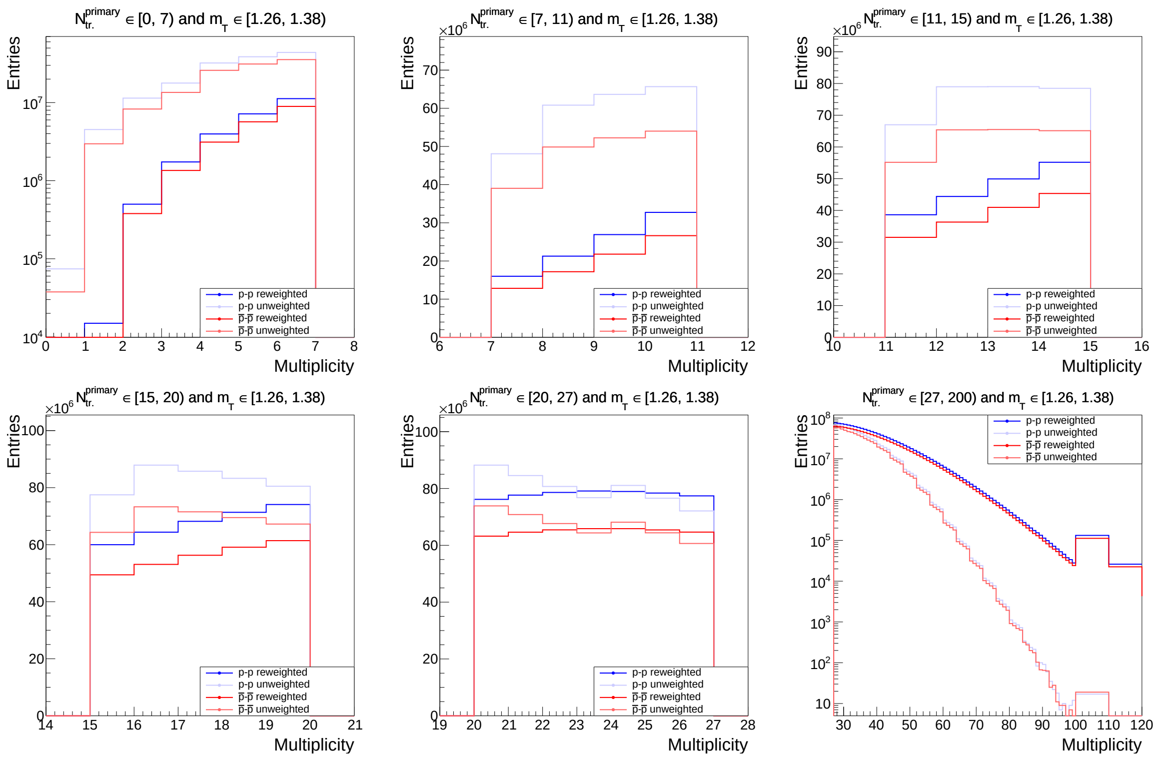


Figure 9.16: Reweighted and unweighted multiplicity distributions for p–p and p̄–p̄ in blue and red, respectively, for $m_{\mathrm{T}} \in [1.26, 1.38)$ and all multiplicity intervals. More detail in 5.1.2

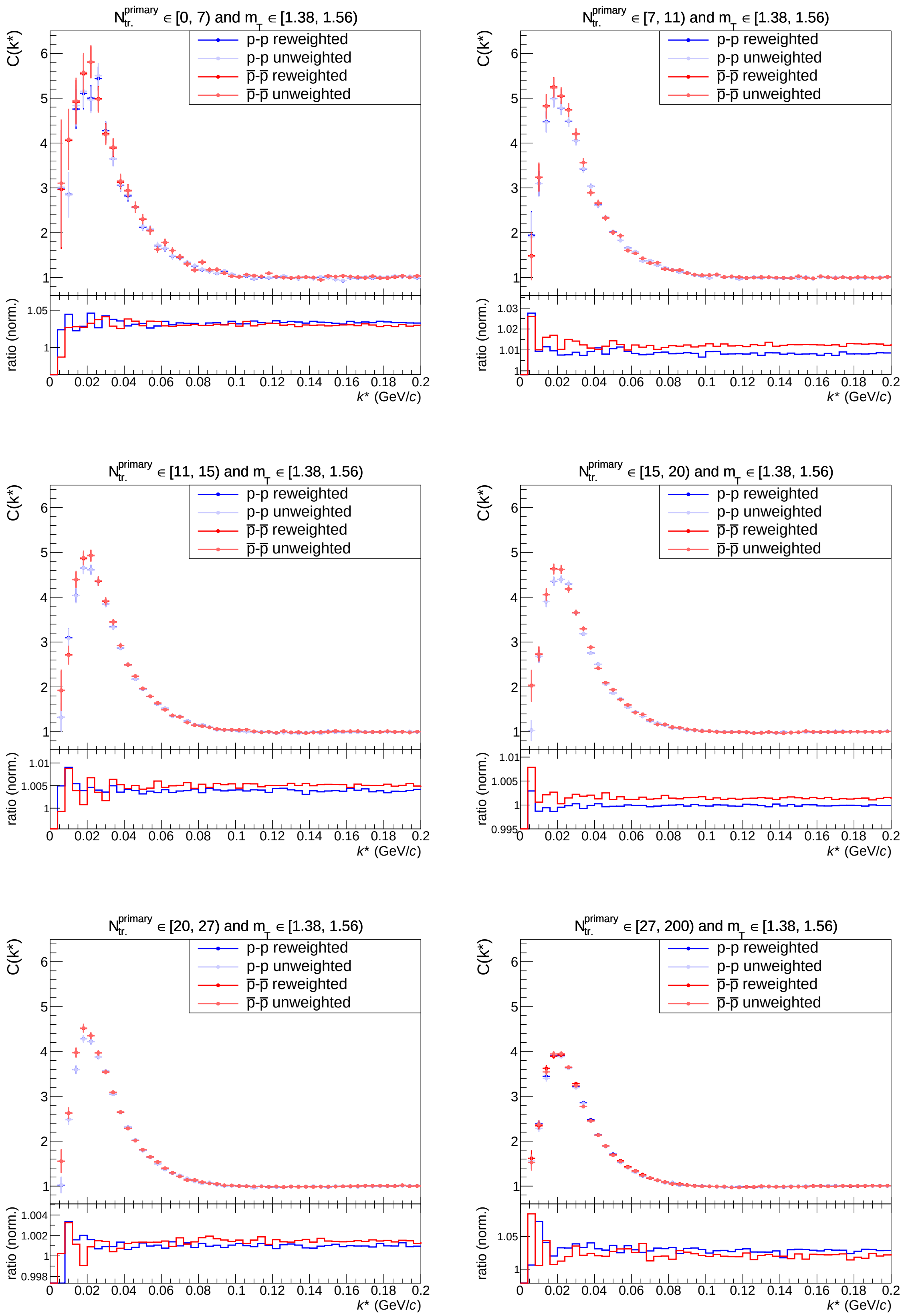


Figure 9.17: Reweighted and unweighted correlation functions (upper panel) and their ratio (lower panel) for p–p and $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ in blue and red, respectively, for $m_{\mathrm{T}} \in [1.38, 1.56)$ and all multiplicity intervals. More detail in 5.1.2

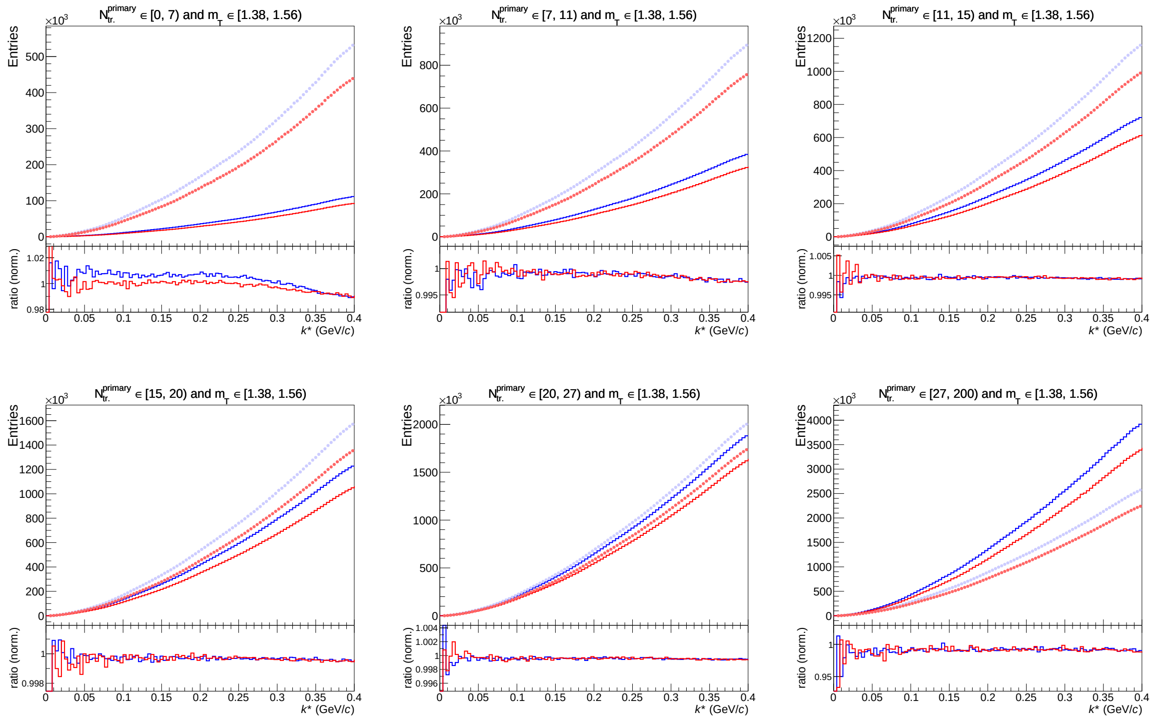


Figure 9.18: Reweighted and unweighted mixed event distributions (upper panel) and their ratio (lower panel) for p–p and $\overline{\text{p}}$–$\overline{\text{p}}$ in blue and red, respectively, for $m_{\mathrm{T}} \in [1.86, 2.21)$ and all multiplicity intervals. More detail in 5.1.2

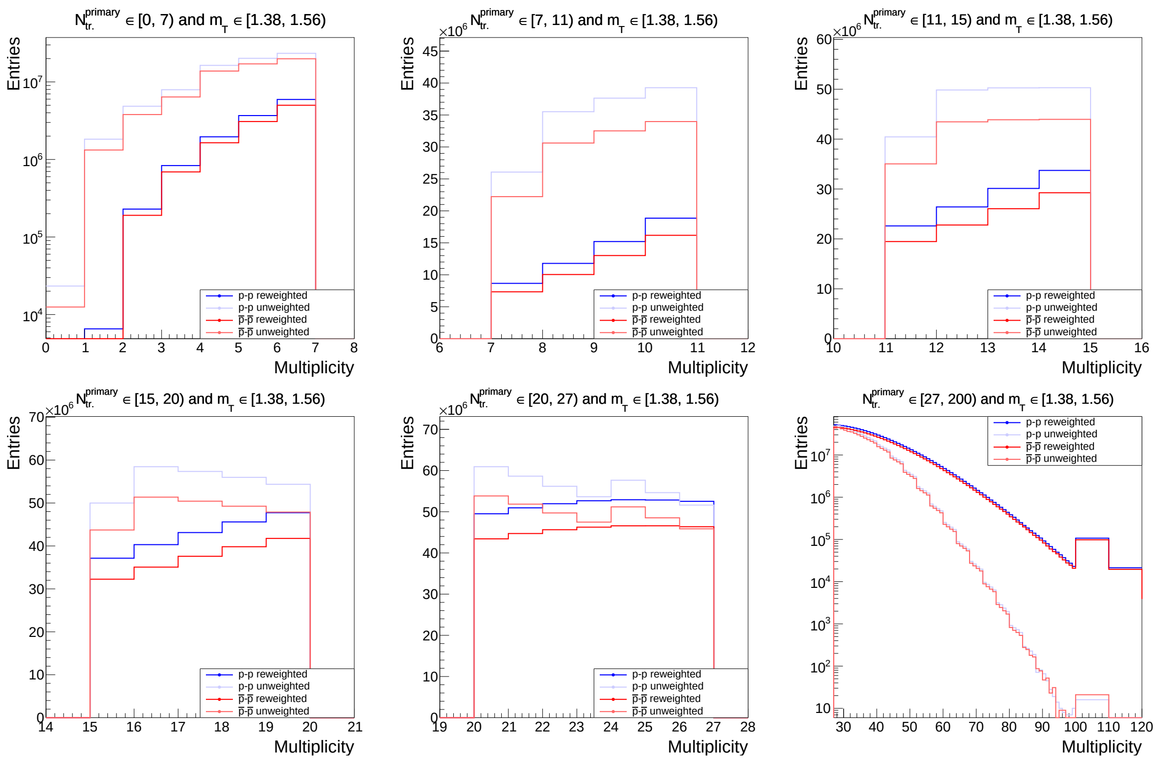


Figure 9.19: Reweighted and unweighted multiplicity distributions for p–p and $\overline{\text{p}}$–$\overline{\text{p}}$ in blue and red, respectively, for $m_{\mathrm{T}} \in [1.26, 1.38)$ and all multiplicity intervals. More detail in 5.1.2

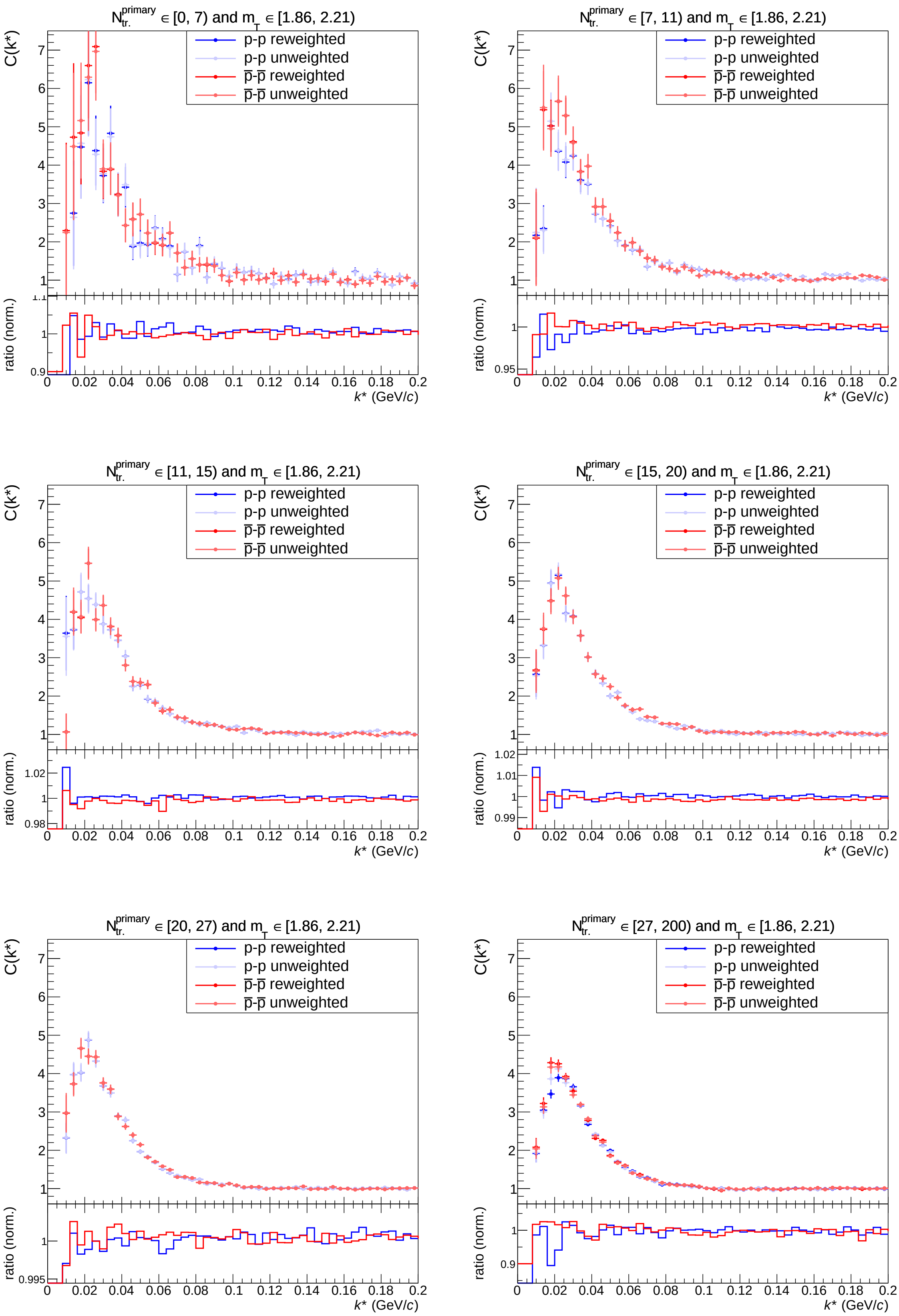


Figure 9.20: Reweighted and unweighted correlation functions (upper panel) and their ratio (lower panel) for p–p and $\overline{\text{p}}$–$\overline{\text{p}}$ in blue and red, respectively, for $m_T \in [1.86, 2.21)$ and all multiplicity intervals. More detail in 5.1.2

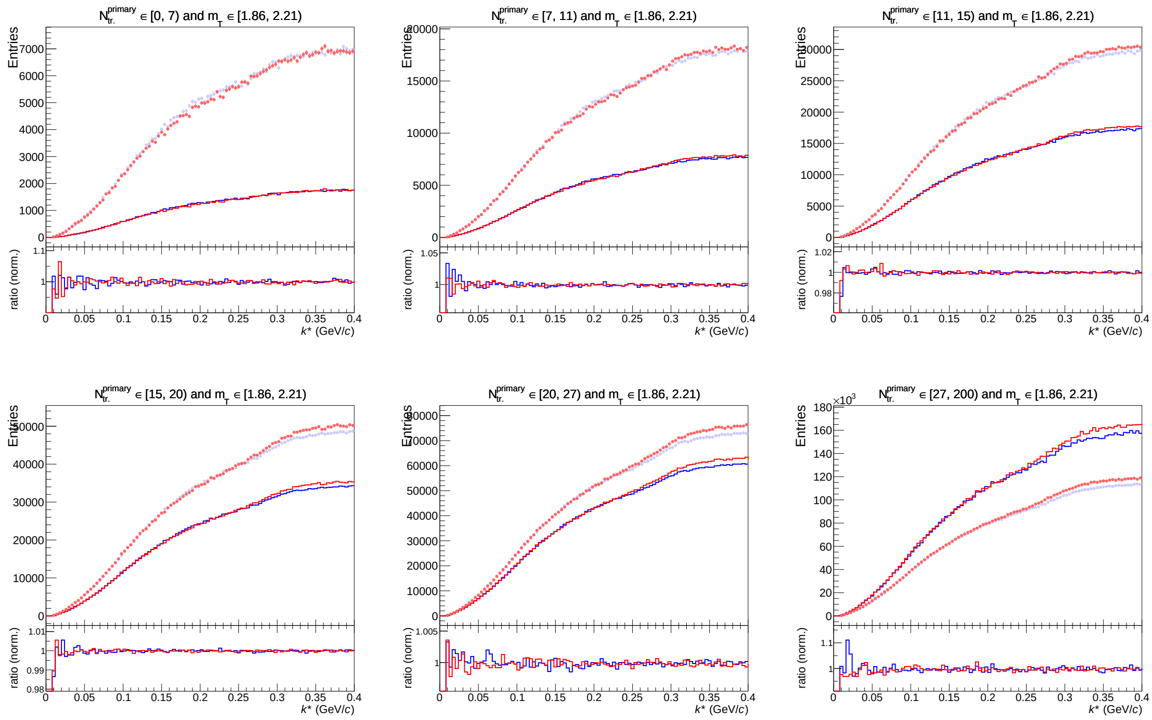


Figure 9.21: Reweighted and unweighted mixed event distributions (upper panel) and their ratio (lower panel) for p–p and $\overline{\text{p}}$–$\overline{\text{p}}$ in blue and red, respectively, for $m_{\text{T}} \in [1.86, 2.21)$ and all multiplicity intervals. More detail in 5.1.2

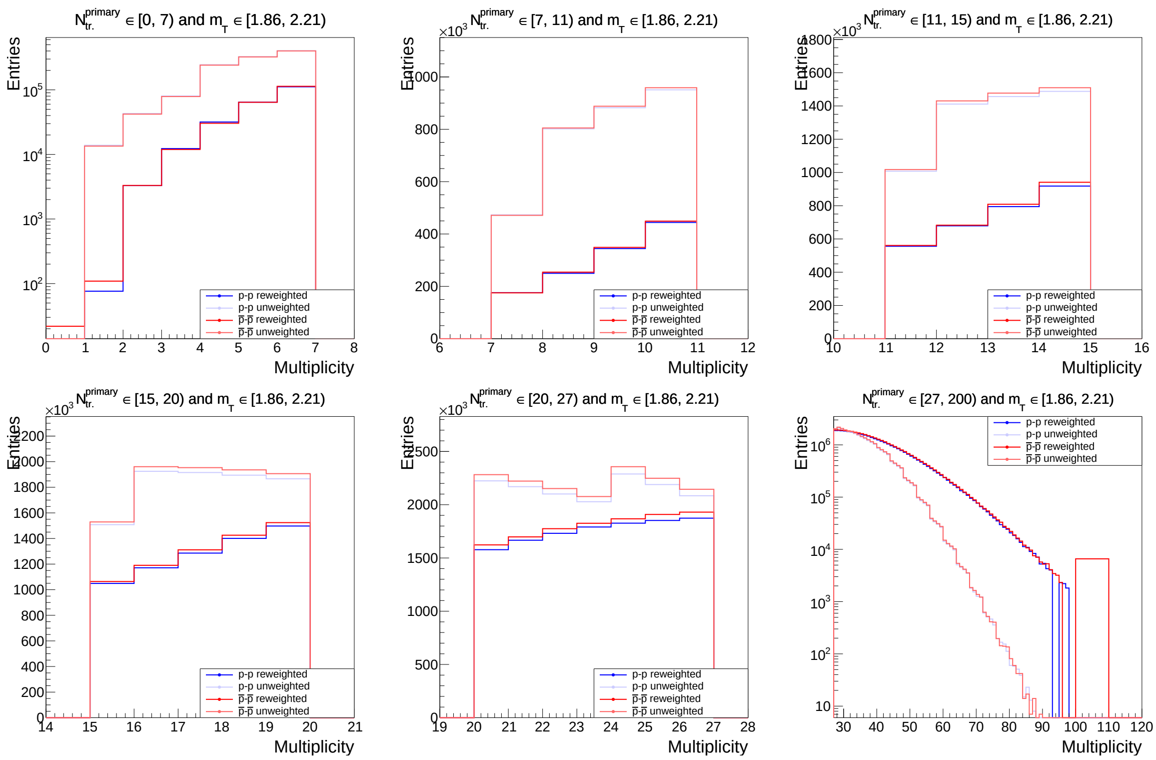


Figure 9.22: Reweighted and unweighted multiplicity distributions for p–p and $\overline{\text{p}}$–$\overline{\text{p}}$ in blue and red, respectively, for $m_{\text{T}} \in [1.86, 2.21)$ and all multiplicity intervals. More detail in 5.1.2

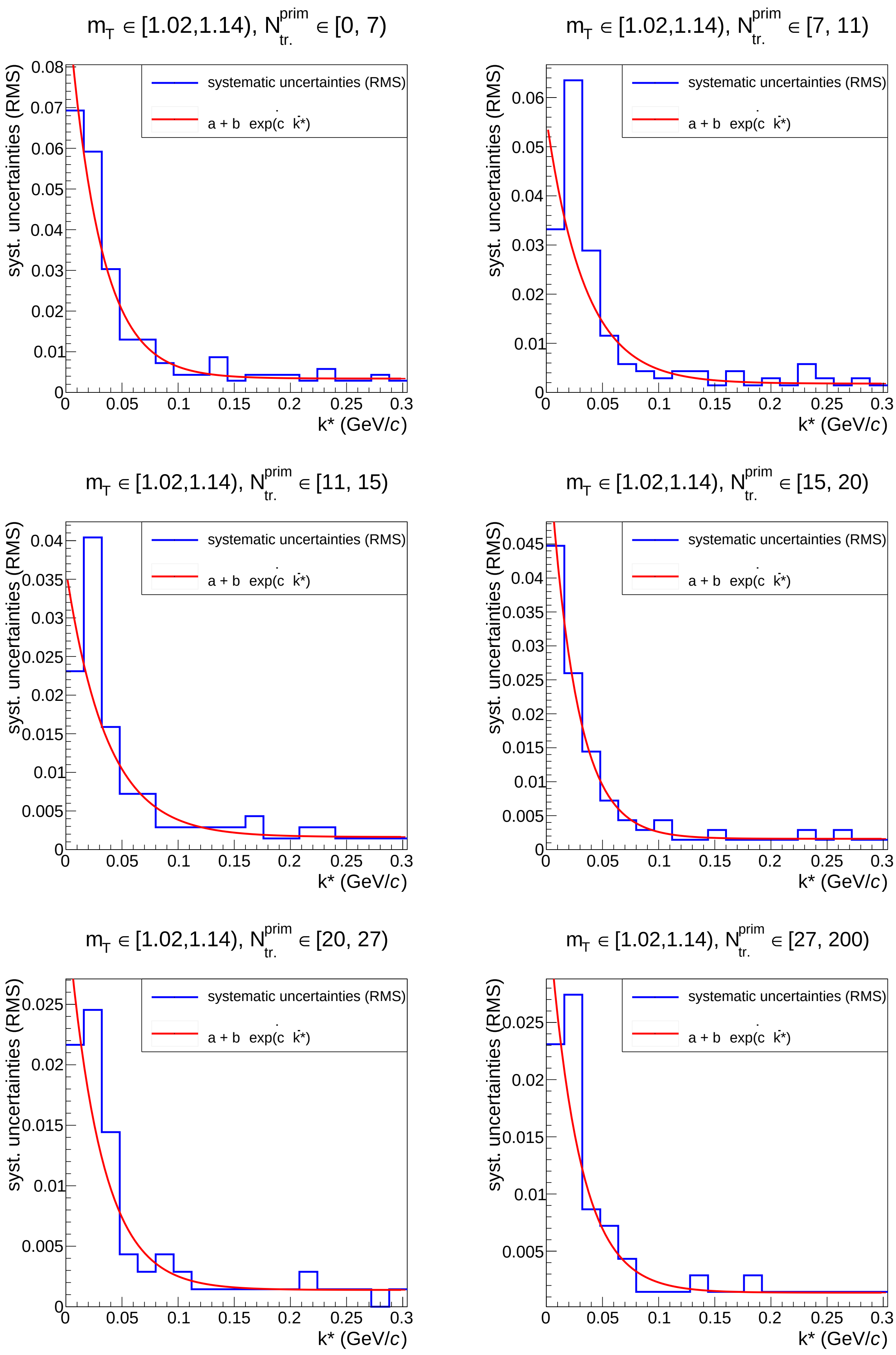


Figure 9.23: Systematic uncertainties of the p–p correlation functions for all multiplicity intervals in the second largest m_T interval ($m_T \in [1.02\,\mathrm{GeV}/c,\ 1.14\,\mathrm{GeV}/c)$)

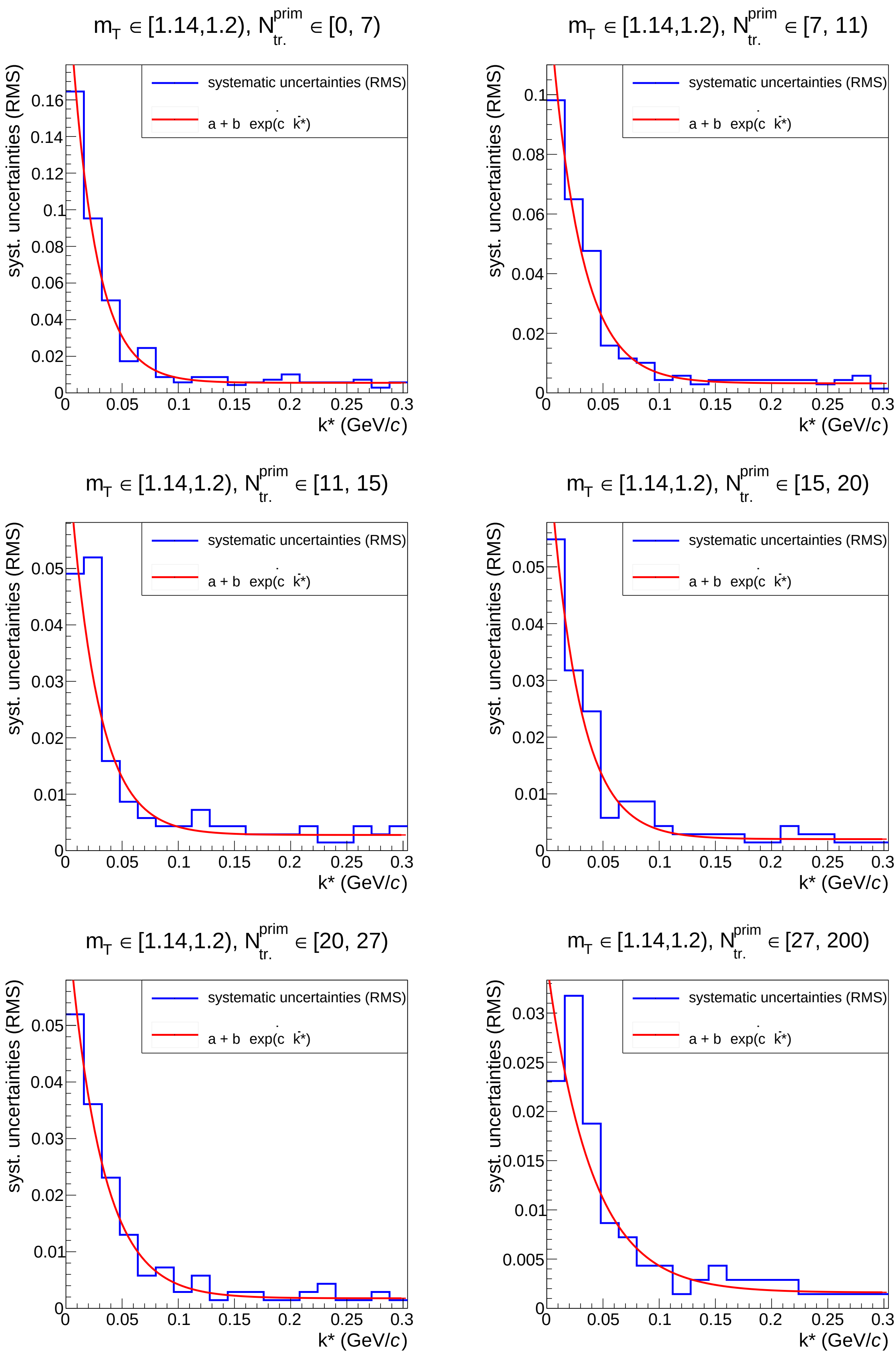


Figure 9.24: Systematic uncertainties of the p–p correlation functions for all multiplicity intervals in the second largest m_T interval ($m_T \in [1.14\,\mathrm{GeV}/c,\ 1.20\,\mathrm{GeV}/c)$)

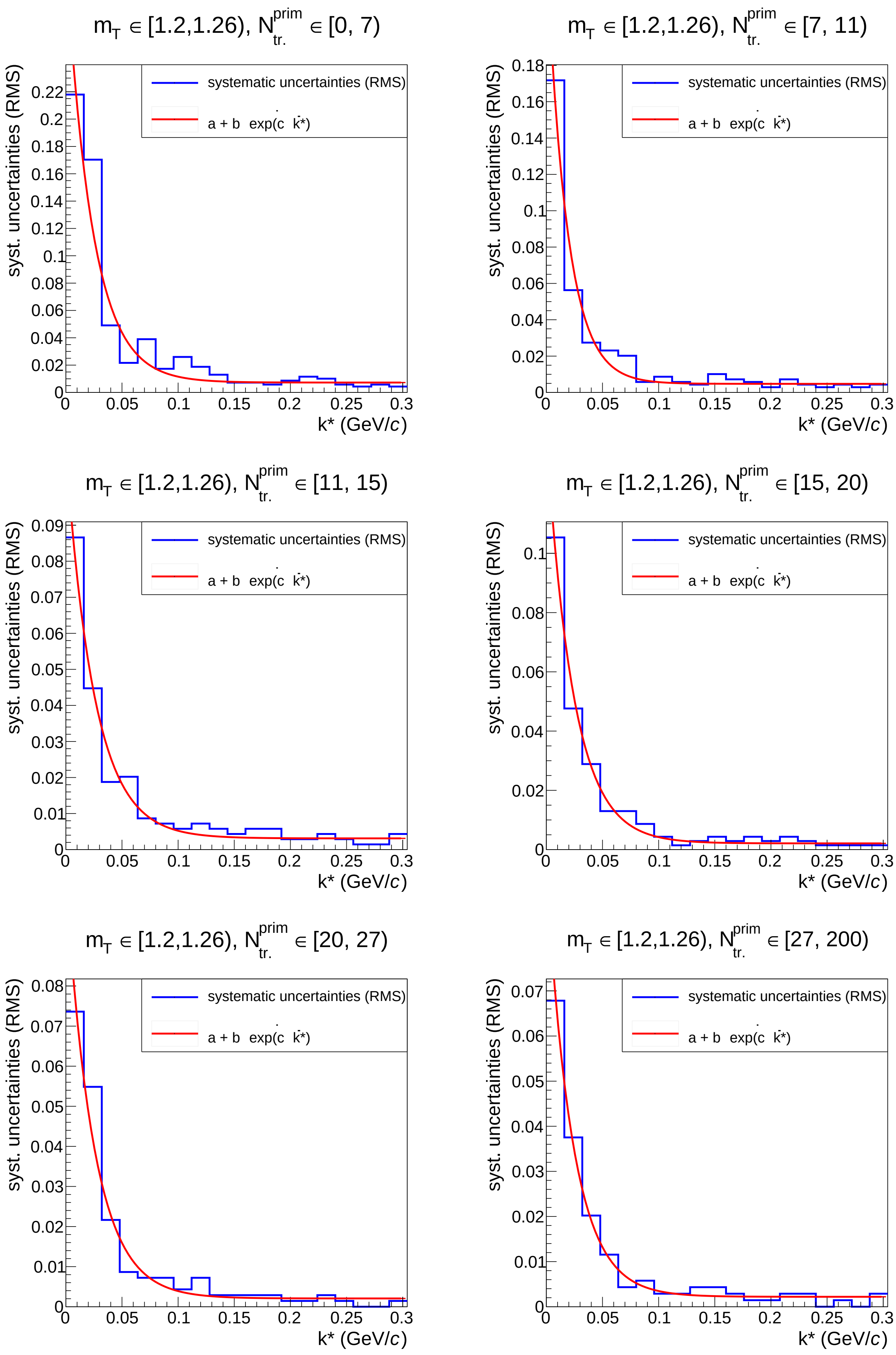


Figure 9.25: Systematic uncertainties of the p–p correlation functions for all multiplicity intervals in the second largest m_T interval ($m_T \in [1.20\,\mathrm{GeV}/c,\ 1.26\,\mathrm{GeV}/c)$)

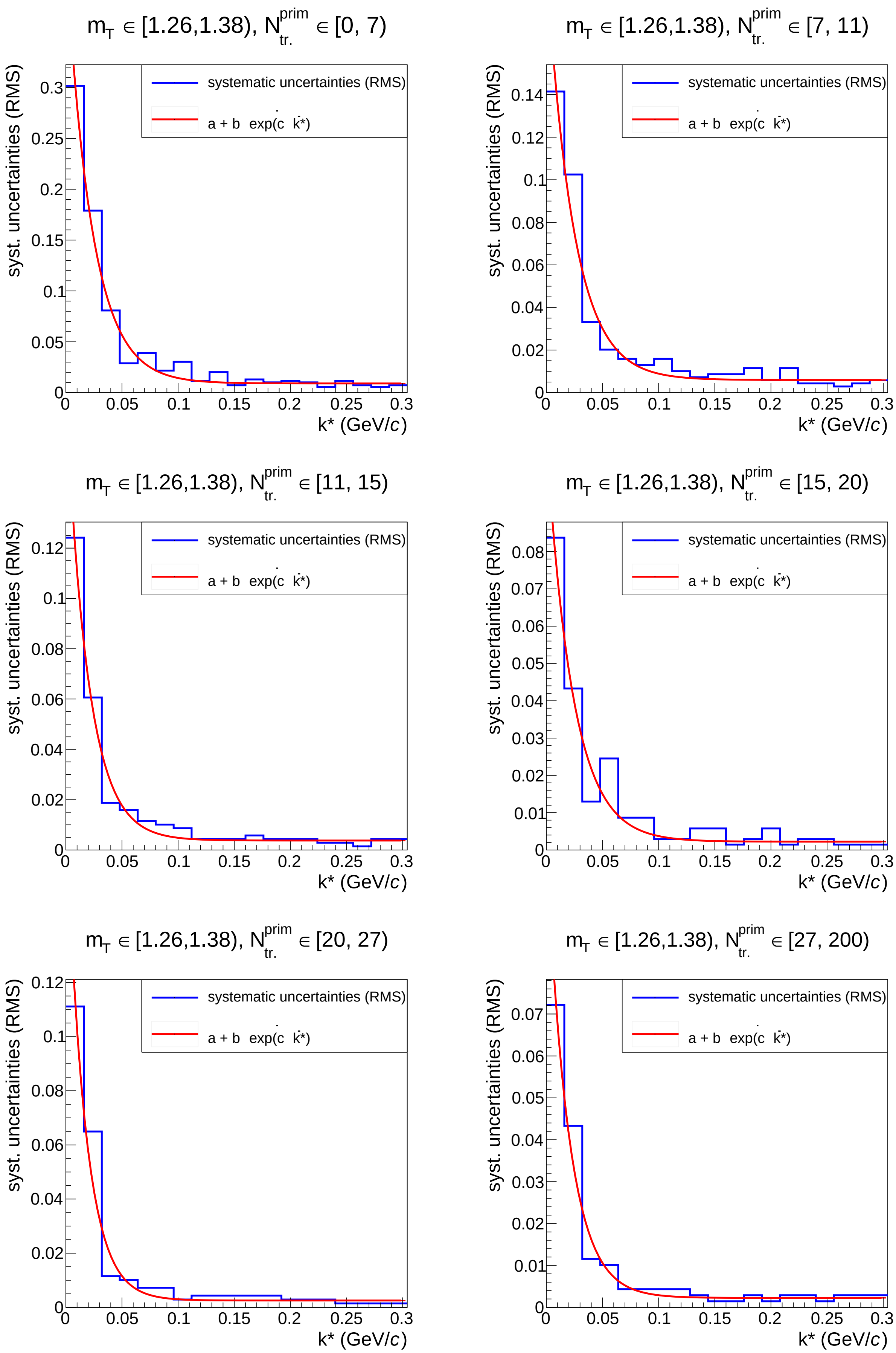


Figure 9.26: Systematic uncertainties of the p–p correlation functions for all multiplicity intervals in the second largest m_T interval ($m_T \in [1.26\,\mathrm{GeV}/c,\ 1.38\,\mathrm{GeV}/c)$)

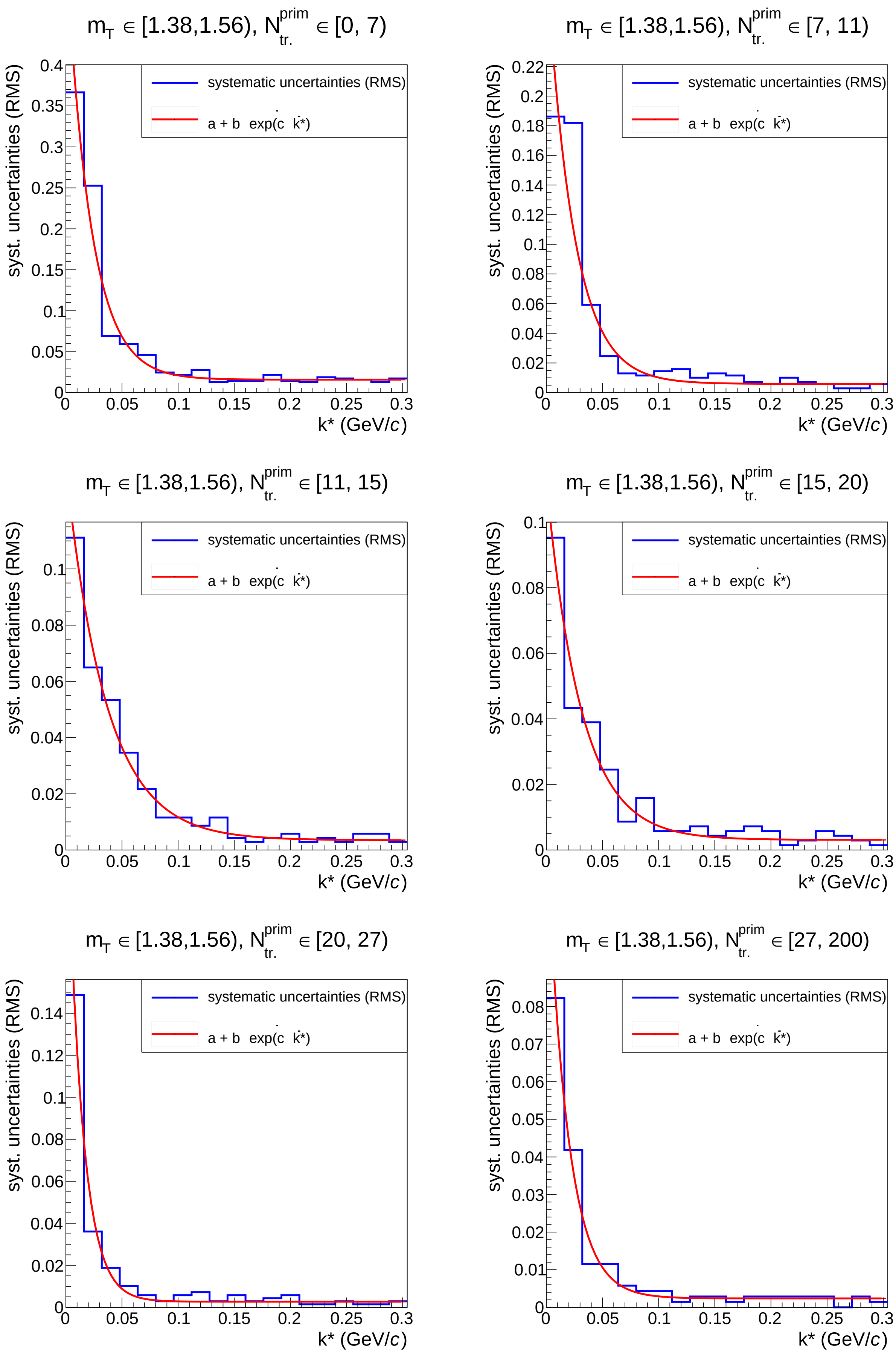


Figure 9.27: Systematic uncertainties of the p–p correlation functions for all multiplicity intervals in the second largest m_T interval ($m_T \in [1.38\,\text{GeV}/c,\ 1.56\,\text{GeV}/c)$)

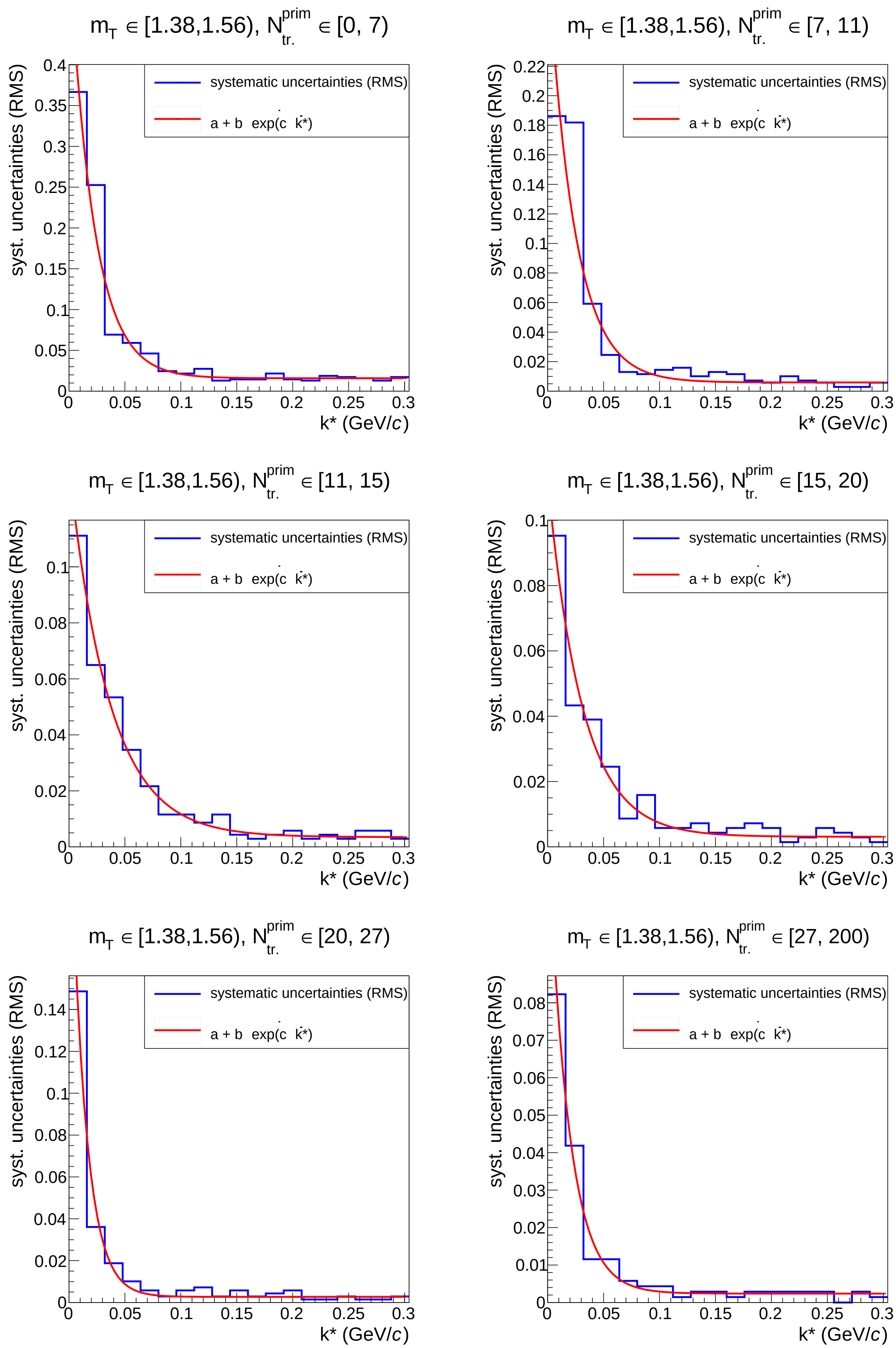


Figure 9.28: Systematic uncertainties of the p–p correlation functions for all multiplicity intervals in the second largest m_T interval ($m_T \in [1.86\,\text{GeV}/c,\ 2.21\,\text{GeV}/c)$)

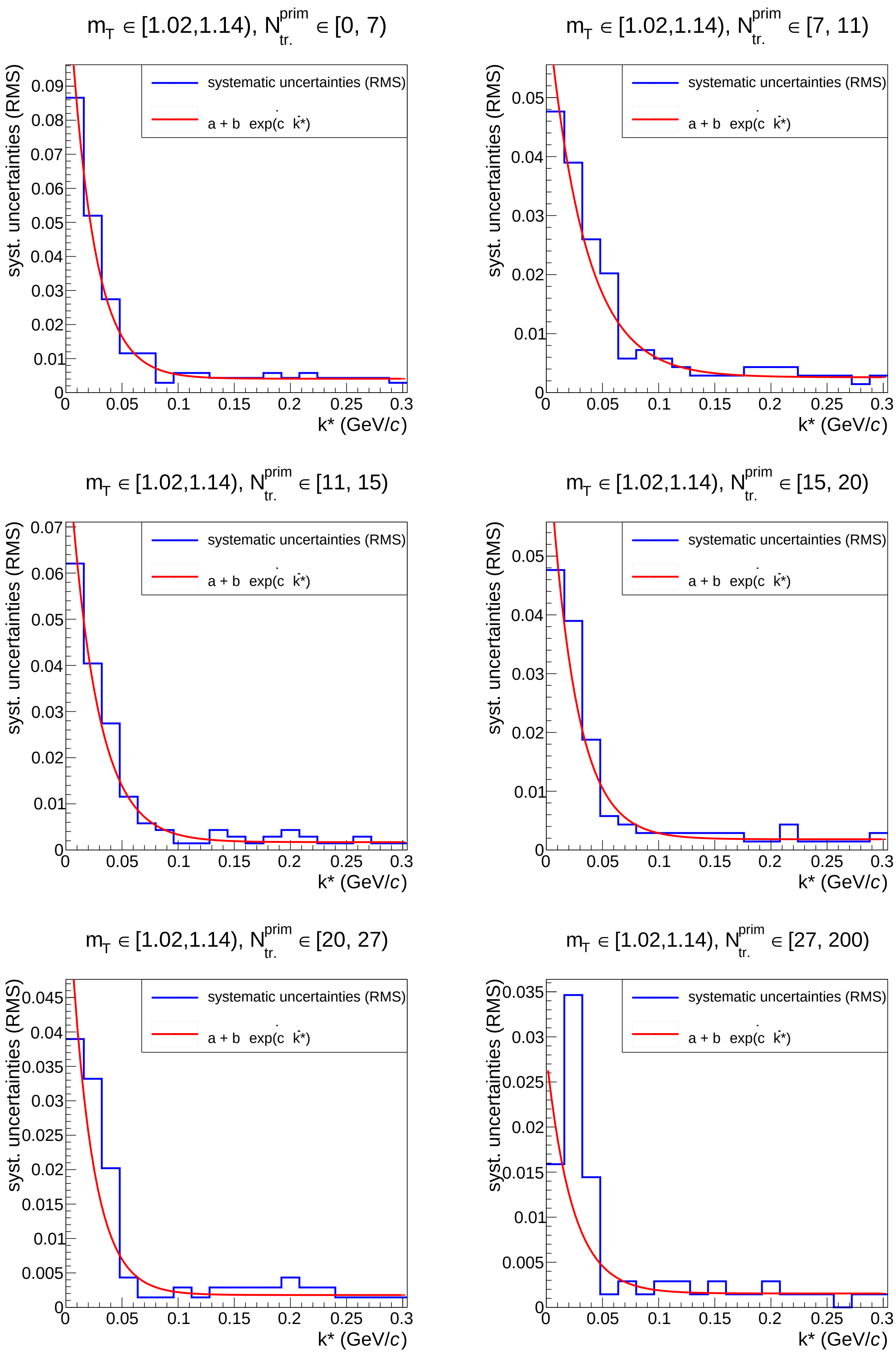


Figure 9.29: Systematic uncertainties of the $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ correlation functions for all multiplicity intervals in the second largest m_{T} interval ($m_{\mathrm{T}} \in [1.02\,\mathrm{GeV}/c,\ 1.14\,\mathrm{GeV}/c)$)

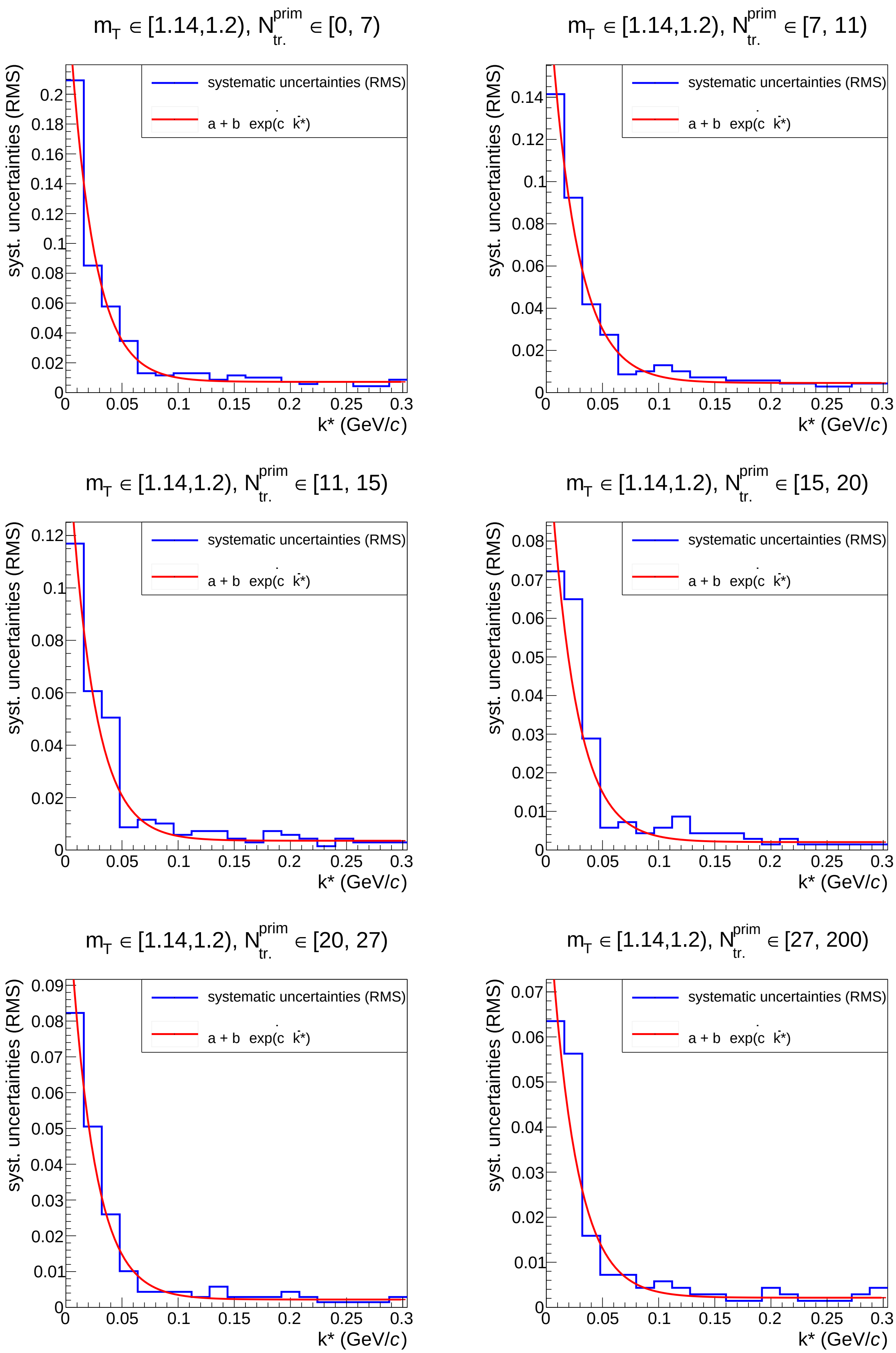


Figure 9.30: Systematic uncertainties of the $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ correlation functions for all multiplicity intervals in the second largest m_{T} interval ($m_{\mathrm{T}} \in [1.14\,\mathrm{GeV}/c,\ 1.20\,\mathrm{GeV}/c)$)

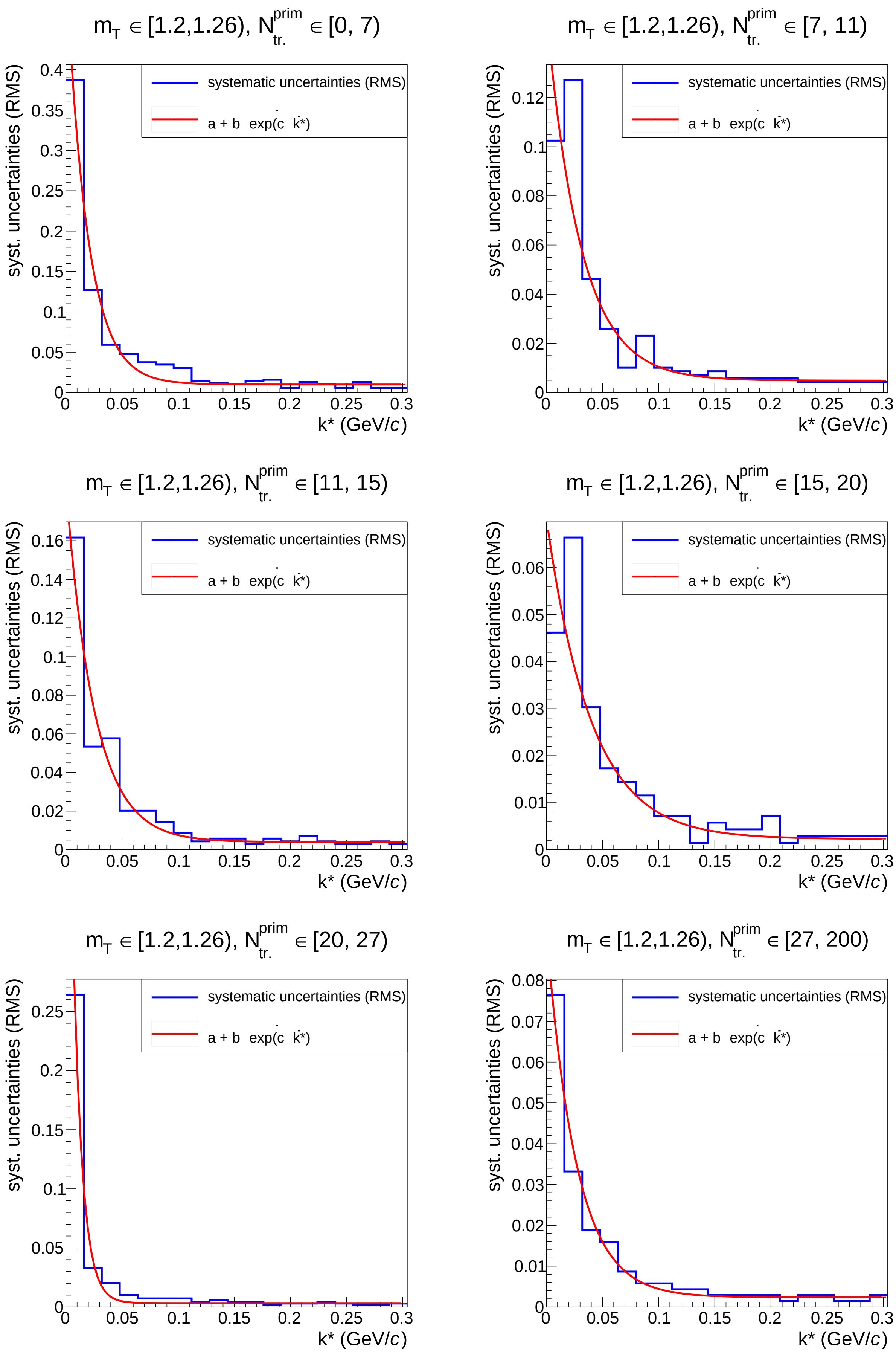


Figure 9.31: Systematic uncertainties of the $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ correlation functions for all multiplicity intervals in the second largest m_{T} interval ($m_{\mathrm{T}} \in [1.20\,\mathrm{GeV}/c,\ 1.26\,\mathrm{GeV}/c)$)

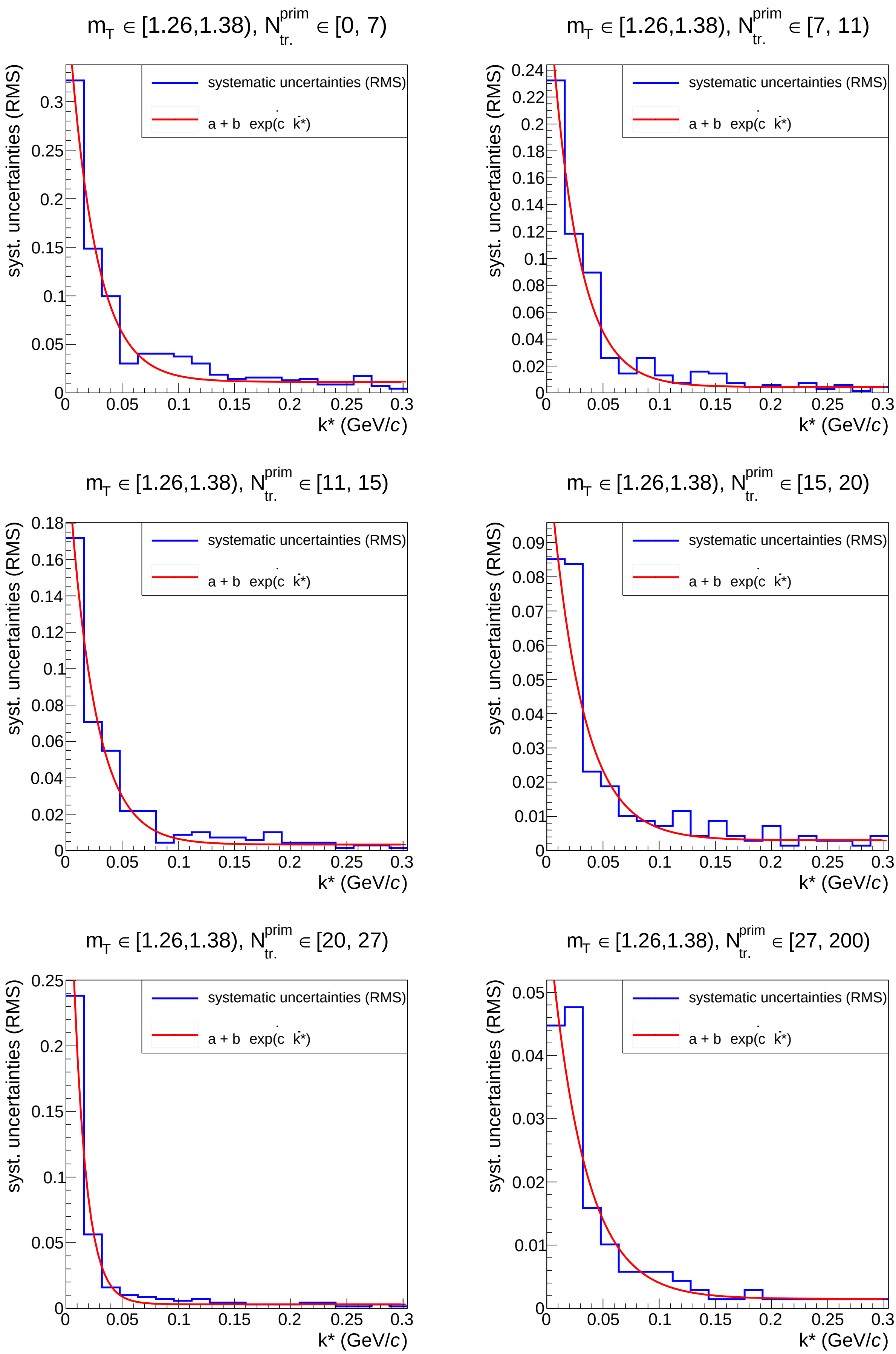


Figure 9.32: Systematic uncertainties of the $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ correlation functions for all multiplicity intervals in the second largest m_T interval ($m_\mathrm{T} \in [1.26\,\mathrm{GeV}/c,\ 1.38\,\mathrm{GeV}/c)$)

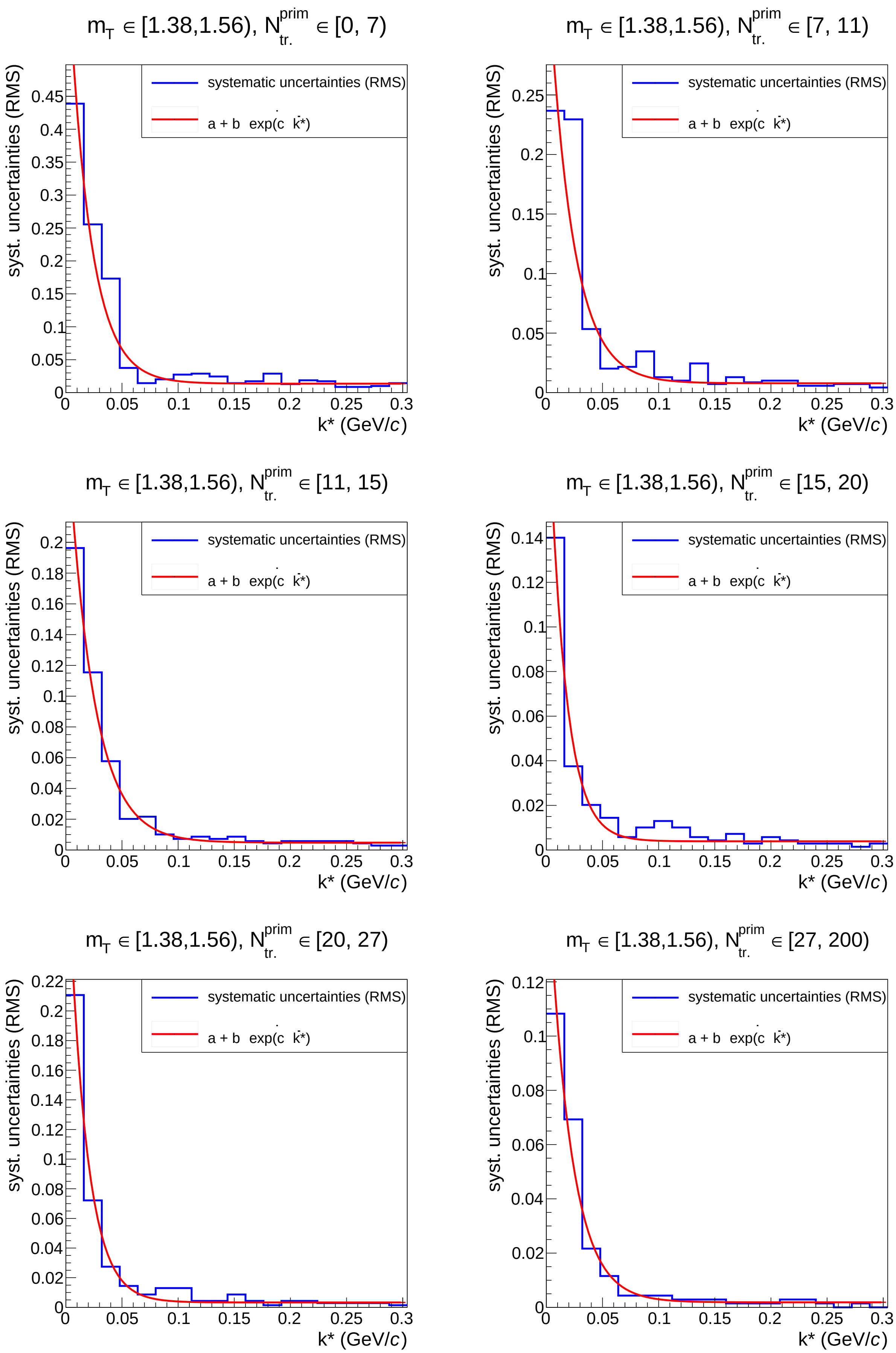


Figure 9.33: Systematic uncertainties of the $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ correlation functions for all multiplicity intervals in the second largest m_{T} interval ($m_{\mathrm{T}} \in [1.38\,\mathrm{GeV}/c,\ 1.56\,\mathrm{GeV}/c)$)

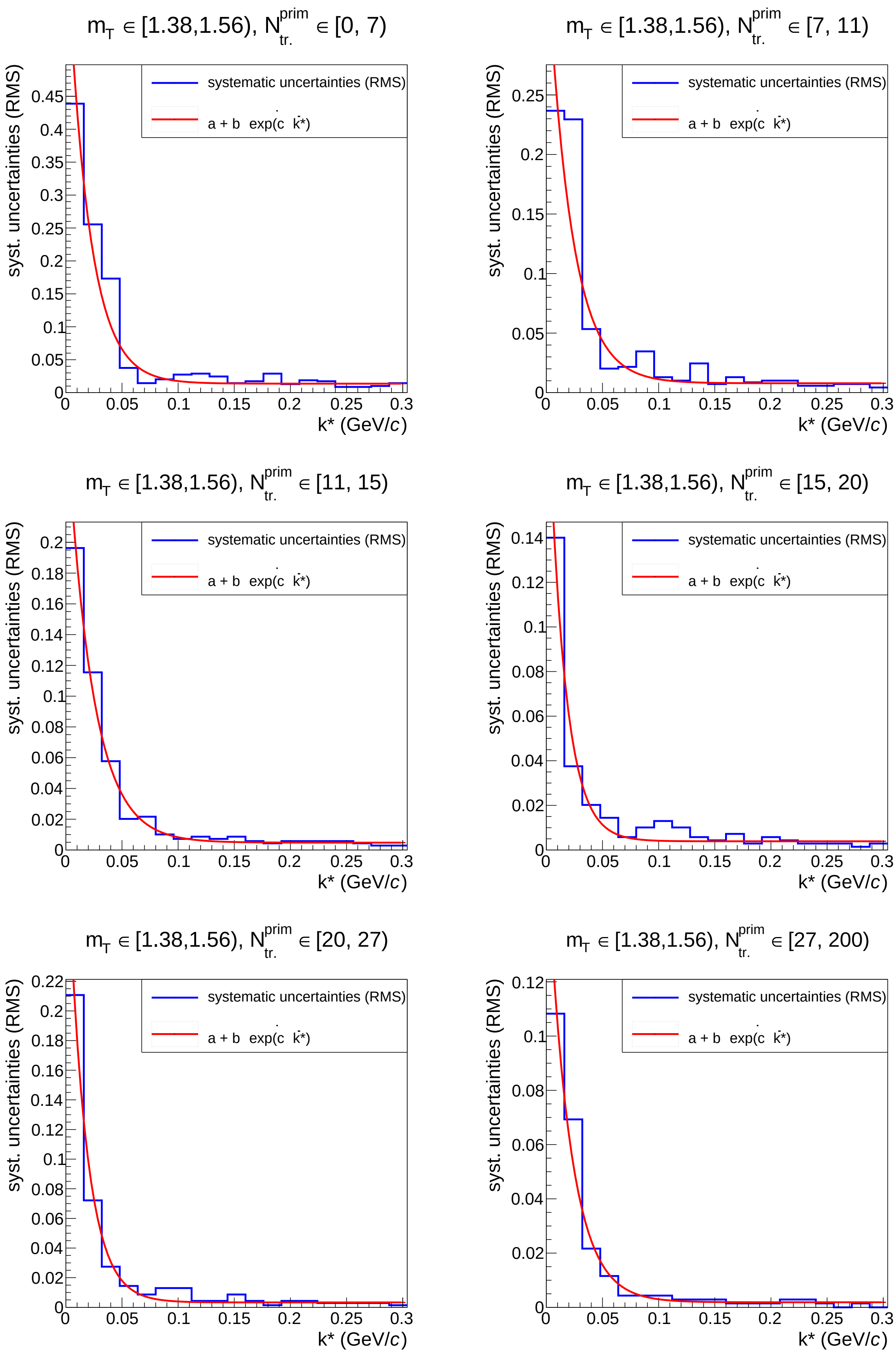


Figure 9.34: Systematic uncertainties of the $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$ correlation functions for all multiplicity intervals in the second largest m_{T} interval ($m_{\mathrm{T}} \in [1.86\,\mathrm{GeV}/c,\ 2.21\,\mathrm{GeV}/c)$)

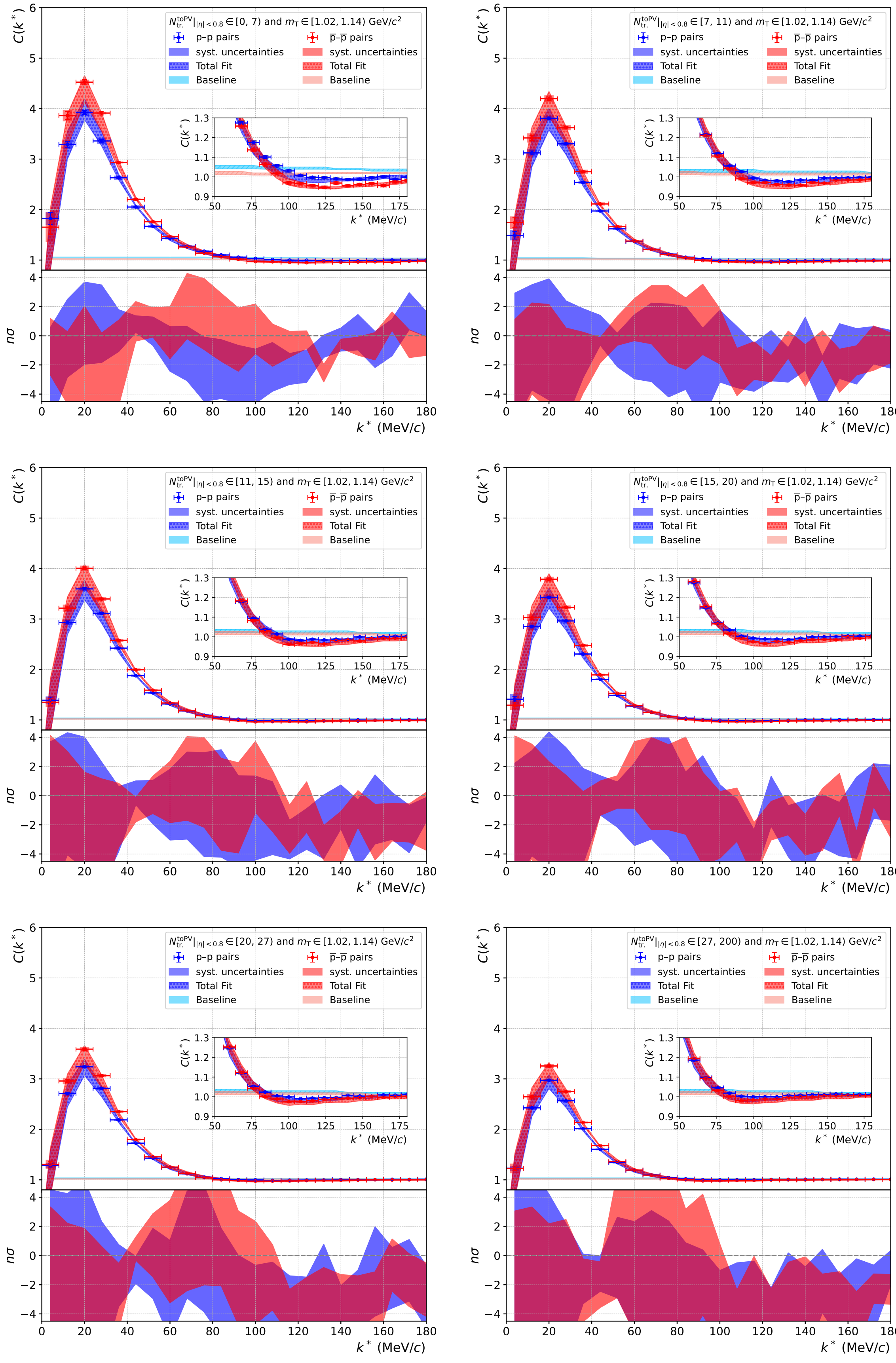


Figure 9.35: Correlation functions for the second largest m_T interval and all multiplicity intervals. The blue plots correspond to the p–p pairs, the red plots to $\overline{\text{p}}$–$\overline{\text{p}}$.

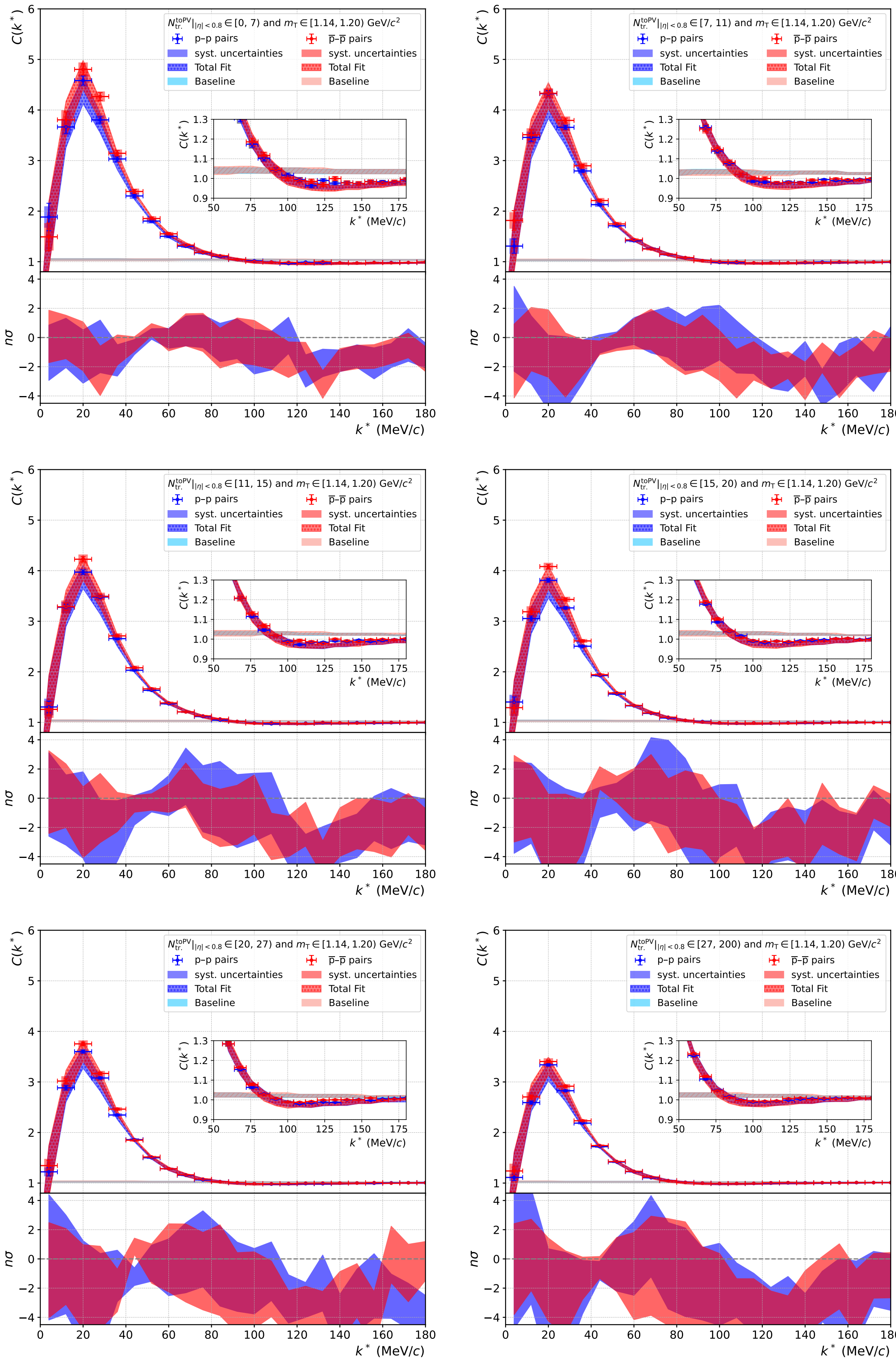


Figure 9.36: Correlation functions for the second largest m_{T} interval and all multiplicity intervals. The blue plots correspond to the p–p pairs, the red plots to $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$.

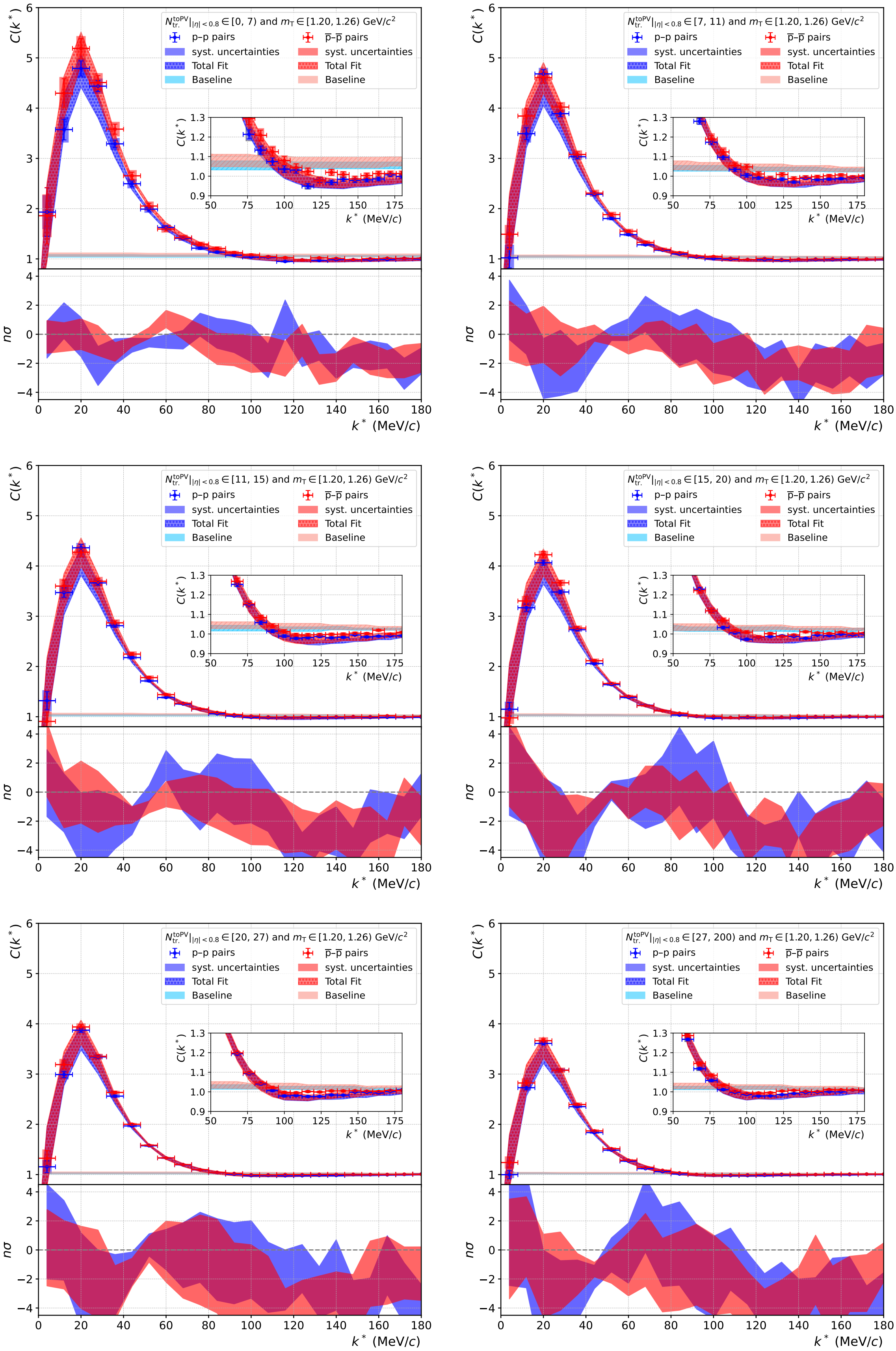


Figure 9.37: Correlation functions for the second largest m_{T} interval and all multiplicity intervals. The blue plots correspond to the p–p pairs, the red plots to $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$.

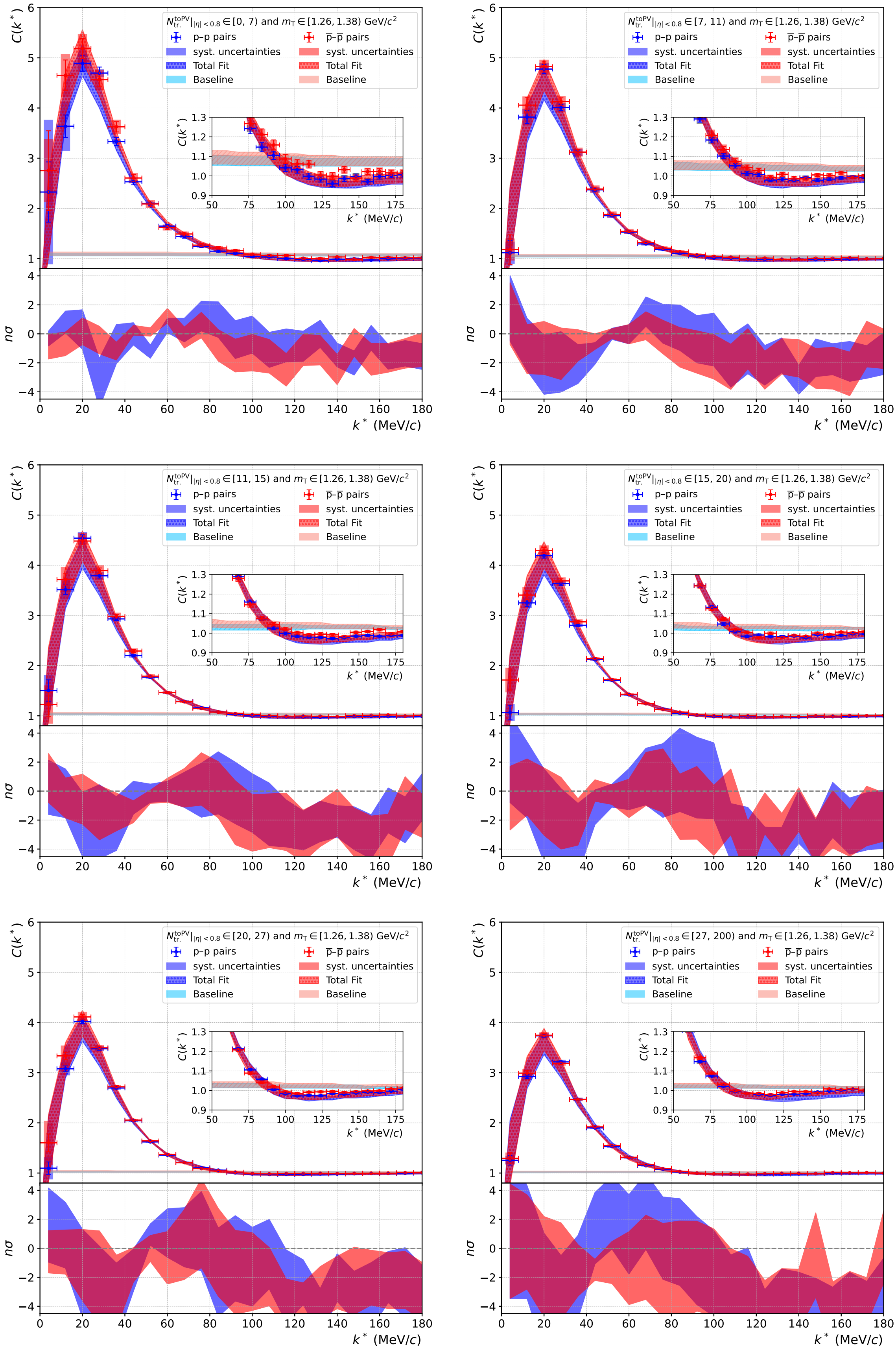


Figure 9.38: Correlation functions for the second largest m_T interval and all multiplicity intervals. The blue plots correspond to the p–p pairs, the red plots to $\overline{\text{p}}$–$\overline{\text{p}}$.

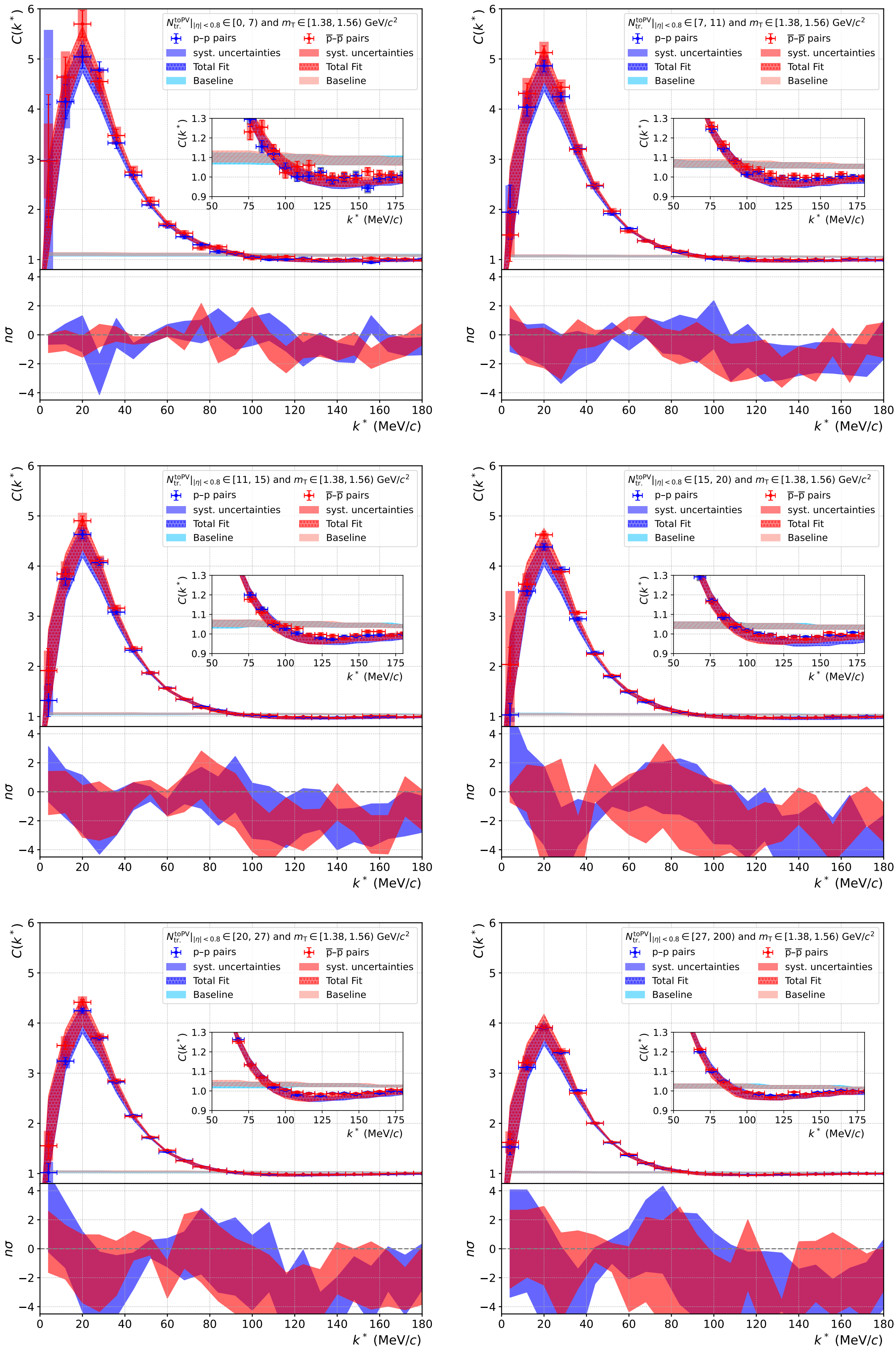


Figure 9.39: Correlation functions for the second largest m_T interval and all multiplicity intervals. The blue plots correspond to the p–p pairs, the red plots to $\overline{p}$–$\overline{p}$.

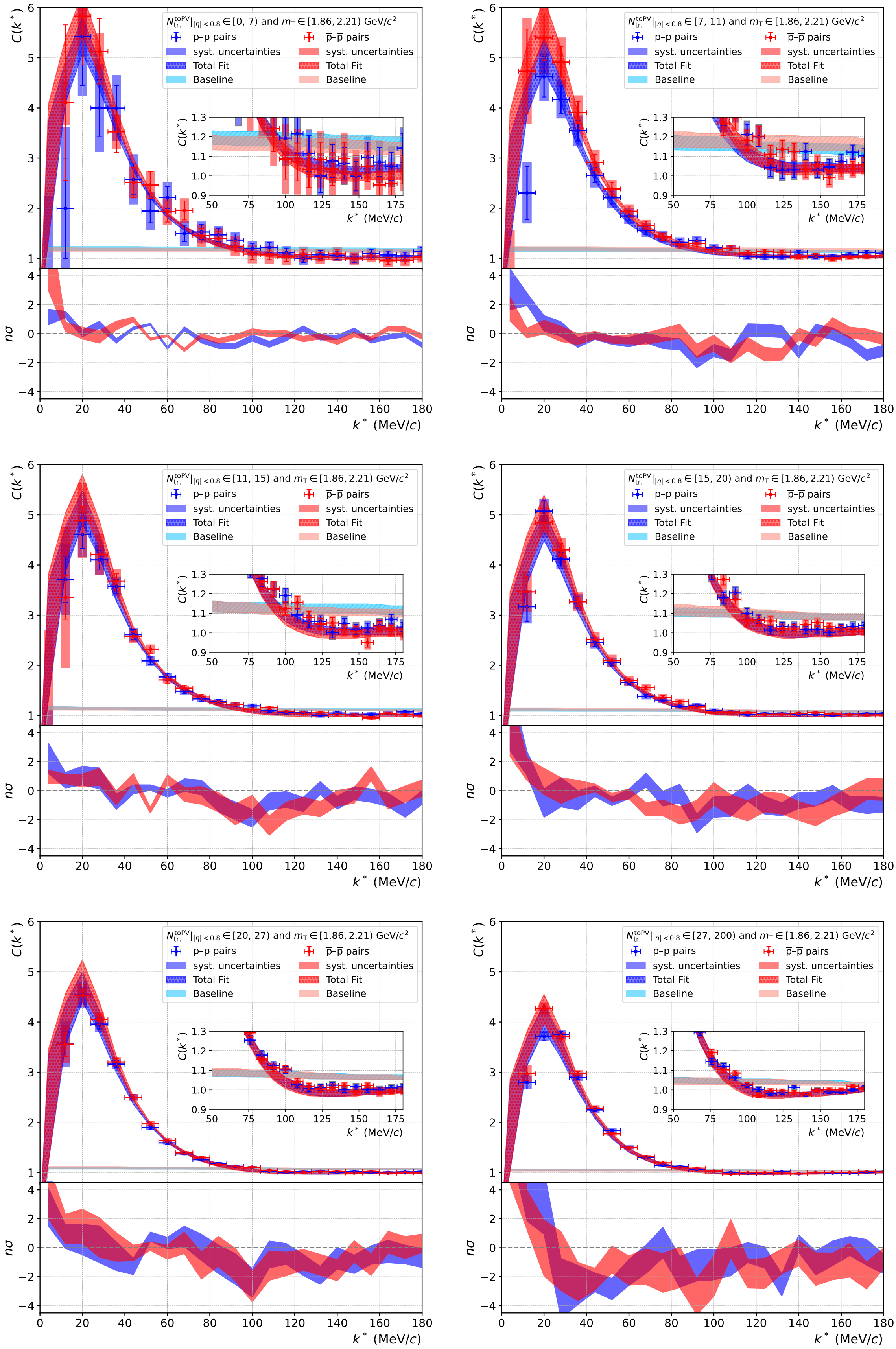


Figure 9.40: Correlation functions for the second largest m_{T} interval and all multiplicity intervals. The blue plots correspond to the p–p pairs, the red plots to $\overline{\mathrm{p}}$–$\overline{\mathrm{p}}$.

Acknowledgments

This marks a milestone in the journey that I started in 2017, by enrolling at TUM to study physics. It was an incredible way up to this point and I am infinitely grateful for all the people the supported and accompanied me. First of all, I want to thank Prof. Laura Fabbietti for her supervision and the many opportunities she offered me in her group over the last years. Her ideas and guidance made this thesis possible and shaped my first steps in physics. This brings me to my direct supervisor and collaborator: Anton Riedel and Jaime Gonzalez, proud members of the pp-task-force. Thanks for all the discussions and the times we spent together trying to make this analysis happen. I learned a lot from you and it was always a charm working together. The same goes for every single other member of the Dense and Strange Hadronic Matter group and especially for Dimitar, Valentina, Laura and Raffa for all their help and encouragements. I want to dedicate my warmest gratitude to Niko and Vincent, who have been my friends from the very first day at university. I cannot think about the past years without also thinking of you and every time I do, it paints a big smile on my face. And finally, all my love goes to my parents, my sister and my brother, and Karim, my longest friend from childhood. You all supported me always and in every way possible way .

List of Figures

List of Tables

Bibliography

[1] R. Hanbury Brown and R. Twiss. "A Test of a new type of stellar interferometer on Sirius". In: *Nature* 178 (1956), pp. 1046–1048. DOI: 10.1038/1781046a0.

[2] R. H. Brown and R. Twiss. "LXXIV. A new type of interferometer for use in radio astronomy". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 45.366 (1954), pp. 663–682. DOI: 10.1080/14786440708520475. eprint: https://doi.org/10.1080/14786440708520475. URL: https://doi.org/10.1080/14786440708520475.

[3] G. Goldhaber, S. Goldhaber, W.-Y. Lee, and A. Pais. "Influence of Bose-Einstein statistics on the anti-proton proton annihilation process". In: *Phys. Rev.* 120 (1960). Ed. by R. M. Weiner, pp. 300–312. DOI: 10.1103/PhysRev.120.300.

[4] M. A. Lisa, S. Pratt, R. Soltz, and U. Wiedemann. "FEMTOSCOPY IN RELATIVISTIC HEAVY ION COLLISIONS: Two Decades of Progress". In: *Annual Review of Nuclear and Particle Science* 55.1 (2005), pp. 357–402. DOI: 10.1146/annurev.nucl.55.090704.151533. eprint: https://doi.org/10.1146/annurev.nucl.55.090704.151533. URL: https://doi.org/10.1146/annurev.nucl.55.090704.151533.

[5] "The ALICE experiment – A journey through QCD". In: (Nov. 2022). arXiv: 2211.04384 [nucl-ex].

[6] E. Shuryak. "Why does the quark gluon plasma at RHIC behave as a nearly ideal fluid?" In: *Prog. Part. Nucl. Phys.* 53 (2004). Ed. by A. Faessler, pp. 273–303. DOI: 10.1016/j.ppnp.2004.02.025. arXiv: hep-ph/0312227.

[7] J. E. Parkkila, A. Onnerstad, S. F. Taghavi, C. Mordasini, A. Bilandzic, M. Virta, and D. J. Kim. "New constraints for QCD matter from improved Bayesian parameter estimation in heavy-ion collisions at LHC". In: *Phys. Lett. B* 835 (2022), p. 137485. DOI: 10.1016/j.physletb.2022.137485. arXiv: 2111.08145 [hep-ph].

[8] S. Acharya et al. "p-p, p-Λ and Λ-Λ correlations studied via femtoscopy in pp reactions at $\sqrt{s} = 7$ TeV". In: *Phys. Rev. C* 99.2 (2019), p. 024001. DOI: 10.1103/PhysRevC.99.024001. arXiv: 1805.12455 [nucl-ex].

[9] S. Acharya et al. "Search for a common baryon source in high-multiplicity pp collisions at the LHC". In: *Phys. Lett. B* 811 (2020), p. 135849. DOI: 10.1016/j.physletb.2020.135849. arXiv: 2004.08018 [nucl-ex].

[10] S. Acharya et al. "Common femtoscopic hadron-emission source in pp collisions at the LHC". In: (Nov. 2023). arXiv: 2311.14527 [hep-ph].

[11] L. Tolos and L. Fabbietti. "Strangeness in Nuclei and Neutron Stars". In: *Prog. Part. Nucl. Phys.* 112 (2020), p. 103770. DOI: 10.1016/j.ppnp.2020.103770. arXiv: 2002.09223 [nucl-ex].

[12] A. Collaboration et al. "Unveiling the strong interaction among hadrons at the LHC". In: *Nature* 588 (2020). [Erratum: Nature 590, E13 (2021)], pp. 232–238. DOI: 10.1038/s41586-020-3001-6. arXiv: 2005.11495 [nucl-ex].

[13] S. Acharya et al. "First Observation of an Attractive Interaction between a Proton and a Cascade Baryon". In: *Phys. Rev. Lett.* 123.11 (2019), p. 112002. DOI: 10.1103/PhysRevLett.123.112002. arXiv: 1904.12198 [nucl-ex].

[14] S. Acharya et al. "First measurement of the Λ-Ξ interaction in proton-proton collisions at the LHC". In: *Phys. Lett. B* 844 (2023), p. 137223. DOI: 10.1016/j.physletb.2022.137223. arXiv: 2204.10258 [nucl-ex].

[15] S. Acharya et al. "Exploring the NΛ–NΣ coupled system with high precision correlation techniques at the LHC". In: *Phys. Lett. B* 833 (2022), p. 137272. DOI: 10.1016/j.physletb.2022.137272. arXiv: 2104.04427 [nucl-ex].

[16] S. Acharya et al. "Accessing the strong interaction between Λ baryons and charged kaons with the femtoscopy technique at the LHC". In: *Phys. Lett. B* 845 (2023), p. 138145. DOI: 10.1016/j.physletb.2023.138145. arXiv: 2305.19093 [nucl-ex].

[17] S. Acharya et al. "Experimental Evidence for an Attractive p-ϕ Interaction". In: *Phys. Rev. Lett.* 127.17 (2021), p. 172301. DOI: 10.1103/PhysRevLett.127.172301. arXiv: 2105.05578 [nucl-ex].

[18] L. Fabbietti, V. Mantovani Sarti, and O. Vazquez Doce. "Study of the Strong Interaction Among Hadrons with Correlations at the LHC". In: *Ann. Rev. Nucl. Part. Sci.* 71 (2021), pp. 377–402. DOI: 10.1146/annurev-nucl-102419-034438. arXiv: 2012.09806 [nucl-ex].

[19] S. Acharya et al. "First study of the two-body scattering involving charm hadrons". In: *Phys. Rev. D* 106.5 (2022), p. 052010. DOI: 10.1103/PhysRevD.106.052010. arXiv: 2201.05352 [nucl-ex].

[20] S. Acharya et al. "Towards the understanding of the genuine three-body interaction for p–p–p and p–p–Λ". In: *Eur. Phys. J. A* 59.7 (2023), p. 145. DOI: 10.1140/epja/s10050-023-00998-6. arXiv: 2206.03344 [nucl-ex].

[21] U. Heinz and B. V. Jacak. "TWO-PARTICLE CORRELATIONS IN RELATIVISTIC HEAVY-ION COLLISIONS". In: *Annual Review of Nuclear and Particle Science* 49.1 (), pp. 529–579. DOI: 10.1146/annurev.nucl.49.1.529. eprint: https://doi.org/10.1146/annurev.nucl.49.1.529. URL: https://doi.org/10.1146/annurev.nucl.49.1.529.

[22] T. Pierog, I. Karpenko, J. M. Katzy, E. Yatsenko, and K. Werner. "EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider". In: *Phys. Rev. C* 92 (3 Sept. 2015), p. 034906. DOI: 10.1103/PhysRevC.92.034906. URL: https://link.aps.org/doi/10.1103/PhysRevC.92.034906.

[23] D. L. Mihaylov, V. M. Sarti, O. W. Arnold, L. Fabbietti, B. Hohlweger, and A. M. Mathis. "A femtoscopic correlation analysis tool using the Schrödinger equation (CATS)". In: *The European Physical Journal C* 78 (5 Feb. 23, 2018), p. 394. DOI: 10.1140/epjc/s10052-018-5859-0.

[24] C. P. Office. *LHC- The guide- faq*. URL: https://cds.cern.ch/record/2809109/files/CERN-Brochure-2021-004-Eng.pdf (visited on 11/19/2023).

[25] "Vol. 10 (2020): High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1". In: 10 (2020). Ed. by I. Béjar Alonso, O. Brüning, P. Fessia, M. Lamont, L. Rossi, and L. Tavian. DOI: 10.23731/CYRM-2020-0010.

[26] W. Busza, K. Rajagopal, and W. van der Schee. "Heavy Ion Collisions: The Big Picture and the Big Questions". In: *Annual Review of Nuclear and Particle Science* 68.1 (2018), pp. 339–376. DOI: 10.1146/annurev-nucl-101917-020852. eprint: https://doi.org/10.1146/annurev-nucl-101917-020852. URL: https://doi.org/10.1146/annurev-nucl-101917-020852.

[27] "ALICE upgrades during the LHC Long Shutdown 2". In: (Feb. 2023). arXiv: 2302.01238 [physics.ins-det].

[28] K. Aamodt et al. "The ALICE experiment at the CERN LHC". In: *JINST* 3 (2008), S08002. DOI: 10.1088/1748-0221/3/08/S08002.

[29] "Letter of intent for ALICE 3: A next-generation heavy-ion experiment at the LHC". In: (Nov. 2022). arXiv: 2211.02491 [physics.ins-det].

[30] M. Mager. "ALPIDE, the Monolithic Active Pixel Sensor for the ALICE ITS upgrade". In: *Nucl. Instrum. Meth. A* 824 (2016). Ed. by M. G. Bisogni, M. Grassi, M. Incagli, R. Paoletti, and G. Signorelli, pp. 434–438. DOI: 10.1016/j.nima.2015.09.057.

[31] G. Aglieri Rinella. "The ALPIDE pixel sensor chip for the upgrade of the ALICE Inner Tracking System". In: *Nucl. Instrum. Meth. A* 845 (2017). Ed. by G. Badurek, T. Bergauer, M. Dragicevic, M. Friedl, M. Jeitler, M. Krammer, J. Schieck, and C. Schwanda, pp. 583–587. DOI: 10.1016/j.nima.2016.05.016.

[32] J. A. et al. "The ALICE TPC, a large 3-dimensional tracking device with fast readout for ultra-high multiplicity events". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 622.1 (Jan. 20, 2010), pp. 316–367. URL: https://www.sciencedirect.com/science/article/pii/S0168900210008910.

[33] A. Deisting, C. Garabatos, A. Szabo, and D. Vranic. "Measurements of ion mobility in argon and neon based gas mixtures". In: *Nucl. Instrum. Meth. A* 845 (2017). Ed. by G. Badurek, T. Bergauer, M. Dragicevic, M. Friedl, M. Jeitler, M. Krammer, J. Schieck, and C. Schwanda, pp. 215–217. DOI: 10.1016/j.nima.2016.06.093. arXiv: 1603.07638 [physics.ins-det].

[34] "Technical Design Report for the Upgrade of the ALICE Time Projection Chamber". In: (Mar. 2014).

[35] J. Adolfsson et al. "The upgrade of the ALICE TPC with GEMs and continuous readout". In: *JINST* 16.03 (2021), P03022. DOI: 10.1088/1748-0221/16/03/P03022. arXiv: 2012.09518 [physics.ins-det].

[36] F. Sauli. "GEM: A new concept for electron amplification in gas detectors". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 386.2 (1997), pp. 531–534. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(96)01172-2. URL: https://www.sciencedirect.com/science/article/pii/S0168900296011722.

[37] R. Veenhof. *Garfield - simulation of gaseous detectors*. URL: https://garfield.web.cern.ch/garfield/.

[38] AliceO2Group. *O2 Analysis Framework Documentation*. URL: https://aliceo2group.github.io/analysis-framework/docs/basics-tasks/Introduction.html (visited on 11/22/2023).

[39] AliceO2Group. *O2 Physics Framework*. URL: https://github.com/AliceO2Group/O2Physics (visited on 04/24/2023).

[40] AliceO2Group. *O2 Physics Framework - FemtoDream*. URL: https://github.com/AliceO2Group/O2Physics/tree/master/PWGCF/FemtoDream (visited on 04/24/2023).

[41] D. L. Mihaylov. "Analysis techniques for femtoscopy and correlation studies in small collision systems and their applications to the investigation of proton-Lambda and Lambda-Lambda interactions with ALICE". en. PhD thesis. Technische Universität München, 2021, p. 217.

[42] R. B. Wiringa, V. G. J. Stoks, and R. Schiavilla. "Accurate nucleon-nucleon potential with charge-independence breaking". In: *Phys. Rev. C* 51 (1 Jan. 1995), pp. 38–51. DOI: 10.1103/PhysRevC.51.38. URL: https://link.aps.org/doi/10.1103/PhysRevC.51.38.

[43] A. Mathis, L. Fabbietti, B. Hohlweger, V. M. Sarti, D. Mihaylov, and O. V. Doce. "p-p, p-Λ and Λ-Λ correlations studied via femtoscopy in pp reactions at $\sqrt{s}$ = 13 TeV". In: *ALICE analysis note* (2018). URL: https://alice-notes.web.cern.ch/node/738.

[44] J. Haidenbauer, U. .-. Meißner, and A. Nogga. "Hyperon–nucleon interaction within chiral effective field theory revisited". In: *Eur. Phys. J. A* 56.3 (2020), p. 91. DOI: 10.1140/epja/s10050-020-00100-4. arXiv: 1906.11681 [nucl-th].

[45] D. Mihaylov and J. González González. "Novel model for particle emission in small collision systems". In: *Eur. Phys. J. C* 83.7 (2023), p. 590. DOI: 10.1140/epjc/s10052-023-11774-7. arXiv: 2305.08441 [hep-ph].

[46] J. Adam et al. "Enhanced production of multi-strange hadrons in high-multiplicity proton-proton collisions". In: *Nature Phys.* 13 (2017), pp. 535–539. DOI: 10.1038/nphys4111. arXiv: 1606.07424 [nucl-ex].

[47] 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". In: *Phys. Rev. Lett.* 116.17 (2016), p. 172301. DOI: 10.1103/PhysRevLett.116.172301. arXiv: 1509.04776 [hep-ex].

[48] A. Kisiel, M. Gałażyn, and P. Bożek. "Pion, kaon, and proton femtoscopy in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}}$=2.76 TeV modeled in (3+1)D hydrodynamics". In: *Phys. Rev. C* 90.6 (2014), p. 064914. DOI: 10.1103/PhysRevC.90.064914. arXiv: 1409.4571 [nucl-th].

[49] W. Zhao, Y. Zhou, K. Murase, and H. Song. "Searching for small droplets of hydrodynamic fluid in proton–proton collisions at the LHC". In: *Eur. Phys. J. C* 80.9 (2020), p. 846. DOI: 10.1140/epjc/s10052-020-8376-x. arXiv: 2001.06742 [nucl-th].