

Università di Torino

Tesi di Dottorato

**Proton-proton femtoscopy in $\Upsilon(nS)$
decays at Belle**

Alessandro Boschetti

Advisors:

Prof. Stefano Giovanni Spataro

Dr. Umberto Tamponi

XVIII cycle

February 13, 2026

Abstract

The study of hadron production in high-energy collisions provides access to the non-perturbative regime of Quantum Chromodynamics, where confinement and hadronization govern the formation of final-state particles. Among the various tools developed to probe this regime, two-particle femtoscopy has proven particularly powerful, enabling the extraction of spatial and temporal scales associated with particle emission through correlations at low relative momentum. While femtoscopic techniques have been extensively applied in heavy-ion and proton–proton collisions, their use in electron–positron annihilation remains comparatively unexplored. The clean experimental environment provided by e^+e^- colliders, together with the well-constrained nature of the nucleon–nucleon interaction, offers an ideal setting in which to study the space–time structure of hadronization without the complicating influence of collective effects or final-state rescattering in the medium.

This thesis presents a detailed femtoscopic analysis of proton–proton pairs produced in e^+e^- annihilation at the $\Upsilon(1S)$ and $\Upsilon(2S)$ resonances, making use of the high-statistics datasets collected by the Belle experiment at KEK. The proton–proton system is particularly advantageous for such studies: its final-state interaction, governed by well-established strong and Coulomb potentials and modified by Fermi–Dirac statistics, acts as a calibrated probe of the emission source. This allows the correlation function to be interpreted primarily in terms of the geometry of hadron production. By exploiting a combination of event-mixing techniques, background suppression strategies, and state-of-the-art theoretical modeling of the interaction kernel, the analysis extracts effective source radii as a function of the pair momentum and event category.

The results provide new experimental constraints on the space–time structure of hadronization in Υ decays. The measured correlation functions exhibit clear signatures of the proton–proton final-state interaction, and their comparison with theoretical calculations enables the determination of source sizes in the femtometer range. The dependence of the extracted radii on the kinematics and production environment sheds light on the mechanism by which baryons are formed in e^+e^- collisions, offering insights complementary to those obtained from meson femtoscopy and from studies in hadronic and nuclear collisions. In addition, the analysis contributes to the broader program of baryon–baryon correlation studies, providing a benchmark against which less precisely known interaction channels may be calibrated.

Overall, this work demonstrates the potential of femtoscopic techniques in e^+e^- annihilation and establishes proton–proton correlations as a sensitive probe of the hadron emission source in the decay of heavy quarkonium states. The methods and results presented here lay the groundwork for future measurements at Belle II, where significantly larger datasets will enable more differential and higher-precision studies of baryon production and strong interactions at the femtometer scale.

Sintesi

Lo studio della produzione adronica nelle collisioni ad alta energia fornisce accesso al regime non perturbativo della Cromodinamica Quantistica, in cui il confinamento e l'adronizzazione governano la formazione delle particelle nello stato finale. Tra gli strumenti sperimentali sviluppati per indagare queste dinamiche, la femtoscopia ha assunto un ruolo di particolare rilievo, consentendo di estrarre le scale spaziali e temporali del processo di emissione tramite correlazioni a basso momento relativo. Sebbene tale tecnica sia stata ampiamente applicata nelle collisioni protone-protone e di ioni pesanti, il suo utilizzo nell'annichilazione elettrone-positrone è rimasto finora relativamente poco esplorato. L'ambiente sperimentale estremamente pulito offerto dai collisionatori e^+e^- , unito alla buona conoscenza dell'interazione nucleone-nucleone, rappresenta un contesto ideale per studiare la struttura spazio-tempo dell'adronizzazione, priva degli effetti collettivi e delle interazioni di medio raggio presenti in sistemi più complessi.

In questa tesi viene presentata un'analisi femtosopica dettagliata delle coppie protone-protone prodotte in annichilazioni e^+e^- alle risonanze $\Upsilon(1S)$ e $\Upsilon(2S)$, utilizzando gli ampi campioni di dati raccolti dall'esperimento Belle presso il centro di ricerca KEK. Il sistema $p - p$ risulta particolarmente vantaggioso poiché la sua interazione di stato finale –che combina la repulsione coulombiana, il potenziale forte e le simmetrie imposte dalla statistica di Fermi-Dirac– può essere modellizzata con precisione, permettendo di interpretare la correlazione sperimentale principalmente in funzione della distribuzione della sorgente di emissione. Attraverso tecniche di event mixing, strategie di soppressione dei fondi e l'impiego di modelli teorici aggiornati per la descrizione dell'interazione a due corpi, l'analisi determina i raggi sorgente efficaci come funzione del momento della coppia e della categoria di evento.

I risultati ottenuti forniscono nuovi vincoli sperimentali sulla struttura spazio-tempo dell'adronizzazione nei decadimenti delle risonanze Υ . Le funzioni di correlazione misurate mostrano con chiarezza l'impronta dell'interazione finale protone-protone; il confronto con i calcoli teorici consente di estrarre le dimensioni della sorgente, dell'ordine del femtometro. La dipendenza del raggio della sorgente da variabili cinematiche e caratteristiche dell'evento offre indicazioni rilevanti sui meccanismi di formazione dei barioni nelle collisioni e^+e^- , integrando i risultati derivati da misure femtoscopiche di mesoni e dalle collisioni adroniche e nucleari.

Nel complesso, questo lavoro dimostra l'efficacia della femtoscopia nell'annichilazione elettrone-positrone e stabilisce le correlazioni protone-protone come un sensibile strumento per sondare la sorgente di emissione adronica nei decadimenti degli stati di quarkonio pesante. I metodi e i risultati presentati gettano le basi per future misure presso Belle II, dove dataset significativamente più estesi permetteranno studi più differenziati e di maggiore precisione sulle interazioni forti e sulla formazione dei barioni alla scala del femtometro.

Contents

1	Introduction	1
1.1	The landscape of hadron physics in the non-perturbative QCD regime . .	1
1.2	A clean environment for hadron studies	2
1.3	Dynamics of the proton-proton system	3
1.4	Evolution of the femtoscopy technique	4
2	Strong interaction models and hadronization	7
2.1	Modeling strong interactions between hadrons	8
2.1.1	Potential models	9
2.1.2	Effective field theories	11
2.1.3	Coupled-channel dynamics	12
2.1.4	Lattice QCD	13
2.2	Hadronization	14
2.2.1	Lund string model	14
2.2.2	Cluster model	16
2.2.3	Statistical hadronization	17
2.2.4	Coalescence model	20
2.3	Hadronic decays of the narrow $\Upsilon(nS)$ states	22
2.4	The femtoscopy technique	23
2.4.1	Hanbury Brown–Twiss intensity interferometry	23
2.4.2	Two-particle correlation function	24
2.4.3	Hanbury Brown–Twiss radii	25
2.4.4	Koonin–Pratt formalism	27
2.4.5	Nuclear scattering parameters	29
2.4.6	Extracting parameters from the correlation function	30
2.4.7	The CATS framework	31
2.4.8	Using femtoscopy to decipher the nature of exotic hadrons	32
2.4.9	Recent experimental results	33
3	The Belle experiment	41
3.1	The KEKB B-factory	42
3.2	The Belle detector	43
3.2.1	Silicon Vertex Detector	45
3.2.2	Central Drift Chamber	45
3.2.3	Time-of-Flight detector	48
3.2.4	Aerogel Cherenkov Counter	50
3.2.5	Electromagnetic Calorimeter	51
3.2.6	K_L and muon detector	52

3.2.7	Trigger system	52
3.3	Data production	55
4	Software framework	57
4.1	Overview	57
4.2	Belle and Belle II software ecosystems	57
4.2.1	The Belle data format and legacy software	57
4.2.2	The basf2 framework	58
4.3	Motivation for a unified framework	58
4.4	The B2BII conversion chain	59
4.4.1	Motivation and role of B2BII	59
4.4.2	Concept and architecture of B2BII	59
4.5	Track reconstruction in the Belle detector	60
4.5.1	Helix parameterization of charged tracks	60
4.5.2	Track finding strategy	62
4.5.3	Track fitting	62
4.5.4	Tracking performance	63
4.6	Analysis workflow in basf2	64
4.6.1	Structure of the basf2 processing path	64
5	Candidate selection	67
5.1	Dataset	67
5.1.1	Event skimming criteria	67
5.2	Pre-selection	68
5.2.1	Analysis of the pre-selected sample	69
5.3	Candidate selection	71
5.3.1	SVD hit requirement	71
5.3.2	ACC photo-electron requirement	74
5.3.3	TOF-based selection	76
5.3.4	Momentum-dependent dE/dx selection	77
5.4	Removal of clone tracks	79
5.4.1	Detector information for proton candidates	79
5.4.2	Inclusive $\Upsilon(2S) \rightarrow p\Lambda + X$ decays	82
6	Prompt proton identification	85
6.1	Two-dimensional model for prompt proton identification	85
6.1.1	Momentum binning and charge separation	85
6.1.2	Structure of the two-dimensional probability density function	86
6.1.3	Model for the time-of-flight mass distribution	86
6.1.4	Model for the transverse distance distribution	88
6.1.5	Initialization and fit stability	88
6.1.6	Likelihood definition and fit procedure	89
6.2	Fit strategy	90
6.2.1	One-dimensional fit to dr	90
6.2.2	One-dimensional fit to m_{TOF}	92
6.2.3	Definition of the 2D model	93
6.2.4	Definition of the signal region	94
6.2.5	Fit quality	100
6.3	Prompt proton fraction	100

6.3.1	Momentum dependence of fitted yields and fractions	100
6.3.2	Kernel smoothing of the prompt–proton fraction	101
6.3.3	Use in correlation–function construction	103
7	Building the correlation function	105
7.1	Same-event and mixed-event pair construction	105
7.1.1	Same-event pairs	105
7.1.2	Mixed-event pairs	106
7.1.3	Definition of the correlation function	110
7.2	Results at generator level	110
7.2.1	Acceptance effects	113
7.2.2	EMCIC	118
7.2.3	Linear approximation of the EMCIC baseline	125
7.3	Results in the reconstructed MC samples	126
8	Systematic uncertainties	129
8.1	Event shape binning	129
8.2	Selection on $N_{\text{hits}}^{\text{CDC}}$	132
8.3	Momentum-dependent dE/dx selection	132
8.4	Tracking momentum resolution	132
8.4.1	MC study	133
8.4.2	Momentum resolution using $D \rightarrow K\pi\pi$ decays	136
8.5	Fit model	139
9	Results	141
9.1	Results on the data samples	141
9.1.1	Results in the $\Upsilon(2S)$ data sample	142
9.1.2	Results in the $\Upsilon(1S)$ data sample	144
9.2	Discussion of the results	147
9.3	The CATS framework: a toy study	148
9.3.1	Toy generation and fit of the pp correlation function	148
10	Summary and conclusions	151

Chapter 1

Introduction

1.1 The landscape of hadron physics in the non-perturbative QCD regime

Understanding the dynamics of hadrons remains one of the most compelling challenges in contemporary particle physics. Although the Standard Model provides a remarkably successful framework for the strong interaction through Quantum Chromodynamics (QCD), the regime in which this theory becomes most intriguing—namely, low energies and large distances—is also the regime where the dynamics is governed by non-perturbative effects. In this domain the strong coupling is large, perturbative expansions fail, and quarks and gluons manifest only indirectly through bound states and emergent phenomena. The hadrons observed in experiments, from the simplest mesons and baryons to multiquark candidates and exotic configurations, arise from a rich interplay of confinement, chiral symmetry breaking, and non-perturbative gluonic dynamics. As a result, low-energy hadron physics occupies a unique position: it is both governed by a well-defined fundamental theory and yet it is profoundly difficult to compute from first principles.

Over the past decades, considerable progress has been achieved through a combination of theoretical innovation and high-quality experimental data. Effective field theories, lattice QCD calculations, and phenomenological models have each provided essential insights into specific aspects of hadron structure and interactions. Nevertheless, many key questions remain open. How do hadrons form from quarks and gluons in the process of hadronization? What determines the spectrum and properties of hadronic resonances? How do nucleons interact at the femtometer scale, and how does this interaction depend on spin, isospin, and internal structure? These questions illustrate the broader challenge of connecting the fundamental degrees of freedom of QCD to the observable properties of hadronic matter.

Experiments continue to play a decisive role in this endeavour. Precision studies of hadron production, spectroscopy, and scattering provide indispensable constraints for theoretical approaches and often reveal patterns not anticipated by existing models. In particular, the ability to probe the strong interaction at different distance scales by varying the kinematic conditions of a reaction has been central to uncovering the layered complexity of QCD. High-energy collisions allow access to partonic degrees of freedom and asymptotic behavior, while low-energy processes and final-state interactions reveal the more intricate, non-perturbative structure of hadrons themselves.

In this context, the study of hadron physics at low energies remains a vibrant and

evolving field. It serves as a bridge between fundamental theory and experiment, linking the microscopic constituents of matter to the emergent phenomena that govern their observable behavior. The exploration of nucleon–nucleon interactions, the characterization of hadronic sources, and the detailed investigation of hadronization mechanisms are all threads of this broader narrative. As new experimental facilities continue to deliver increasingly precise data, and as theoretical tools become more sophisticated, the opportunity to deepen our understanding of the strong force at the femtometer scale has never been greater.

The work presented in this thesis is situated within this landscape. By examining the correlations between protons produced in electron–positron annihilation, a unique window into the interplay between hadronization and nucleon–nucleon dynamics is accessed. Before arriving at these specific questions, it is essential to outline the broader experimental setting that makes such studies possible. This leads naturally to a discussion of electron–positron collisions as a clean and versatile environment for probing the strong interaction.

1.2 A clean environment for hadron studies

Electron–positron collisions provide one of the cleanest systems for exploring the strong interaction. Unlike hadron colliders, where the initial state consists of extended composite objects with complex partonic structures, the e^+e^- system is elementary and fully characterized by quantum electrodynamics. The annihilation of an electron and a positron proceeds through a well-defined pointlike interaction, typically mediated by a virtual photon or, at higher energies, by the Z boson. As a consequence, the initial conditions of the reaction are known with exceptional precision: the four-momentum of the initial state is fixed, the reaction plane is not obscured by remnants of colliding hadrons, and backgrounds arising from beam structure or partonic interactions are minimal. These features make e^+e^- annihilation uniquely well suited for studying hadronic processes in a clean and controlled environment.

Over the past decades, experiments at e^+e^- colliders such as LEP, BaBar, Belle, and now Belle II have played a central role in shaping our understanding of hadron physics. Their ability to operate at specific center-of-mass energies has allowed detailed studies of quarkonium systems, fragmentation processes, electroweak observables, and rare decays. In particular, the energy regime around the bottomonium resonances—the $\Upsilon(nS)$ —has been especially fruitful. These resonances provide an abundant source of hadrons produced in a well-defined dynamical context: in particular, narrow bottomonium — $\Upsilon(1, 2, 3S)$ —is a heavy quark–antiquark bound state that annihilates predominantly into gluons, which subsequently form hadrons through the process of hadronization. The resulting events exhibit moderate multiplicities, relatively low background contamination, and kinematic configurations that lend themselves naturally to detailed correlation studies.

Another advantage of the e^+e^- environment is the absence of complicating phenomena such as multiple parton interactions or color exchanges between overlapping proton remnants, which are common in hadron collisions. This leads to a significant conceptual simplification: the hadronic final state arises from a single hard interaction, and the space–time structure of particle production is dominated by fragmentation rather than by collective or thermalized dynamics. As a result, the mechanisms governing baryon

production and the geometric properties of the hadron-emitting source can be examined with a clarity not always achievable in more complex collision systems.

For studies of baryon–baryon correlations, the advantages are particularly pronounced. The moderate particle multiplicity of e^+e^- events makes easier to suppress backgrounds, while the controlled topology of the events enables efficient reconstruction of proton pairs over a broad range of momenta. Furthermore, the simplicity of the initial state facilitates the interpretation of low-relative-momentum correlations, which are especially sensitive to the interplay between final-state interactions and the spatial distribution of the emitting source.

In this sense, e^+e^- colliders offer a unique window into the strong interaction at the femtometer scale. They allow us to isolate the physics of hadronization from that of the initial collision, and to observe the emergence of baryonic degrees of freedom under exceptionally clean conditions. The study of proton–proton correlations in this environment is therefore particularly informative, as it provides sensitivity to nucleon–nucleon interactions while simultaneously probing the mechanisms responsible for baryon formation.

The next step in this introduction is to examine more closely the structure and dynamics of the proton–proton system itself, and to outline the physical motivations for studying its correlations in the low-relative-momentum regime.

1.3 Dynamics of the proton-proton system

Over many decades, nucleon–nucleon scattering experiments and bound-state properties have been used to constrain proton–proton and proton–neutron potentials with high precision. Modern descriptions, whether phenomenological or based on effective field theories, reproduce a wide range of low-energy observables and provide a reliable representation of the strong and electromagnetic components of the p–p interaction. In particular, the interplay between short-range repulsion, intermediate-range attraction, and the repulsive Coulomb force between two positively charged protons is, at the level relevant for femtosopic analyses, reasonably well established.

By contrast, the space–time structure of hadron emission in high-energy collisions is much less constrained. When a pair of protons is produced in an electron–positron annihilation event, their relative momentum distribution at low values is shaped by two distinct ingredients: the final-state interaction between the protons and the distribution of their relative emission points, often referred to as the source function. The final-state interaction, governed by the known nucleon–nucleon potential and electromagnetic repulsion, acts as a “filter” that converts information about the geometry of the emitting source into observable correlations in momentum space. From this perspective, the proton–proton interaction is not the primary unknown of the problem; instead, it is a well-characterized tool that can be used to probe the properties of the hadron-emitting source.

An additional ingredient arises from quantum statistics. Protons are spin- $\frac{1}{2}$ fermions, and the antisymmetrization of their two-particle wave function, according to Fermi–Dirac statistics, influences the correlation pattern at small relative momentum. For identical fermions in a spin-symmetric state, the spatial part of the wave function must be antisymmetric, which leads to a suppression of pairs with small relative separation in coordinate space and, correspondingly, to characteristic structures in momentum space. In the proton–proton case, the observed correlation function thus reflects the combined effect of Fermi–Dirac statistics, Coulomb repulsion, and the strong nuclear interaction. Each of

these contributions is reasonably well modeled, and their interplay can be calculated with small theoretical uncertainties.

This situation makes the p–p system particularly valuable in the context of femtoscopy in electron–positron annihilation. Since the interaction kernel is known, deviations in the measured correlation function can be traced back, to a large extent, to the properties of the emission source rather than to uncertainties in the potential itself. In other words, proton–proton correlations offer a way to “invert” the problem: instead of using a known source to extract an unknown interaction, one can use a known interaction to study an unknown source. The central objective then becomes the characterization of the source distribution—the effective size, shape, and possible dependence on kinematic variables—associated with hadron emission in Υ decays.

The focus of this thesis is thus to exploit our existing knowledge of the p–p interaction in order to extract information about the hadron emission source in e^+e^- annihilation, specifically in events involving Υ decays. By confronting measured correlation functions with theoretical calculations that incorporate Fermi–Dirac statistics, Coulomb repulsion, and strong final-state interactions, one can infer characteristic length scales of the source and investigate how they depend on the kinematic properties of the produced protons. In this way, the proton–proton system serves as a calibrated probe of the space–time structure of hadronization in a clean experimental environment.

Having established this role, the next step is to place such studies within the broader historical and conceptual development of femtoscopy, tracing how techniques originally conceived in other fields have been adapted to probe the femtometer-scale structure of particle-emitting sources in high-energy collisions.

1.4 Evolution of the femtoscopy technique

The idea that correlations between identical particles can reveal information about the spatial structure of their source has an origin far outside particle physics. In the 1950s, Hanbury Brown and Twiss demonstrated that the intensity correlations of photons detected from distant stars could be used to determine their angular size, inaugurating what became known as intensity interferometry [1, 2]. Their technique relied on the quantum statistical properties of identical bosons: photons emitted independently from an extended source exhibit characteristic correlations in their detection rates that depend on the source geometry. This insight, initially surprising even to some of the founders of quantum optics, opened the door to extract spatial information from momentum-space observables without the need for direct imaging.

It was only a few years later that this concept found its way into the field of high-energy physics. Goldhaber, Lee, and Pais realized that the same statistical principles governing photon correlations could apply to hadrons produced in particle collisions [3]. Their pioneering work on identical pion pairs revealed a momentum-space enhancement at small relative momenta, now known as the Bose–Einstein or Hanbury Brown–Twiss (HBT) effect. This enhancement was understood as the result of the symmetrization of the two-particle wave function, combined with the finite size of the emission region. For the first time, the space–time extent of hadron production could be inferred from experimentally accessible observables.

As accelerator energies increased, and as detectors became more sophisticated, femtoscopy emerged as a powerful tool in nuclear and particle physics [4]. In heavy-ion

collisions, where the creation of a hot and dense medium resembling the quark–gluon plasma is expected, two-particle correlations became central to characterizing the collective expansion and freeze-out geometry of the system. Measurements of pion, kaon, and proton correlations played a major role in establishing the hydrodynamic behavior of the medium and in determining its spatial and temporal evolution. In these large and complex systems, femtoscopy evolved into a precision method for probing the dynamics of strongly interacting matter under extreme conditions.

The technique later expanded beyond the study of identical bosons. The development of theoretical frameworks incorporating final-state interactions allowed femtoscopy to be applied to non-identical particles and to fermions, whose correlations are shaped not only by quantum statistics but also by Coulomb and strong forces [5]. In this broader context, two-particle correlations serve as a natural probe of both the source distribution and the interaction potential. This dual sensitivity has enabled, for example, the extraction of scattering parameters for hyperon–nucleon and hyperon–hyperon systems that are otherwise difficult to access experimentally [6]. The method has thus become a bridge between nuclear physics, particle physics, and the study of exotic states of matter.

In more recent years, interest has grown in applying femtoscopy to small collision systems such as proton–proton, proton–nucleus, and electron–positron interactions [7]. These environments lack the collective expansion characteristic of heavy-ion collisions and offer a cleaner, more elementary setting in which to study the space–time characteristics of hadron production. The reduced complexity of small systems makes them ideal for isolating the effects of hadronization and final-state interactions without the confounding influence of a thermal medium or long-range correlations. In particular, baryon–baryon femtoscopy in such systems provides a means of probing the strong interaction at low energies using particles created in high-energy reactions, a connection that has attracted significant theoretical and experimental attention.

Within this evolving landscape, electron–positron annihilation occupies a distinctive position. The absence of initial-state hadrons eliminates many of the complications present in proton-induced reactions, while the fragmentation of quarks and gluons provides a well-defined setting for studying the formation of hadrons at the femtometer scale. Applying femtoscopy to this environment allows one to address questions that are difficult to pose elsewhere. What is the effective space–time extent of hadron emission in a system without collective expansion? How do different hadron species emerge from the same underlying fragmentation process? And how do final-state interactions between baryons translate the geometry of the source into observable patterns in the correlation function?

The path from astronomical interferometry to the study of hadronic sources in e^+e^- annihilation illustrates the conceptual power and adaptability of femtoscopy. Techniques originally devised to measure the diameter of distant stars have become tools for probing the quantum structure of matter at distances more than twenty orders of magnitude smaller. In the present work, femtoscopic methods are used not to map the sky but to investigate the space–time geometry of hadronization in Υ decays, leveraging the known proton–proton interaction to extract information about the size of the emitting source.

Chapter 2

Strong interaction models and hadronization

The discussion so far has outlined the broad physical motivations for studying proton–proton correlations in electron–positron annihilation and has traced the conceptual development of femtoscopy from its origins to its application in modern collider experiments. This qualitative perspective provides the necessary context, but a more detailed understanding of the phenomena addressed in this thesis requires a closer examination of the theoretical and experimental tools that make such studies possible. In particular, the extraction of source properties from correlation measurements relies on the interplay between three tightly connected ingredients: the modeling of the strong interaction between hadrons, the description of the hadronization process through which protons are formed, and the formalism that links these elements to observable correlation functions.

A first aspect concerns the treatment of final-state interactions. While the proton–proton potential is relatively well constrained, different theoretical approaches exist for describing the strong dynamics between hadrons, especially when extending femtoscopic studies to systems involving hyperons or other baryons. Understanding the assumptions and limitations of these interaction models is essential, both for interpreting existing results and for placing the present analysis in the broader context of ongoing efforts to characterize baryon–baryon forces.

A second ingredient is the mechanism by which hadrons emerge from the quarks and gluons produced in electron–positron annihilation or in the decay of heavy quarkonium states. The space–time geometry of this hadronization process directly determines the emission source probed by femtoscopic techniques. Different models—ranging from string fragmentation approaches such as the Lund model, to statistical hadronization, to coalescence-based scenarios—offer distinct perspectives on how baryons are produced and how their initial separations in coordinate space are established. These phenomenological frameworks provide the conceptual basis for interpreting the source sizes extracted from data and for understanding their dependence on momentum and event topology.

Finally, the quantitative extraction of source parameters relies on a well-developed methodology that connects theoretical inputs to measurable observables. The femtoscopy formalism, implemented through correlation functions, wave-function evaluation techniques, and parametrized source models, has been refined through decades of experimental work in systems ranging from heavy-ion collisions to small-scale reactions. Recent measurements from collaborations such as ALICE, STAR, and earlier e^+e^- experiments illustrate both the capabilities and the subtleties of the method, highlighting the need for

careful treatment of backgrounds, purity corrections, and the contributions of different final-state effects.

The remainder of this chapter is devoted to these foundational elements. The discussion begins with hadronic interaction models relevant for femtoscopy, proceeds to an overview of hadronization mechanisms and their implications for baryon production, and concludes with a technical exposition of the femtoscopic formalism and a survey of recent experimental results. Together, these components form the theoretical and methodological groundwork upon which the analysis presented in later chapters is built.

2.1 Modeling strong interactions between hadrons

The strong interaction, one of the four fundamental forces of nature, governs the behavior of quarks and gluons, the elementary constituents of hadronic matter. Its underlying theory, Quantum Chromodynamics (QCD), is a non-Abelian gauge theory based on the symmetry group $SU(3)_{\text{color}}$. Over the past decades, QCD has provided a remarkably successful description of high-energy processes, where the property of asymptotic freedom allows perturbative calculations to yield precise predictions that have been extensively confirmed in collider experiments.

At low energies, however, the situation is strikingly different. The same property that makes QCD predictive at high momentum transfers — its running coupling α_S — becomes a major obstacle as the coupling grows strong in the infrared. In this non-perturbative regime, phenomena such as confinement and the spontaneous breaking of chiral symmetry emerge, shaping the observable spectrum of hadrons and their interactions. While the formulation of QCD on a discretized space-time lattice (Lattice QCD) and Effective Field Theories (EFTs) have made great progress in addressing these challenges, many aspects of hadronic physics remain elusive.

Understanding hadron–hadron interactions at low energies is not only essential for completing the picture of strong interactions, but also has profound implications for nuclear structure, astrophysical environments such as neutron stars, and the formation of exotic states of matter. Despite significant theoretical and experimental progress, a fully unified description of hadron dynamics in the non-perturbative domain is still lacking, making it one of the central open problems in contemporary physics.

Over the years, a variety of complementary approaches have been developed to tackle the problem of low-energy QCD and to gain insight into hadron–hadron interactions. One of the earliest and most intuitive strategies is the construction of Potential Models (see Sec. 2.1.1). These frameworks aim to capture the essential features of the strong force in terms of effective interactions between hadrons, often inspired by analogies with atomic or nuclear physics. While phenomenological in nature, they have proven remarkably successful in reproducing spectra and bound-state properties [8, 9].

Another powerful line of investigation is provided by Effective Field Theory (EFT) (Sec. 2.1.2), which systematically exploits the separation of scales inherent to hadronic systems. Chiral Effective Field Theory, for example, incorporates the consequences of spontaneous chiral symmetry breaking and has become a cornerstone for describing low-energy pion–nucleon and nucleon–nucleon interactions [10, 11]. Such frameworks allow for controlled expansions, providing a bridge between the symmetries of QCD and experimentally measurable observables.

A further layer of complexity arises in systems where multiple scattering channels

are strongly coupled, requiring the treatment of Coupled-Channel Dynamics (Sec. 2.1.3). This approach is particularly relevant in the study of resonances and exotic hadronic states, where the interplay between open and closed channels shapes the observed spectrum [12, 13]. By accounting for these coupled effects, one can gain a deeper understanding of the emergence of resonant structures and the role of multi-quark configurations or hadronic molecules.

In parallel, major progress has been achieved through first-principles calculations in Lattice formulation of Quantum Chromodynamics (Lattice QCD) (Sec. 2.1.4). By discretizing spacetime on a finite grid and performing large-scale numerical simulations, this method provides direct access to the non-perturbative dynamics of quarks and gluons. Lattice QCD has yielded important results for hadron spectroscopy, form factors, and scattering parameters [14, 15], offering invaluable benchmarks for effective models and phenomenological approaches. Despite its successes, lattice computations face challenges such as finite-volume effects and the treatment of multi-hadron systems, which continue to motivate ongoing developments.

Together, these approaches highlight both the progress achieved and the challenges that remain in understanding strong interactions at low energies. They provide the theoretical foundation upon which the present work is built, and they frame the open questions that motivate the studies developed in the following chapters.

2.1.1 Potential models

One of the most traditional approaches to describe strong interactions between hadrons is the use of potential models. These are inspired by the successful description of the nucleon–nucleon interaction and are usually based on the exchange of mesons as effective degrees of freedom. The simplest case is the one-boson exchange (OBE) model [16], in which a generic single meson ϕ mediates the interaction between two hadrons. In momentum space the potential takes the form

$$V_\phi(\mathbf{q}) = -\frac{g_\phi^2}{\mathbf{q}^2 + m_\phi^2},$$

which, after Fourier transformation, leads to a Yukawa-type potential in coordinate space,

$$V_\phi(r) = -\frac{g_\phi^2}{4\pi} \frac{e^{-m_\phi r}}{r}.$$

Here m_ϕ is the mass of the exchanged meson, which determines the range of the force, g_ϕ is the hadron–meson coupling constant, q is the exchanged momentum and r is the radial distance between the hadrons. Pions, being the lightest mesons, generate the longest-range part of the interaction, while heavier mesons such as the ρ , σ , and ω control the intermediate and short-distance behavior.

In the case of nucleons, the one-pion exchange potential (OPEP) illustrates how spin and isospin degrees of freedom enter. The central and tensor components of the pion exchange force can be written schematically as

$$V_\pi(r) \sim (\boldsymbol{\sigma}_1 \cdot \boldsymbol{\sigma}_2) Y(r) + S_{12} T(r),$$

where $\boldsymbol{\sigma}_i$ are Pauli matrices acting on the nucleon spins, S_{12} is the tensor operator, and the functions $Y(r)$ and $T(r)$ describe the radial dependence of the central and tensor

forces. This tensor force is crucial for describing phenomena such as the S – D wave mixing in the deuteron.

Phenomenological analyses of the nucleon–nucleon interaction show a characteristic pattern: a strong repulsive core at short distances, a medium-range attraction, and a long-range Yukawa tail. In the OBE picture, the repulsive core is often attributed to the ω -meson exchange, the attraction to the scalar σ exchange, and the long-range part to the pion exchange. This interplay explains both the binding of the deuteron and the saturation properties of the nuclear matter: at intermediate distances the attractive component allows two nucleons to form a bound state, while the short-range repulsion prevents them from collapsing into an overly compact configuration; when many nucleons are present, the same balance limits the strength of the interaction and leads to a stable equilibrium density.

The same framework can be generalized to systems of heavy hadrons, where meson exchange between charmed or bottom mesons and baryons may generate molecular states. For instance, the interaction between D and $\bar{D}^*$ mesons, relevant for the $X(3872)$, or between $\bar{D}^{(*)}$ and $\Sigma_c^{(*)}$, relevant for the hidden-charm pentaquarks, can be modeled in terms of π , ρ , σ , and ω exchanges, supplemented by contact terms constrained by heavy-quark spin and flavor symmetries. In these cases, the effective potential has the generic form

$$V(r) = \sum_{\phi} C_{\phi} \Gamma_{\phi} \frac{e^{-m_{\phi} r}}{r},$$

where the coefficients C_{ϕ} weight the relative strength of each exchanged meson in a given interaction channel, while Γ_{ϕ} represents the corresponding operator structure acting on the spin and isospin degrees of freedom of the nucleons.

Once such a potential is specified, the two-body problem reduces to solving the Schrödinger equation

$$\left[-\frac{\nabla^2}{2\mu} + V(r) \right] \psi(\mathbf{r}) = E\psi(\mathbf{r}),$$

with μ the reduced mass of the system. The appearance of bound or resonant solutions signals the possible existence of hadronic molecules.

A widely used realistic nucleon–nucleon interaction is the Argonne $v18$ (AV18) potential [17]. It belongs to the class of phenomenological potentials constructed to reproduce a large set of NN scattering data with very high precision. The AV18 potential includes:

- **Electromagnetic interactions:** one- and two-photon exchange, vacuum polarization, and magnetic moment terms, ensuring the correct treatment of long-range Coulomb effects in pp scattering.
- **One-pion exchange:** governing the long-range part of the nuclear force,

$$V_{\pi}(r) \propto \frac{e^{-m_{\pi} r}}{r},$$

with the appropriate spin–isospin dependence.

- **Intermediate and short-range parts:** represented by a sum of operator structures, covering central, tensor, spin–orbit, and quadratic momentum-dependent terms. In total, AV18 employs 18 operators, hence its name.

The full potential can be written schematically as

$$V_{NN} = \sum_{p=1}^{18} V_p(r) O_p,$$

where O_p are the independent spin–isospin operators (including 1, $\vec{\sigma}_1 \cdot \vec{\sigma}_2$, S_{12} , $\vec{L} \cdot \vec{S}$, and the corresponding isospin operators).

Because of its accurate description of pp scattering data at low energies, the Argonne $v18$ potential is commonly used in femtoscopic studies and in calculations where a realistic baseline for the pp interaction is required.

Potential models are therefore not derived directly from QCD, but they provide an intuitive and phenomenologically rich description of hadron–hadron interactions. They have been essential in nuclear physics and now serve as an important tool in the study of exotic hadrons, where they offer a concrete mechanism for the formation of molecular states.

2.1.2 Effective field theories

While potential models provide an intuitive picture of hadron–hadron forces in terms of meson exchange, a more systematic framework is offered by effective field theories (EFTs). The central idea of EFT is to exploit the separation of scales in QCD: at low energies, the relevant degrees of freedom are hadrons rather than quarks and gluons, and their interactions can be organized in a controlled expansion in powers of momentum and light quark masses.

For systems involving light quarks, the most successful example is the chiral effective field theory (ChEFT), which is founded on the approximated chiral symmetry of QCD and its spontaneous breaking. In ChEFT, the interactions among pions and nucleons are described by an effective Lagrangian that is systematically expanded in powers of external momenta and pion masses [18]. At leading order, this includes terms such as

$$\mathcal{L}_{\text{ChEFT}} = \bar{N} (i\gamma^\mu \partial_\mu - m_N) N + \frac{g_A}{2f_\pi} \bar{N} \gamma^\mu \gamma_5 \tau^a N \partial_\mu \pi^a + \dots, \quad (2.1)$$

where $N = (p, n)^T$ denotes the nucleon isospin doublet field and $\bar{N} = N^\dagger \gamma^0$ its Dirac adjoint. The matrices γ^μ are the Dirac gamma matrices acting in spinor space, and γ_5 is the chirality matrix. Furthermore, f_π is the pion decay constant, g_A is the axial-vector coupling of the nucleon, and τ^a are the Pauli matrices acting in isospin space. Higher-order corrections include multi-pion interactions and short-range contact terms, whose coefficients—low-energy constants (LECs)—must be fixed by data or lattice QCD [19]. This framework allows for a controlled expansion in powers of Q/Λ_χ , where Q is a typical momentum scale and $\Lambda_\chi \sim 1 \text{ GeV}$ is the chiral symmetry breaking scale.

When heavy quarks are involved, such as in systems containing charm or bottom quarks, chiral symmetry alone is not sufficient. Instead, heavy-quark symmetry becomes a guiding principle: in the limit of infinite heavy-quark mass, spin and flavor of the heavy quark decouple from the dynamics. This leads to Heavy Hadron Chiral Perturbation Theory (HHChPT), which merges heavy-quark symmetry with chiral symmetry, and has been applied to describe interactions among heavy mesons and baryons.

Another important EFT approach is applicable to systems near two-hadron thresholds, where the scattering length is anomalously large—much greater than the range of

the interaction. In such cases, a pionless or "halo" EFT is appropriate: the dynamics can be captured by contact interactions alone, such as

$$\mathcal{L}_{\text{EFT}} = C_0(H^\dagger H)(H^\dagger H) + C_2(H^\dagger \nabla H)(H^\dagger \nabla H) + \dots,$$

where H is the hadron field. The large scattering length arises from tuning C_0 , while higher-order terms introduce systematic corrections.

These EFT frameworks have proven powerful in the study of exotic hadrons. For instance, the $X(3872)$, lying extremely close to the $D^0 \bar{D}^{*0}$ threshold, is often described using a nonrelativistic EFT tailored to such shallow bound states ("X-EFT") [20]. In this treatment, the $X(3872)$ behaves like a universal shallow molecule.

In this way, EFTs provide a powerful theoretical framework that combines model-independence with systematic error estimates. They respect the essential symmetries of QCD while parametrizing unknown short-distance physics, making them indispensable in the study of both conventional and exotic hadrons.

2.1.3 Coupled-channel dynamics

Many hadrons, especially those near open-flavor thresholds, cannot be understood in terms of a single hadronic channel but require a simultaneous treatment of multiple scattering and decay channels. This approach naturally explains the appearance of dynamically generated resonances, such as the $\Lambda(1405)$ in the $\bar{K}N-\pi\Sigma$ system or the $X(3872)$ in the $D\bar{D}^*-J/\psi\pi\pi$ system.

The theoretical foundation of coupled-channel dynamics is based on unitarity and analyticity of the scattering amplitude. If V denotes the interaction kernel between two hadrons, the full scattering matrix T can be obtained by solving an integral equation, such as the Bethe–Salpeter or Lippmann–Schwinger equation. In its algebraic form, often used in the chiral unitary approach, the resummed amplitude reads

$$T = \frac{V}{1 - VG},$$

where G is the two-hadron loop function that encodes the propagation of intermediate states [21, 22, 12]. The poles of T in the complex energy plane correspond to bound states or resonances, while the unitarity condition ensures that the imaginary part of T is consistent with open decay channels.

This formalism has been extensively applied in meson–baryon and meson–meson systems. A classic case is the $\Lambda(1405)$, which emerges as a dynamically generated state from $\bar{K}N$ interactions strongly coupled to $\pi\Sigma$ channels [22]. The framework has also been used in the heavy quark sector, where near-threshold exotic states such as the $Z_c(3900)$, $Z_{cs}(3985)$ and the hidden-charm pentaquarks $P_c(4312)$, $P_c(4440)$, and $P_c(4457)$ can be interpreted as hadronic molecules dynamically generated by coupled-channel effects [23, 24].

An important advantage of coupled-channel methods is their ability to describe line shapes observed in experiments. Near threshold, the mixing of open channels often produces asymmetric distributions and cusp-like structures, which are well reproduced by unitarized amplitudes. This has been essential, for example, in disentangling the nature of the $X(3872)$, where the interplay between neutral and charged $D\bar{D}^*$ channels and their coupling to charmonium-like decay modes plays a decisive role [25].

In summary, coupled-channel dynamics provides a natural and powerful explanation for many unconventional hadrons. By respecting unitarity and allowing multiple thresholds to communicate, this approach captures phenomena beyond the scope of simple

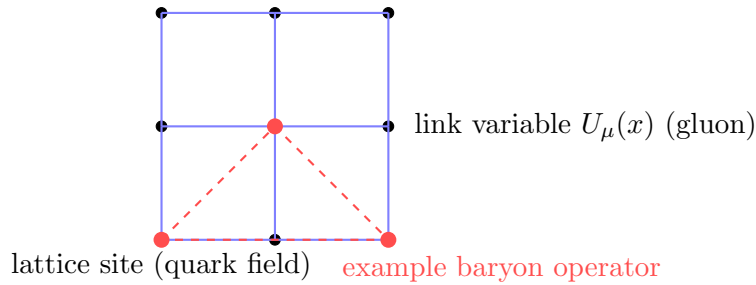


Figure 2.1: Cartoon of lattice QCD: quark fields live on lattice sites (dots), gluon fields are represented by $SU(3)$ link variables $U_\mu(x)$ (blue lines). Gauge-invariant hadron operators are built by combining quark fields at different sites, connected by gluon links. Correlation functions of such operators provide access to hadron masses and matrix elements.

potential models or single-channel treatments, and it remains crucial in modern hadron spectroscopy.

2.1.4 Lattice QCD

Lattice QCD is the most direct first-principles approach to study the non-perturbative regime of quantum chromodynamics. The central idea is to replace continuous spacetime by a discrete four-dimensional Euclidean lattice. Quarks live on the lattice sites, while gluons are represented by link variables $U_\mu(x) \in SU(3)$ that connect neighboring sites. In this way, the QCD partition function becomes a finite-dimensional path integral that can be evaluated numerically:

$$Z = \int \mathcal{D}U_\mu(x) \mathcal{D}\bar{\psi}(x) \mathcal{D}\psi(x) \exp[-S_G[U] - S_F[\bar{\psi}, \psi, U]] .$$

Here S_G is the discretized Yang–Mills action for the gluons (for example Wilson’s plaquette action [26]), and S_F is the discretized fermion action for the quarks. Expectation values of operators are obtained using Monte Carlo importance sampling, i.e. by generating an ensemble of representative field configurations and averaging over them. The basic ingredients of lattice QCD and the construction of gauge-invariant hadronic operators are illustrated schematically in Fig. 2.1.

Physical information is extracted from correlation functions of suitably chosen interpolating operators. For a hadron H , one defines a two-point function

$$C_H(t) = \sum_{\mathbf{x}} \langle O_H(\mathbf{x}, t) O_H^\dagger(\mathbf{0}, 0) \rangle .$$

At large Euclidean time t , this correlator falls off exponentially,

$$C_H(t) \sim e^{-M_H t} ,$$

so that the mass M_H of the hadron can be read off from the slope in a logarithmic plot. This basic strategy allows lattice QCD to compute the spectrum of ground-state hadrons, their excitations, and matrix elements relevant for structure and decays.

Over the past decades, lattice QCD has evolved from early calculations at heavy pion masses and coarse lattices to modern simulations with nearly physical quark masses, large volumes, and fine spacings. This progress has enabled increasingly precise determinations

of hadron masses, form factors, and matrix elements, providing stringent tests of QCD and valuable input for phenomenology. Comprehensive reviews of these developments can be found in [27, 28].

2.2 Hadronization

Quarks and gluons cannot be observed as free particles, but only bound as color-neutral hadrons. The reason why is yet to be understood, since this property does not arise as a natural consequence of the QCD theory. Nonetheless, there is wide agreement on the phenomenology of confinement: as energy flows with intent to break a quark-antiquark pair, the formation of color-electric flux tubes among the quark and the antiquark results in the fragmentation of the flux tubes into newly produced mesons and baryons. This process is called *hadronization*. In the following, a brief review of the various models used to describe this phenomenon is provided, with emphasis on recent results concerning the space-time properties of the process.

2.2.1 Lund string model

The Lund string model [29, 30] is one of the most successful phenomenological frameworks to describe hadronization. The key assumption is that confinement can be represented by a stringlike flux tube of constant tension $\kappa \simeq 1 \text{ GeV/fm}$ connecting color charges. The linear confining potential between a quark and an antiquark reads

$$V(r) \simeq \kappa r, \quad (2.2)$$

which dominates at large separations and is supported both by spectroscopy (Regge trajectories) and by lattice QCD results.

As the $q\bar{q}$ pair separates, the potential energy stored in the string increases until it becomes energetically favorable to create a new $q'\bar{q}'$ pair from the vacuum. This process “breaks” the string into two independent pieces, each of which forms hadrons. By iteration, a high-mass string fragments into a cascade of hadrons, naturally generating the approximately flat rapidity distributions observed in e^+e^- annihilations.

The distribution of momentum fractions carried by successive hadrons is encoded in the Lund symmetric fragmentation function,

$$f(z) \propto \frac{1}{z} (1-z)^a \exp\left(-\frac{bm_{\perp}^2}{z}\right), \quad (2.3)$$

where z is the light-cone momentum fraction taken by the hadron, $m_{\perp}^2 = m_h^2 + p_{\perp}^2$ its squared transverse mass, and a, b are tunable parameters. This functional form ensures left-right symmetry of the fragmentation chain, boost invariance, and reproduces the scaling properties of inclusive spectra. Transverse momenta are generated by a tunneling mechanism, leading to a Gaussian suppression factor

$$P(p_{\perp}) \sim \exp\left(-\frac{\pi p_{\perp}^2}{\kappa}\right), \quad (2.4)$$

which also explains the strong suppression of heavy quark pair production in string breaking.

The model has been extensively implemented in event generators such as PYTHIA, and tuned to e^+e^- , ep and pp data, where it provides a remarkably universal description of hadron production.

While the Lund model has traditionally been formulated in momentum space, recent work has extended it to provide a consistent *space-time picture* of hadronization [31]. This is motivated both by theoretical considerations (causality and uncertainty relations) and by experimental observables sensitive to the geometry and time scales of hadron production, such as femtoscopy.

In $e^+e^- \rightarrow q\bar{q}$ annihilation, the separating partons move approximately on the light cone, pulling out a relativistic string world sheet, as shown in Fig. 2.2.

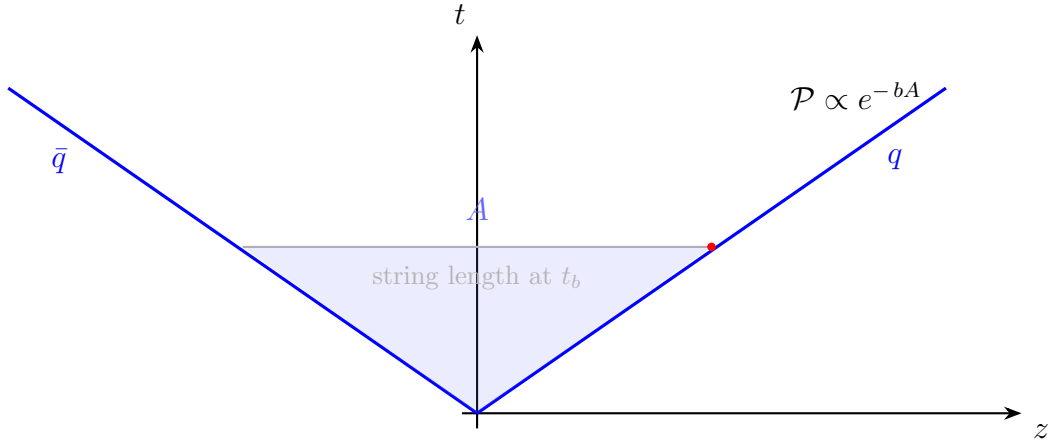


Figure 2.2: Minimal space-time sketch of a string stretched between a quark and an antiquark. The shaded triangle is the world-sheet area A swept out before a break (red dot). In the Lund model, hadronization histories are weighted by the area law $\mathcal{P} \propto e^{-bA}$.

The breaking vertices, where new $q\bar{q}$ pairs are produced, are localized in this world sheet and can be assigned coordinates in both energy-momentum and space-time. The two pictures are related by the linearity of the string force:

$$\frac{dE}{dz} = \frac{dp_z}{dt} = \kappa. \quad (2.5)$$

Here E and p_z are the energy and longitudinal momentum of the quarks, z and t denote their longitudinal position and time, and κ is the string tension. The constancy of κ reflects the linear confining potential, leading to a direct correspondence between the space-time evolution of the string and the flow of energy and momentum along it.

The probability $\mathcal{P}$ for a given hadronization history is proportional to an “area law”,

$$\mathcal{P} \sim \exp(-bA), \quad (2.6)$$

where A is the area of the string world sheet before breaking. The parameter b controls how strongly large areas are suppressed: a larger value of b favors earlier breaks (smaller A), while a smaller value allows later breaks. It is the same parameter that appears in the Lund symmetric fragmentation function in Eq. (2.3), where it governs the suppression of large transverse masses. This connects the fragmentation function with the distribution of production vertices in space-time. Phenomenological fits to e^+e^- and hadronic data typically find b of order $0.5\text{--}1 \text{ GeV}^{-2}$, with the precise value depending on the event-generator tune.

A natural outcome is that hadron production is not instantaneous: the first hadrons typically appear after a proper time $\tau \sim 1$ fm, and production then proceeds roughly along hyperbolae of constant invariant time, $\tau^2 = t^2 - z^2$. Slow hadrons are produced earlier in the central region, while faster hadrons appear later and further away, consistent with causality and boost invariance. Different prescriptions (“early”, “middle” or “late”) can be used to define the effective production point of a hadron, but all give results consistent within hadronic uncertainties.

This picture allows one to assign approximate production vertices to each hadron, providing a microscopic space–time description of the hadronization process. Notably, it was shown that hadrons are typically produced within a space–time region of order ~ 1 fm, consistent with confinement scales. Furthermore, the ordering of string breaks correlates with the longitudinal momentum ordering of hadrons, preserving causality.

These developments bridge the gap between the momentum-space formulation of the Lund model and the space–time picture necessary for phenomenological applications, thereby opening the door to studies of hadronization in dense QCD matter and femtoscopy. The space–time structure of string fragmentation in the Lund model—assigning production vertices in both energy–momentum and space–time—was first developed by Ferreres-Solé and Sjöstrand (2018) and has since been implemented in the C++ 8 series of PYTHIA, including the current version 8.3 [32, 33]. This extension enables phenomenological studies of hadron densities, production times, and space–time correlations in both elementary and heavy-ion collisions.

2.2.2 Cluster model

The cluster model [34] provides a simple, local picture of hadronization rooted in QCD preconfinement: after a perturbative parton shower has evolved down to a cutoff momentum $Q_0 \sim 1$ GeV, the colour topology ensures that nearby partons combine into colour singlet objects of limited invariant mass, independent of the hard scale. These color singlets, called *clusters*, are then interpreted as precursors of hadrons.

The construction proceeds as follows. Partons are assigned effective constituent masses, and each gluon is forced to split nonperturbatively, $g \rightarrow q\bar{q}$, so that only quarks and antiquarks remain. Neighbouring colour-connected partons are then paired to form clusters, typically of the type $q\bar{q}$. The distribution of cluster masses is universal and falls off rapidly above a few GeV, reflecting preconfinement. Clusters with masses close to those of physical hadrons are directly identified with them, while heavier clusters undergo isotropic two-body decays,

$$C \rightarrow H_1 + H_2, \tag{2.7}$$

with phase space and simple flavour weights governing the choice of hadrons. Strangeness suppression and diquark production are implemented by phenomenological probabilities.

This minimal setup already captures the key features: locality, approximate universality, and simple hadron spectra. It has formed the backbone of hadronization in the HERWIG event generator for decades.

More recent developments have extended the model in several directions. Colour reconnection mechanisms have been added to allow alternative pairings of colour lines, improving agreement with underlying-event and minimum-bias data at the LHC. Baryonic reconnection models enhance baryon yields in soft events. Spacetime-aware reconnection schemes incorporate geometric information in hadronization, offering a better description

of collective effects. In addition, refinements of the cluster-splitting dynamics and dedicated treatments of heavy-flavour hadronization have improved phenomenology. These extensions preserve the original preconfinement picture, while providing the flexibility needed to match the precision of modern collider data.

The original cluster model is formulated entirely in momentum space and does not assign explicit space-time coordinates to hadronization [34]. Nevertheless, the infrared cutoff $Q_0 \sim 1$ GeV suggests a characteristic formation time and length of order

$$\tau_{\text{had}} \sim \frac{1}{Q_0} \sim 0.2 \text{ fm}/c,$$

indicating a compact hadronization region. Preconfinement further implies that colour-connected partons forming a cluster are not only close in momentum but also in coordinate space, giving a local picture of hadronization. In contrast to the string model, no extended flux tube arises, and clusters decay promptly into hadrons. More recently, “space-time aware” colour reconnection models in HERWIG have begun to attach production vertices to partons and minimize a boost-invariant distance measure, thus providing a first dynamical implementation of hadronization geometry within the cluster framework [35].

2.2.3 Statistical hadronization

The first use of a statistical approach to describe the production of hadrons in high-energy collisions dates back to the 1950s and 1960s [36, 37]. The statistical hadronization model (SHM) assumes that, as a consequence of strong-interaction dynamics, the outcome of a high-energy collision can be described in terms of a set of color-singlet clusters (or fireballs). Each cluster is characterized by mass, volume, internal quantum numbers, and momentum, and their distributions are determined by the dynamics of the preceding partonic stage.

The key feature of SHM is that clusters are assigned a finite volume, allowing the use of statistical mechanics and thermodynamic quantities such as temperature. Importantly, this does not imply thermalization by rescattering after hadron formation. Instead, it is an intrinsic property of hadronization: all allowed quantum states of the system are assumed to be equally probable (microcanonical postulate). Early formulations [38, 39] expressed the SHM in a canonical framework, with clusters described by a temperature T and volume V , rather than fixed energy and momentum. Later work [40] showed that an ensemble of many small clusters can be effectively replaced by a single *Equivalent Global Cluster* (EGC), allowing canonical thermodynamics to be applied consistently at the global level. A schematic illustration of the statistical hadronization picture, from the formation of color-singlet clusters to their replacement by an Equivalent Global Cluster (EGC) and subsequent hadron emission, is shown in Fig. 2.3.

Formally, the number of accessible hadronic states for a system with total four-momentum P , conserved charges $\mathbf{Q}$ (baryon number, electric charge, strangeness, etc.),

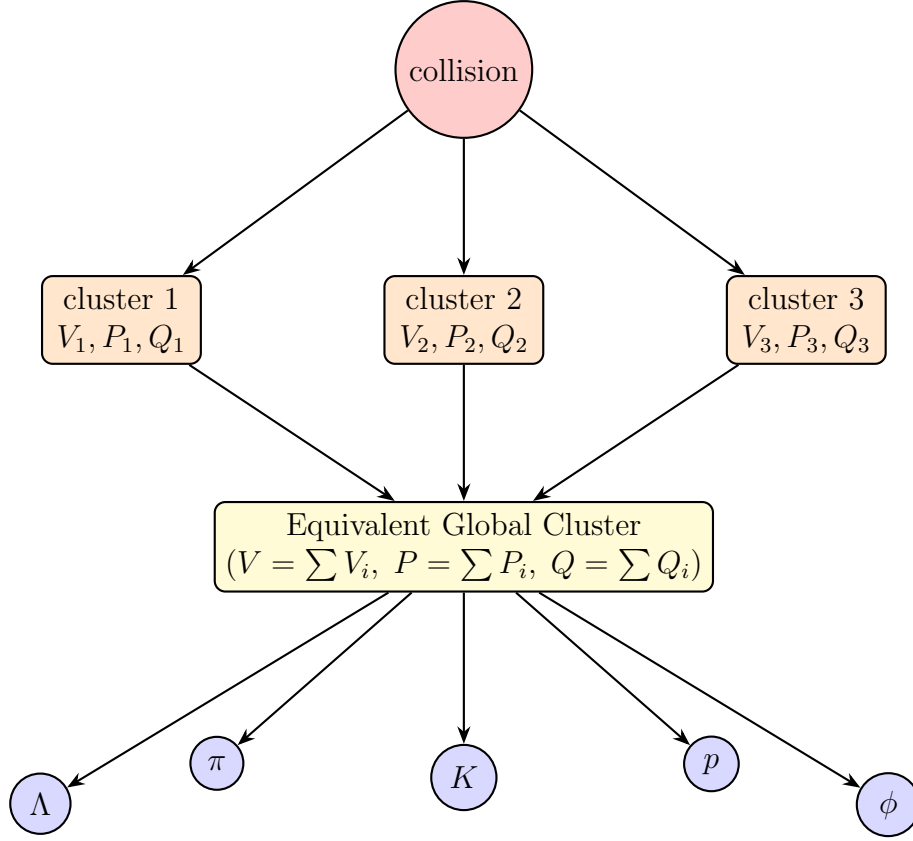


Figure 2.3: Schematic representation of the Statistical Hadronization Model. A high-energy collision produces color-singlet clusters with given volume V_i , four-momentum P_i , and conserved charges Q_i . These can be replaced by an Equivalent Global Cluster (EGC) with total V , P , and Q . Hadrons are then produced statistically from the EGC according to available phase space and conservation laws.

and volume V , is given in the microcanonical ensemble by

$$\Omega(P, \mathbf{Q}, V) = \lim_{\varepsilon \rightarrow 0} \frac{1}{(2\pi)^{4+n}} \int_{-\infty-i\varepsilon}^{+\infty-i\varepsilon} dx^0 \int d^3\mathbf{x} \int_{-\pi}^{\pi} d^n\phi \exp \left[i(P \cdot x + \mathbf{Q} \cdot \phi) + F(x, \phi) \right], \quad (2.8)$$

$$F(x, \phi) = \frac{V}{(2\pi)^3} \sum_j (2J_j + 1) \int d^3p \log(1 \pm e^{-i(p \cdot x + q_j \cdot \phi)})^{\pm 1}, \quad (2.9)$$

where the $\pm$ signs distinguish fermions and bosons. Here, Ω counts all possible multi-hadron states consistent with the exact conservation of energy–momentum and charges. The function $F(x, \phi)$ encodes the single-particle contributions of species j , weighted by spin degeneracy $2J_j + 1$.

In the limit of large energy P^0 and volume V , the microcanonical density of states reduces to the canonical partition function

$$Z(\beta, \mathbf{Q}, V) = \int d^4P e^{-\beta \cdot P} \Omega(P, \mathbf{Q}, V), \quad \beta^\mu = (1/T, \mathbf{0}), \quad (2.10)$$

so that a temperature T emerges as a well-defined parameter. This connects the statistical

formulation of hadronization with standard thermodynamics: the system behaves as if it were in thermal equilibrium at temperature T .

From this canonical partition function one can compute mean primary multiplicities. For a hadron species j , the average yield is

$$\langle n_j \rangle_{\text{primary}} = \frac{V T (2J_j + 1)}{2\pi^2} \sum_{K=0}^{\infty} \frac{e^{-\langle s\bar{s} \rangle} \langle s\bar{s} \rangle^K}{K!} \sum_{n=1}^{\infty} (\mp 1)^{n+1} \frac{m_j^2}{n} K_2\left(\frac{nm_j}{T}\right) \frac{Z(\mathbf{Q} - q_j)}{Z(\mathbf{Q})}. \quad (2.11)$$

This expression can be understood term by term:

- m_j and J_j are the mass and spin of hadron j ; the prefactor encodes its phase-space weight.
- The inner sum ($n = 1, 2, \dots$) implements Bose–Einstein or Fermi–Dirac statistics. In the Boltzmann limit one keeps only $n = 1$.
- K_2 is the modified Bessel function of the second kind, which arises from integrating over the thermal momentum distribution.
- The ratio $Z(\mathbf{Q} - q_j)/Z(\mathbf{Q})$ enforces exact conservation of discrete charges.
- The outer Poisson sum describes fluctuations in the number of $s\bar{s}$ pairs, which is a way of modeling the observed suppression of strangeness production.

Measured multiplicities must also include the decay of unstable resonances. Thus, the final average multiplicity is obtained as

$$\langle n_j \rangle_{\text{final}} = \langle n_j \rangle_{\text{primary}} + \sum_k \langle n_k \rangle_{\text{primary}} \text{Br}(k \rightarrow j), \quad (2.12)$$

where the second term accounts for all feed-down contributions via branching ratios.

The SHM has been successfully applied to inclusive hadron multiplicities measured in e^+e^- annihilation at LEP energies ($\sqrt{s} = 91.2$ GeV) as well as at lower energies ($\sqrt{s} = 14, 22, 29$, and 35 GeV). The data sets include a broad spectrum of hadrons: charged and neutral pions, kaons, protons, Λ , ϕ , and heavier strange baryons, as compiled by the LEP collaborations (ALEPH, DELPHI, L3, OPAL) and earlier experiments at PETRA and PEP. The main findings can be summarized as follows:

- The hadronization temperature extracted from all e^+e^- data is essentially universal, $T \simeq 165\text{--}170$ MeV, in agreement with results from hadronic and nuclear collisions.
- The strangeness suppression factor γ_s is consistently $\gamma_s \simeq 0.65$ in e^+e^- , indicating incomplete equilibration of strange quarks in elementary reactions.
- The effective cluster volume V grows with center-of-mass energy. At LEP energies (91.2 GeV), one finds $V \simeq 21 \text{ fm}^3$, while at lower energies ($\sqrt{s} \sim 14\text{--}35$ GeV) volumes are $V \simeq 8\text{--}14 \text{ fm}^3$.
- The growth of V with $\sqrt{s}$ reflects the increasing multiplicity and entropy of the hadronizing system, while T remains constant.

$\sqrt{s}$ [GeV]	Experiment(s)	V [fm ³]
14	TASSO, MARK II	8.2 ± 1.0
22	TASSO	10.5 ± 1.2
29	HRS, MARK II	12.0 ± 1.3
35	TASSO	13.6 ± 1.5
91.2	ALEPH, DELPHI, L3, OPAL	21.0 ± 2.0

Table 2.1: Fitted cluster volumes in e^+e^- collisions from SHM analysis.

The fitted EGC volumes obtained from statistical-model analysis are summarized in Table 2.1. Errors correspond to those quoted in the fits and are typically dominated by systematic uncertainties.

In summary, e^+e^- annihilation data across a wide range of energies are consistently described by the statistical hadronization model with a *universal* hadronization temperature $T \simeq 165$ MeV. The only significant energy dependence appears in the effective volume parameter V , which increases with $\sqrt{s}$ in line with the overall rise of multiplicity. This provides strong evidence that hadronization is a universal statistical process governed by a fixed temperature, with V acting as the extensive scaling parameter of the system.

2.2.4 Coalescence model

The coalescence (or recombination) model provides a widely used framework for describing hadron formation in high-energy nuclear collisions [41, 42]. In this picture, hadrons are not produced by the independent fragmentation of single partons, but rather by nearby quarks in phase space combining into color-singlet bound states. This mechanism becomes relevant when the density of partons at hadronization is sufficiently large that quarks can “find” partners within their spatial and momentum range. A schematic picture describing the Wigner-overlap criterion used in the coalescence formalism is shown in figure 2.4.

The essential ingredient of the model is the overlap between the quark phase-space distribution at freeze-out and the Wigner function of the hadron wave function. To make this explicit, one separates the center-of-mass and relative coordinates of the n constituent quarks. Let the total hadron momentum be $\mathbf{P} = \sum_i \mathbf{p}_i$, and define

$$\mathbf{R} = \sum_{i=1}^n \frac{m_i}{M} \mathbf{r}_i, \quad \boldsymbol{\rho}_i = \mathbf{r}_i - \mathbf{R}, \quad \mathbf{k}_i = \mathbf{p}_i - \frac{m_i}{M} \mathbf{P},$$

with $M = \sum_i m_i$. Here $\mathbf{R}$ and $\mathbf{P}$ denote the hadron center-of-mass position and momentum, while $\{\boldsymbol{\rho}_i, \mathbf{k}_i\}$ describe the internal degrees of freedom. The hadron Wigner function $W_H(\{\boldsymbol{\rho}_i, \mathbf{k}_i\})$ encodes the spatial and momentum-space structure of the hadron wave function, normalized to unity.

In this formalism, the invariant yield of hadron species H is

$$E \frac{d^3 N_H}{d^3 P} = g_H \int d^3 \mathbf{R} \prod_{i=1}^n \left[\frac{d^3 \boldsymbol{\rho}_i d^3 \mathbf{k}_i}{(2\pi)^3} \right] \delta^{(3)}\left(\sum_j \boldsymbol{\rho}_j\right) \delta^{(3)}\left(\sum_j \mathbf{k}_j\right) \times \left[\prod_{i=1}^n f_q\left(\mathbf{R} + \boldsymbol{\rho}_i, \frac{m_i}{M} \mathbf{P} + \mathbf{k}_i\right) \right] W_H(\{\boldsymbol{\rho}_i, \mathbf{k}_i\}). \quad (2.13)$$

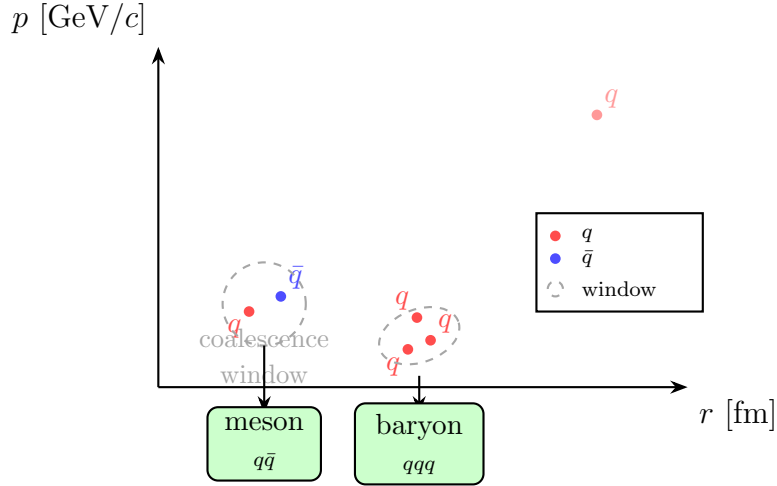


Figure 2.4: Quarks shown as points in the (r, p) phase space. When quarks (or a quark and an antiquark) are close in both position and momentum, they fall inside a schematic “coalescence window” (dashed curve) and recombine into a hadron: a meson ($q\bar{q}$) or a baryon (qqq). This visualizes the Wigner-overlap criterion used in the coalescence formalism.

where g_H is the statistical degeneracy of the hadron and $f_q(\mathbf{r}, \mathbf{p})$ is the quark phase-space distribution at freeze-out. The two δ -functions ensure that internal coordinates and momenta are defined relative to the hadron center-of-mass.

This formulation naturally explains a number of features observed in heavy-ion collisions. For instance, the enhanced baryon-to-meson ratios at intermediate transverse momentum arise because forming a baryon requires three quarks to be close in phase space, whereas a meson requires only two. The larger phase-space volume available for baryon formation at intermediate p_T leads to the observed baryon enhancement relative to expectations from fragmentation.

In non-central heavy-ion collisions, the initial spatial anisotropy of the overlap region between the two nuclei leads to an anisotropic collective expansion of the produced medium, resulting in an anisotropy in the transverse momentum distribution of emitted particles. This momentum-space anisotropy is commonly quantified by the elliptic flow coefficient v_2 , defined as the second moment of the azimuthal momentum distribution. Within the coalescence model, this collective behavior at the partonic level naturally leads to a scaling of v_2 with the number of constituent quarks.

If quarks already carry an anisotropic collective flow, then the hadron elliptic flow approximately follows

$$v_2^H(p_T) \simeq n_q v_2^q\left(\frac{p_T}{n_q}\right), \quad (2.14)$$

where n_q is the number of valence quarks in the hadron. This “constituent-quark scaling” has been confirmed experimentally at RHIC and the LHC [43, 44, 45].

It should be emphasized that coalescence is not generally invoked for hadronization in elementary processes such as e^+e^- annihilations. In such systems, the parton density is too low for quarks to find nearby partners in phase space, and independent fragmentation models such as the Lund string [29] or cluster model [34] successfully describe the data. Nonetheless, there exist particular cases in which elementary reactions can temporarily generate higher parton densities. A notable example is provided by the hadronic decays of the narrow $\Upsilon(nS)$ states, which predominantly proceed through three-gluon emission.

The resulting gluon cascades increase the local parton density before hadronization, potentially making coalescence mechanisms relevant. This point is revisited in detail in the following section.

In summary, the coalescence model provides a simple yet powerful tool to connect the quark-level dynamics of the deconfined medium to observed hadron spectra and flow observables in heavy-ion collisions, thereby complementing fragmentation-based approaches to hadronization.

2.3 Hadronic decays of the narrow $\Upsilon(nS)$ states

The three narrow vector bottomonia $\Upsilon(nS)$, with $n = 1, 2, 3$, lie below the open- $B\bar{B}$ threshold and therefore decay predominantly via annihilation into light hadrons or via transitions to lower bottomonium states. The hadronic annihilation modes proceed through three gluons (ggg), a gluon-gluon-photon ($gg\gamma$), or a virtual photon ($\gamma^* \rightarrow q\bar{q}$) followed by hadronization. The corresponding partial widths scale, to leading order, as

$$\Gamma(\Upsilon \rightarrow ggg) \propto \alpha_s^3, \quad \Gamma(\Upsilon \rightarrow gg\gamma) \propto \alpha \alpha_s^2, \quad \Gamma(\Upsilon \rightarrow \gamma^* \rightarrow q\bar{q}) \propto \alpha^2, \quad (2.15)$$

which explains the observed hierarchy $\mathcal{B}(ggg) \gg \mathcal{B}(gg\gamma) \gtrsim \mathcal{B}(\gamma^* \rightarrow q\bar{q})$ in the $\Upsilon(1S)$, with typical PDG averages $\mathcal{B}(\Upsilon(1S) \rightarrow ggg) \simeq 82\%$ and $\mathcal{B}(\Upsilon(1S) \rightarrow gg\gamma) \simeq 2\text{--}3\%$ [46]. Similar hadronic patterns hold for the $\Upsilon(2S)$ and $\Upsilon(3S)$, though their total widths include sizable hadronic transitions (e.g. $\Upsilon(2S, 3S) \rightarrow \Upsilon(1S) \pi\pi$) that compete with direct annihilation [47, 48]. The diagrams representing these decay patterns are shown in Fig. 2.5.

A particularly sensitive probe of hadronization dynamics is the production of light nuclei and antinuclei. The coalescence picture predicts that an antideuteron ($\bar{d}$) forms when an antiproton and antineutron emerge nearby in phase space from the same event; qualitatively, environments with higher baryon-antibaryon phase-space density (as expected in gluon-rich ggg decays) should enhance $\bar{d}$ production relative to fragmenting $q\bar{q}$ jets. This expectation is corroborated experimentally. CLEO first established that $\bar{d}$ production is enhanced in $\Upsilon(nS)$ ($n = 1, 2, 3$) decays relative to the nearby continuum $e^+e^- \rightarrow q\bar{q}$ at $\sqrt{s} \sim 10$ GeV, reporting branching ratios at the $\text{few} \times 10^{-5}$ level and setting only upper limits for the continuum [49]. More recently, BABAR measured

$$\mathcal{B}(\Upsilon(1S) \rightarrow \bar{d}X) = (2.81 \pm 0.49 \text{ (stat)} {}^{+0.20}_{-0.24} \text{ (syst)}) \times 10^{-5}, \quad (2.16)$$

$$\mathcal{B}(\Upsilon(2S) \rightarrow \bar{d}X) = (2.64 \pm 0.11 \text{ (stat)} {}^{+0.26}_{-0.21} \text{ (syst)}) \times 10^{-5}, \quad (2.17)$$

$$\mathcal{B}(\Upsilon(3S) \rightarrow \bar{d}X) = (2.33 \pm 0.15 \text{ (stat)} {}^{+0.31}_{-0.28} \text{ (syst)}) \times 10^{-5}, \quad (2.18)$$

and, at the continuum energy $\sqrt{s} \simeq 10.58$ GeV,

$$\sigma(e^+e^- \rightarrow \bar{d}X) = (9.63 \pm 0.41 \text{ (stat)} {}^{+1.17}_{-1.01} \text{ (syst)}) \text{ fb}, \quad (2.19)$$

confirming the relative enhancement in the gluon-rich bottomonium environment [50]. Together, these measurements demonstrate that the initial partonic configuration (ggg vs. $\gamma^* \rightarrow q\bar{q}$) has a sizeable impact on final-state cluster formation and light-nucleus coalescence in e^+e^- collisions at ~ 10 GeV.

For detailed and regularly updated branching-fraction tables and mode inventories, the reader can refer to the PDG particle listings for $\Upsilon(1S)$, $\Upsilon(2S)$ and $\Upsilon(3S)$ [46, 47, 48].

The Belle experiment has accumulated large datasets at the bottomonium resonances and in the nearby continuum regions. According to Table 1 of Ref. [51], the integrated luminosities and the corresponding number of events are as follows.

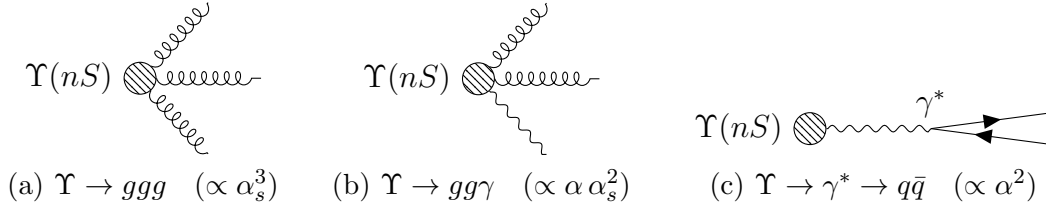


Figure 2.5: Leading hadronic annihilation topologies of the narrow $\Upsilon(nS)$ states: (a) three gluons (ggg), (b) two gluons plus a photon ($gg\gamma$), and (c) a virtual photon converting to a light quark pair ($\gamma^* \rightarrow q\bar{q}$). The indicated scaling with the couplings α_s and α explains the observed hierarchy of partial widths discussed in the text.

At the $\Upsilon(1S)$ resonance, Belle collected an integrated luminosity of 5.7 fb^{-1} , corresponding to approximately 1.0×10^8 $\Upsilon(1S)$ decays. At the $\Upsilon(2S)$, the dataset amounts to 24.9 fb^{-1} , yielding about 1.6×10^8 decays. In addition, continuum data were taken about 60 MeV below each resonance, with integrated luminosities of 1.7 fb^{-1} in both cases. Assuming a typical hadronic cross section $\sigma(e^+e^- \rightarrow q\bar{q}) \sim 3 \text{ nb}$ at these energies, this corresponds to about 5×10^6 continuum events per sample.

A summary is provided in Table 2.2.

Table 2.2: Belle datasets at $\Upsilon(nS)$, $n = 1, 2, 3$, including resonance decays and continuum events. Resonance yields are taken from Ref. [51]; continuum yields are estimated using $\sigma(e^+e^- \rightarrow q\bar{q}) \sim 3 \text{ nb}$.

Dataset	Luminosity [fb^{-1}]	Resonance decays	Continuum $q\bar{q}$ events
$\Upsilon(1S)$	5.7	$\sim 1.0 \times 10^8$	$\sim 1.7 \times 10^7$
$\Upsilon(2S)$	24.9	$\sim 1.6 \times 10^8$	$\sim 7.5 \times 10^7$
$\Upsilon(3S)$	2.9	$\sim 1.1 \times 10^7$	$\sim 8.7 \times 10^6$

2.4 The femtoscopy technique

Clearly, the spatio-temporal structure of the particle emitting source in relativistic collisions cannot be accessed directly due to the microscopic space and time scales involved ($\sim 10^{-14} \text{ m}$ and $\sim 10^{-22} \text{ s}$). Instead, one exploits correlations among particles in the final state, a technique generically termed *femtосcopy* [4]. This method descends conceptually from the Hanbury Brown–Twiss (HBT) intensity interferometry of photons emitted by stars [52], but has been adapted to strongly interacting, short-lived hadronic sources where final-state interactions and collective expansion play crucial roles. In modern usage, femtoscopy encompasses not only identical-particle Bose–Einstein correlations, but also correlations induced by Coulomb and strong final-state interactions, as well as coalescence analyses.

2.4.1 Hanbury Brown–Twiss intensity interferometry

The Hanbury Brown–Twiss (HBT) effect, or intensity interferometry, is a technique used to extract information about the spatial and temporal structure of a source by studying correlations in the detected intensity of emitted radiation. Unlike traditional interferom-

etry, which relies on measuring the phase coherence of waves, HBT relies on measuring *intensity correlations*, that is, coincidences between detections at different points or times.

The method was introduced in the 1950s by Hanbury Brown and Twiss, who used it to determine the angular size of stars [53]. By placing two detectors at different positions and recording the correlated fluctuations in the received light intensity, they could infer the spatial extent of the emitting source. The discovery that such correlations reveal geometrical information was a major breakthrough and laid the foundation of modern intensity interferometry.

The basic observable is the normalized second-order correlation function

$$g^{(2)}(\mathbf{r}_1, \mathbf{r}_2; \tau) = \frac{\langle I(\mathbf{r}_1, t) I(\mathbf{r}_2, t + \tau) \rangle}{\langle I(\mathbf{r}_1, t) \rangle \langle I(\mathbf{r}_2, t) \rangle},$$

which measures the probability of detecting two particles (or photons) at spacetime points $(\mathbf{r}_1, t)$ and $(\mathbf{r}_2, t + \tau)$, normalized to the product of single detection probabilities. For chaotic or thermal sources, this function is larger than unity at small relative separations, a phenomenon known as *bunching*. The width of the correlation peak reflects the coherence length of the source and can be related, via Fourier transforms, to the source size.

In quantum terms, the effect arises because two indistinguishable bosons (such as photons or pions) can be exchanged without altering the overall quantum amplitude. The interference between these two indistinguishable processes increases the joint detection probability, leading to the observed correlation peak [54]. For fermions, the Pauli exclusion principle instead leads to a suppression, or *antibunching*, at small relative separations.

The power of HBT interferometry lies in its generality: it does not require control of phases or coherent sources, and it can be applied to any chaotic emitter. This has made it a versatile tool, first in astronomy, then in quantum optics, and later in nuclear and particle physics. In relativistic heavy-ion collisions, HBT analysis of identical particles (typically pions) provides information on the size and lifetime of the particle-emitting region at thermal freeze-out [55, 56]. By fitting the correlation function, one can extract so-called “HBT radii,” which characterize the spatial extension and temporal evolution of the source.

2.4.2 Two-particle correlation function

The basic experimental observable in femtoscopy is the *two-particle correlation function*. It is defined as the ratio of the measured two-particle distribution to the product of single-particle distributions:

$$C(\mathbf{p}_1, \mathbf{p}_2) = \frac{\frac{dN_{12}}{d^3p_1 d^3p_2}}{\frac{dN_1}{d^3p_1} \frac{dN_2}{d^3p_2}}, \quad (2.20)$$

where dN_{12} is the inclusive yield of particle pairs, and dN_1, dN_2 are the single-particle spectra. By construction, $C = 1$ corresponds to uncorrelated emission, while deviations from unity signal the presence of correlations.

The two-particle correlation function is sensitive to the space–time separation of the emission points through two main mechanisms:

- **Quantum statistical effects:** for identical bosons (e.g. pions), the indistinguishability of the emission processes leads to an enhanced probability of detecting the particles close together in momentum space (Bose–Einstein enhancement). For identical fermions, the Pauli exclusion principle produces a suppression at small relative momentum.
- **Final-state interactions:** the mutual Coulomb repulsion (or attraction) and possible strong interactions between the two particles distort the correlation function, adding sensitivity to the interaction potential as well as to the source geometry.

In summary, the two-particle correlation function provides a direct link between measurable momentum-space observables and the underlying space–time and interaction dynamics of the emitting source, forming the starting point for femtoscopic analyses.

More formally, in the absence of final-state interactions the two-particle correlation function can be written schematically as

$$C(\vec{q}, \vec{K}) = 1 + \frac{\left| \int d^4x S(x, K) e^{iq \cdot x} \right|^2}{\left| \int d^4x S(x, K) \right|^2}, \quad (2.21)$$

where $q = p_1 - p_2$ and $K = (p_1 + p_2)/2$ denote the relative and average four-momenta of the particle pair, and $S(x, K)$ is the emission (source) function, describing the probability to emit a particle with momentum K from the space–time point $x = (t, \vec{x})$.

The phase factor $q \cdot x = q^0 t - \vec{q} \cdot \vec{x}$ explicitly couples the relative energy q^0 to the emission time difference. Since experimentally accessible correlations arise predominantly at small relative momentum, $|\vec{q}| \approx 0$, the corresponding energy difference $q^0 = E_1 - E_2$ is also small for on-shell particles. As a consequence, the time-dependent phase oscillates slowly and the correlation function becomes weakly sensitive to relative emission times.

This observation motivates the equal–time approximation, in which the temporal dependence of the source is integrated out and the correlation function is expressed in terms of an effective three–dimensional spatial emission function.

2.4.3 Hanbury Brown–Twiss radii

A central goal of femtoscopy is to extract information about the space–time structure of the particle-emitting source from the experimentally measured two-particle correlation function. In practice, this correlation function is often analyzed under the equal-time approximation and the simplifying assumption that the effective spatial emission region can be approximated by a Gaussian distribution.

Following these assumptions, the correlation function can be parameterized as [4]

$$C(\mathbf{k}^*, \mathbf{P}) = 1 + \lambda(\mathbf{P}) \exp \left[-R_o^2(\mathbf{P}) k_o^{*2} - R_s^2(\mathbf{P}) k_s^{*2} - R_l^2(\mathbf{P}) k_l^{*2} - 2R_{ol}^2(\mathbf{P}) k_o^* k_l^* \right], \quad (2.22)$$

where the relevant momenta are defined as:

- $\mathbf{k}^* = \frac{1}{2}(\mathbf{p}_1 - \mathbf{p}_2)$: the relative momentum of the pair in the pair rest frame,
- $\mathbf{P} = \frac{1}{2}(\mathbf{p}_1 + \mathbf{p}_2)$: the average momentum of the pair.

To facilitate interpretation, $\mathbf{k}^*$ is decomposed in the so-called “out–side–long” (OSL) coordinate system, depicted in Fig 2.6:

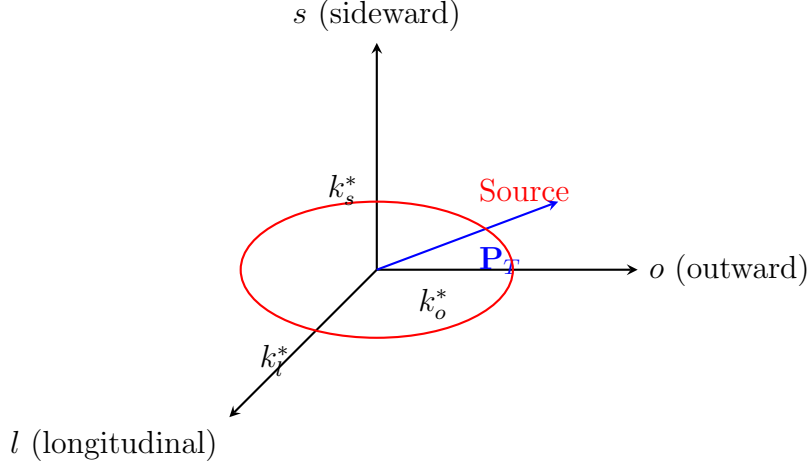


Figure 2.6: Schematic representation of the “out–side–long” (OSL) coordinate system used in femtoscopy analyses. The average pair momentum $\mathbf{P}$ defines the outward (o) direction in the transverse plane. The sideward (s) axis is orthogonal to it, while the longitudinal (l) axis follows the beam direction.

- the *outward* component (k_o^*) is defined along the transverse component of $\mathbf{P}$,
- the *sideward* component (k_s^*) is perpendicular to o in the transverse plane,
- the *longitudinal* component (k_l^*) is aligned with the beam axis.

The parameters R_o , R_s , and R_l are known as the *HBT radii*, while R_{ol} is a cross-term that becomes important in asymmetric or non-central collisions. These radii are not literal “sizes” of the entire fireball, but rather measure the widths of the so-called *region of homogeneity* — the portion of the source that actually emits particles with momentum $\mathbf{P}$. Different momenta probe different regions, so the $\mathbf{P}$ -dependence of the radii carries essential information on the dynamics of the system.

From a theoretical perspective, the HBT radii can be expressed in terms of the variances of the emission function $S(x, P)$, which describes the probability of emitting a particle with momentum P from space-time point x :

$$R_s^2(\mathbf{P}) = \langle \tilde{x}_s^2 \rangle, \quad (2.23)$$

$$R_o^2(\mathbf{P}) = \langle (\tilde{x}_o - \beta_T \tilde{t})^2 \rangle, \quad (2.24)$$

$$R_l^2(\mathbf{P}) = \langle (\tilde{x}_l - \beta_L \tilde{t})^2 \rangle. \quad (2.25)$$

Here $\tilde{x}_i = x_i - \langle x_i \rangle$ and $\tilde{t} = t - \langle t \rangle$ denote fluctuations around the average emission point, while β_T and β_L are the transverse and longitudinal components of the pair velocity.

These expressions allow us to interpret each radius:

- R_s reflects the transverse geometric size of the emission region;
- R_o combines spatial size with the emission duration $\Delta\tau$, since a longer emission time effectively increases the outward separation;
- R_l measures the extent of the source along the beam direction, strongly influenced by its longitudinal expansion;

- R_{ol} encodes correlations between outward and longitudinal directions, and becomes important in azimuthally asymmetric systems.

A schematic representation of the correlation function for two identical bosons is shown in Fig. 2.7.

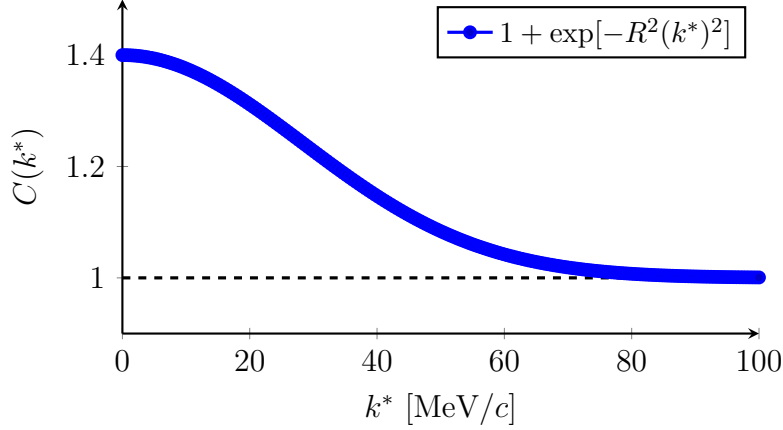


Figure 2.7: Schematic two-particle correlation function as a function of the relative momentum k^* . The enhancement at small k^* reflects Bose–Einstein correlations of identical particles. The width of the peak is inversely related to the HBT radii, which characterize the size of the homogeneity region in high-energy collisions.

In summary, the HBT radii extracted from the correlation function provide a quantitative link between experimental observables and the underlying space–time dynamics of the source. Their dependence on the average pair momentum $\mathbf{P}$ has been extensively exploited in high-energy hadronic and nuclear collisions, where they reveal signatures of collective flow and the expansion of the quark–gluon plasma [4].

2.4.4 Koonin–Pratt formalism

In the previous subsection, the HBT radii were introduced as a convenient Gaussian parametrization of the measured two-particle correlation function. While extremely useful, this parametrization is not fundamental: it does not explicitly show how the correlation function arises from the space–time distribution of emission points and from the quantum and final-state interactions between particles. A more rigorous connection is provided by the Koonin–Pratt formalism [57, 58].

Neglecting higher-order symmetrization and assuming independent emission, the correlation function for a pair of identical fermions can be written as

$$C(\mathbf{k}^*, \mathbf{P}) = 1 + \int d^4x_1 d^4x_2 S(x_1, \mathbf{P}) S(x_2, \mathbf{P}) \left[\left| \Psi_{\mathbf{k}^*}^{(+)}(x_1 - x_2) \right|^2 - 1 \right], \quad (2.26)$$

where $S(x, \mathbf{P})$ is the source function, i.e. the probability density for emitting a fermion with momentum $\mathbf{P}$ from space–time point x , and $\Psi_{\mathbf{k}^*}^{(+)}$ is the antisymmetric relative two-particle wave function including quantum statistics and final-state interactions (FSI)¹.

¹The superscript “+” in $\Psi^{(+)}$ denotes the choice of outgoing-wave boundary conditions in the scattering solution, as is standard in scattering theory. It is not related to Bose or Fermi symmetry; (anti)symmetrization is implemented separately depending on whether the particles are bosons or fermions.

Equation (2.26) makes explicit that the correlation function is sensitive not to the full source, but to the so-called *region of homogeneity*: the sub-region of the emission source that contributes to particles of a given average momentum $\mathbf{P}$ [59]. A schematic illustration is shown in Fig. 2.8.

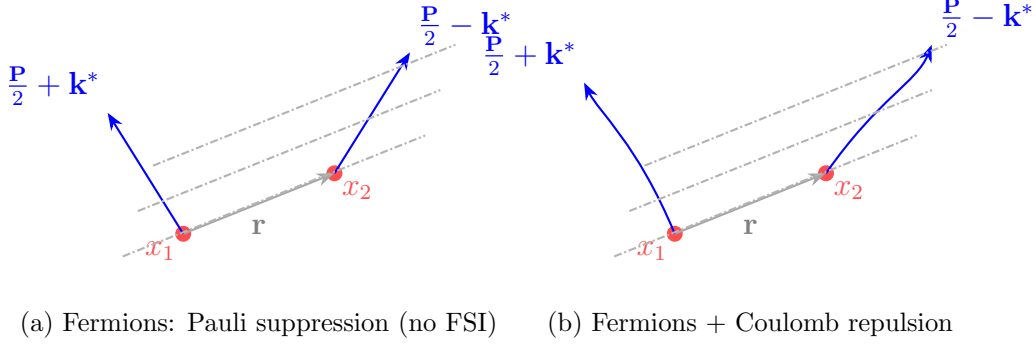


Figure 2.8: Two identical fermions emitted from points x_1 and x_2 (separation $\mathbf{r}$). **(a)** Without final-state interactions (FSI), antisymmetrization (Pauli principle) suppresses the correlation at small $\mathbf{k}^*$. **(b)** With Coulomb repulsion, the relative wave function is further modified, pushing trajectories apart and enhancing the suppression. In both panels, the average pair momentum is $\mathbf{P}/2$ and the relative momentum is $\mathbf{k}^*$.

The single-particle spectrum is obtained in the same framework by integrating the emission function $s(p, x)$ over space-time:

$$E \frac{dN}{d^3p} = \int d^4x s(p, x). \quad (2.27)$$

For a pair of fermions, the two-particle distribution generalizes to

$$E_a E_b \frac{dN_{ab}}{d^3p_a d^3p_b} = \int d^4x_a d^4x_b s(p_a, x_a) s(p_b, x_b) \left| \Psi_{\mathbf{k}^*}^{(+)}(x_a - x_b) \right|^2, \quad (2.28)$$

which reduces to Eq. (2.26) when expressed in terms of the source function $S(x, \mathbf{P})$. The factorization in Eq. (2.28) assumes that particles are emitted independently; any correlated or coherent emission would invalidate this approximation.

For identical fermions without FSI, the relative wave function is antisymmetric:

$$\Psi_{\mathbf{k}^*}^{(+)}(\mathbf{r}) = \frac{1}{\sqrt{2}} (e^{i\mathbf{k}^* \cdot \mathbf{r}} - e^{-i\mathbf{k}^* \cdot \mathbf{r}}),$$

so that the interference term in Eq. (2.28) leads to a *suppression* of the correlation function at small relative momentum. This Pauli blocking is the fermionic analogue of the Bose–Einstein enhancement for bosons.

Final-state interactions further modify the kernel $|\Psi_{\mathbf{k}^*}^{(+)}|^2$. Coulomb repulsion enhances the suppression for like-charged fermion pairs, whereas Coulomb attraction weakens it for oppositely charged pairs. At short distances, the strong nuclear force introduces additional modifications, including resonant enhancements that can provide direct sensitivity to scattering parameters of hadron–hadron interactions [4]. Thus, FSIs represent both a complication—since they distort the pure Pauli signal—and an opportunity, as they encode valuable dynamical information.

Several approximations underlie the connection between measured correlations and source geometry: the smoothness and equal-time approximations, the neglect of higher-order symmetrization, and the assumption of independent emission. In practice, resonance decays and collective expansion further blur the relation between femtoscopic radii and the global source size. Nevertheless, the Koonin–Pratt formalism provides the rigorous theoretical framework from which femtoscopic analyses of identical fermion pairs can be understood.

2.4.5 Nuclear scattering parameters

The Koonin–Pratt framework shows that the measured two-particle correlation function is built from the relative two-body wave function folded with the emission source. At small relative momentum, $k^* \equiv |\mathbf{k}^*|$, the strong nuclear interaction modifies the correlation in a particularly sensitive way. This makes femtoscopy a powerful tool for extracting the low-energy scattering properties of hadrons directly from experimental data.

At low k^* , the interaction is dominated by the s -wave ($L = 0$) component. Its behavior can be expressed through the *effective-range expansion* of the phase shift $\delta_0(k^*)$:

$$k^* \cot \delta_0(k^*) = -\frac{1}{a_0} + \frac{1}{2}d_0 k^{*2} + \mathcal{O}(k^{*4}). \quad (2.29)$$

Here a_0 is the *scattering length*, which quantifies the strength and sign of the interaction in the zero-momentum limit, and d_0 (often denoted r_0) is the *effective range*, which encodes finite-size corrections due to the spatial extent of the interaction region.

It is often convenient to recast this information in terms of the s -wave scattering amplitude,

$$f(k^*) = \frac{1}{k^* \cot \delta_0(k^*) - i k^*} = \frac{1}{-a_0^{-1} + \frac{1}{2}d_0 k^{*2} - i k^*}, \quad (2.30)$$

which is the quantity directly entering the kernels used in femtoscopy analyses.

The physical meaning of a_0 is particularly intuitive:

- $a_0 > 0$ usually corresponds to an attractive interaction. A very large positive a_0 indicates the presence of a near-threshold pole in the scattering amplitude, associated with a shallow bound state or resonance.
- $a_0 < 0$ reflects an overall repulsive interaction, although small negative values can also appear in weakly attractive systems that do not form bound states.

The effective range d_0 provides complementary information by characterizing how the interaction deviates from a purely zero-range form.

In femtoscopy, one typically fits experimental correlation functions with theoretical kernels that fold the source distribution with the scattering amplitude of Eq. (2.30). In this way, the analysis yields not only the femtoscopic source size but also the fundamental low-energy scattering parameters a_0 and d_0 . This dual sensitivity makes femtoscopy an invaluable tool for probing hadronic interactions, especially in cases where no direct scattering data exist.

2.4.6 Extracting parameters from the correlation function

Building on the Koonin–Pratt formalism, one often works with a simplified form of the correlation function in the pair rest frame. Under the equal-time and smoothness approximations, and neglecting higher-order symmetrization, it can be written as

$$C(k^*) = 1 + \int d^3r S(\mathbf{r}) (|\phi_{k^*}(\mathbf{r})|^2 - 1), \quad (2.31)$$

where $S(\mathbf{r})$ is the source distribution as a function of the distance between emission points, and $\phi_{k^*}(\mathbf{r})$ is the relative two-particle wave function including symmetrization and strong final-state interactions.

For source sizes much larger than the range of the nuclear interaction, the kernel depends on the s -wave scattering amplitude, which in the effective-range expansion takes the form introduced in equation 2.30. The relative wave function can then be approximated by

$$\phi_{k^*}(\mathbf{r}) \simeq e^{i\mathbf{k}^* \cdot \mathbf{r}} + \frac{f(k^*)}{r} e^{ik^* r}, \quad (2.32)$$

so that the correlation function acquires an explicit dependence on the scattering parameters a_0 and d_0 .

In practice, femtoscopic analyses proceed as follows:

1. Experimentally, the two-particle correlation function $C(k^*)$ is constructed from particle pairs at small relative momentum.
2. A parametrized source function $S(\mathbf{r})$ (most commonly Gaussian) is assumed, characterized by a radius parameter R .
3. The model correlation function, Eq. (2.31), is evaluated using the scattering amplitude (2.30) and the approximate wave function (2.32).
4. The experimental $C(k^*)$ is fitted by floating a_0 , d_0 , and R .

This procedure allows the simultaneous determination of the femtoscopic source size and of the low-energy scattering parameters. In this way, femtoscopy serves as both a space–time microscope of the emitting source and a precision tool for hadronic interaction studies, particularly valuable in cases where direct scattering data are unavailable.

A widely used analytic implementation of Eqs. (2.31)–(2.32) is the Lednický–Lyuboshits (LL) formalism [5, 60]. The LL model provides closed expressions for the correlation function under the assumption of a Gaussian source and incorporates both quantum statistics and final-state interactions via the scattering amplitude. Because of its simplicity and speed, the LL formalism has been the standard tool for femtoscopic analyses over the past decades.

It should be emphasized that the approximation leading to Eq. (2.32) assumes that the source radius R is much larger than the range of the strong interaction potential. This condition is usually satisfied in heavy-ion collisions, where $R \sim 3\text{--}6$ fm, but may break down for smaller systems such as pp , pA , or e^+e^- collisions, where the emitting source can be of order ~ 1 fm or less. In such cases, part of the emission occurs within the range of the nuclear force, and the simple asymptotic form of the scattering wave function is no longer valid.

	Lednický–Lyuboshits (LL)	CATS (numerical)
Method	Analytic approximation using the scattering amplitude $f(k^*)$ and a Gaussian source	Full numerical solution of the Schrödinger equation with strong + Coulomb potentials
Assumptions	Source radius $R \gg$ range of nuclear interaction; Gaussian $S(\mathbf{r})$	No restriction on source size; arbitrary $S(\mathbf{r})$ allowed
Strengths	Simple closed-form expressions; very fast; widely used in heavy-ion analyses	Model-independent treatment of finite-range effects; accurate for small sources
Limitations	Accuracy decreases for small systems ($R \lesssim 2\text{--}3$ fm); cannot treat complex source shapes	Computationally more expensive; requires realistic potentials as input
Applications	Heavy-ion collisions with $R \sim 3\text{--}6$ fm	Small systems (pp, pA, e^+e^-) with $R \sim 1$ fm; precision analyses

Table 2.3: Comparison of the analytic Lednický–Lyuboshits (LL) formalism and the numerical CATS framework for femtoscopy. LL is efficient and accurate for large sources, while CATS provides a full wave-function treatment needed for small systems or precision extractions of scattering parameters.

When this happens, the correlation function must be computed using the full two-body wave function obtained by solving the Schrödinger equation with the relevant strong and Coulomb potentials. While the LL formalism remains highly useful for large sources ($R \gtrsim 2\text{--}3$ fm), modern numerical frameworks such as CATS (Correlation Analysis Tool using the Schrödinger equation) have been developed to explicitly incorporate finite-range effects and realistic interaction potentials [61, 62]. These methods are particularly important for femtoscopy studies in small systems, where precision requires going beyond the large-source approximation. A schematic comparison of the LL and CATS approaches is given in Table 2.3.

2.4.7 The CATS framework

To overcome the limitations of analytic approximations such as the Lednický–Lyuboshits model, the dedicated CATS software has been developed at the Technische Universität München (TUM) [61, 62]. The key idea is to compute the two-particle correlation function directly from the numerical solution of the Schrödinger equation, thereby including the full structure of the interaction potential without relying on large-source assumptions.

For a pair of (anti-)protons, the measured correlation function $C_{\text{data}}(k^*)$ is compared with a theoretical correlation function $C_{\text{th}}(k^*; r_0)$ obtained from the convolution

$$C_{\text{th}}(k^*; r_0) = \int d^3\mathbf{r} S(\mathbf{r}; r_0) |\psi_{k^*}(\mathbf{r})|^2, \quad (2.33)$$

where $S(\mathbf{r}; r_0)$ is the normalized emission source and $\psi_{k^*}(\mathbf{r})$ is the relative two-particle wave function including both quantum statistics and final-state interactions. For simplic-

ity, $S(\mathbf{r})$ is often taken as a spherically symmetric Gaussian,

$$S(\mathbf{r}; r_0) = \frac{1}{(4\pi r_0^2)^{3/2}} \exp\left(-\frac{r^2}{4r_0^2}\right), \quad (2.34)$$

with a single parameter r_0 that characterizes the effective source size.

The relative wave function $\psi_{k^*}(\mathbf{r})$ is obtained by solving the radial Schrödinger equation with reduced mass μ ,

$$-\frac{\hbar^2}{2\mu} \frac{d^2 u_{k^*\ell}}{dr^2} + \left[V(r) + \frac{\hbar^2 \ell(\ell+1)}{2\mu r^2} \right] u_{k^*\ell}(r) = \frac{\hbar^2 k^{*2}}{2\mu} u_{k^*\ell}(r), \quad u_{k^*\ell}(r) = r R_{k^*\ell}(r), \quad (2.35)$$

where $V(r)$ is the strong nucleon–nucleon potential (e.g. Argonne *v18*), and the Coulomb interaction is included explicitly. The full wave function is reconstructed as a partial-wave sum, with antisymmetrization enforced for identical fermions and statistical weights applied to the spin–isospin channels.

Being I the iso-spin quantum number, S the spin, and L and J the angular momenta, the observable correlation function is then obtained as a weighted sum over all allowed quantum channels,

$$C(k^*) = \sum_{I,S,L,J} w(I, S, L, J) C_{I,S,L,J}(k^*), \quad (2.36)$$

where the weights w follow from spin multiplicities and angular-momentum degeneracies. For example, in the proton–proton system, the Pauli principle forbids some channels (such as the ${}^3S_1$ state), which is automatically taken into account.

The CATS framework thus provides a rigorous way to calculate the two-particle correlation function by solving the Schrödinger equation with realistic interaction potentials. This is particularly important for small sources ($r_0 \lesssim 1$ fm), where a substantial fraction of pairs is emitted at short distances inside the range of the nuclear interaction. In this regime, approximate asymptotic approaches fail, while CATS retains full sensitivity to the potential, allowing precision extractions of both femtoscopic source sizes and low-energy scattering parameters.

2.4.8 Using femtoscopy to decipher the nature of exotic hadrons

An application of the femtoscopy technique is presented here. Since many exotic candidates, such as the $X(3872)$, $T_{cc}(3875)$, $Z_c(3900)$, or hidden-charm pentaquarks, lie close to two-hadron thresholds, their properties are strongly influenced by hadron–hadron dynamics. Femtoscopic correlation functions measured in high-energy pp , pA , AA and, in principle, e^+e^- collisions can provide access to these dynamics and thereby help clarify the internal structure of such states.

A typical application is to extract scattering parameters of hadron pairs that may form molecular states. For example, the $D\bar{D}^*$ system relevant to the $X(3872)$ or the DD^* system relevant to the $T_{cc}(3875)$ can be studied via their correlation functions. The low-momentum behavior of the correlation function is sensitive to the scattering amplitude 2.30, so that the fitted values of the scattering length a_0 and effective range d_0 can be directly compared with the expectations for a shallow bound state. A large and positive a_0 accompanied by a small effective range supports the interpretation of a weakly bound hadronic molecule, while the absence of such features would argue against it.

The review of Liu *et al.* [63] highlights several concrete examples. For instance:

- The $D_s^*(2317)$ and related $D^{(*)}K$ systems: correlation functions can constrain the DK scattering parameters, testing the molecular hypothesis for these states.
- The $T_{cc}(3875)$ and $X(3872)$: femtoscopy provides sensitivity to the DD^* and $D\bar{D}^*$ interactions, essential for confirming their near-threshold molecular nature.
- The $Z_c(3900)$ and $Z_{cs}(3985)$: by analyzing $D\bar{D}^*$ and $D_s\bar{D}^*$ correlations, one can test whether these tetraquark candidates are dynamically generated from meson–meson interactions.
- The hidden-charm pentaquarks $P_c(4312)$, $P_c(4440)$, and $P_c(4457)$: femtoscopy of $\bar{D}^{(*)}\Sigma_c^{(*)}$ systems offers a rare probe into the heavy meson–baryon interactions that likely underlie their structure.

In summary, femtoscopic analyses of hadron–hadron correlations provide direct access to the effective interaction parameters that determine whether an exotic state is predominantly molecular. This method therefore complements spectroscopy and lattice QCD, and plays a decisive role in unraveling the nature of the rapidly growing spectrum of exotic hadrons.

2.4.9 Recent experimental results

In this section, the highlights of the most recent results in the femtoscopy sector are reviewed. The discussion begins with results obtained using heavy-ion collision data, for which the smoothness assumption is applicable. This includes the pp and $\bar{p}\bar{p}$ correlations measured in heavy ion collisions at STAR and ALICE.

Measurement of the $\bar{p}\bar{p}$ correlation function at STAR

In 2015, the STAR Collaboration at the Relativistic Heavy Ion Collider (RHIC) has reported the first direct measurement of the antiproton–antiproton interaction [64]. Using a dataset of $\sim 5 \times 10^8$ Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV, momentum correlations among $\bar{p}\bar{p}$ pairs were studied with the femtoscopy technique. The resulting correlation function for $\bar{p}\bar{p}$ pairs exhibits an enhancement at small relative momentum k^* , indicating that the interaction is *attractive*, as in the proton–proton case. The measured $C(k^*)$ is shown in Fig. 2.9.

Fitting with the Lednický–Lyuboshitz model [65] yields the s -wave scattering parameters for the $\bar{p}\bar{p}$ interaction:

$$f_0 = 7.41 \pm 0.19 \text{ (stat.)} \pm 0.36 \text{ (sys.) fm}, \quad (2.37)$$

$$d_0 = 2.14 \pm 0.27 \text{ (stat.)} \pm 1.34 \text{ (sys.) fm}, \quad (2.38)$$

where f_0 is the scattering length and d_0 the effective range. These values are consistent, within uncertainties, with those for proton–proton interactions ($f_0 \approx 7.8$ fm, $d_0 \approx 2.8$ fm) [66], thus supporting the invariance of strong interactions under charge conjugation, parity, and time reversal (CPT).

In addition to scattering parameters, the analysis also extracted the effective Gaussian source radii of emission points for (anti)protons. The fitted source size for protons is

$$R_{pp} = 2.75 \pm 0.01 \text{ (stat.)} \pm 0.04 \text{ (sys.) fm}, \quad (2.39)$$

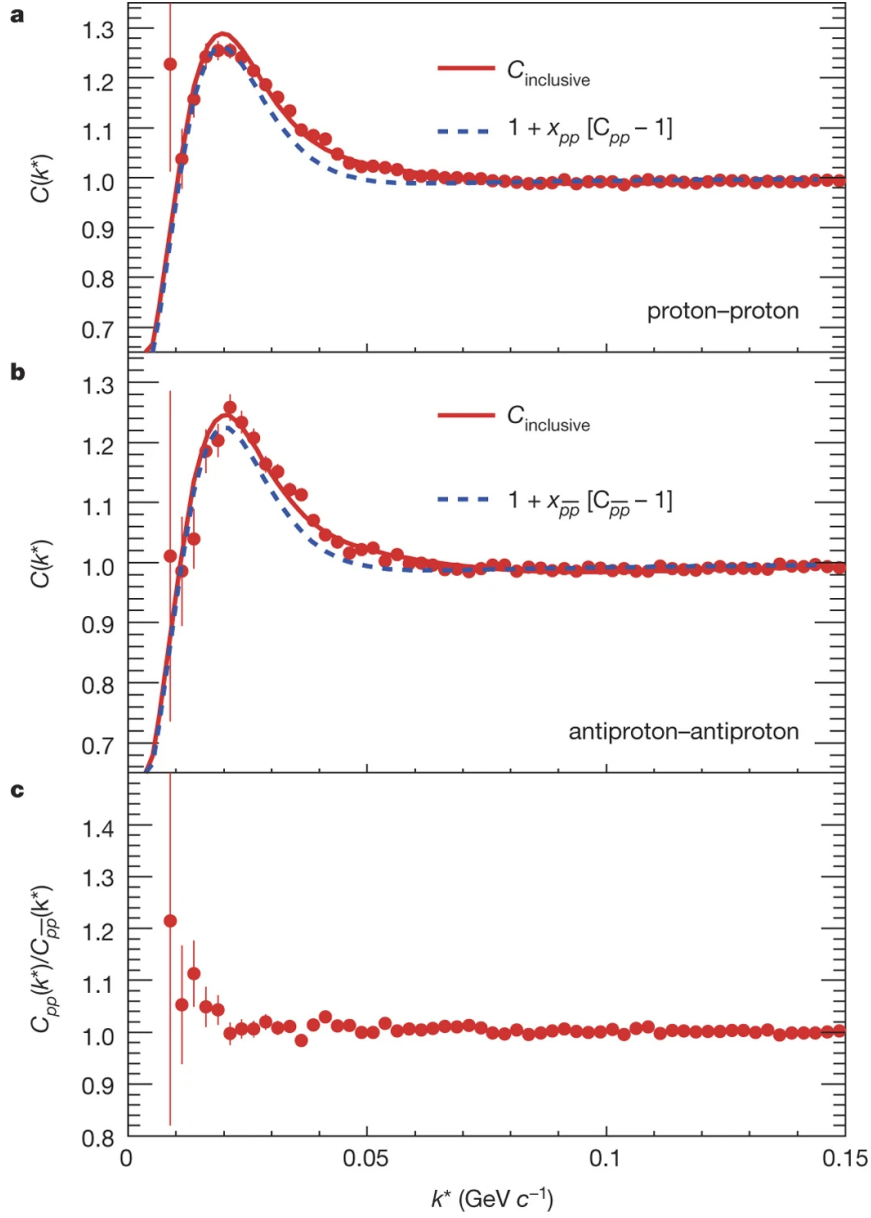


Figure 2.9: Correlation functions for proton–proton pairs (a) and antiproton–antiproton pairs (b). The ratio of the former to the latter is shown in c. Image taken from Ref. [64]. Errors are statistical only.

while for antiprotons it is

$$R_{\bar{p}\bar{p}} = 2.80 \pm 0.02 \text{ (stat.)} \pm 0.03 \text{ (sys.) fm.} \quad (2.40)$$

The agreement of these radii further corroborates the equivalence of proton and antiproton emission dynamics in relativistic heavy-ion collisions.

In summary, this measurement demonstrates that the antiproton–antiproton force is attractive, provides the first quantitative determination of its scattering length and effective range, and confirms consistency with the corresponding proton–proton values. This constitutes a direct verification of CPT symmetry in the baryonic sector of the strong interaction, and offers essential input for theoretical modeling of antinuclear structure and

for searches of antinuclei in cosmic environments.

Correlation between strange baryons

The STAR Collaboration has provided the first statistically significant measurement of the $\Lambda\Lambda$ interaction in heavy-ion collisions [67]. Using ~ 1 billion Au+Au collision events at $\sqrt{s_{NN}} = 200$ GeV recorded at RHIC, they employed two-particle intensity interferometry (femtoscopy) to construct the correlation function for $\Lambda\Lambda$ pairs. This technique, pioneered in nuclear collisions [65], allows access to the space-time characteristics of the emitting source as well as to the strong final-state interactions between hyperons.

The measured $\Lambda\Lambda$ correlation function shows an enhancement at small relative momentum, indicating an attractive interaction. By fitting the data with the Lednický–Lyuboshitz analytical model [65], the collaboration extracted the scattering parameters of the $\Lambda\Lambda$ system. The best-fit values correspond to a scattering length

$$a_{\Lambda\Lambda} = -1.10 \pm 0.37 \text{ (stat.)} \pm 0.30 \text{ (sys.) fm}, \quad (2.41)$$

and an effective range

$$r_{\Lambda\Lambda} = 8.52 \pm 2.56 \text{ (stat.)} \pm 2.09 \text{ (sys.) fm}. \quad (2.42)$$

These results imply a moderately attractive but weak $\Lambda\Lambda$ interaction, insufficient to produce a deeply bound dihyperon state (the H -dibaryon hypothesized by Jaffe [68]).

Furthermore, the extracted interaction strength is consistent with hypernuclear binding energies measured in double- Λ hypernuclei [69], and in qualitative agreement with recent lattice QCD calculations [70]. The femtoscopic source size was also obtained, yielding an effective Gaussian radius of $R_{\Lambda\Lambda} \approx 3$ fm, consistent with the size extracted from proton–proton and Λ –proton correlations in the same collision system.

The significance of this measurement extends beyond nuclear structure. A quantitative determination of hyperon interactions is essential for modeling hyperon-rich matter in neutron star interiors. In particular, the moderate attraction reported by STAR constrains the onset density of hyperons and helps address the so-called “hyperon puzzle” in neutron star equations of state [71].

In summary, the STAR Collaboration’s analysis of Au+Au collisions at RHIC represents the first direct experimental extraction of $\Lambda\Lambda$ scattering parameters from heavy-ion collisions. The observed attractive interaction rules out a deeply bound H -dibaryon, supports consistency with hypernuclear and lattice QCD results, and provides crucial input to the physics of dense hadronic matter.

Scaling of radii in high-multiplicity pp collisions

In 2011, ALICE has performed a three-dimensional two-pion femtoscopy of $\sqrt{s} = 0.9$ and 7 TeV pp events, differential in event multiplicity and pair transverse momentum k_T [72]. The measured two-pion correlation functions are shown in Fig. 2.10, while the extracted femtoscopic radii and their scaling behavior are summarized in Fig. 2.11.

The extracted HBT radii ($R_{\text{out}}, R_{\text{side}}, R_{\text{long}}$) exhibit three robust features: (i) at fixed k_T , all radii increase approximately linearly with the cube root of the mid-rapidity charged-particle density, $\langle dN_{\text{ch}}/d\eta \rangle^{1/3}$; (ii) at fixed multiplicity, the radii decrease with k_T (most clearly for R_{long} , and—with increasing multiplicity—also for $R_{\text{out}}, R_{\text{side}}$), signaling stronger space–momentum correlations; and (iii) the results are largely independent of $\sqrt{s}$

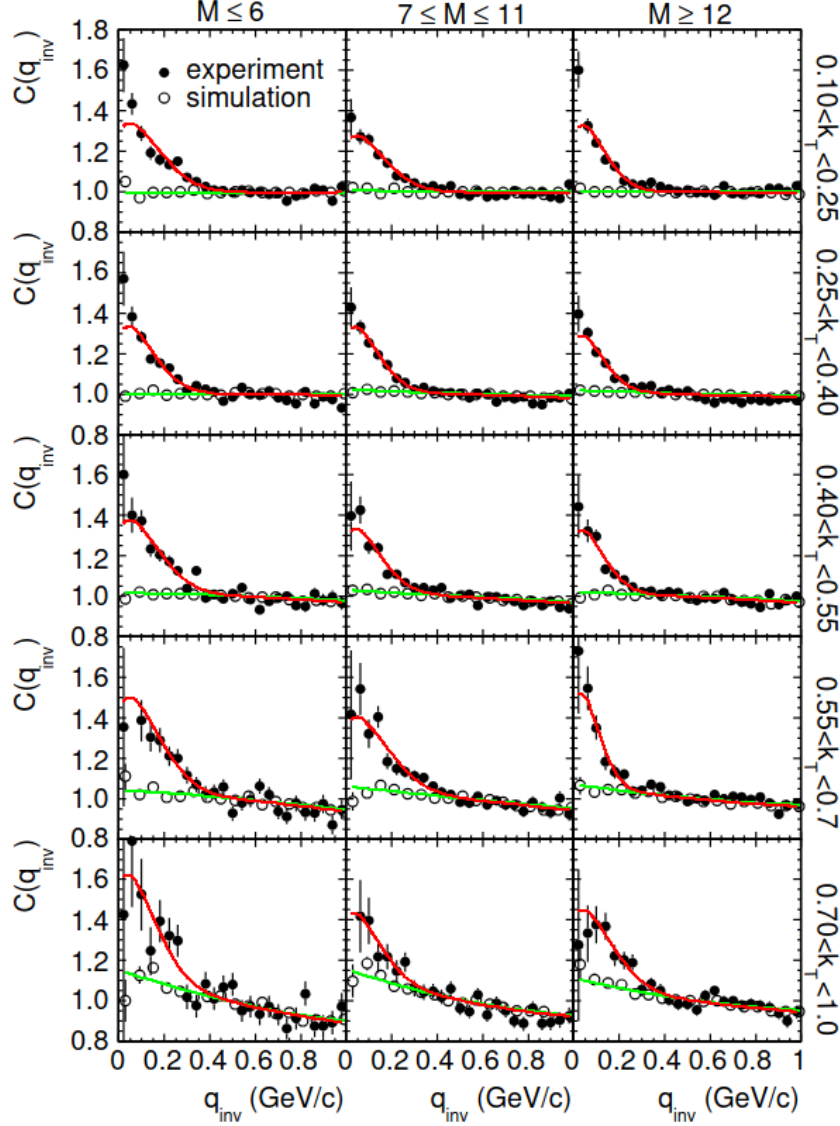
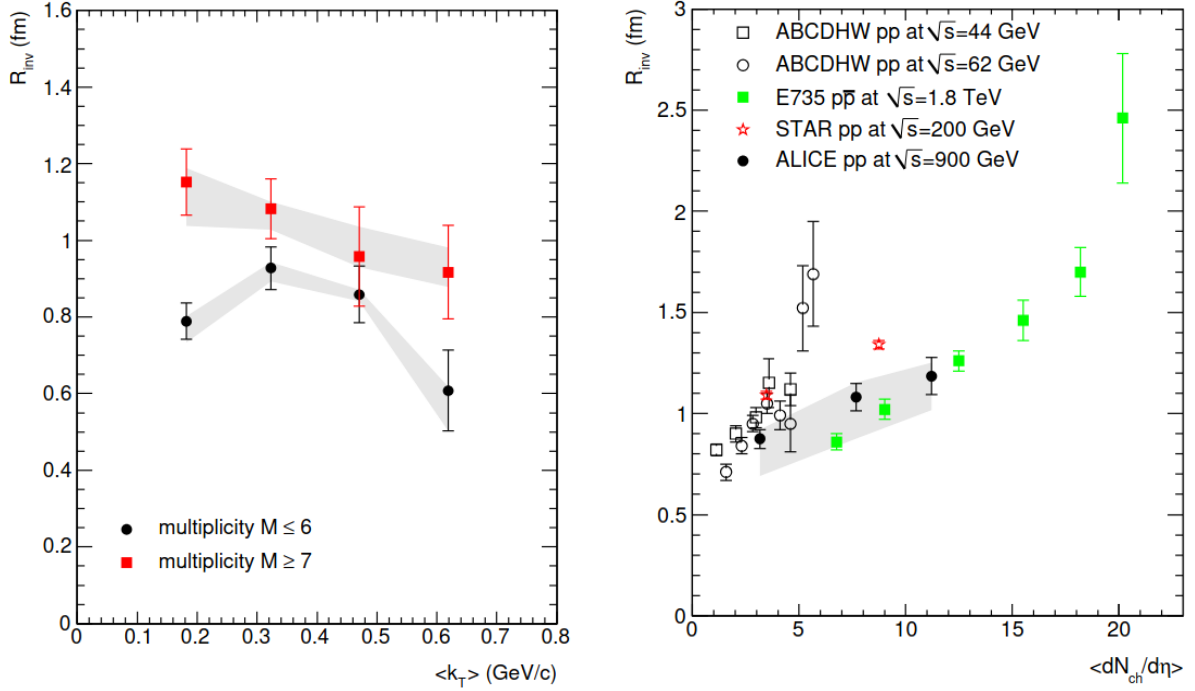


Figure 2.10: Two-pion Bose–Einstein correlation function measured in pp collisions (ALICE). $q_{\text{inv}} = 2k^*$ is the relative momentum between the two pions. The points show the experimental correlation function in bins of event multiplicity and pair transverse momentum, together with the corresponding reference (simulation) distribution and the fit curves. The figure is taken from Ref. [72].

once multiplicity and k_T are held fixed. Together, these trends suggest the development of collective-like dynamics in high-multiplicity pp events.

This scaling persists separately for each k_T bin, as visible in Fig. 2.11a, and is clearest for R_{side} and R_{long} . The slope in R_{out} shows a stronger k_T dependence at low multiplicity, but also tends toward a negative k_T slope as multiplicity rises. These multiplicity trends, illustrated in Fig. 2.11b, are consistent with an increasing homogeneity volume as particle production grows.

In heavy-ion (AA) data at RHIC/SPS, the radii from different systems and energies were found to fall on nearly the same straight lines when plotted vs. $\langle dN_{\text{ch}}/d\eta \rangle^{1/3}$ at fixed k_T (the so-called “universal” multiplicity scaling). ALICE shows that while pp radii do scale linearly with $\langle dN_{\text{ch}}/d\eta \rangle^{1/3}$, the *slope and offset* differ systematically from the AA



(a) Radius parameter as a function of $\langle k_T \rangle$ for different multiplicity classes.

(b) Radius parameter as a function of event multiplicity (charged-particle density).

Figure 2.11: Scaling of femtoscopic radius parameters extracted from two-pion Bose–Einstein correlations in pp collisions (ALICE), shown as a function of (a) the mean pair transverse momentum $\langle k_T \rangle$ and (b) the event multiplicity. The figure is taken from Ref. [72].

lines at the same k_T . Hence the heavy-ion “universal scaling” does *not* extend to pp —an explicit violation driven by system-specific physics rather than multiplicity alone.

Since pp and AA collisions can achieve comparable $\langle dN_{ch}/d\eta \rangle$ via very different initial-state geometries and particle-production mechanisms, the distinct $\langle dN_{ch}/d\eta \rangle^{1/3}$ slopes imply a geometric imprint on the final homogeneity lengths. ALICE argues that high-multiplicity pp events stem from a single, compact (proton-sized) production zone with subsequent rescattering, whereas in AA the same multiplicity arises from many nucleon–nucleon scatterings occurring at separated space–time points—naturally yielding larger radii at fixed multiplicity. Therefore, initial geometry must be accounted for in any scaling arguments.

The observed negative k_T slopes in R_{out} and R_{side} at higher multiplicity mirror the heavy-ion features often attributed to collective flow, lending support to space–momentum correlations in pp at high multiplicity. Earlier femtoscopy across systems hinted at similar multiplicity/ k_T patterns, but ALICE provides the first direct pp – AA comparison at overlapping multiplicities, clarifying where “universal” behavior breaks down and highlighting the role of system size/shape.

Femtосcopy in e^+e^- annihilation

A dedicated study of two-particle correlations in hadronic Z decays at LEP was performed by the ALEPH Collaboration using identified pp , $p\bar{p}$ and $K_S^0 K_S^0$ pairs [73]. The correlation function was measured as a function of the Lorentz-invariant relative momentum Q , and

interpreted in terms of quantum-statistical (Fermi–Dirac or Bose–Einstein) effects.

For pp and $\bar{p}\bar{p}$ pairs, ALEPH observes a depletion of pairs at low relative momentum, extending up to $Q \sim 3$ GeV, which is interpreted as a manifestation of Fermi–Dirac correlations. The measured correlation function for pp and $\bar{p}\bar{p}$ pairs is shown in Fig. 2.12.

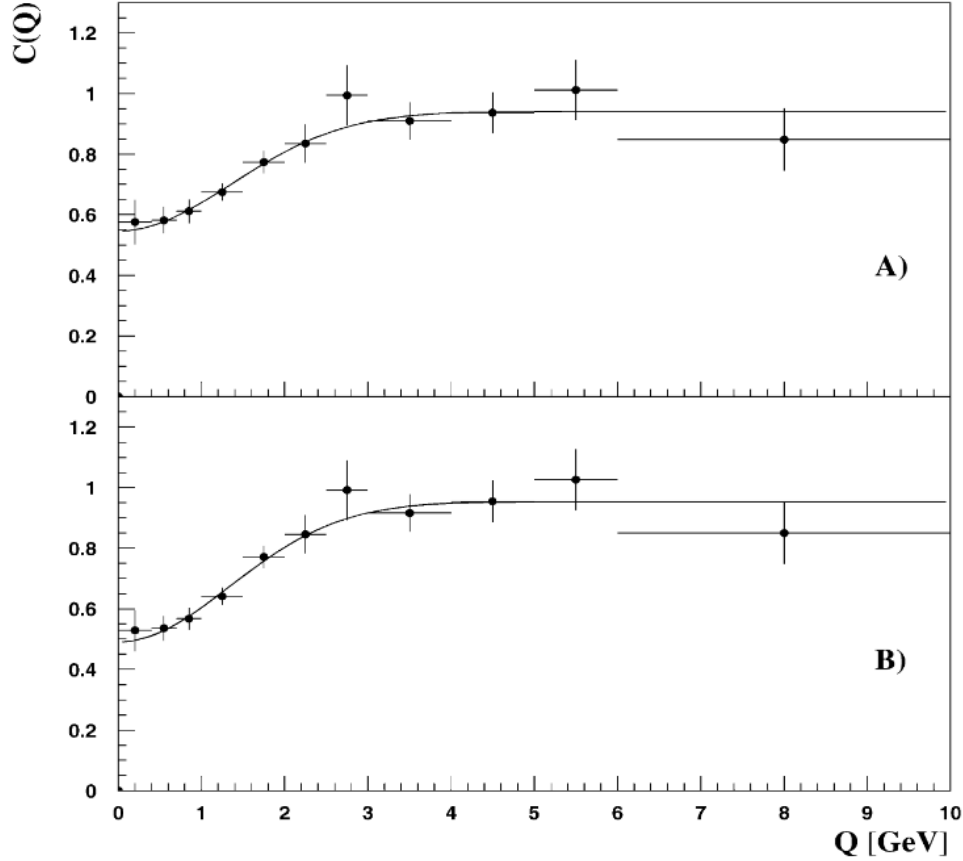


Figure 2.12: Correlation function $C(Q)$ for (A) pp and (B) $\bar{p}\bar{p}$ pairs in hadronic Z decays measured by the ALEPH Collaboration as a function of the Lorentz-invariant relative momentum Q . The solid curves represent Goldhaber-type fits performed over the full range $0 < Q < 10$ GeV. The figure is taken from Ref. [73].

For $K_S^0 K_S^0$ pairs, a low- Q enhancement is reported and attributed to Bose–Einstein symmetrization. The analysis employs both mixed-event and Monte Carlo–based reference samples and uses a double-ratio technique to mitigate biases associated with the choice of reference.

The measured correlation functions are fitted with a Goldhaber-type parameterization, assuming a Gaussian source and, in most cases, neglecting additional long-range terms. For baryon–baryon pairs, ALEPH reports an effective Gaussian radius of order $R \simeq 0.1$ fm, significantly smaller than typical hadronic length scales. While the qualitative observation of a low- Q depletion is robust, the quantitative interpretation of this scale requires careful scrutiny.

A first critical aspect of the analysis is the exceptionally large fit range in relative momentum. The correlation function for pp and $\bar{p}\bar{p}$ pairs is constructed and normalized over the interval $0 < Q < 10$ GeV, with the normalization fixed at very large values of Q . In this regime, non-femtoscopic correlations induced by global constraints—most notably

energy and momentum conservation and the jet-like structure of e^+e^- events—are known to generate long-range correlations in relative momentum. Such effects do not vanish at a few hundred MeV, but extend well into the multi-GeV region. Fitting the correlation function over such a wide range without an explicit model of these contributions inevitably couples the extracted femtoscopic parameters to the assumed baseline behavior.

A second limitation concerns the treatment of the baseline itself. Although ALEPH introduces a polynomial term to account for residual correlations, the limited statistical precision of the pp and $\bar{p}\bar{p}$ samples forces the final fits to be performed with these terms fixed to zero. As a result, any deviation of the correlation function from unity at low Q is effectively absorbed into the quantum-statistical component of the fit. In this configuration, the extracted radius is not solely determined by the genuine short-range correlation, but is instead highly sensitive to the implicit assumption that all long-range correlations have been removed by the reference construction.

The reference sample construction itself represents an additional source of systematic uncertainty. Event mixing enlarges the accessible phase space and removes not only quantum-statistical correlations, but also physical correlations associated with event topology. The double-ratio procedure relies on Monte Carlo simulations to compensate for these distortions. However, the same analysis explicitly shows that the Monte Carlo does not perfectly reproduce the mixed-event reference, particularly for the $K_S^0 K_S^0$ channel. This mismatch indicates that residual non-femtoscopic structures can survive in the final correlation function, especially in low-statistics channels such as *baryon–baryon* pairs.

Finally, the role of conservation-law-induced correlations is not explicitly addressed. Energy–momentum conservation is expected to produce an anti-correlation at low relative momentum in systems with limited multiplicity, such as individual jets in hadronic Z decays. In the absence of a dedicated treatment of these effects, the observed depletion in the pp and $\bar{p}\bar{p}$ correlation function cannot be uniquely attributed to Fermi–Dirac statistics. Instead, the extracted radius should be regarded as an effective parameter, strongly influenced by the chosen fit range and baseline assumptions.

In summary, the ALEPH measurement demonstrates the experimental accessibility of two-baryon correlations in e^+e^- annihilation and highlights the importance of reference-sample systematics. However, the combination of a very broad Q range, limited statistical precision, and the absence of an explicit modeling of energy–momentum conservation-induced correlations (EMCICs) suggests that the reported sub-femtometer source size for pp and $\bar{p}\bar{p}$ pairs is likely biased by non-femtoscopic effects. These limitations motivate a renewed investigation using higher-statistics samples and a controlled treatment of long-range correlations.

Chapter 3

The Belle experiment

In this chapter, the experimental apparatus used by the Belle experiment throughout its operational period is described. The data analyzed in this thesis were produced in e^+e^- collisions at the KEKB collider, located in Tsukuba, Japan. Belle collected data from 1999 to 2010, accumulating over 1 ab^{-1} of integrated luminosity across the $\Upsilon(nS)$ resonance region. A schematic view of the facility is shown in Fig. 3.1. The main features of KEKB are summarized in Sec. 3.1.

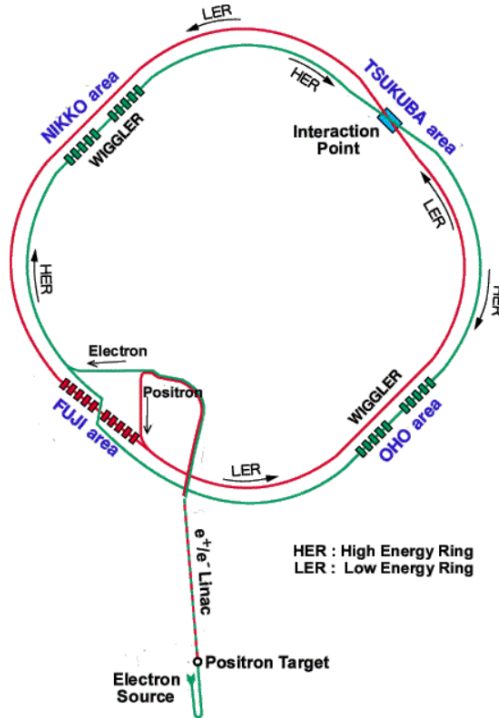


Figure 3.1: Schematic layout of the KEKB collider facility, where asymmetric-energy e^+e^- collisions were produced for the Belle experiment.

The reconstruction of events and the study of physics processes are made possible by the information provided by the Belle detector, a complex general-purpose spectrometer surrounding the interaction point. An overview of the detector and its subsystems is presented in Sec. 3.2.

3.1 The KEKB B-factory

The primary goal of the Belle experiment was to study time-dependent CP violation in the B meson system. To achieve this, the KEKB collider was designed to meet a set of stringent requirements. The most promising way to observe time-dependent CP violation is through measurements of the relative phase between the amplitudes of the decays $B^0 \rightarrow J/\psi K_S^0$ and $\bar{B}^0 \rightarrow J/\psi K_S^0$ [74]. Since these channels have very small branching fractions ($\sim 10^{-4}$), the collider had to deliver extremely high luminosity. KEKB reached a peak instantaneous luminosity of about $\mathcal{L} \sim 10^{34} \text{ cm}^{-2}\text{s}^{-1}$, a world record at the time. It operated from 1999 until June 2010, providing over a decade of high-quality data.

A key design choice was to tune the beam energy to the $\Upsilon(4S)$ resonance, which decays almost exclusively to $B\bar{B}$ pairs. The $\Upsilon(4S)$ lies just 20 MeV above the $B\bar{B}$ threshold, so in its rest frame the two B mesons are produced with negligible kinetic energy. This would make the determination of the time difference between their decays very difficult, since the B lifetime is short ($\tau \approx 1.5 \text{ ps}$) and the decay length in the laboratory frame would be only a few tens of micrometers. To overcome this limitation, KEKB adopted asymmetric beam energies, boosting the $B\bar{B}$ system in the laboratory frame and thereby increasing the measurable decay length. Electrons were stored in the High Energy Ring (HER) at $E_{\text{HER}} = 8.0 \text{ GeV}$, and positrons in the Low Energy Ring (LER) at $E_{\text{LER}} = 3.5 \text{ GeV}$. The center-of-mass energy is defined as $\sqrt{s}$, where

$$s = (p_{e^+}^\mu + p_{e^-}^\mu)^2,$$

with p^μ the four-momentum of the beams. Collisions occurred at a single interaction point (IP), surrounded by the Belle detector.

Because of the large branching fraction of $\Upsilon(4S) \rightarrow B\bar{B}$ ($\approx 96\%$), KEKB is classified as a *B-factory*. Another B-factory was the PEP-II collider at SLAC, where the BaBar detector operated. Beyond their core CP violation programs, B-factories proved to be invaluable across a wide range of particle-physics topics. All quarkonium states with $J^{PC} = 1^{--}$ can be directly produced in e^+e^- annihilation, either by tuning the beam energy directly to the resonance or through Initial State Radiation (ISR). In the latter case, one of the incoming leptons emits a hard photon γ_{ISR} , reducing the effective center-of-mass energy to

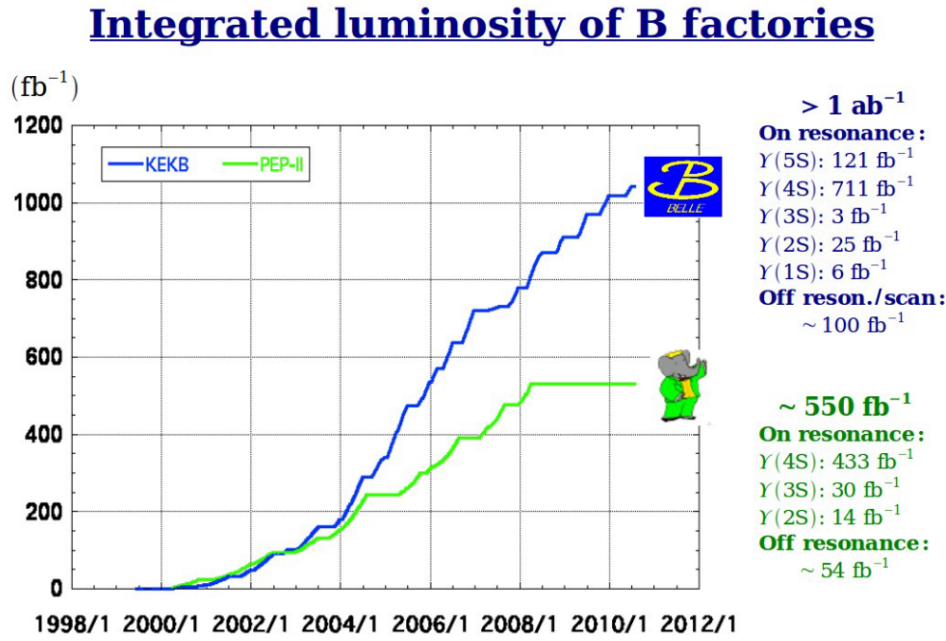
$$\sqrt{s'} = \sqrt{s} - E_{\gamma_{\text{ISR}}}.$$

This mechanism has allowed the study of both bottomonium and charmonium vector states.

Figure 3.2 shows the integrated luminosity collected by Belle and BaBar as a function of time. The Belle dataset exceeds 1 ab^{-1} , including more than 100 fb^{-1} at the $\Upsilon(10860)$ resonance, previously named $\Upsilon(5S)$. This enabled unique measurements, such as the observation of the anomalous dipion transitions at the $\Upsilon(10860)$ and the discovery of the exotic Z_b states. A summary of Belle's integrated luminosities at different resonances is given in Table 3.1.

Table 3.1: Summary of the integrated luminosities collected by Belle at various resonances.

Resonance	On-peak luminosity (fb^{-1})	Off-peak luminosity (fb^{-1})
$\Upsilon(1S)$	5.7	1.8
$\Upsilon(2S)$	24.9	1.7
$\Upsilon(3S)$	2.9	0.3
$\Upsilon(4S)$	711	89.4
$\Upsilon(10860)$	121.4	1.7
Scan	—	27.6

Figure 3.2: Integrated luminosities of the Belle and BaBar experiments as a function of time. Belle accumulated more than 1 ab^{-1} in total, including unique samples above the $\Upsilon(4S)$.

3.2 The Belle detector

The Belle detector was a large-solid-angle magnetic spectrometer covering about 91% of the full 4π solid angle. A superconducting solenoid provided a magnetic field of $B = 1.5 \text{ T}$ inside a cylindrical volume 4.4 m long and 3.4 m in diameter. The flux was returned through an iron yoke that surrounded the detector and also served as the absorber for the outermost muon system.

The beam pipe around the interaction point was made of beryllium in order to minimize the material budget, thanks to its short radiation length X_0 . This choice reduced multiple scattering and energy loss of final-state particles before entering the tracking systems.

The detector was composed of several concentric subdetectors, shown in Fig. 3.3 and listed here from the innermost to the outermost layers:

- **Silicon Vertex Detector (SVD):** consisted of four layers of double-sided silicon

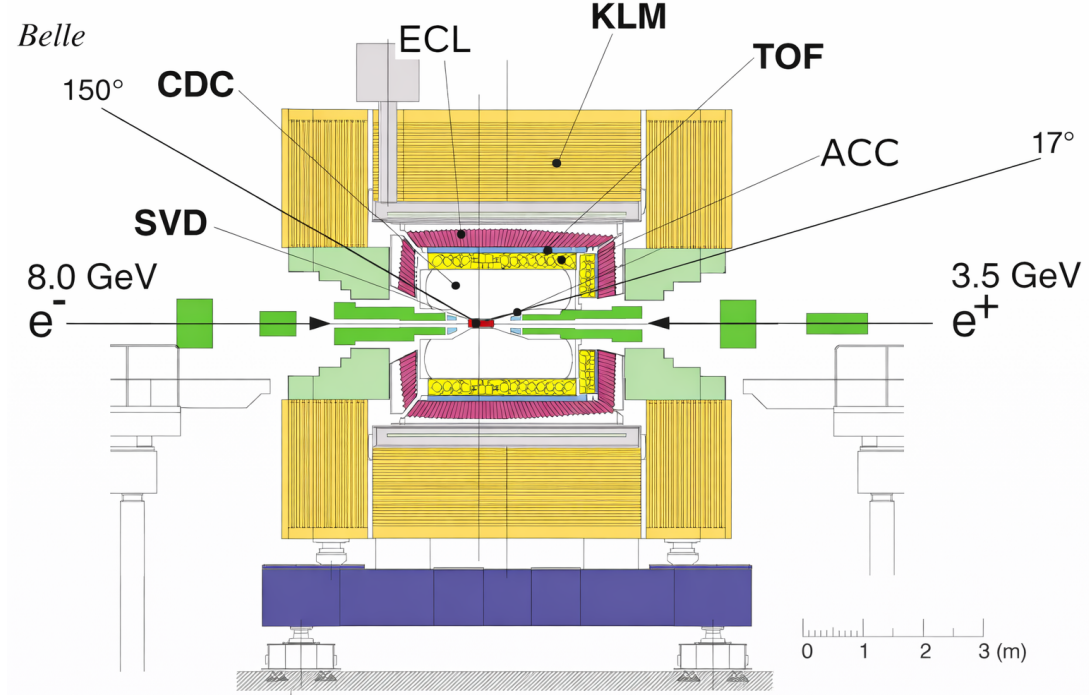


Figure 3.3: Side view of the Belle detector. The major subsystems are indicated.

strip detectors in its final configuration. Its primary role was the precise reconstruction of decay vertices, though it also contributed to charged-particle tracking.

- **Central Drift Chamber (CDC):** a large cylindrical chamber filled with a low- Z gas mixture, used to reconstruct charged-particle tracks and to measure their specific ionization dE/dx for particle identification.
- **Time-of-Flight system (TOF):** an array of plastic scintillator counters surrounding the CDC. It provided timing information for charged particles with a resolution of about 100 ps, contributing to particle identification and trigger decisions.
- **Aerogel Cherenkov Counter (ACC):** an array of threshold Cherenkov detectors using silica aerogel radiators. Together with the TOF and CDC, it formed the main particle identification system, particularly for separating kaons from pions over a wide momentum range.
- **Electromagnetic Calorimeter (ECL):** comprised of CsI(Tl) crystals that measured electromagnetic showers with high energy and position resolution. It was crucial for identifying electrons and photons and for reconstructing neutral mesons such as π^0 and η .
- **K_L and Muon detector (KLM):** located in the iron flux return outside the solenoid, this system identified muons and detected neutral long-lived kaons (K_L) for momenta above about 600 MeV/ c .
- **Trigger system:** a multi-level trigger architecture combining information from several subdetectors, including the CDC, TOF, ECL, and KLM. It selected events of interest in real time, reducing the event rate to a level suitable for data acquisition and offline processing.

A detailed description of each subsystem is provided in the following sections. The main reference for this overview is the KEK Belle Technical Design Report [75].

3.2.1 Silicon Vertex Detector

To achieve precise measurements of decay vertices, the silicon vertex detector (SVD) was installed directly outside the beam pipe. Over the lifetime of Belle, three versions of the SVD were employed.

The first version, **SVD1**, consisted of three layers of double-sided silicon strip detectors (DSSD). A DSSD is a p - n diode sandwiched between two sets of segmented electrodes, each connected to independent readout electronics. When a charged particle traverses the silicon, it generates electron-hole pairs that drift along the electric field to the electrodes, allowing reconstruction of the hit position. In double-sided detectors, the strips on one side are oriented orthogonally to those on the other, so that both the r and ϕ coordinates can be measured within a single layer.

Being located only a few centimeters from the interaction point (IP) and built from solid-state sensors, the SVD and its readout were subject to significant radiation damage. One of the main limitations of SVD1 was the low radiation tolerance of its readout electronics [76]. Furthermore, its polar acceptance ($23^\circ < \theta < 139^\circ$) was smaller than that of the central drift chamber (CDC). To address these shortcomings, the first upgrade, **SVD1.5**, introduced improved radiation-tolerant readout electronics.

During the 2003 Belle shutdown, the detector was further upgraded to **SVD2**. This version featured four layers and extended its angular coverage to match that of the CDC. In addition, the innermost layer was placed closer to the IP, which significantly improved the vertex resolution.

The impact-parameter resolution depends on the so-called pseudo momentum of the track, defined as

$$\tilde{p} = p \beta \sin^{3/2} \theta,$$

which incorporates the momentum and angular dependence into a single variable. The measured resolution for the r - ϕ and z coordinates is shown in Fig. 3.4 for both SVD1 and SVD2. The improvement of SVD2 is especially significant for low-momentum tracks.

All the data analyzed in this thesis were collected with the SVD2 configuration. Its layout is shown in Fig. 3.5.

3.2.2 Central Drift Chamber

The central drift chamber (CDC) was the primary tracking device of Belle, responsible for reconstructing charged-particle trajectories and providing particle identification through specific energy loss measurements (dE/dx).

The operating principle of a drift chamber relies on the collection of charge carriers released by the ionization of gas molecules traversed by charged particles. A set of thin wires inside the chamber provides a nearly radial electric field: anode wires collect the electrons, while alternating cathode wires shape the field lines. Close to the anodes, the strong field accelerates electrons, producing secondary ionization and thereby amplifying the signal through avalanche multiplication. Away from the wires, the drift velocity of electrons saturates at a constant value, so the position of the original ionization can be reconstructed by measuring the time difference between a reference signal and the arrival of the drifting charge.

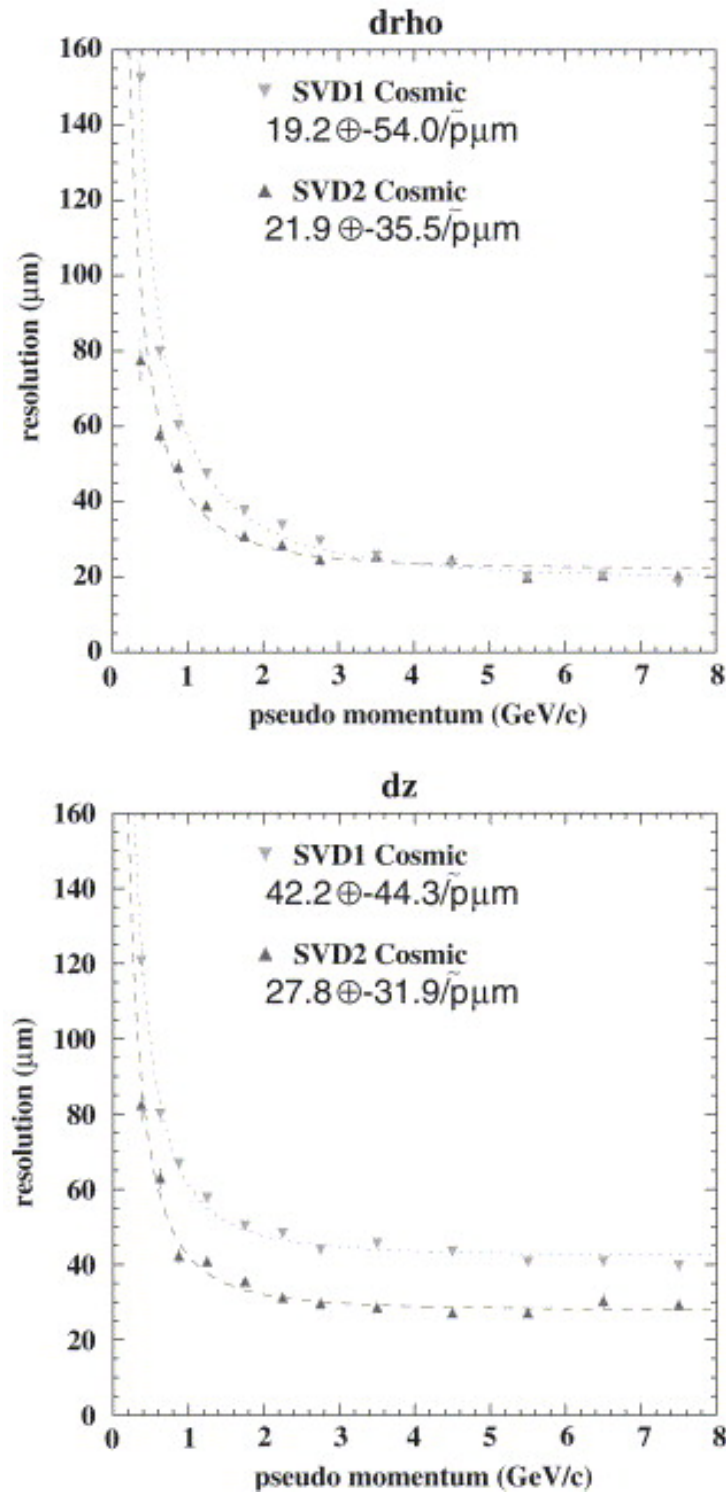


Figure 3.4: Impact-parameter resolution in the r - ϕ and z coordinates as a function of track pseudo momentum. The upgrade from SVD1 to SVD2 yielded a marked improvement, especially for low-momentum tracks. Figures taken from Ref. [76].

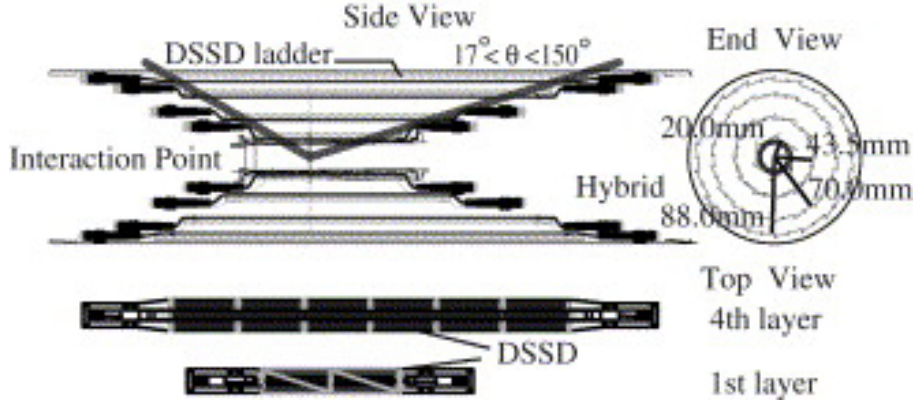


Figure 3.5: Schematic layout of the upgraded SVD2, featuring four layers and extended angular acceptance.

The structure of the Belle CDC is shown in Fig. 3.6. It was built asymmetrically along the z axis in order to maximize acceptance, with the innermost section designed in a conical shape to approach the interaction point as closely as possible. This configuration allowed efficient tracking of charged particles with transverse momentum down to 50 MeV/ c , covering the polar angular range $17^\circ < \theta < 150^\circ$.

A crucial design requirement of the CDC was to minimize multiple scattering in order to achieve good momentum resolution, while retaining a dE/dx resolution suitable for particle identification. This balance required the choice of a low- Z gas mixture, but not so low as to compromise ionization statistics. The final configuration employed a 50% He – 50% C₂H₆ mixture, which provided an optimal trade-off between momentum resolution and dE/dx performance.

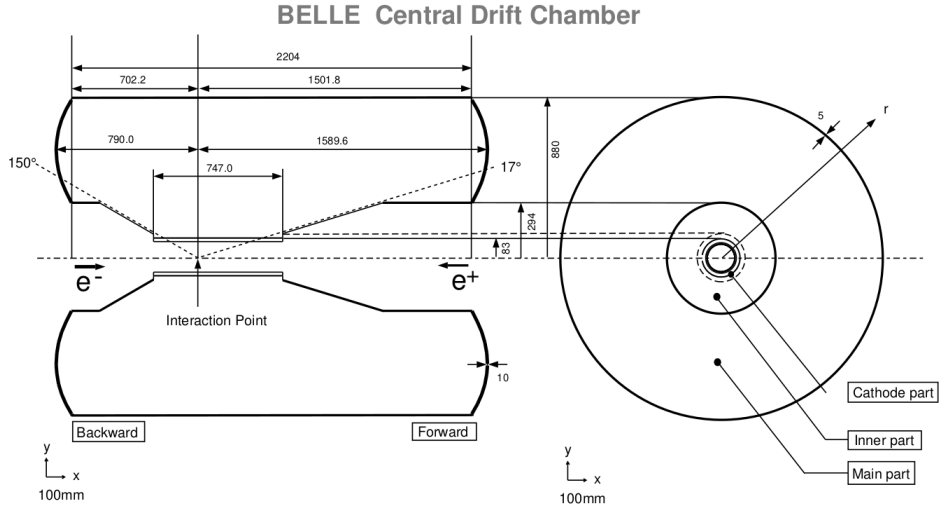


Figure 3.6: Structure of the Belle central drift chamber (CDC), which provided tracking and dE/dx measurements over a wide angular acceptance.

By combining the drift times from many layers of sense wires, the chamber reconstructed precise space points along the trajectories of charged particles. In the presence of the 1.5 T solenoidal magnetic field, these trajectories were curved, allowing the determination of particle momenta from the measured curvature. Together with the silicon vertex detector, the CDC provided the core tracking capability of Belle, yielding excellent

momentum resolution and robust pattern recognition across a wide kinematic range. The transverse momentum dependence of the p_t resolution is shown in Fig. 3.7.

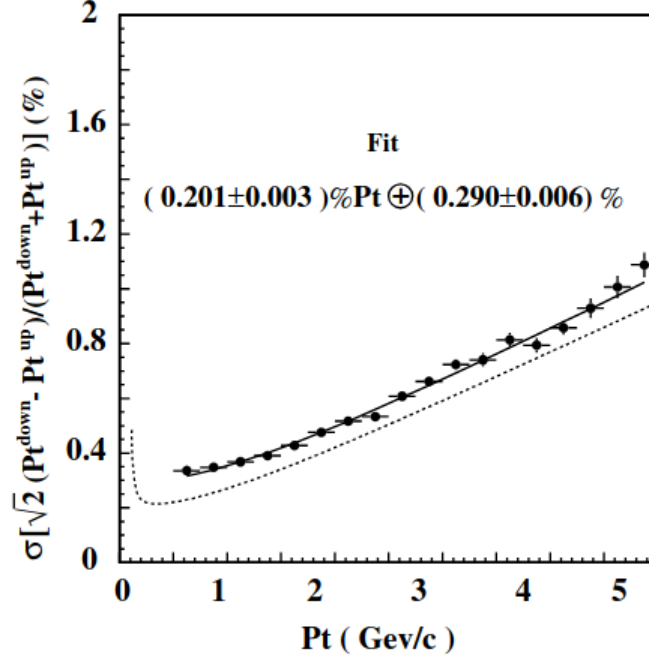


Figure 3.7: Transverse-momentum resolution of the Belle CDC as a function of p_t (cosmic-ray data). Figure adapted from Ref. [77].

The tracking performance of the CDC is particularly important for our femtoscopy analysis. Since such studies rely on the precise reconstruction of the relative momentum between pairs of charged particles, any degradation in momentum resolution directly impacts the measured correlation function at small k^* . A detailed discussion of the tracking resolution and its impact on femtoscopy analyses is presented in sections 4.5.4 and 8.4.

3.2.3 Time-of-Flight detector

A straightforward way to identify a particle is to determine its mass. Once the momentum p has been measured, one needs an additional kinematic observable to extract the mass, based on the relation

$$p = \beta\gamma mc, \quad (3.1)$$

where β is the velocity in units of c and γ is the Lorentz factor.

The principle of a time-of-flight (TOF) detector is to measure β from the time Δt taken by a particle to travel a known distance L . For a given momentum p , the measured flight time is related to the mass m through relativistic kinematics:

$$\frac{\Delta t}{L} = \frac{1}{c} \sqrt{1 + \frac{m^2 c^2}{p^2}} \approx \frac{1}{c} \left(1 + \frac{m^2 c^2}{2p^2} \right), \quad (3.2)$$

where the approximation holds for $p \gg mc$.

Charged particles crossing matter can excite its molecules; if de-excitation occurs via the emission of visible photons, the material is called a scintillator. When the scintillator

is coupled to a photon sensor such as a photomultiplier tube (PMT), the system forms a scintillation counter, providing precise timing information.

The Belle TOF system consisted of plastic scintillator counters, chosen for their excellent time resolution (~ 100 ps) and mechanical robustness. Their rapid timing made the TOF suitable not only for particle identification, but also as part of the level-1 trigger system (see Sec. 3.2.7).

The detector comprised 64 modules, each made of two trapezoidal TOF counters accompanied by a more coarsely segmented thick scintillator counter (TSC). Positioned at a radius of 1.2 m from the interaction point, the system covered the polar angular range $34^\circ < \theta < 120^\circ$. The dimensions of a TOF module are shown in Fig. 3.8.

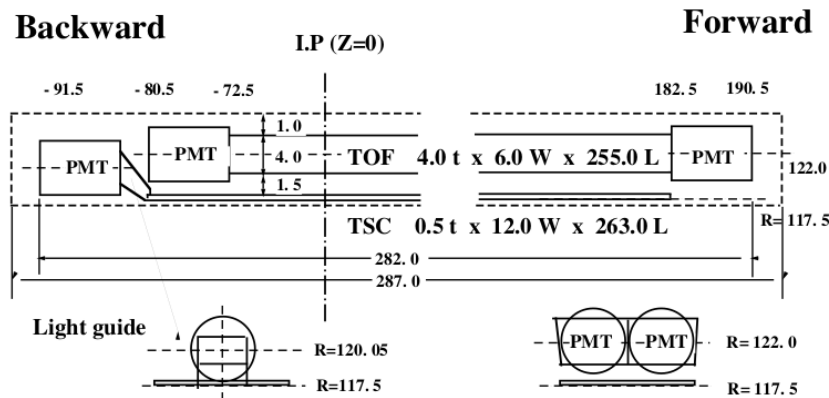


Figure 3.8: Dimensions of a TOF module.

An essential aspect of the Belle TOF system was the careful calibration of timing offsets and detector response. The calibration was performed primarily using high-momentum muons from dimuon events [78], since their trajectories are nearly straight and minimally affected by hadronic interactions. For each PMT, the measured raw time was corrected for three main effects: the signal propagation delay along the scintillator bar (a function of the hit position z), the dependence on the pulse height (time-walk correction), and a constant electronic offset specific to each channel. A two-dimensional fit was used to determine these corrections simultaneously, after which a higher-order polynomial was applied to remove residual z -dependent variations. Typical calibration constants achieved a residual spread of less than 20 ps per channel, and an overall time resolution for muon pairs of about 100–110 ps.

For hadrons, however, additional complications arise. The tracking algorithm in the CDC assumes a pion mass when extrapolating trajectories to the TOF counters. While this approximation is adequate for high-momentum kaons and protons, at low momentum it leads to systematic shifts in the predicted flight time. Moreover, hadronic interactions in the material between the CDC and the TOF (such as in the ACC region) can distort trajectories and energy deposits, degrading the timing performance. As a result, empirical momentum-dependent corrections were developed separately for pions, kaons, and protons, reducing systematic shifts to below ~ 30 ps.

Despite these corrections, the intrinsic time resolution achieved for pions and kaons (~ 120 – 130 ps) was consistently better than that for protons (~ 150 – 170 ps). This difference is largely due to the heavier proton mass, which results in lower velocities at a given momentum, stronger sensitivity to trajectory extrapolation uncertainties, and a higher probability of hadronic interactions before reaching the TOF system. In addition,

protons experience larger energy loss while traversing the beam pipe and inner tracking detectors (SVD and CDC), leading to further velocity reduction and an increased spread in the measured time-of-flight. The better timing resolution for lighter hadrons translated into more effective K/π separation, while proton identification relied more heavily on complementary detectors (CDC dE/dx and the ACC). However, in this thesis a new proton identification is developed from scratch, making extensive use of the time-of-flight information, as described in Chapters 5 and 6.

3.2.4 Aerogel Cherenkov Counter

Another method for particle identification is the detection of Cherenkov radiation. When a charged particle traverses a medium with refractive index n , it emits Cherenkov light if its velocity exceeds the phase velocity of light in that medium, i.e. if $\beta \geq 1/n$. The radiation is emitted at a fixed angle

$$\theta_C = \arccos\left(\frac{1}{\beta n}\right),$$

with respect to the particle's trajectory. The velocity threshold $\beta_t = 1/n$ defines the condition for radiation to occur. At fixed momentum, by selecting an appropriate refractive index, only particles with mass below a certain value will produce Cherenkov photons. Thus, a threshold counter based on this principle can discriminate between particle species.

The Aerogel Cherenkov Counter (ACC) system in Belle was optimized for π/K separation in the intermediate momentum region, where neither dE/dx information from the CDC nor timing from the TOF alone provided sufficient discrimination. It consisted of 960 aerogel modules with refractive indices tuned between 1.01 and 1.03 depending on the polar angle, to maximize the separation power across the acceptance. In practice, the refractive indices of the aerogel modules was set at production time by controlling the material density through the fabrication process.

The modules were arranged to cover both the barrel and forward endcap regions, as illustrated in Fig. 3.9. Each module contained a block of silica aerogel radiator coupled to

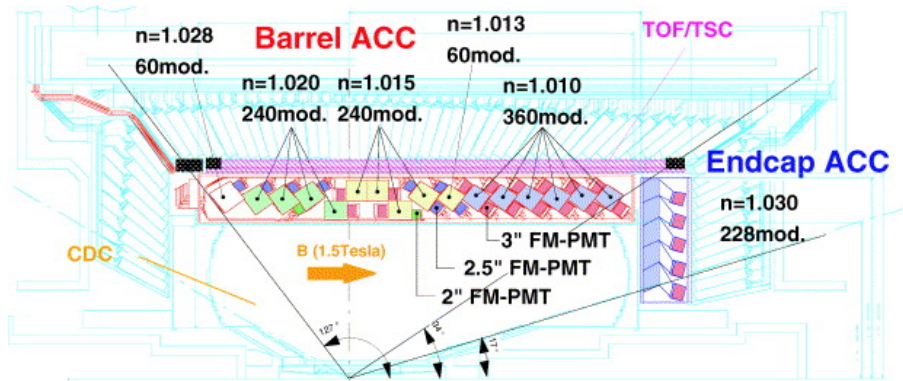


Figure 3.9: Layout of the Belle ACC, covering the barrel and forward endcap regions. It provided π/K separation in the intermediate momentum range not accessible to the CDC dE/dx and TOF systems.

photomultiplier tubes, which collected the Cherenkov photons and converted them into electrical signals.

3.2.5 Electromagnetic Calorimeter

The primary purpose of the electromagnetic calorimeter (ECL) was the measurement of photon energies, but it also contributed to electron identification and the reconstruction of neutral mesons. The system consisted of 8736 CsI(Tl) crystals arranged in a barrel section and two annular endcaps, as shown in Fig. 3.10. Together, they covered about 91% of the solid angle, with a small loss of acceptance ($\sim 3\%$) due to gaps required for cables and support structures. Each crystal was oriented to approximately point towards the interaction point. In order to further extend the forward coverage, an additional subsystem, the Extreme Forward Calorimeter (EFC), was constructed.

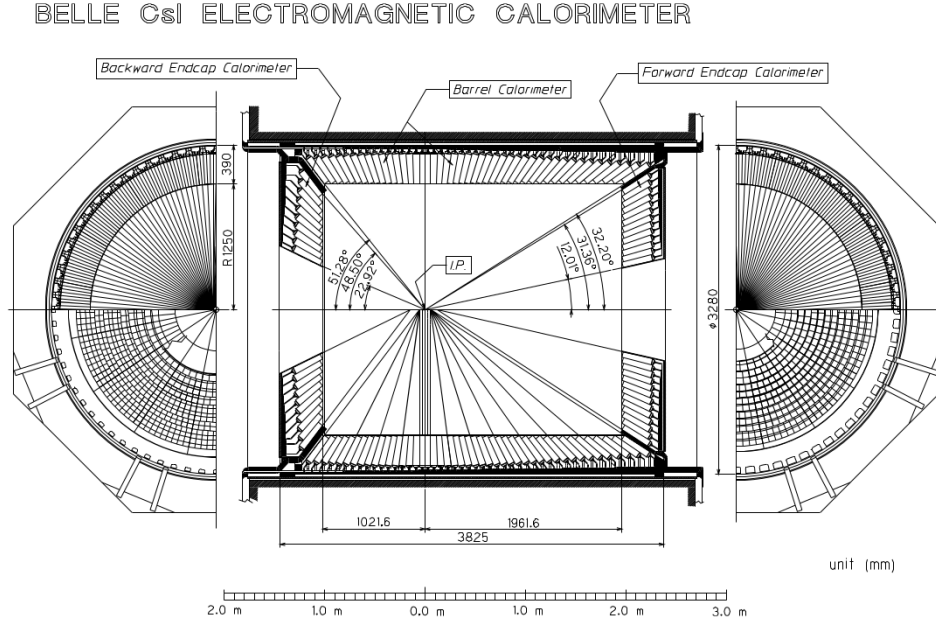


Figure 3.10: Layout of the Belle ECL, consisting of 8736 CsI(Tl) crystals in the barrel and endcaps, covering about 91% of the full solid angle.

Cesium iodide activated with thallium (CsI(Tl)) is an inorganic scintillator. In a pure CsI crystal, energy deposited by an incident particle can excite electrons into the conduction band, but the subsequent recombination is inefficient and emits light outside the visible range ($\lambda \sim 300$ nm), poorly suited for photodetection. Doping with thallium introduces impurity levels within the band gap, enabling efficient de-excitation with visible photon emission. This increases the light yield from about 2000 photons/MeV for pure CsI to roughly 65,000 photons/MeV for CsI(Tl) [79].

High-energy photons ($E \gtrsim 100$ MeV) interacting with the calorimeter predominantly convert to e^+e^- pairs. The resulting electrons and positrons radiate bremsstrahlung photons and occasionally produce δ rays, giving rise to an electromagnetic shower [80]. To contain more than 95% of the incident photon's energy, at least 10–15 radiation lengths (X_0) are required. The Belle crystals were 30 cm deep, corresponding to $16.2 X_0$, ensuring good shower containment. Laterally, multiple scattering broadens the shower; the transverse dimensions of the crystals (typically 6×6 cm², varying with polar angle) were chosen so that about 80% of the shower energy was deposited in a single crystal, while the remainder spreads into neighbors to form clusters.

Each crystal was read out by two silicon PIN photodiodes connected to charge-sensitive preamplifiers. The preamplifier output was processed by shaping electronics,

and the resulting signal was split into two streams: one for the trigger and the other for data acquisition.

The precision in energy measurement is characterized by the detector resolution, defined as the relative spread in response to monochromatic beams. At Belle, the resolution was determined using electron, photon, and pion test beams, by summing the energy deposit over matrices of 3×3 or 5×5 crystals. The resolution can be parametrized as

$$\frac{\sigma_E}{E} = \frac{0.0066}{E} \oplus \frac{1.53}{E^{1/4}} \oplus 1.18\%, \quad (3.3)$$

for 3×3 clusters, and

$$\frac{\sigma_E}{E} = \frac{0.066}{E} \oplus \frac{0.81}{E^{1/4}} \oplus 1.34\%, \quad (3.4)$$

for 5×5 clusters, with E expressed in GeV.

These design choices and performance parameters made the Belle ECL a highly effective subsystem for photon and electron measurements, and a cornerstone of many analyses relying on neutral final states.

3.2.6 K_L and muon detector

The K_L and muon (KLM) detector was designed to identify long-lived neutral kaons and muons with high efficiency and purity. It was built as a sampling detector, consisting of alternating layers of resistive plate chambers (RPCs) and iron plates, arranged in a barrel and two endcap regions. The iron served both as a flux return yoke for the solenoid and as absorber material, while the RPCs provided the active detection layers.

Long-lived neutral kaons typically undergo hadronic interactions with the iron nuclei, producing secondary showers of charged particles that could be detected by the RPCs. In contrast, muons do not interact strongly with matter and traverse the detector with a minimal deflection, leaving long penetrating tracks. This difference in penetration depth and scattering provided a natural handle for distinguishing muons from hadrons.

Each RPC consisted of two parallel plate electrodes separated by a thin gas-filled gap and operated at a typical voltage of about 8 kV. Under these conditions, the detectors worked in streamer mode: an ionizing particle passing through the gas triggered a localized avalanche, which induced a discharge across the plates. The electrodes were made of high-resistivity glass ($> 10^{10} \Omega \text{ cm}$), which limited the discharge current and protected the readout electronics. Signals were collected on external pickup strips, arranged in orthogonal directions to provide both spatial and timing information. In the KLM configuration, two RPC layers were combined with orthogonal strip readout, ensuring redundancy and good spatial resolution.

The performance of the system is summarized in Fig. 3.12a and Fig. 3.12b. Muon detection efficiency exceeded 90% for $p > 2 \text{ GeV}/c$, while the probability of misidentifying charged pions or kaons as muons (fake rate) was kept below 2%. The latter was studied using control samples of $K_S^0 \rightarrow \pi^+\pi^-$ decays collected in e^+e^- collisions.

3.2.7 Trigger system

In an e^+e^- collider, bunch crossings occur at high frequency, but only a small fraction of them results in interactions of physical interest. Most crossings produce no interaction, or yield final states that are not relevant for the Belle physics program (e.g. elastic Bhabha

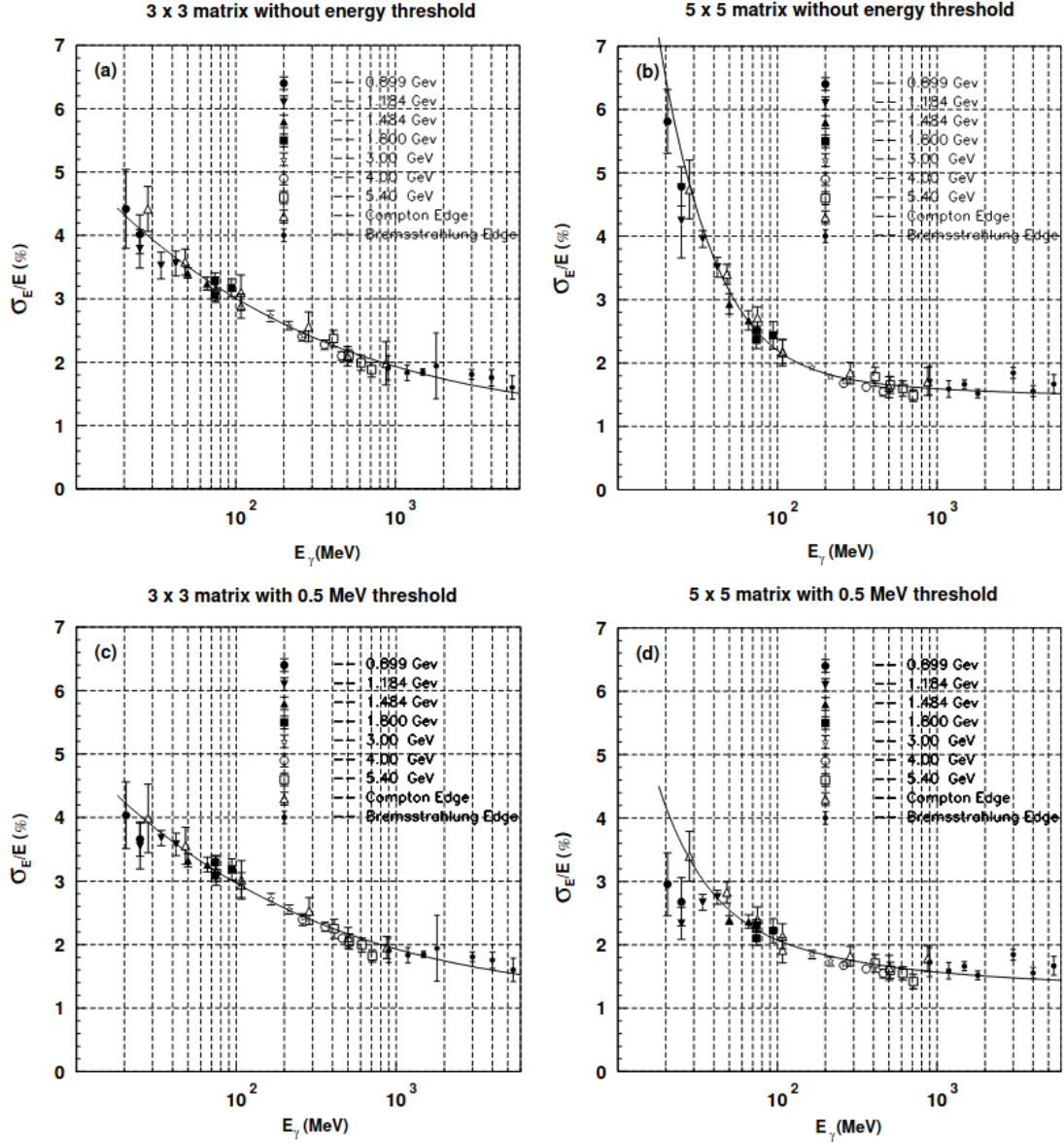


Figure 3.11: Relative energy resolution of the Belle electromagnetic calorimeter (ECL) as a function of photon energy, shown for different cluster sizes (3×3 and 5×5 crystal matrices) and for two per-crystal energy-threshold configurations (no threshold and a 0.5 MeV threshold). The curves represent the corresponding parametrizations of σ_E/E used to describe the measured performance.

scattering). Moreover, the high beam currents required for KEKB's luminosity lead to significant machine background rates of order ~ 100 Hz. To ensure efficient data collection within the tolerance of the data acquisition system (DAQ), which could sustain at most about 500 Hz, events had to be triggered and filtered in real time. The trigger had to achieve two key goals: maximize efficiency for physics signals of interest while strongly suppressing background.

The Belle trigger system consisted of two main stages: a hardware-based Level-1 trigger and a software-based Level-3 trigger. The Level-1 trigger operated with fixed latency ($2.2 \mu\text{s}$) and was built from sub-detector trigger logics coordinated by a central unit, the Global Decision Logic (GDL). The GDL combined signals from all subsystems

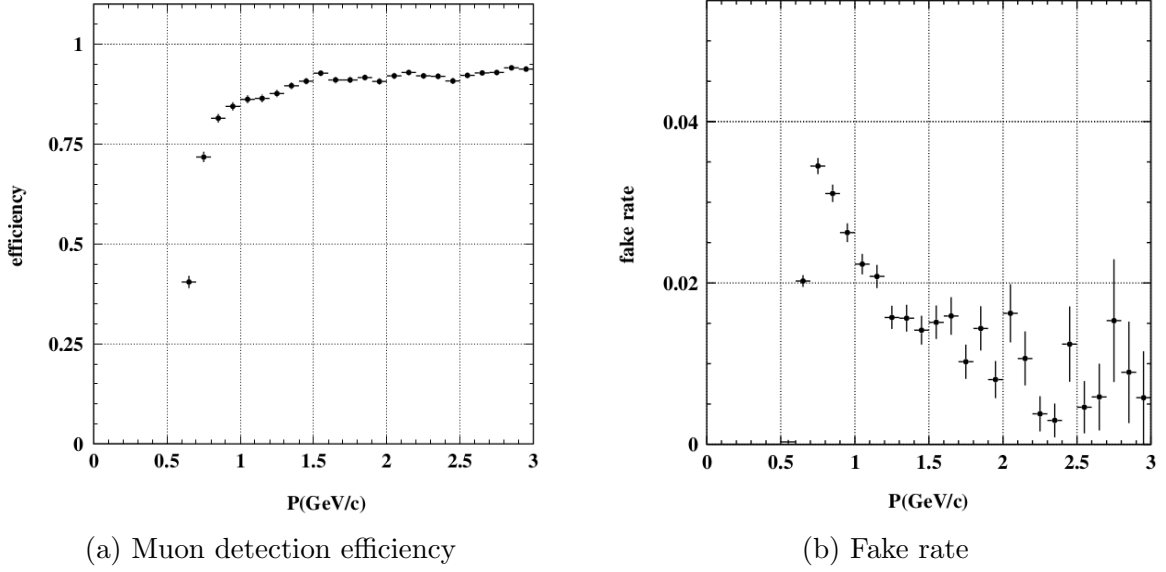


Figure 3.12: Performance of the KLM detector as a function of momentum. (a) Muon detection efficiency. (b) Fake rate, i.e. misidentification probability of pions/kaons as muons.

to classify the event type and issue a trigger decision.

Two broad classes of sub-detector triggers were implemented:

- **Track triggers**, which used the Central Drift Chamber (CDC) to identify charged tracks originating from the interaction point and to reject beam-related backgrounds.
- **Energy triggers**, which relied on the Electromagnetic Calorimeter (ECL). The calorimeter was segmented into trigger cells, each corresponding to a block of 4×4 adjacent crystals. Energy deposits in these cells were combined to identify electromagnetic showers consistent with photons or electrons.

A schematic overview of the Level-1 trigger logic is shown in Fig. 3.13. The combination of CDC- and ECL-based decisions provided a robust balance between efficiency for hadronic $B\bar{B}$ events and suppression of QED and beam background processes. Events passing Level-1 were forwarded to higher-level processing, where additional software-based filters (Level-3) refined the selection using more detailed event reconstruction.

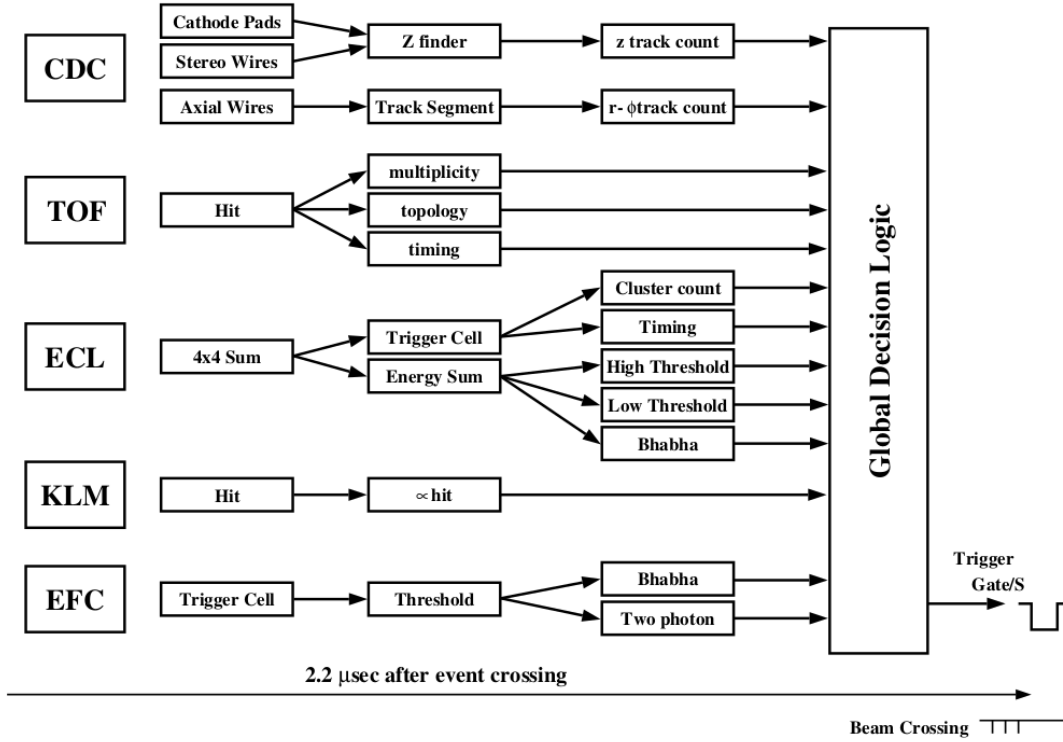


Figure 3.13: Schematic view of the Belle Level-1 trigger system. Sub-detector trigger logics (CDC and ECL) were combined by the Global Decision Logic to issue real-time event decisions with fixed latency.

3.3 Data production

After an event passed the hardware-based Level-1 trigger, it was processed by the software triggers and the data acquisition (DAQ) system, which transformed raw detector signals into formats suitable for offline analysis. This system was designed to reduce the high beam-related background while preserving essentially all events of physics interest.

In the Belle trigger architecture, no dedicated Level-2 trigger stage was implemented between Level-1 and the software-based trigger.

The first stage after Level-1 was therefore the **Level-3 trigger**, implemented in software and running on an online computing farm. It employed fast tracking algorithms using CDC information, complemented by calorimeter data, to reject events from beam backgrounds such as Bhabha scattering or beam-gas interactions [81]. The online farm consisted of six VME crates, each containing multiple processors dedicated to event filtering and temporary storage. The combined system provided a throughput of about 15 MB/s, corresponding to a maximum event rate of ~ 500 Hz and a typical event size of 30 kB.

Events passing Level-3 were transferred to a tape library for permanent storage. Before archiving, the raw data underwent online reconstruction, producing **Data Summary Tables (DSTs)**: compact formats that contained only the reconstructed information relevant to physics analyses, such as tracks, clusters, and particle-identification variables.

To further reduce beam-induced backgrounds, Belle employed a software-based **Level-4 filter** immediately before full reconstruction. This stage targeted events originating from off-momentum beam $e^\pm$ undergoing bremsstrahlung or Coulomb scattering with the

residual gas in the beam pipe. The Level-4 selection required at least 4 GeV of visible energy in the ECL and at least one charged track with $p_T > 300$ MeV/ c originating from near the interaction point. These requirements efficiently suppressed background while retaining more than 98% of hadronic $B\bar{B}$ events. Accepted events were then processed through the full reconstruction chain.

The DST format served as the starting point for all Belle physics analyses. Given the large data volume, collaborators typically produced reduced samples, or *skims*, by applying loose selections tailored to the physics channel under study. These skims preserved high efficiency while greatly reducing dataset size, enabling more manageable and focused analyses. Hadronic events are selected using the HadronBJ skim, which corresponds to the requirements listed in Sec. 5.1.1.

Chapter 4

Software framework

4.1 Overview

The analysis presented in this thesis is based on a software workflow that transforms the original Belle data, recorded in the early 2000s at KEKB, into the high-level objects used to construct the proton–proton correlation function. This workflow integrates the Belle II Analysis Software Framework (basf2) with a dedicated conversion layer, B2BII, which maps the Belle data model into the Belle II format. By adopting this approach, the entire analysis—from low-level event processing to the construction of final histograms and fit results—can be performed within a single, modern, and actively supported framework.

The starting point of the workflow is the Belle mini Data Summary Table (mDST) format, which encodes reconstructed tracks, clusters, particle candidates, and event-level information. These data are first processed by the B2BII conversion modules, which reinterpret the Belle objects. The converted samples can then be processed seamlessly by basf2, enabling the use of its reconstruction utilities, particle-identification tools, selection modules, and ntuple writer.

Within basf2, dedicated steering scripts implement the event and track selections, the construction of intermediate ntuples, and, where appropriate, the first step of pair building. The resulting ntuples form the input to the offline analysis layer, which is based on ROOT and a set of custom C++ and Python tools.

By structuring the analysis in this modular way, each stage remains reproducible, clearly documented, and easy to validate independently. Moreover, using a unified framework ensures consistent data handling across samples, simplifies large-scale processing on computing clusters, and facilitates long-term maintainability of the analysis code.

4.2 Belle and Belle II software ecosystems

4.2.1 The Belle data format and legacy software

The Belle experiment employed a dedicated software framework, BASF, to reconstruct and analyse events produced at KEKB. The reconstructed information was distributed primarily in the mini-DST (mDST) format, a compact representation containing charged tracks, calorimeter clusters, particle-identification likelihoods, and event-level metadata. Although BASF served the experiment throughout its operation, it predates many of the modern design principles and tools that now underpin large-scale HEP software. Its data

model, coding conventions, and build environment differ substantially from those of contemporary frameworks, and maintaining or extending the Belle codebase is increasingly impractical.

For the purposes of this thesis, the Belle reconstruction is treated as a fixed input. No attempt is made to rerun or modify the original reconstruction chain. Instead, the analysis is built upon a conversion layer that enables Belle data to be processed using the Belle II software infrastructure.

4.2.2 The basf2 framework

The Belle II Analysis Software Framework [82] (basf2) is a modern, modular, C++17-based environment designed to support all stages of event simulation, reconstruction, and physics analysis for the Belle II experiment. Its architecture is built around a sequence of modules that operate on a transient DataStore, allowing users to construct flexible “processing paths” controlled by Python steering scripts.

Key features of basf2 relevant to this analysis include:

- Modular event processing: each physics or technical operation—reading input, applying selections, building particle candidates, writing ntuples—is performed by a dedicated module, ensuring transparency and reproducibility.
- Unified data model: reconstructed objects (tracks, clusters, particle candidates) follow a consistent set of C++ classes, independent of the underlying data origin.
- Python-based steering: analysis logic is expressed in compact Python scripts, enabling rapid configuration, parameter passing, and job submission.
- Integrated conditions and geometry management: global tags provide a reproducible description of detector conditions, calibrations, and alignment constants.
- Extensive analysis tooling: basf2 provides ready-to-use modules for particle identification, event selection, vertexing, and ntuple writing, as well as bindings to ROOT for statistical analysis.

Because B2BII converts Belle events into the Belle II data model, the converted samples can be processed by basf2 without special treatment. This allows the analysis to exploit the full suite of basf2 tools—even for Belle data—and ensures consistency with the workflows used in Belle II.

4.3 Motivation for a unified framework

Using basf2 to analyse Belle data offers several practical advantages:

- Long-term maintainability: the Belle II framework is actively supported, documented, and continuously developed, whereas Belle’s original software is frozen.
- Modern software environment: basf2 provides a clean C++/Python interface, improved memory management, and contemporary HPC compatibility.
- Consistency across datasets: both simulated and real data samples can be treated uniformly, simplifying validation and reducing source-specific systematic effects.

- Integration with custom tools: custom C++ classes, Python analysis modules, and external packages (e.g., for correlation-function fitting) can be integrated naturally into the basf2 pipeline.

For these reasons, basf2 serves as the central software environment for this thesis, with B2BII providing the essential bridge that maps Belle data into a format compatible with the framework.

4.4 The B2BII conversion chain

4.4.1 Motivation and role of B2BII

Belle data were originally reconstructed using the Belle-specific BASF software and stored in a data model that predates the modern object-oriented architecture of Belle II. To analyse Belle data within the Belle II software ecosystem, a dedicated compatibility layer—B2BII (“Belle-to-Belle II”)—was developed. B2BII enables Belle events to be interpreted as Belle II events by translating the Belle mDST structures into Belle II’s standardized data model. With this conversion step, Belle data can be processed using basf2 modules as if they had been produced by the Belle II experiment.

This functionality is essential for the present analysis. It allows all subsequent steps to be performed within basf2, using the same tools, configuration machinery, and software environment used for Belle II studies.

4.4.2 Concept and architecture of B2BII

B2BII is implemented as a set of basf2 modules written in C++ and Python. These modules read Belle mDST files, interpret their contents, and populate the Belle II DataStore with corresponding objects. Although the detector technologies of Belle and Belle II differ, many concepts—tracks, calorimeter clusters, PID likelihoods, event-level quantities—are sufficiently similar to allow a well-defined mapping.

The conversion involves several key steps:

- Event metadata: run number, event number, trigger information, and global quantities are copied into Belle II equivalents.
- Charged tracks: Belle’s track parameters and covariance matrices are translated into Belle II’s Track and TrackFitResult objects, maintaining associations to hit patterns and PID sources where available.
- Neutral clusters: ECL clusters are mapped to Belle II calorimeter objects, preserving energies, positions, and shape variables.
- Composite candidates: when present in mDST (e.g. K_S^0 , Λ^0), these are reproduced as Belle II Particle objects with proper mother–daughter relations.

Finally, the populated DataStore is written out as a Belle II mDST, which is indistinguishable—at the level of the basf2 user interface—from a sample produced by the Belle II reconstruction chain.

For this thesis, the conversion was performed using the official B2BII modules distributed with basf2. Dedicated steering scripts were written to:

- Read Belle mDST files from the experiment’s data repository;
- Configure the appropriate B2BII conversion modules;
- Apply minimal event-level filtering (e.g. hadronic-event preselection, where applicable)

The conversion was performed once per Belle data sample and the resulting files served as the starting point for all subsequent analysis steps. This approach ensures consistency and avoids repeated I/O and conversion overhead when processing large datasets in parallel on computing clusters.

Where necessary, global-tag settings matching Belle data-taking conditions were applied. Although many Belle II conditions are irrelevant for converted samples (e.g. detector alignment constants), the consistency of the conditions database ensures a reproducible and well-defined processing environment.

4.5 Track reconstruction in the Belle detector

Charged-particle trajectories in the Belle experiment are reconstructed using a dedicated tracking algorithm developed for the Belle detector and optimized for the moderate track multiplicities typical of e^+e^- collisions at the $\Upsilon(nS)$ resonances. The reconstruction relies primarily on information from the CDC sub-detector, complemented by measurements from the SVD at small radii.

The tracking strategy follows a sequential approach in which track finding and track fitting are treated as conceptually distinct steps. Pattern recognition is performed mainly in the CDC, where the large number of measurement layers and low detector occupancy allow for efficient global track reconstruction. Once track candidates are identified, they are refined through helix fitting in a uniform solenoidal magnetic field, with optional attachment of SVD hits to improve vertex resolution.

4.5.1 Helix parameterization of charged tracks

The helix is described by the parameter set

$$(d_0, \phi_0, \kappa, z_0, \tan \lambda),$$

where:

- d_0 is the signed transverse impact parameter, defined as the distance of closest approach to the beam axis in the transverse plane;
- ϕ_0 is the azimuthal angle of the momentum vector at the POCA;
- κ is the signed curvature of the trajectory in the transverse plane;
- z_0 is the longitudinal coordinate of the POCA;
- λ is the dip angle, defined as the angle between the momentum vector and the transverse plane.

The sign of the curvature κ reflects the particle charge and the direction of the magnetic field. In a magnetic field of magnitude B , the curvature is related to the transverse momentum p_T by

$$\kappa = \frac{qB}{p_T}, \quad (4.1)$$

where q is the particle charge. The longitudinal momentum component is related to the dip angle λ through

$$\tan \lambda = \frac{p_z}{p_T}. \quad (4.2)$$

Using the arc length s measured from the POCA along the trajectory, the position of the particle can be expressed parametrically as

$$x(s) = x_0 + \frac{1}{\kappa} [\sin(\phi_0 + \kappa s) - \sin \phi_0], \quad (4.3)$$

$$y(s) = y_0 - \frac{1}{\kappa} [\cos(\phi_0 + \kappa s) - \cos \phi_0], \quad (4.4)$$

$$z(s) = z_0 + s \tan \lambda, \quad (4.5)$$

where (x_0, y_0, z_0) denotes the position of the POCA. The transverse motion corresponds to a circle of radius $R = 1/|\kappa|$, while the longitudinal motion is linear in the arc length, resulting in a three-dimensional helical trajectory.

This parametrization is used internally in the Belle tracking software for both track fitting and extrapolation and provides a consistent description of charged-particle motion throughout the detector volume.

A schematic illustration of the helix parameterization used in the Belle tracking is shown in Fig. 4.1. The figure illustrates the definition of the main helix parameters in the transverse and longitudinal projections, including the transverse impact parameter d_0 , the azimuthal angle ϕ_0 , the curvature radius R , and the longitudinal parameter z_0 .

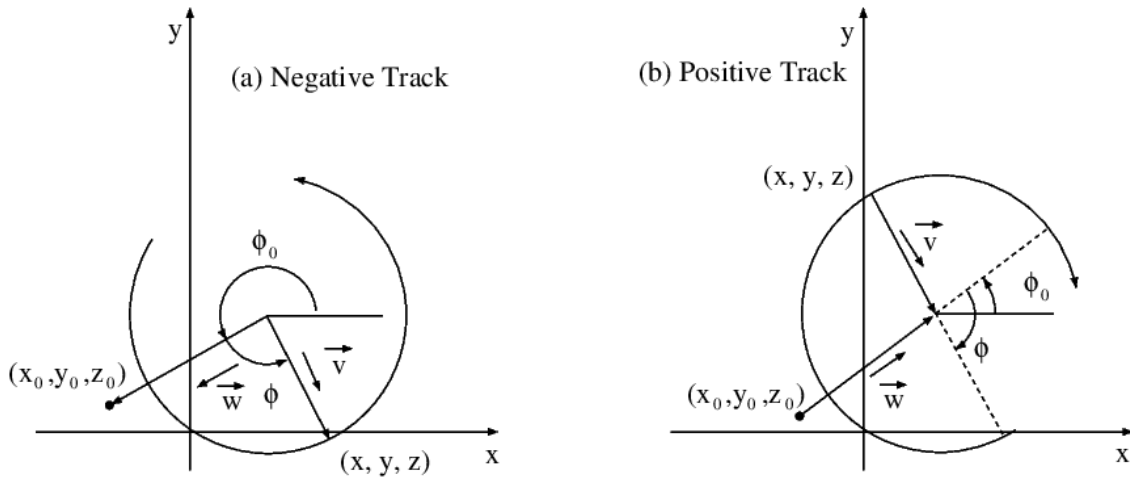


Figure 4.1: Schematic representation of the helix parameterization used in the Belle tracking in the transverse (x - y) plane. The trajectories of negatively charged (left) and positively charged (right) particles are shown, illustrating the sign convention for the curvature and the definition of the azimuthal angle ϕ_0 at the point of closest approach (x_0, y_0, z_0) . The momentum direction and auxiliary vectors used to define the helix parameters are also indicated.

4.5.2 Track finding strategy

The track finding algorithm is designed for the low-occupancy environment characteristic of e^+e^- collisions at the B factories and follows a global pattern-recognition strategy based on the geometry of charged-particle motion in a uniform magnetic field.

In the transverse (x - y) plane, the projection of a charged-particle trajectory is a circle,

$$(x - x_c)^2 + (y - y_c)^2 = R^2, \quad (4.6)$$

where (x_c, y_c) is the center of curvature and R is the radius, related to the transverse momentum through

$$R = \frac{p_T}{|q|B}. \quad (4.7)$$

To simplify the identification of circular trajectories, a conformal transformation is applied to the hit coordinates in the transverse plane. For a hit at position (x, y) , the transformed coordinates (u, v) are defined as

$$u = \frac{x}{x^2 + y^2}, \quad v = \frac{y}{x^2 + y^2}. \quad (4.8)$$

Under this transformation, circles passing sufficiently close to the interaction region are approximately mapped into straight lines in the (u, v) plane. As a result, track candidates can be identified as linear correlations among transformed hit coordinates.

The track-finding procedure starts by forming local hit segments within individual axial superlayers of the CDC. Segments consistent with a common linear pattern in the conformal plane are combined across multiple superlayers to form transverse track candidates. Geometrical consistency criteria are applied to suppress fake and duplicate candidates.

Once a transverse track candidate is identified, hits from the stereo layers are associated to the trajectory. The stereo layers provide sensitivity to the longitudinal coordinate through their tilted wire geometry. Using the measured drift distances and stereo angles, the z coordinate along the trajectory can be reconstructed, allowing determination of the full three-dimensional helix parameters.

The resulting three-dimensional track candidates are then passed to the track-fitting stage, where the helix parameters are refined and track quality requirements are applied. Candidates failing basic criteria, such as a minimum number of associated hits or acceptable fit quality, are rejected.

4.5.3 Track fitting

After the track-finding stage, the parameters of each charged-particle trajectory are refined through a global helix fit using the hit information from the Central Drift Chamber (CDC), and from the silicon vertex detector (SVD) when available. The goal of the track fit is to determine the optimal set of helix parameters and their covariance matrix by minimizing the residuals between the measured hit positions and the predicted helical trajectory in the magnetic field.

The trajectory is described by a five-parameter helix model, defined at the point of closest approach to the interaction region. Given a set of helix parameters $\mathbf{h}$, the expected position of the trajectory at the location of a detector hit is obtained by propagating the helix through the detector volume. For each associated hit i , a residual $r_i(\mathbf{h})$ is defined as

the difference between the measured drift distance and the distance of closest approach between the helix and the corresponding sense wire.

The track fit proceeds by minimizing a global χ^2 function,

$$\chi^2(\mathbf{h}) = \sum_i \frac{r_i^2(\mathbf{h})}{\sigma_i^2}, \quad (4.9)$$

where σ_i denotes the uncertainty associated with the i -th hit, including contributions from drift-time resolution and detector geometry. The minimization yields the best-fit helix parameters $\hat{\mathbf{h}}$ and the corresponding covariance matrix, obtained from the inverse of the Hessian of the χ^2 function at the minimum.

The fit accounts for the helical motion of the particle in the uniform solenoidal magnetic field and incorporates the full three-dimensional geometry of the CDC, including the stereo wire configuration. For tracks with associated SVD hits, the fit is extended to include these measurements, significantly improving the determination of the impact parameters and the vertex resolution.

Quality criteria based on the number of associated hits, the χ^2 value, and the estimated parameter uncertainties are applied to reject poorly reconstructed or spurious track candidates. The resulting fitted tracks provide the input for subsequent reconstruction stages, including particle identification, vertex fitting, and the construction of composite particles.

4.5.4 Tracking performance

The tracking efficiency is evaluated using a data-driven method based on exclusive charmed meson decays, which provide clean and well-constrained control samples. This approach allows the tracking performance to be measured directly from data with minimal dependence on Monte Carlo modeling.

The method exploits fully reconstructed hadronic decays of charged and neutral D mesons, such as

$$D^0 \rightarrow K^- \pi^+, \quad D^+ \rightarrow K^- \pi^+ \pi^+,$$

produced in continuum $e^+e^- \rightarrow c\bar{c}$ events. These decay modes are selected because of their large branching fractions, low background levels, and well-defined final-state topologies.

Tracking efficiency is determined using a “partial reconstruction” technique. A subset of the final-state particles in the D decay is reconstructed and used to infer the presence and kinematic properties of the remaining charged particle. The efficiency is then defined as the probability to reconstruct this remaining track, given that the rest of the decay is successfully reconstructed and satisfies all selection criteria. This procedure effectively measures the single-track reconstruction efficiency in data.

The efficiency is evaluated as a function of relevant kinematic variables, including the laboratory momentum and polar angle of the missing track. Separate studies are performed for different particle species and charge combinations to verify the stability of the result. Background contributions are estimated and subtracted using sideband techniques in the reconstructed D -meson mass distributions.

The results reported in Ref. [83] show that the tracking efficiency for charged particles in the Belle detector exceeds approximately 95% over most of the accessible momentum and angular ranges. For momenta above a few hundred MeV/ c , the efficiency is high and

exhibits only a weak dependence on kinematics. At lower momenta, a gradual decrease in efficiency is observed, consistent with the increased curvature of tracks and reduced hit coverage in the CDC.

The same method is applied to Monte Carlo simulated events, and the ratio of efficiencies measured in data and simulation is found to be consistent with unity within a few percent across the studied phase space. These comparisons are used to derive systematic uncertainties associated with tracking efficiency in Belle physics analyses.

Overall, the Belle tracking algorithm demonstrates high and stable performance, as validated by control-channel studies using exclusive D -meson decays.

4.6 Analysis workflow in basf2

This section describes the analysis workflow implemented within the Belle II Analysis Software Framework (basf2), which is used to process the converted Belle data after the B2BII translation. The workflow is defined through Python steering scripts and consists of a modular sequence of processing steps that apply event- and track-level selections, construct the relevant physics objects, and produce the ntuples used in the offline proton–proton correlation analysis.

4.6.1 Structure of the basf2 processing path

In basf2, a job is defined by a *processing path*, i.e. an ordered sequence of modules that operate on a shared event DataStore. Each module performs a well-defined task and reads from or writes to the DataStore, enabling a flexible and modular data-processing architecture. A schematic overview of this workflow is shown in Fig. 4.2.

For this analysis, a typical processing path includes the following stages:

- Input modules, responsible for reading the B2BII-converted mDST files and initializing the event DataStore.
- Event-level preselection, optionally applied to suppress beam-related backgrounds and unphysical events.
- Track-level selection, where quality requirements such as impact parameters and hit counts are imposed on reconstructed tracks.
- Optional candidate reconstruction, used in dedicated validation studies involving composite particles such as K_S^0 or Λ^0 .
- Ntuple writing, in which the selected tracks and event-level quantities are stored in a compact ROOT format for subsequent offline analysis.

In addition to event-level data stored in the DataStore, basf2 relies on a dedicated Conditions Database to provide detector conditions, calibration constants, alignment parameters, and other time-dependent configuration data. These conditions are retrieved from a centralized Conditions Database Server and made available to the reconstruction and analysis modules through the DBStore interface.

The DBStore provides a uniform and transparent access mechanism to conditions data, ensuring that each module automatically receives the correct set of parameters corresponding to the run number, experiment configuration, and validity interval of the

processed event. This design decouples detector calibration and alignment information from the event data itself and guarantees reproducibility and consistency across different processing campaigns and computing environments.

The processing path is expressed in a concise Python steering script, allowing the same configuration to be executed either locally or on distributed batch systems with minimal modification. The modular design of basf2 ensures that individual processing steps can be developed, validated, and updated independently, providing a transparent and reproducible analysis workflow.

A schematic illustration of the modular processing model adopted in basf2 is shown in Fig. 4.2, highlighting the execution flow of modules along a processing path, the data flow through the event DataStore, and the interaction with the conditions database.

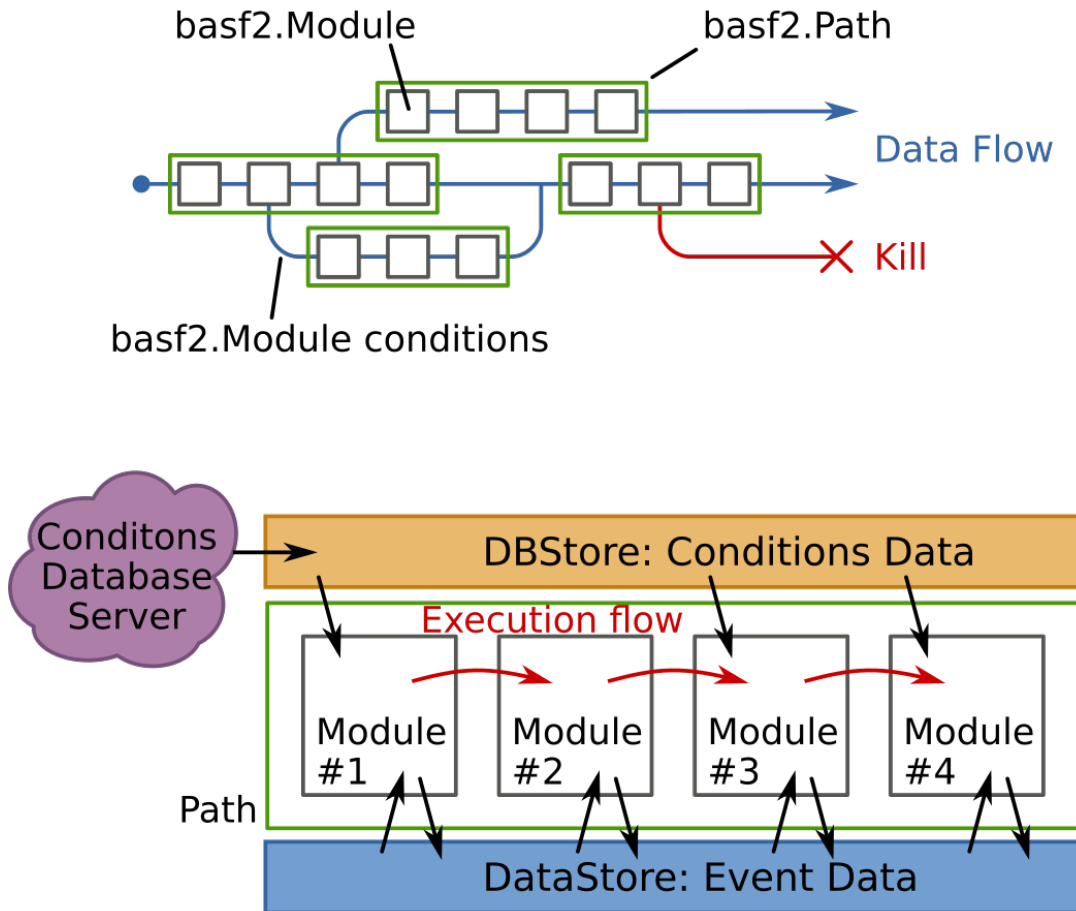


Figure 4.2: Schematic overview of the basf2 modular processing workflow. Events are processed by an ordered sequence of modules along a processing path, defining the execution flow. Modules exchange event-level information via the shared DataStore, while detector conditions and calibration constants are accessed through the conditions database. Conditional paths allow events to be filtered dynamically during processing.

Chapter 5

Candidate selection

This chapter describes the selection of charged-particle candidates used in the analysis. The goal of the selection is to define a clean and well-controlled sample of proton and antiproton tracks while retaining high efficiency, and minimizing biases in the correlation function.

5.1 Dataset

The analysis is based on data collected by the Belle experiment at the $\Upsilon(2S)$ and $\Upsilon(1S)$ resonances. The $\Upsilon(3S)$ sample could be used, but the luminosity integrated in this energy region is too low to allow a statistically significant result. The integrated luminosities of the samples used in this work are $\mathcal{L}_{\Upsilon(2S)} = 24.9 \text{ fb}^{-1}$ and $\mathcal{L}_{\Upsilon(1S)} = 5.7 \text{ fb}^{-1}$, respectively, as summarized in Table 2.2.

Monte Carlo simulated samples are used to validate the selection criteria, study detector effects, and estimate background contributions. Signal events are generated using **EvtGen** [84], while the detector response is simulated with **GEANT3**. The same simulation and reconstruction chain is applied consistently to both the $\Upsilon(2S)$ and $\Upsilon(1S)$ samples.

For each resonance, the signal Monte Carlo sample is produced as a single *stream*, where the term “stream” denotes a simulated dataset corresponding to the same integrated luminosity as the real data. This choice allows for a direct comparison between data and Monte Carlo samples without additional luminosity scaling.

In addition to the signal samples, Monte Carlo samples of background processes are employed. These include continuum light- and charm-quark pair production ($e^+e^- \rightarrow q\bar{q}$ with $q = u, d, s, c$) as well as $\tau^+\tau^-$ pair production. The subsequent τ decays are handled by **TAUOLA** and radiative corrections are simulated with **PHOTOS** [85, 86]. The continuum $e^+e^- \rightarrow q\bar{q}$ and the $e^+e^- \rightarrow \tau^+\tau^-$ processes are generated using **KKMC**. The quark fragmentation is simulated using **PYTHIA 8**. Each background sample is generated with an effective luminosity corresponding to one stream, matching the integrated luminosity of the data samples. These background simulations are used to assess residual non-resonant contributions and to validate the robustness of the candidate selection.

5.1.1 Event skimming criteria

The DST datasets used in this analysis include a wide range of physics processes and non-collision backgrounds, such as QED interactions, $\tau^+\tau^-$ production, two-photon processes, and beam-induced events. Since the analysis targets hadronic final states, a preliminary

event-level selection (“skimming”) is applied to enhance the fraction of such events while retaining a very high signal efficiency.

At Belle, this is achieved through standardized skims based on global event properties, track multiplicities, and calorimetric information. The baseline selection adopted in this work is the **HadronB** skim, designed to select generic hadronic events. For studies explicitly involving charmonium production, the **HadronJ** skim is additionally employed.

The definitions, selection criteria, and performance of these skims are documented in Belle Note 390 [87].

HadronB skim

The HadronB skim selects generic hadronic e^+e^- events while efficiently suppressing backgrounds from QED, $\tau^+\tau^-$, two-photon, and beam-related processes. The selection relies on loose requirements on charged-track quality, event vertex consistency, track multiplicity, visible energy, and global momentum balance.

In particular, events are required to contain at least three good charged tracks originating from the interaction region, a reconstructed primary vertex consistent with the nominal collision point, sufficient visible energy relative to the center-of-mass energy, and a balanced longitudinal momentum. Additional constraints on the total calorimetric energy and its angular distribution further suppress low-multiplicity and forward-peaked backgrounds.

Monte Carlo studies reported in Ref. [87] show that the HadronB skim retains approximately 99% of hadronic $B\bar{B}$ events, indicating that the skim introduces a negligible bias for the observables considered in this analysis.

HadronJ skim

The HadronJ skim is designed for analyses targeting events containing a J/ψ meson. It applies all HadronB requirements and additionally requires the presence of at least one oppositely charged track pair with momentum above 0.8 GeV/ c and an invariant mass compatible with a J/ψ candidate.

When the combined HadronBJ skim is used, the HadronJ criteria act only as an event tag: events that do not contain a J/ψ candidate are retained with the same efficiency as in the HadronB skim and are not rejected.

5.2 Pre-selection

Charged tracks are reconstructed in a fully simulated Monte Carlo (MC) sample of $\Upsilon(2S) \rightarrow$ hadrons. In order to reduce the processing load without introducing biases in kinematic distributions, a random pre-scale of 1% is applied at the *event* level.

On top of the HadronBJ skim, a set of deliberately loose, detector-driven requirements is imposed, using only tracking and time-of-flight (TOF) information:

- **Transverse momentum:** $p_T > 280$ MeV/ c . This threshold matches the TOF geometrical and kinematic acceptance and suppresses strongly curling tracks.
- **Impact parameters:**

- $|dz| < 5$ cm, where dz is the *longitudinal impact parameter*, defined as the distance along the z axis between the track and the nominal interaction point at the point of closest approach to the beam line;
- $dr < 1$ cm, where dr is the *transverse impact parameter*, defined as the radial distance in the r - ϕ plane between the track helix and the beam axis at the point of closest approach.
- **CDC fiducial acceptance:** the track polar angle is required to lie within the nominal CDC coverage (cf. Sec. 3.2.2), ensuring robust pattern recognition and reliable momentum measurement.
- **TOF data quality:** a valid TOF time measurement with good status bits (well-calibrated channel, no saturation or overflow), suitable for a mass hypothesis, is required.
- **Broad TOF mass window:** the TOF mass is computed from the measured momentum p , path length L , and flight time Δt according to

$$\beta \equiv \frac{v}{c} = \frac{L}{c \Delta t}, \quad m_{\text{TOF}} = \frac{p}{c} \sqrt{\beta^{-2} - 1},$$

and the requirement

$$0.700 < m_{\text{TOF}} < 1.700 \text{ GeV}/c^2$$

is imposed. The m_{TOF} distribution measured on simulated data, together with the defined selection, is shown in Fig. 5.1. This intentionally wide window is centered around the proton mass and retains essentially all proton (and antiproton) candidates, while rejecting clear outliers originating from mistimed or non-hadronic tracks.

These loose criteria provide a seed sample for the subsequent selection step, with 96.7% efficiency.

5.2.1 Analysis of the pre-selected sample

As a first step following the loose preselection described in Sec. 5.2, the spatial origin of reconstructed tracks in the Monte Carlo sample is examined. Since the simulation provides access to the true production vertex of each particle, a direct distinction can be made between *primary* and *non-primary* tracks:

- **Primary:** particles produced in the primary e^+e^- annihilation or in subsequent hadronization or decay processes of short-lived resonances, prior to any interaction with detector materials.
- **Non-primary:** particles originating from secondary interactions with the detector material, such as hadronic interactions, photon conversions, or long-lived weak decays occurring at displaced vertices, e.g. $\Lambda^0 \rightarrow p\pi$.

Within this classification, four mutually exclusive categories are defined:

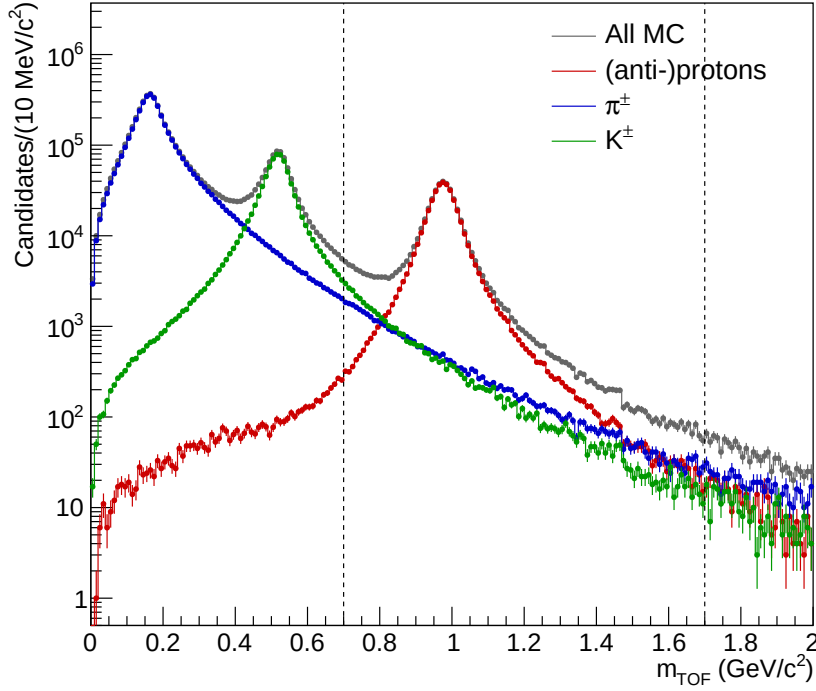


Figure 5.1: Distribution of the TOF mass observed on simulated data. The black dashed lines enclose the selected region.

1. *Primary protons and antiprotons* — the genuine signal of interest for femtoscopic studies, produced promptly at the interaction point.
2. *Protons and antiprotons from hyperon decays* — secondary baryons originating from weak decays of strange hyperons (e.g. Λ , Σ), representing the dominant feed-down contribution.
3. *Protons and antiprotons from material interactions* — secondary baryons produced in hadronic interactions with the detector material, typically characterized by large displacements from the interaction point.
4. *All remaining particles* — true charged particles other than protons and antiprotons (predominantly pions and kaons), originating either from the interaction point or from secondary processes in the detector material, and included here solely for an initial qualitative visualization of background contributions.

To visualize the spatial origin of these components, the (x, y) coordinates of the simulated production vertices are inspected separately for each of the four categories. Genuine primary particles are expected to cluster tightly around the interaction region centered on the beam spot, whereas non-primary particles are expected to exhibit broader and more structured spatial distributions, including ring-like features corresponding to the beam pipe and inner detector layers. This comparison provides a validation of the pre-selection, enables a quantitative assessment of background contamination, and motivates the impact-parameter requirements that are refined in later stages of the analysis.

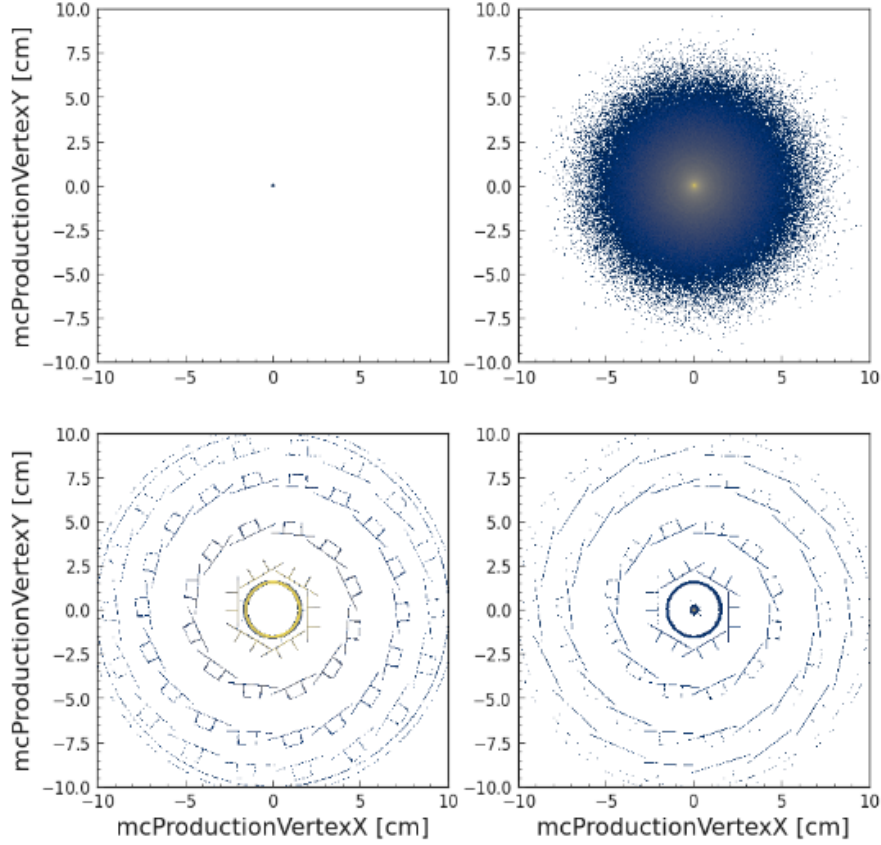


Figure 5.2: MC truth-matched (x, y) production coordinates for the signal and main background components. The upper left pad shows the signal (prompt p and $\bar{p}$). The upper right pad shows p and $\bar{p}$ produced in hyperon decays. The lower left pad shows p and $\bar{p}$ produced in the detector. The lower right pad shows the remaining background, consisting mainly of pions and kaons.

5.3 Candidate selection

5.3.1 SVD hit requirement

From the inspection of the simulated production vertices (Sec. 5.2.1), it is observed that non-primary tracks often arise from secondary interactions in the beam pipe or in the inner detector layers. Such tracks typically leave fewer hits in the Silicon Vertex Detector (SVD) than prompt tracks originating from the interaction point. A requirement on the number of associated SVD hits, $N_{\text{hits}}^{\text{SVD}}$, is therefore introduced to enhance the purity of primary protons. Since the SVD consists of double-sided silicon sensors, each crossed layer contributes two hits; consequently, $N_{\text{hits}}^{\text{SVD}}$ always takes even values.

Since the goal is not to optimize a signal significance but rather to improve the purity of the prompt-proton sample for femtoscopy, while maintaining a high efficiency, no figure-of-merit maximization is performed.

The values $N_{\text{hits}}^{\text{SVD}} \in \{3, 5, 7, 9\}$ are considered, and for each threshold the following quantities are determined:

- the *efficiency*, defined as the fraction of primary (anti-)protons retained by the selection relative to the preselected sample;

- the *purity*, defined as the fraction of selected candidates that are true primary (anti-)protons.

Figure 5.3 shows the efficiency–purity curve and the signal and background distributions of the $N_{\text{hits}}^{\text{SVD}}$ observable.

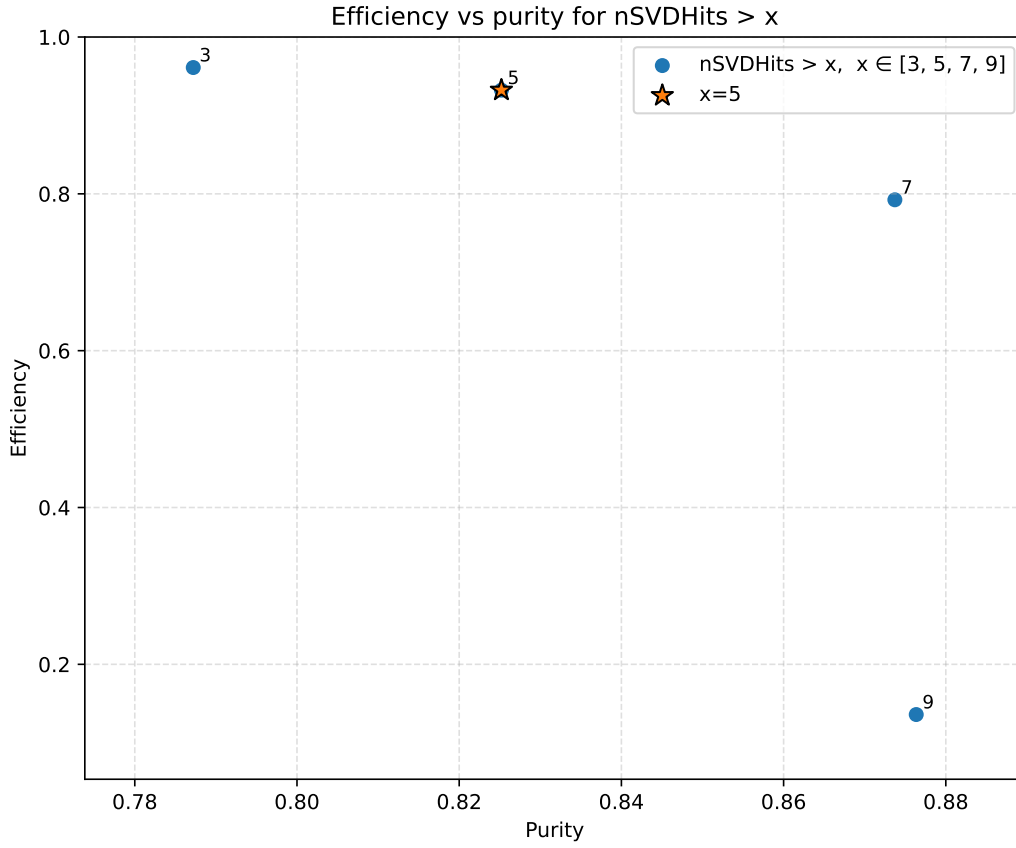
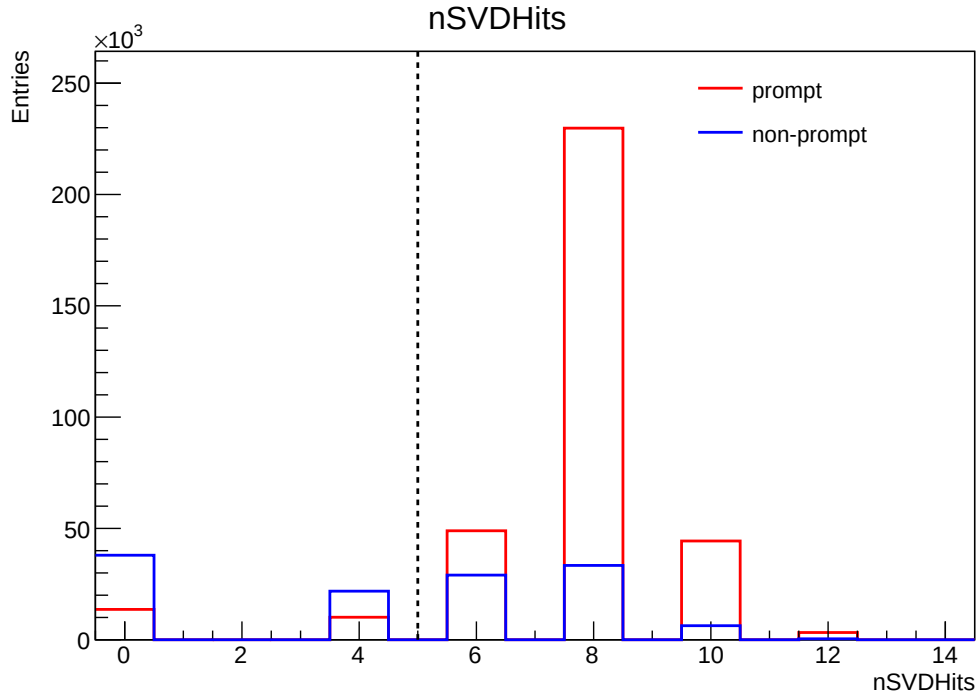
(a) Efficiency–purity curve for the N_{hits}^{SVD} requirement.(b) Signal and background N_{hits}^{SVD} distributions.

Figure 5.3: Optimization of the N_{hits}^{SVD} selection. Top: efficiency–purity trade-off as the cut is scanned. Bottom: signal and background distributions, with the adopted cut shown as a vertical dashed line.

The results indicate that requiring more than five SVD hits provides the best compromise between efficiency retention and suppression of non-primary contamination. The baseline selection criterion for proton and antiproton candidates is therefore defined as

$$N_{\text{hits}}^{\text{SVD}} > 5.$$

5.3.2 ACC photo-electron requirement

To further suppress contamination from non-proton tracks, the response of the Aerogel Cherenkov Counter (ACC) is exploited (see Sec. 3.2.4). For protons in the relevant momentum range, the velocity is typically below the Cherenkov threshold, and the number of photo-electrons ($N_{\text{p.e.}}^{\text{ACC}}$) is therefore expected to be small, whereas lighter hadrons (pions and kaons) produce larger signals.

The $N_{\text{p.e.}}^{\text{ACC}}$ distributions are studied for the four classes defined in Sec. 5.2.1, and the efficiency and purity of the proton selection are evaluated for a series of *upper* thresholds scanned in unit steps:

$$N_{\text{p.e.}}^{\text{ACC}} < t, \quad t \in \{0, 1, 2, \dots, 20\}.$$

Here, the efficiency is defined as the fraction of primary (anti-)protons retained by the selection, while the purity is defined as the fraction of selected candidates that are true primaries.

Based on this scan, the requirement

$$N_{\text{p.e.}}^{\text{ACC}} < 5$$

is adopted as the default ACC-based selection, as it provides the best compromise between efficiency and purity for the preselected sample. The corresponding efficiency–purity curve and the $N_{\text{p.e.}}^{\text{ACC}}$ distributions for signal and background Monte Carlo are shown in Fig. 5.4.

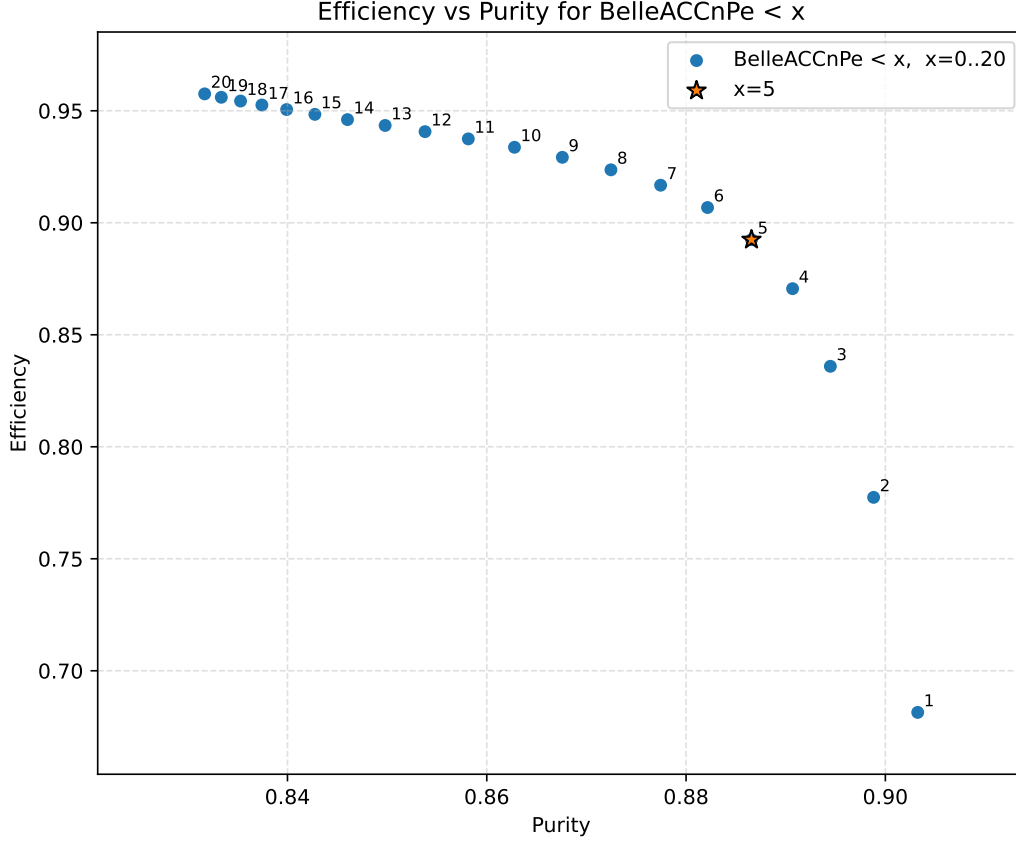
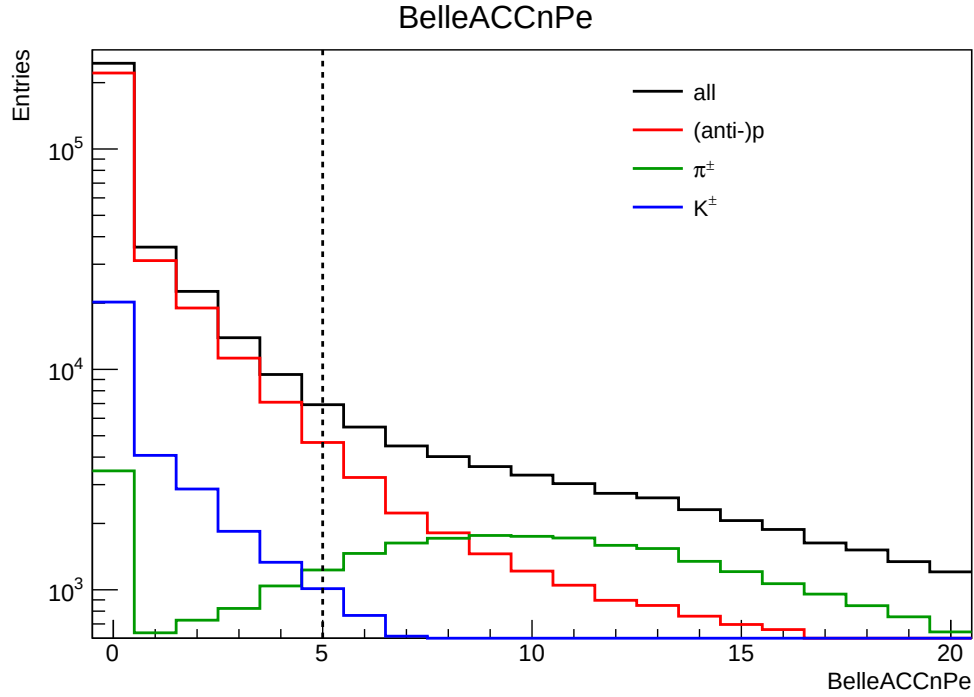
(a) Efficiency–purity curve for the $N_{\text{p.e.}}^{\text{ACC}}$ requirement.(b) Signal and background $N_{\text{p.e.}}^{\text{ACC}}$ distributions.

Figure 5.4: Optimization of the $N_{\text{p.e.}}^{\text{ACC}}$ selection. Top: efficiency–purity trade-off as the cut is scanned. Bottom: signal and background distributions, with the adopted cut shown as a vertical dashed line.

5.3.3 TOF-based selection

After reducing the contamination from non-proton and non-primary tracks through the SVD and ACC requirements, the initial 1% pre-scale is removed and the analysis is continued using the *full* $\Upsilon(2S)$ MC sample. This ensures that the subsequent selection steps are performed on the highest available statistics.

The quality of the TOF information is investigated using experimental data collected in Belle experiment 71. In this dataset, the reconstructed TOF mass m_{TOF} exhibits unexpected non-Gaussian tails that are not reproduced by the MC simulation. To diagnose this behavior, the correlation with the velocity estimate from TOF is examined,

$$\beta_{\text{TOF}} = \frac{l_{\text{TOF}}}{ct_{\text{TOF}}},$$

where l_{TOF} is the measured path length and t_{TOF} the measured time of flight of the particle.

The two-dimensional $(m_{\text{TOF}}, \beta_{\text{TOF}})$ distribution measured on real data reveals a clear structure at nearly constant $\beta_{\text{TOF}} \simeq 0.34$, as shown in Fig. 5.5a. Since such a feature has no physical origin, it is attributed to spurious TOF hits that were incorrectly flagged as valid during data taking. These entries degrade the quality of the m_{TOF} distribution and, if not removed, would bias the particle-identification selection.

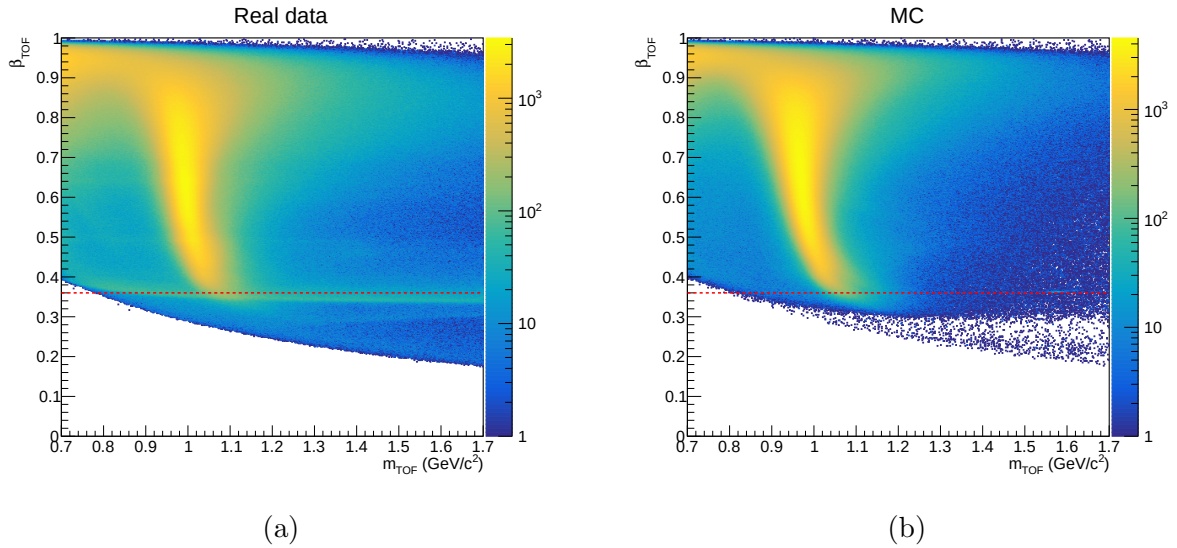


Figure 5.5: Candidates' TOF velocity β_{TOF} versus TOF mass m_{TOF} for selected real data (a) and MC data (b). The red dashed lines represent the selection $\beta_{\text{TOF}} > 0.36$.

To suppress this artifact, the requirement

$$\beta_{\text{TOF}} > 0.36$$

is imposed, which effectively eliminates the spurious structure while preserving genuine proton and antiproton candidates. This additional quality requirement significantly improves the agreement between simulation and data and is therefore adopted as part of the baseline selection.

Furthermore, by studying the distribution of m_{TOF} as a function of the track's momentum p in real data, it is observed that proton candidates with reliable TOF response

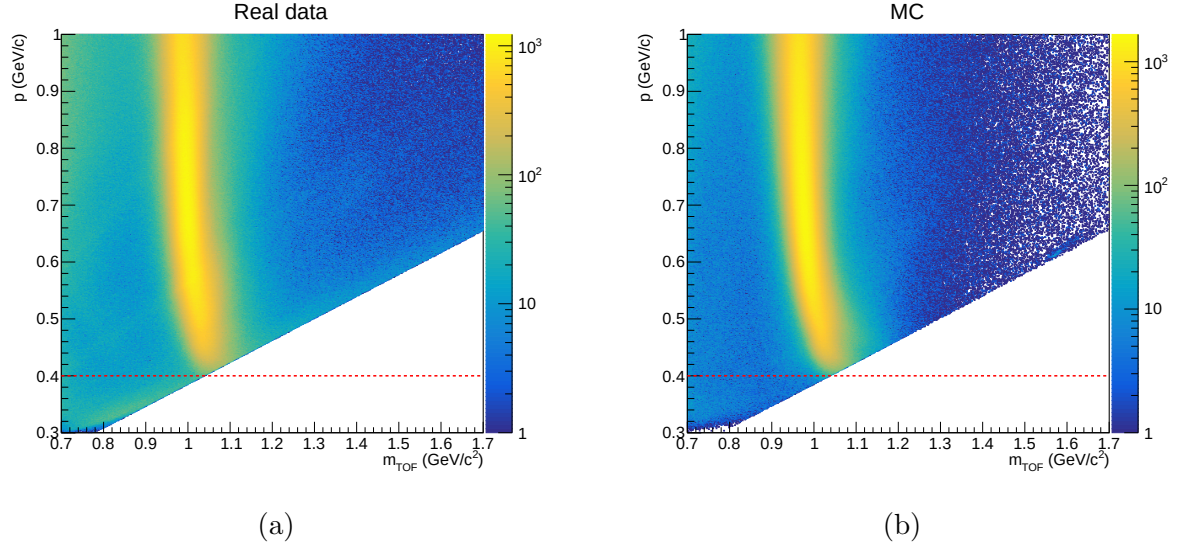


Figure 5.6: Candidates' momentum vs m_{TOF} distribution for selected real data (a) and MC (b). The red dashed lines represent the selection $p > 400 \text{ MeV}/c$.

populate the region above $p > 400 \text{ MeV}/c$. The distributions are shown in figure 5.6 for the selected $\Upsilon(2S)$ data and MC samples. To ensure a robust selection, the additional requirement

$$p > 400 \text{ MeV}/c$$

is imposed.

5.3.4 Momentum-dependent dE/dx selection

The final step of the baseline selection makes use of the ionization energy loss measured in the SVD and CDC. The specific energy loss dE/dx follows the Bethe–Bloch curve, which exhibits a clear dependence on particle momentum. This property allows dE/dx to be exploited as an additional handle to separate protons and antiprotons from lighter hadrons and leptons. The momentum-dependent dE/dx distributions for $\Upsilon(2S)$ MC and data samples are shown in Fig. 5.7. In these distributions, the structure observed in data is broader than in MC, reflecting residual detector effects and imperfect modeling of the ionization response, and an additional band corresponding to deutons is clearly visible in data, while it is absent in the simulated sample.

To construct a practical selection, the following procedure is adopted:

- the track momentum is scanned in the range $0.4 \text{ GeV}/c < p < 1.4 \text{ GeV}/c$, dividing it into bins of $20 \text{ MeV}/c$ width;
- for each momentum bin, the one-dimensional distribution of dE/dx values obtained from the MC sample is examined;
- a lower and an upper cut on dE/dx are manually defined in each bin so as to maximize the purity of (anti-)proton candidates while retaining good efficiency.

The resulting dE/dx selection follows the characteristic proton band in the $(p, dE/dx)$ plane and efficiently rejects background from pions and kaons. The selected data are shown in Fig. 5.8.

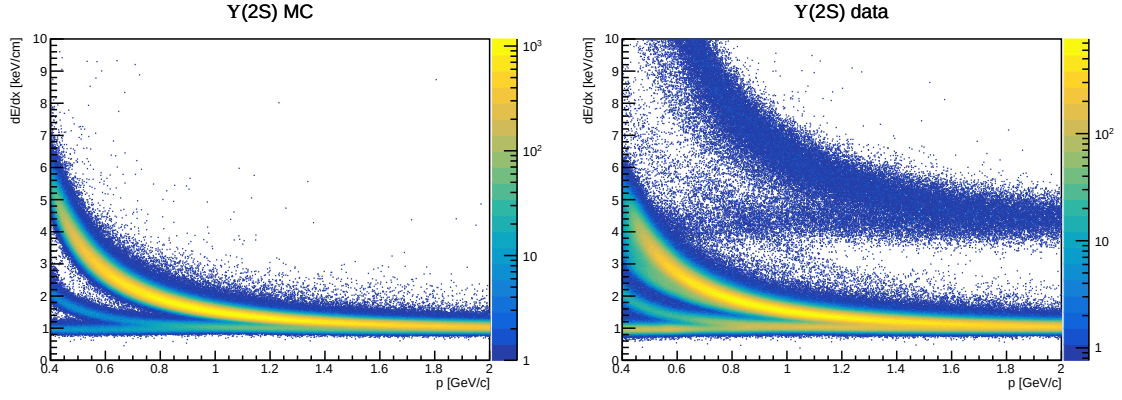


Figure 5.7: Energy deposited per unit length as a function of momentum for proton candidates reconstructed in the $\Upsilon(2S)$ MC and data samples.

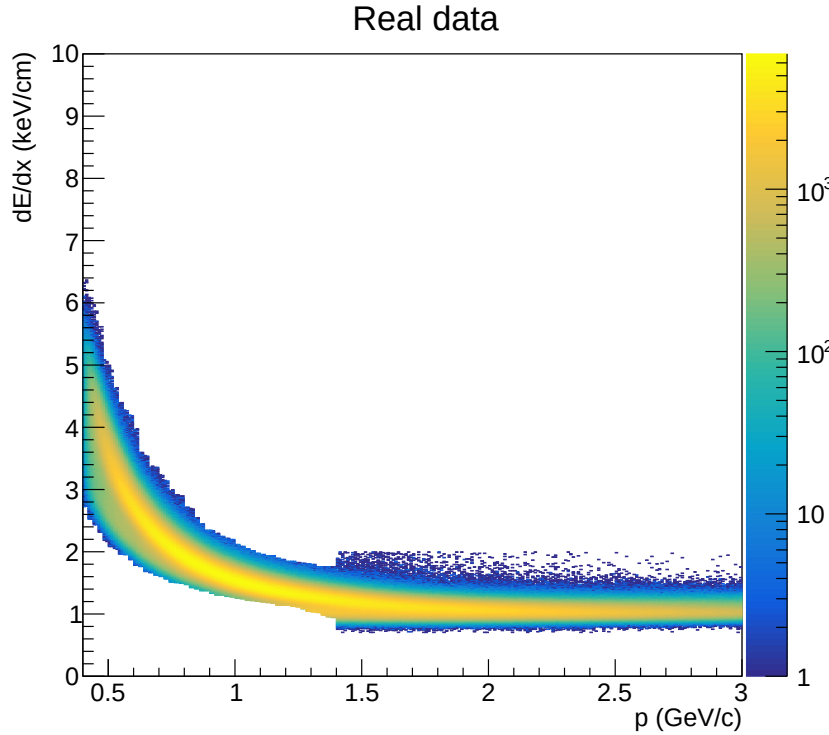


Figure 5.8: Two-dimensional distribution showing the applied momentum-dependent dE/dx selection on real $\Upsilon(2S)$ data.

This selection is particularly effective in the low-momentum region, where particle separation is largest. At higher momenta the bands overlap more strongly and the discrimination power decreases accordingly; however, the complementary TOF and ACC requirements mitigate this loss.

These dE/dx -based cuts, together with the previously defined TOF, SVD, and ACC requirements, complete the baseline selection of proton and antiproton candidates.

5.4 Removal of clone tracks

The presence of clone tracks, i.e. multiple reconstructed tracks originating from the same physical particle, constitutes a potentially significant source of background in femtoscopic analyses. In particular, clone tracks can lead to artificial enhancements of the correlation function at low relative momentum. Since femtoscopic observables are especially sensitive to pairs with small values of k^* , even a small fraction of clone tracks can produce sizable and non-physical correlations if not properly suppressed.

Pairs containing at least one clone proton track are therefore expected to contribute spurious structures to the measured proton–proton correlation function, particularly in the low- k^* region. For this reason, it is essential to estimate the magnitude of this background and to suppress it as effectively as possible.

In order to address this issue, two complementary methods are employed and their results are compared as a consistency check. In the first method, described in Section 5.4.1, proton candidates reconstructed in $\Upsilon(2S)$ MC and data samples are analyzed using information from the CDC and the ECL to identify and suppress clone tracks based on their detector response. In the second method, described in Section 5.4.2, the decay $\Upsilon(2S) \rightarrow p\Lambda \rightarrow pp\pi$ is reconstructed and the $M(pp\pi)$ invariant-mass distribution is examined for the presence of a spurious enhancement near threshold, which would signal the contribution of fake or clone proton tracks.

The agreement between the two approaches provides a robust validation of the clone-removal procedure.

5.4.1 Detector information for proton candidates

Pairs of tracks (labeled p_0 and p_1) are reconstructed using the full $\Upsilon(2S)$ dataset and the corresponding run-dependent MC stream. The tracks are selected according to the criteria described above, with the exception that the $N_{\text{hits}}^{\text{SVD}}$ requirement is not applied in this step in order to deliberately enrich the sample with clone candidates.

In MC, clone tracks are recognized using the `genParticleID` variable. This variable stores the index of the associated `MCParticle` object within the `MCParticleArray` and thus provides a direct link between a reconstructed track and the corresponding simulated particle. The `MCParticle` and `MCParticleArray` classes are part of the basf2 data format and store the generator-level information of all simulated particles in an event. Since each `MCParticle` uniquely represents a generated particle in the event record, tracks sharing the same `genParticleID` are identified as originating from the same simulated particle and are therefore classified as clone tracks. If the two proton candidates forming a pair share the same `genParticleID`, they are identified as clone tracks, corresponding to multiple reconstructed tracks associated with the same simulated particle.

In experimental data, where generator-level information is not available, the presence of clone tracks is investigated using detector-level observables. Clone tracks originate from multiple reconstructions of the same charged-particle trajectory and therefore tend to share detector information with the corresponding genuine track. As a consequence, such tracks may exhibit correlated responses in both the tracking detectors and the calorimeter.

To characterize these effects, observables related to tracking and calorimetry are studied. In the central drift chamber (CDC), clone tracks are expected to share a significant fraction of the same hits, since they correspond to nearly identical trajectories recon-

structed from the same ionization pattern. This results in atypical values of the number of associated CDC hits ($N_{\text{hits}}^{\text{CDC}}$), which can be used as a sensitive handle to identify clone candidates.

In addition, clone tracks may be associated with the same electromagnetic calorimeter (ECL) cluster. In such cases, the measured calorimetric energy is shared between the two reconstructed tracks, leading to characteristic distortions in the energy–momentum correlation. The E_{ECL} versus momentum distribution is therefore used to identify pairs containing a clone track, as discussed in detail in Section 5.4.1.

CDC information

In Fig. 5.9 the distributions of the number of CDC hits, $N_{\text{hits}}^{\text{CDC}}$, for the two tracks forming a pair are shown for both data and Monte Carlo samples. The dominant population corresponds to pairs in which both tracks exhibit a large number of CDC hits, as expected for well-reconstructed proton tracks traversing the full CDC volume, which typically accumulate of the order of fifty hits. In addition to this main population, two secondary structures are visible, centered around $(x, y) \approx (15, 25)$ and $(x, y) \approx (25, 15)$, which do not follow the regular hit multiplicity pattern of genuine track pairs.

These structures are interpreted as signatures of clone tracks. In such cases, the hits produced by a single charged particle are not consistently associated to a unique reconstructed trajectory but are instead split between two reconstructed tracks. As a consequence, one of the two tracks retains a large fraction of the CDC hits, while the other is reconstructed with a significantly reduced number of hits. This asymmetric sharing of detector hits gives rise to the off-diagonal structures observed in the $N_{\text{hits}}^{\text{CDC}}$ correlation plane and provides a clear experimental signature of the presence of clone tracks.

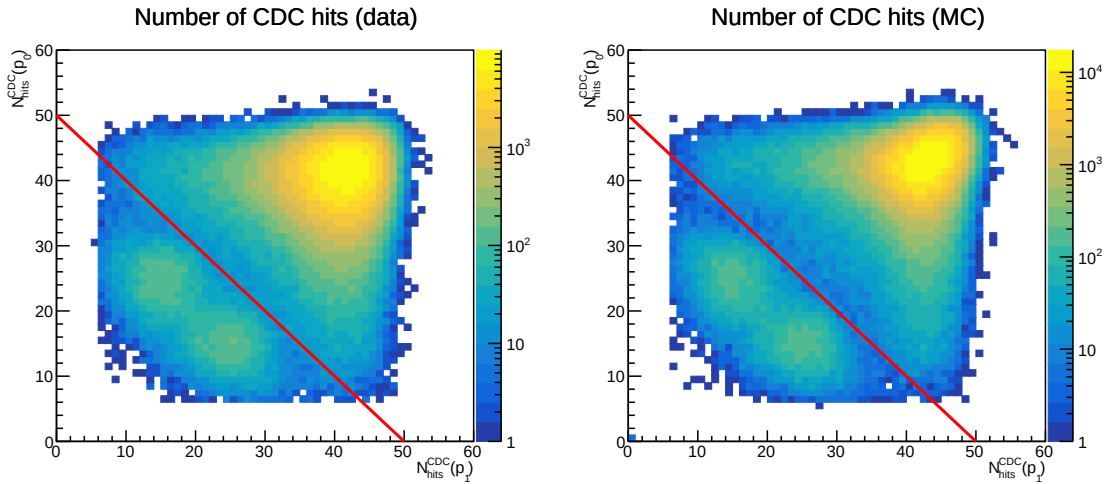


Figure 5.9: Distributions of the number of CDC hits in experimental data (left) and MC (right). The red line corresponds to the equation $N_{\text{hits}}^{\text{CDC}}(p_0) + N_{\text{hits}}^{\text{CDC}}(p_1) = 50$.

In Monte Carlo samples, clone tracks are identified by requiring that the two proton candidates forming a pair share the same `genParticleID`, i.e. `genParticleID(p0) = genParticleID(p1)`. The corresponding distribution in the $N_{\text{hits}}^{\text{CDC}}(p_0)$ versus $N_{\text{hits}}^{\text{CDC}}(p_1)$ plane shows that such clone pairs predominantly populate the region characterized by a low total number of CDC hits, with $N_{\text{hits}}^{\text{CDC}}(p_0) + N_{\text{hits}}^{\text{CDC}}(p_1) < 50$, which is the maximum

number of hits created by a single particle. Guided by this observation, a clone-enriched sample is selected in experimental data by applying the same requirement on the sum of CDC hits, $N_{\text{hits}}^{\text{CDC}}(p_0) + N_{\text{hits}}^{\text{CDC}}(p_1) < 50$, which allows the characteristics of clone tracks to be studied in the absence of Monte Carlo truth information. The resulting samples are shown in Fig. 5.10.

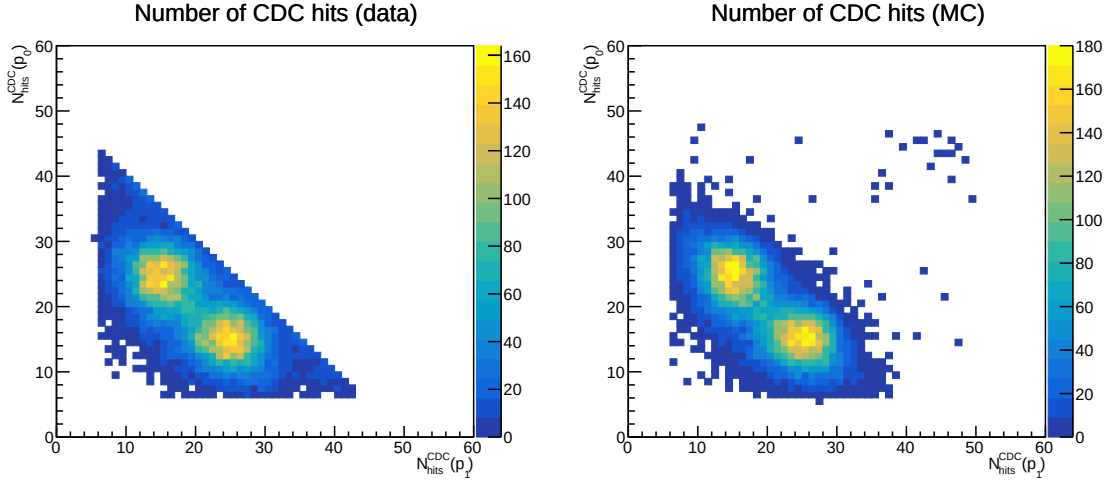


Figure 5.10: Distributions of the number of CDC hits for clones in experimental data (left) and MC (right).

To suppress fake and clone proton tracks, a minimum requirement is imposed on the number of hits in the SVD and in the CDC for each individual track, namely $N_{\text{hits}}^{\text{SVD}} > 5$ and $N_{\text{hits}}^{\text{CDC}} \geq 30$ for both candidates in the pair. The requirement on the number of SVD hits is the same as that introduced earlier in the track-quality selection (see Section 5.3.1) and is applied here to ensure a consistent definition of well-reconstructed tracks throughout the analysis.

The requirement on the number of CDC hits is instead chosen to explicitly suppress clone tracks and is therefore relatively stringent. Studies based on Monte Carlo samples show that clone tracks are characterized by a reduced number of CDC hits for at least one of the reconstructed trajectories, reflecting the splitting of detector hits between multiple track candidates. In contrast, genuine proton tracks traversing the full CDC volume typically accumulate a substantially larger number of hits. The threshold $N_{\text{hits}}^{\text{CDC}} \geq 30$ is selected to efficiently reject clone tracks while preserving the bulk of the signal, without attempting a fine numerical optimization. The sufficiency of this selection for removing clone proton tracks is verified using experimental data, as described in Secs. 5.4.1 and 5.4.2.

ECL information

To further study the contamination from clone tracks, pairs are selected for which the two tracks share the same ECL cluster. Both tracks are required to have positive charge in order to avoid complications arising from the complex interactions of antiprotons in the ECL crystals. In addition, the requirement $k^* < 300 \text{ MeV}/c$ is imposed. The deposited energy E_{ECL} is studied as a function of momentum. The distribution obtained after dropping the selections on $N_{\text{hits}}^{\text{SVD}}$ and $N_{\text{hits}}^{\text{CDC}}$, as described in Sec. 5.4.1, is shown in Fig. 5.11.

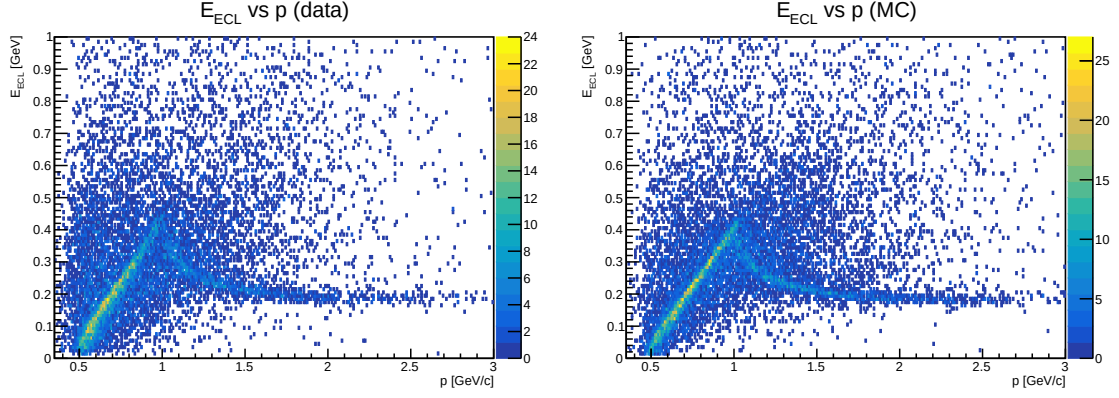


Figure 5.11: Cluster energy vs momentum for pairs of tracks sharing the same cluster, satisfying $k^* < 300 \text{ MeV}/c$ and the loose selection described in Sec. 5.4.1.

For the majority of candidates, the E_{ECL} versus p distribution follows the curve expected for a single proton. This indicates that most of the selected pairs contain one clone track, with both reconstructed trajectories being associated to the same calorimetric energy deposit produced by a single proton. In Fig. 5.12, the same distributions are shown after requiring $N_{\text{hits}}^{\text{SVD}} > 5$ and $N_{\text{hits}}^{\text{CDC}} \geq 30$ for both p_0 and p_1 .

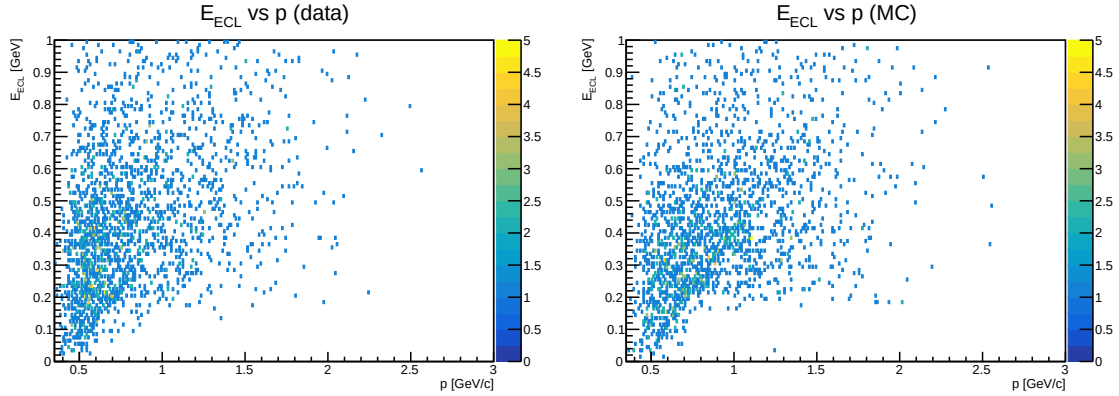


Figure 5.12: Cluster energy vs momentum for pairs of tracks sharing the same cluster, after applying the stringent requirements $N_{\text{hits}}^{\text{SVD}} > 5$ and $N_{\text{hits}}^{\text{CDC}} \geq 30$ to both p_0 and p_1 .

After the application of the stricter selection, no structure consistent with that expected from a single proton track is observed, and the measured energy loss lies above the corresponding curve. In addition, after applying these requirements, no candidate satisfying `genParticleID(p0)==genParticleID(p1)` remains in the MC sample. It is therefore concluded that the combined requirements on $N_{\text{hits}}^{\text{SVD}}$ and $N_{\text{hits}}^{\text{CDC}}$ are sufficient to suppress the clone-track background while preserving a high signal efficiency.

5.4.2 Inclusive $\Upsilon(2S) \rightarrow p\Lambda + X$ decays

To perform a more robust test, candidate $p\Lambda^0$ pairs (and charge conjugates) are reconstructed in the $\Upsilon(2S)$ data sample. Proton candidates are selected as described above, except for the $N_{\text{hits}}^{\text{SVD}}$, $N_{\text{hits}}^{\text{CDC}}$, and momentum-dependent dE/dx requirements, which are relaxed for this study.

The Λ^0 baryons are reconstructed through the decay channel $\Lambda^0 \rightarrow p\pi^-$ and selected using a set of standard Belle criteria, encoded in the selection flag `goodBelleLambda`. The requirement `goodBelleLambda==2` corresponds to a tight selection optimized for purity and is based on a combination of topological and kinematic constraints. In particular, it requires a well-reconstructed displaced decay vertex significantly separated from the interaction point, a small distance of closest approach between the proton and pion tracks at the decay vertex, and a reconstructed Λ^0 invariant mass compatible with the nominal value within a narrow window. Additional requirements are imposed on the pointing angle between the reconstructed Λ^0 momentum and the vector connecting the interaction point to the decay vertex, ensuring consistency with a long-lived neutral particle decaying downstream of the primary vertex. This selection follows the standard prescription adopted in Belle analyses and is described in detail in Belle Note 684 [88].

Moreover, the coordinates x and y of the reconstructed Λ^0 decay vertex are required to lie within a circle of radius 3 cm centered on the interaction point. This additional requirement ensures that the proton and pion daughter tracks traverse all layers of the SVD, providing well-defined tracking information for the subsequent analysis.

In the presence of clone proton tracks, a characteristic peak appears in the $M(pp\pi)$ invariant mass distribution. When the same physical proton is reconstructed multiple times and combined with a correctly reconstructed Λ^0 candidate, the resulting $pp\pi$ system exhibits an artificial accumulation of events near the kinematic threshold. This produces a sharp peak at threshold, which therefore serves as a clear experimental signature of clone contamination in the selected dataset. The results for the $M(pp\pi)$ distribution before and after applying the selections $N_{\text{hits}}^{\text{SVD}} > 5$ and $N_{\text{hits}}^{\text{CDC}} \geq 30$ for proton candidates are compared, as shown in Fig. 5.13. A sharp peak at threshold is observed with the loose selection. This peak disappears after applying the stringent selection mentioned above.

It is therefore concluded that the applied selection correctly removes the background arising from clone proton tracks in the analyzed samples.

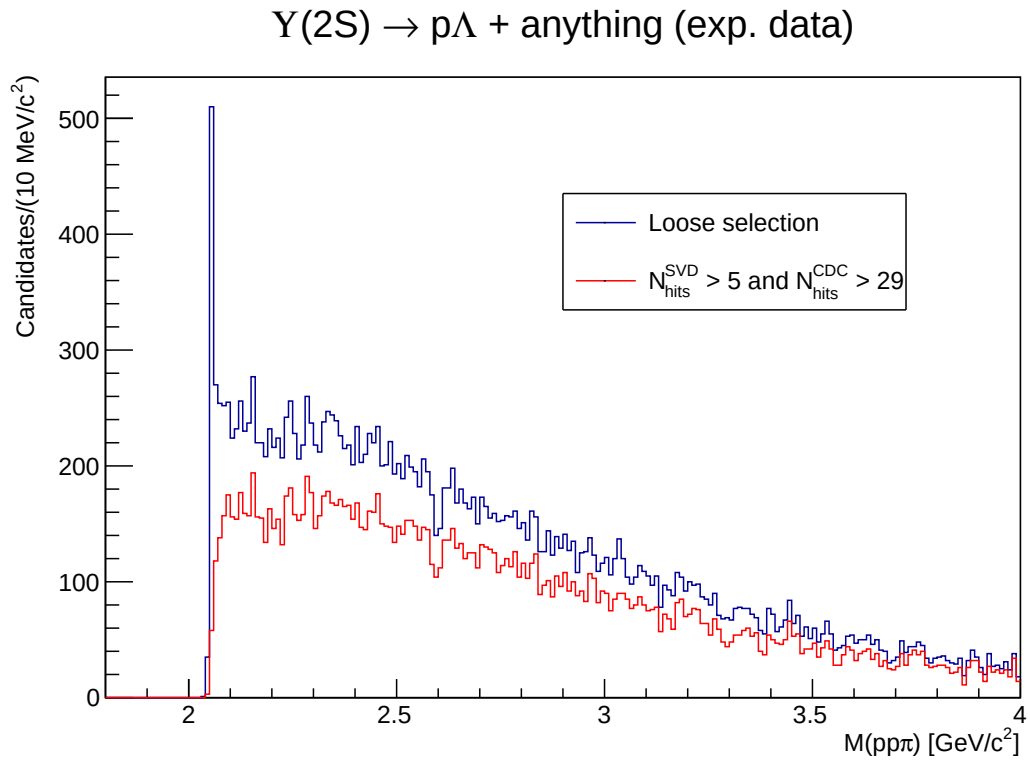


Figure 5.13: Invariant mass of $p\Lambda$ pairs produced in $\Upsilon(2S)$ decays under two different selections. The blue histogram corresponds to the loose selection described in Sec. 5.4.2. The red histogram results after adding the selections on $N_{\text{hits}}^{\text{SVD}}$ and $N_{\text{hits}}^{\text{CDC}}$.

Chapter 6

Prompt proton identification

To provide a precise measurement of the correlation among pp and $\bar{p}\bar{p}$ pairs produced in $\Upsilon(nS)$ decays, prompt (anti)protons must be identified with high purity. The dr variable is used to distinguish promptly produced tracks from tracks originating from secondary vertices. The latter category includes tracks generated via hyperon decays and interactions with the detector material. Protons are separated from pions and kaons using the time-of-flight mass (m_{TOF}) variable. The fraction of prompt protons is estimated by performing a joint fit to the m_{TOF} and dr distributions of single proton candidates. To this end, a two-dimensional model is constructed following the procedure described in the next sections. The fits are performed in bins of the candidate momentum. Positively and negatively charged tracks are treated separately, as their m_{TOF} distributions exhibit small differences.

The prompt proton fraction (f_p^{prompt}) will be used as a weight for each candidate when pairing the tracks in both the same and mixed events. To reduce the statistical fluctuations on f_p^{prompt} , the (p, f_p^{prompt}) data points are smoothed using a non-parametric kernel regression based on the Nadaraya–Watson estimator. The smoothed dependence is obtained as a locally weighted average of the measured points, without imposing a specific functional form. The associated uncertainty is evaluated from the effective weights entering the kernel estimator. This procedure is described in detail in Section 6.3.2.

6.1 Two-dimensional model for prompt proton identification

This section describes the statistical model used to extract the prompt-proton component from the selected sample, together with the adopted fit strategy. An unbinned two-dimensional maximum likelihood fit is performed in the space defined by m_{TOF} and dr . The fits are performed separately in bins of the single-particle momentum and for each charge sign. While the fits in different momentum bins are statistically independent, the fitted parameters obtained in a given bin are used to initialize the fit in the subsequent bin, in order to improve convergence and stability.

6.1.1 Momentum binning and charge separation

The fit is performed in adaptive momentum intervals, chosen to ensure sufficient statistical precision while preserving sensitivity to the momentum dependence of the detector

response and background composition. The momentum bins are defined such that each interval contains approximately a fixed number of reconstructed (anti)proton candidates:

- $\sim 10\,000$ candidates for $0.4 < p < 1.5$ GeV/ c ,
- $\sim 20\,000$ candidates for $1.5 < p < 2.0$ GeV/ c ,
- $\sim 40\,000$ candidates for $p > 2.0$ GeV/ c .

This adaptive binning strategy guarantees stable fits over the full momentum range relevant for the femtoscopy analysis, while maintaining comparable statistical power across bins.

Positive and negative tracks are fitted independently. This choice reflects the presence of charge-asymmetric effects in the detector response, most notably in the time-of-flight system and in the reconstruction of low-momentum tracks, and avoids introducing artificial constraints between proton and antiproton samples.

6.1.2 Structure of the two-dimensional probability density function

The full probability density function (PDF) used to describe the data in a given momentum bin and charge category is constructed as the product of two one-dimensional PDFs,

$$\mathcal{P}(m_{\text{TOF}}, dr) = \mathcal{P}_{\text{TOF}}(m_{\text{TOF}}) \times \mathcal{P}_{dr}(dr), \quad (6.1)$$

where correlations between m_{TOF} and dr are assumed to be negligible after the applied selection requirements. Each component is itself a weighted sum of signal and background contributions, as detailed below.

The validity of the factorized form is verified by studying the joint distribution of m_{TOF} and dr in both data and simulated events. Figure 6.1 shows the two-dimensional density m_{TOF} versus dr , together with the corresponding x -profiles (mean $\langle m_{\text{TOF}} \rangle$ in bins of dr). The two-dimensional distributions exhibit the expected narrow proton band in m_{TOF} with no visible diagonal structures, indicating the absence of a strong correlation. Quantitatively, the linear (Pearson) correlation coefficient is found to be at the few-percent level (below 7%) in both data and simulation, which is consistent with a weak dependence.

The profile curves in Fig. 6.1 display a small and smooth variation of $\langle m_{\text{TOF}} \rangle$ as a function of dr . A localized structure is observed at very small dr (below a few 10^{-2} cm), compatible with detector- and reconstruction-driven effects for tracks with the tightest vertex constraint. At larger dr , the mild drift of $\langle m_{\text{TOF}} \rangle$ remains limited compared to the intrinsic width of the m_{TOF} distribution and does not produce pronounced two-dimensional features. These observations support the use of the factorized PDF as an adequate description for the prompt-fraction extraction, while residual effects are treated as part of the systematic uncertainty evaluation.

6.1.3 Model for the time-of-flight mass distribution

The m_{TOF} distribution is used to separate protons and antiprotons from lighter hadrons. The signal shape for true protons is obtained from simulated events and modeled using a non-parametric kernel estimation technique. Specifically, a RooKeysPdf is constructed

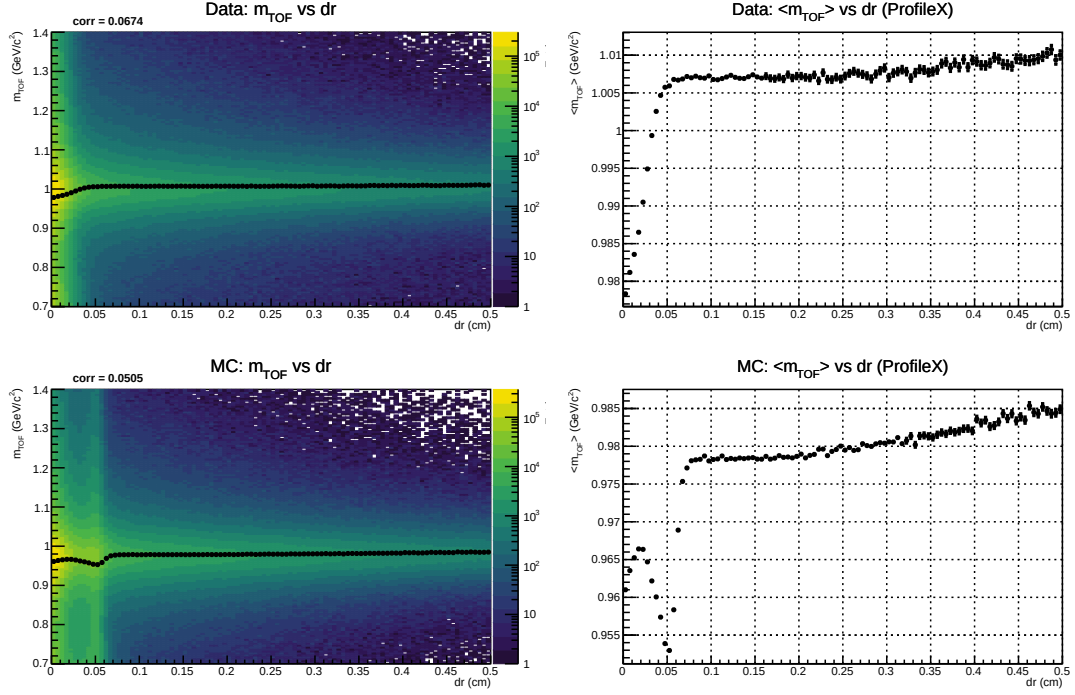


Figure 6.1: Validation of the factorization assumption in the two-dimensional PDF. Left panels: two-dimensional distributions of m_{TOF} versus dr for data (top) and simulation (bottom), shown with a logarithmic color scale. The overlaid points represent the x -profile, i.e. the mean value $\langle m_{\text{TOF}} \rangle$ in bins of dr . Right panels: the corresponding profiles displayed alone to highlight possible trends. The measured Pearson correlation coefficients are at the few-percent level (below 7%), and no pronounced two-dimensional structures are observed.

from Monte Carlo tracks with $|\text{PDG}| = 2212$, providing a smooth template that accurately reproduces the intrinsic shape of the proton m_{TOF} distribution.

To account for residual differences between data and simulation, the Monte Carlo template is convolved with a resolution function. For the proton candidates, the resolution is described by a bifurcated Gaussian, allowing for asymmetric detector resolution effects. The same resolution function is not sufficient to describe the distribution of antiproton candidates: for candidates with negative charge, then, the resolution function is defined as a sum of two bifurcated Gaussian. The parameters of these resolution functions are determined from data and subsequently fixed when building the full two-dimensional model.

The background contribution in m_{TOF} is modeled as the sum of two exponential components, corresponding predominantly to pions and kaons. The relative fraction of these two background components is left floating in the fit, while their exponential slopes are determined from data. The total m_{TOF} PDF is then given by an extended sum of the signal and background PDFs, with independent yields for each component. The m_{TOF} PDF is described in detail in Sec. 6.2.2. An example fit of this one-dimensional model is shown in Fig. 6.2.

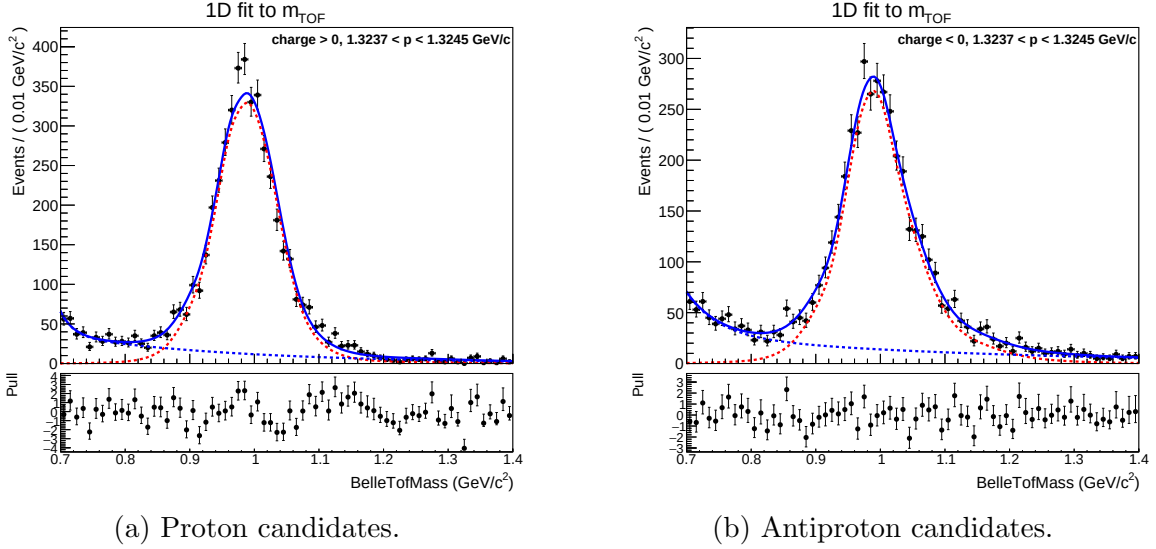


Figure 6.2: Example fits of the one-dimensional m_{TOF} model for proton (left) and antiproton (right) candidates. The dashed red curve corresponds to the signal component, while the blue dashed curve corresponds to the background. The solid blue curve shows the total PDF.

6.1.4 Model for the transverse distance distribution

The dr observable provides sensitivity to the production origin of the reconstructed track. Based on Monte Carlo truth information, the dr distribution is decomposed into three physically distinct components:

- prompt protons, originating from the primary interaction or short-lived resonances;
- protons from weak decays of strange baryons (hyperons);
- residual contributions from other sources.

The prompt component is characterized by a narrow peak around $dr = 0$, reflecting the detector resolution. It is modeled as the sum of a Gaussian core, with the mean fixed at zero, and two exponential tails that account for non-Gaussian resolution effects.

The hyperon component exhibits a broader dr distribution due to the displaced decay vertices of strange baryons. This contribution is modeled as the product of a sum of two exponentials and a second-order polynomial, allowing sufficient flexibility to reproduce the shape observed in simulation.

The remaining component is described by a single exponential function, corresponding to a nearly flat distribution within the fitted dr range. The details of the one-dimensional dr PDF are described in section 6.2.1. An example fit of this model is shown in figure 6.3.

All dr model parameters are determined from simulated samples without continuum background contributions and are fixed when fitting the data.

6.1.5 Initialization and fit stability

To ensure smooth behavior across adjacent momentum bins and improve fit stability, the fitted parameters from a given momentum interval are used as initial values for the

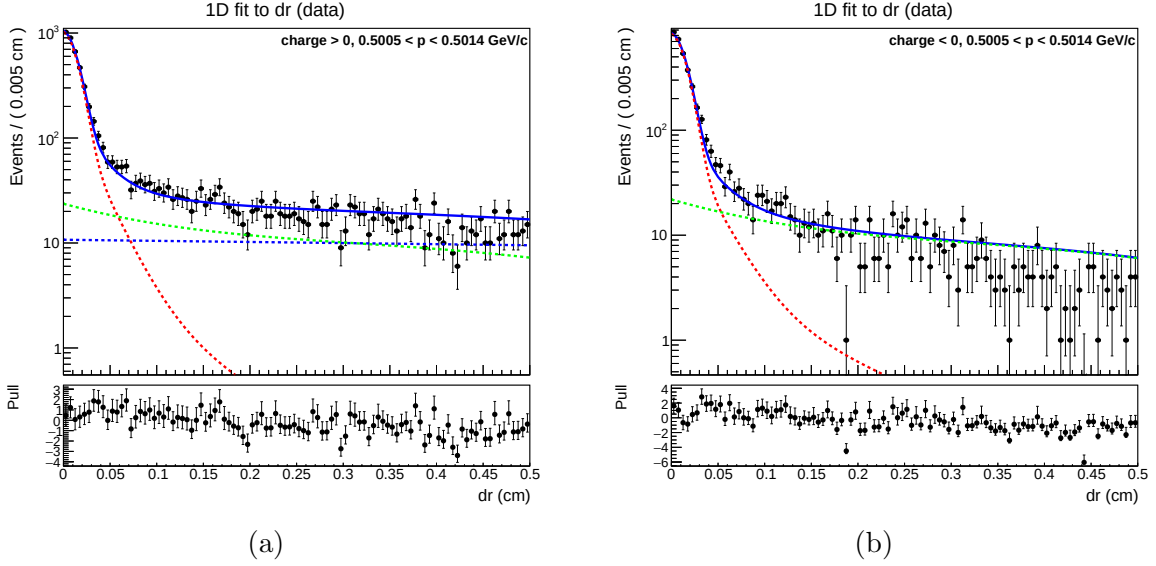


Figure 6.3: One-dimensional fits to the dr distribution in data for positive (a) and negative (b) tracks. The dashed curves indicate the individual analytic components used in the fit: the prompt contribution is shown in red, the hyperon component in green, and the remaining non-prompt contribution in blue.

subsequent bin. In addition, in momentum bins where the number of candidates approximately doubles, the initial values of the signal and background yields are scaled accordingly. This strategy significantly reduces the probability of fit failures and guarantees consistent convergence over the full momentum range.

The resulting two-dimensional model provides a robust and flexible description of the data, enabling the simultaneous determination of proton identity and promptness with minimal bias. The extraction and validation of the prompt proton fractions obtained from this model are discussed in the following section.

6.1.6 Likelihood definition and fit procedure

For each momentum bin and charge category, the model parameters are determined through an extended unbinned maximum-likelihood fit. Given a sample of N reconstructed candidates characterized by the observables (m_{TOF}, dr) , the likelihood function is defined as

$$\mathcal{L} = \exp\left(-\sum_k N_k\right) \prod_{i=1}^N \left[\sum_k N_k \mathcal{P}_k(m_{\text{TOF}}^i, dr^i) \right], \quad (6.2)$$

where the index k runs over the different components of the model (signal and backgrounds), N_k denotes the expected yield of component k , and $\mathcal{P}_k$ is the corresponding normalized two-dimensional probability density function.

As discussed in the previous section, each component PDF factorizes into the product of one-dimensional distributions,

$$\mathcal{P}_k(m_{\text{TOF}}, dr) = \mathcal{P}_{\text{TOF}}^{(k)}(m_{\text{TOF}}) \times \mathcal{P}_{dr}^{(k)}(dr), \quad (6.3)$$

which allows the independent modeling of particle identification and production origin.

6.2 Fit strategy

Before constructing the two-dimensional model, one-dimensional fits to the dr and m_{TOF} distributions are performed separately to obtain initial estimates of the model parameters.

6.2.1 One-dimensional fit to dr

The modeling of the dr distribution is described in this section. In order to define an analytic model for the different physical contributions, Monte Carlo truth information is used to classify reconstructed tracks according to their origin.

In the basf2 software, the variable `mcPrimary` identifies particles produced directly at the interaction point. Tracks with `mcPrimary==1` correspond to prompt particles, while tracks with `mcPrimary==0` originate from secondary decays or interactions with detector material. Moreover, the variable `genMotherPDG` stores the PDG code of the mother particle of the associated generated particle and allows the origin of non-prompt tracks to be classified.

Using this information, the fit to the dr distribution is first performed on three mutually exclusive components of the selected Monte Carlo sample:

- *Prompt tracks*: tracks satisfying `mcPrimary==1`, corresponding to particles produced directly at the interaction point.
- *Tracks from hyperon decays*: non-prompt tracks with a hyperon mother, identified by `mcPrimary==0` and $3000 \leq |\text{genMotherPDG}| \leq 4000$.
- *Tracks from other sources*: non-prompt tracks that do not originate from hyperon decays, including contributions from material interactions and other secondary processes.

The prompt component is modeled as the sum of a Gaussian distribution, with the mean fixed at $dr = 0$, and two exponential probability density functions that describe the non-Gaussian tail:

$$f_{\text{prompt}}(dr) = (1 - f_1 - f_2) \times \mathcal{N}(dr | \mu = 0, \sigma_{\text{prompt}}) + f_1 \times e^{-\lambda_1 \times dr} + f_2 \times e^{-\lambda_2 \times dr}. \quad (6.4)$$

The hyperon component is modeled as the sum of a Gaussian distribution, with its mean fixed at $dr = 0$, and two exponential probability density functions:

$$f_{\text{hyp.}}(dr) = (1 - f_3 - f_4) \times \mathcal{N}(dr | \mu = 0, \sigma_{\text{hyp.}}) + f_3 e^{-\lambda_3 dr} + f_4 e^{-\lambda_4 dr}. \quad (6.5)$$

The “other” component is modeled using a single exponential probability density function:

$$f_{\text{other}}(dr) = e^{-\lambda_{\text{other}} \times dr}. \quad (6.6)$$

An example of the fit results is shown in Fig. 6.4.

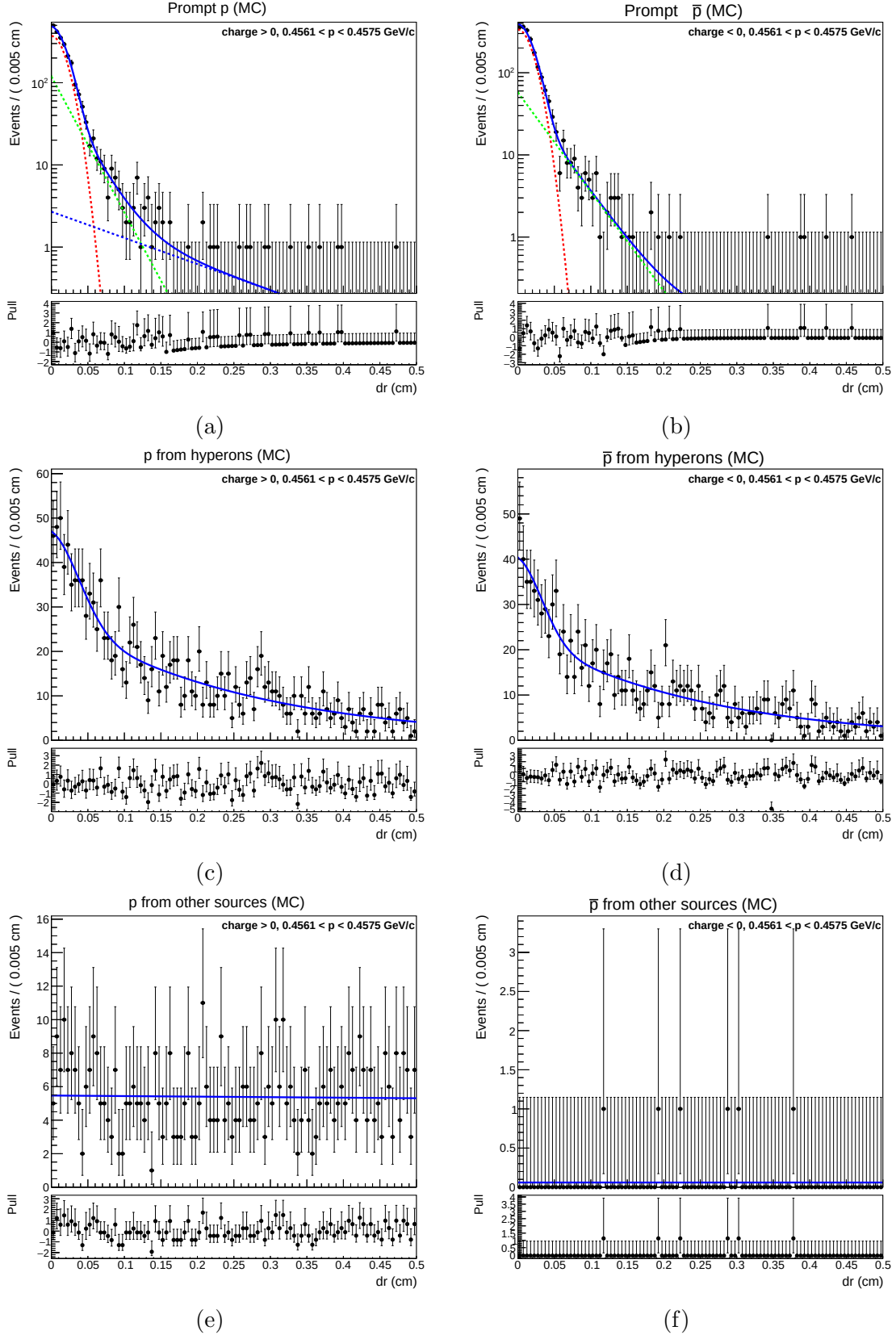


Figure 6.4: Analytic fits to the three MC components of the dr observable in the example momentum bin $456.1 < p < 457.5 \text{ MeV}/c$. Rows correspond to prompt (top), hyperon (middle), and other (bottom) components, while the left and right columns show results for positive and negative tracks, respectively. In panels (a) and (b), the dashed curves represent the three analytic components used to model the prompt contribution, as defined in Eq. (6.4): the Gaussian component is shown in red, while the two exponential components are shown in green and blue.

After convergence, all shape parameters are fixed to their optimized values and the three probability density functions are summed. A Poisson yield is associated with each component, and the resulting model is fitted to the experimental $\Upsilon(2S)$ data. The total PDF is thus

$$Y_{\text{prompt}} f_{\text{prompt}}(dr) + Y_{\text{hyp}} f_{\text{hyp}}(dr) + Y_{\text{other}} f_{\text{other}}(dr). \quad (6.7)$$

After the convergence, the fitted yields are fixed. They will then be used to initialize the 2D model as discussed below. An example fit for positive and negative charges is shown in Figure 6.3.

6.2.2 One-dimensional fit to m_{TOF}

To model the m_{TOF} distribution, a probability density function composed of three components is defined:

- (anti)protons: |mcPDG| == 2212,
- kaons: |mcPDG| == 321,
- pions: |mcPDG| == 211.

The (anti)proton component is described using a non-parametric shape derived from Monte Carlo simulation. In particular, a `RooKeysPdf` [89], denoted as $K_{\text{MC}}(m_{\text{TOF}})$, is constructed from true (anti)proton candidates in simulation and provides a kernel-smoothed estimate of the intrinsic signal line shape.

Detector resolution effects are accounted for through an explicit convolution with an analytical resolution function. For positively charged candidates (protons), the resolution is modeled by a bifurcated Gaussian function R_{bifG} . The corresponding signal probability density function is defined as

$$f_p(m_{\text{TOF}}) = (K_{\text{MC}} * R_{\text{bifG}})(m_{\text{TOF}}) = \int_{-\infty}^{+\infty} K_{\text{MC}}(m') R_{\text{bifG}}(m_{\text{TOF}} - m'; \mu, \sigma_L, \sigma_R) dm', \quad (6.8)$$

where the bifurcated Gaussian is defined as

$$R_{\text{bifG}}(x; \mu, \sigma_L, \sigma_R) = \begin{cases} \frac{1}{\sqrt{2\pi} \sigma_L} \exp\left[-\frac{(x - \mu)^2}{2\sigma_L^2}\right], & x < \mu, \\ \frac{1}{\sqrt{2\pi} \sigma_R} \exp\left[-\frac{(x - \mu)^2}{2\sigma_R^2}\right], & x \geq \mu. \end{cases} \quad (6.9)$$

For negatively charged candidates (antiprotons), the detector response exhibits a more complex structure that cannot be adequately described by a single bifurcated Gaussian. This behavior is primarily attributed to charge-dependent detector effects and to additional non-Gaussian contributions arising from interactions of antiprotons with the detector material, which lead to asymmetric tails and a broader core in the reconstructed m_{TOF} distribution. In this case, the resolution function is modeled as a weighted sum of two bifurcated Gaussian functions,

$$R_{\bar{p}}(x) = f R_{\text{bifG}}(x; \mu_1, \sigma_{L,1}, \sigma_{R,1}) + (1 - f) R_{\text{bifG}}(x; \mu_2, \sigma_{L,2}, \sigma_{R,2}), \quad (6.10)$$

where f is the relative fraction of the two components and is treated as a free parameter of the 1D fit. The antiproton signal PDF is then obtained through the convolution

$$f_{\bar{p}}(m_{\text{TOF}}) = (K_{\text{MC}} * R_{\bar{p}})(m_{\text{TOF}}). \quad (6.11)$$

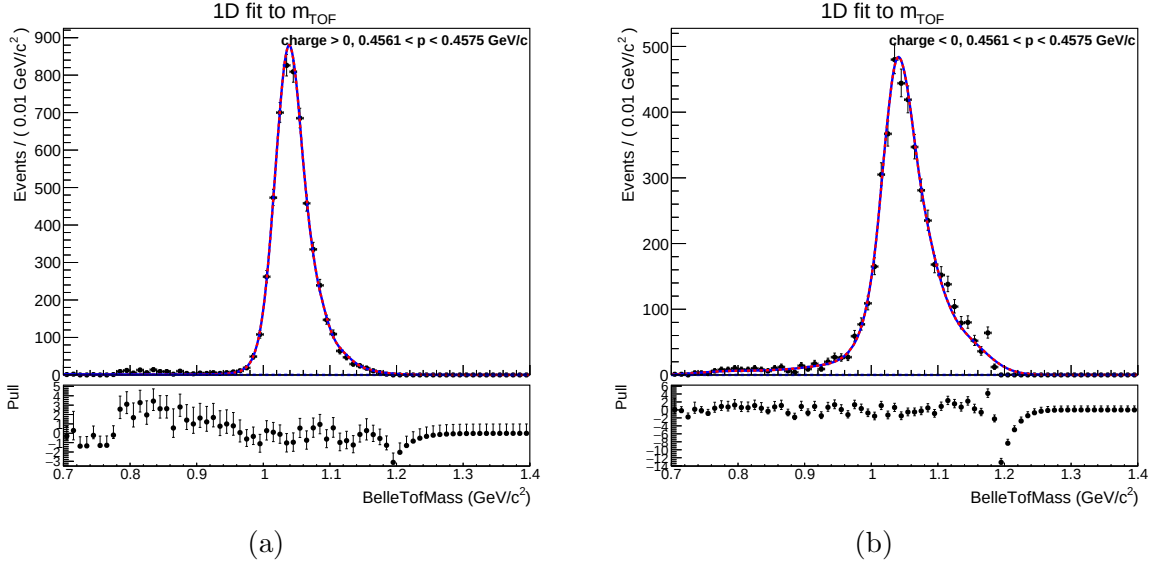


Figure 6.5: Extended UML fits to the m_{TOF} observable on experimental $\Upsilon(2S)$ data for positive (a) and negative (b) tracks. The red histogram is the signal MC template.

This charge-dependent modeling allows the fit to accurately capture differences in the detector resolution between protons and antiprotons while preserving a common intrinsic line shape derived from simulation.

The background components are represented by exponential PDFs:

$$f_{\text{bg}}(m_{\text{TOF}}) = f_{\pi} e^{-\lambda_{\pi} m_{\text{TOF}}} + (1 - f_{\pi}) e^{-\lambda_K m_{\text{TOF}}}, \quad (6.12)$$

where f_{π} is the fraction of the pion component and λ_{π} (λ_K) is the decay constant of the pion (kaon) component. An extended probability density function with Poisson-distributed signal and background yields is constructed:

$$f(m_{\text{TOF}}) = Y_p f_p(m_{\text{TOF}}) + Y_{\pi+K} f_{\text{bg}}(m_{\text{TOF}}), \quad (6.13)$$

and unbinned maximum likelihood fits are performed on the experimental data distributions. After convergence, all shape parameters, including f_{π} , are fixed. An example of such a fit in the momentum bin $456.1 < p < 457.5 \text{ MeV}/c$ is shown in Fig. 6.5.

6.2.3 Definition of the 2D model

The two-dimensional model is constructed using the probability density functions defined for the one-dimensional cases. The relative fractions of the three components of the dr distribution are retained by fixing the corresponding yields Y_{prompt} , $Y_{\text{hyp.}}$, and Y_{other} .

The two-dimensional PDF is composed of four components:

- Prompt (anti)proton:

$$f_{\text{prompt}}(dr) \times f_p(m_{\text{TOF}}),$$

- Non-prompt (anti)proton:

$$\left[\frac{Y_{\text{hyp.}}}{Y_{\text{hyp.}} + Y_{\text{other}}} f_{\text{hyp.}}(dr) + \left(1 - \frac{Y_{\text{hyp.}}}{Y_{\text{hyp.}} + Y_{\text{other}}} \right) f_{\text{other}}(dr) \right] \times f_p(m_{\text{TOF}}),$$

- Prompt, non-(anti)proton:

$$f_{\text{prompt}}(dr) \times f_{\text{bg}}(m_{\text{TOF}}),$$

- Non-prompt, non-(anti)proton:

$$\left[\frac{Y_{\text{hyp.}}}{Y_{\text{hyp.}} + Y_{\text{other}}} f_{\text{hyp.}}(dr) + \left(1 - \frac{Y_{\text{hyp.}}}{Y_{\text{hyp.}} + Y_{\text{other}}} \right) f_{\text{other}}(dr) \right] \times f_{\text{bg}}(m_{\text{TOF}}).$$

The initialization of the component yields in the two-dimensional fit is performed using the results of the one-dimensional fits described previously. In the first momentum bin, the initial yields are defined as

- $Y_{p,\text{init.}}^{\text{prompt}} = Y_p \times Y_{\text{prompt}}/N_{\text{data}},$
- $Y_{p,\text{init.}}^{\text{non-prompt}} = Y_p \times (Y_{\text{hyp.}} + Y_{\text{other}})/N_{\text{data}},$
- $Y_{\pi+K,\text{init.}}^{\text{prompt}} = Y_{\pi+K} \times Y_{\text{prompt}}/N_{\text{data}},$
- $Y_{\pi+K,\text{init.}}^{\text{non-prompt}} = Y_{\pi+K} \times (Y_{\text{hyp.}} + Y_{\text{other}})/N_{\text{data}},$

where Y_p and $Y_{\pi+K}$ are the yields obtained from the one-dimensional m_{TOF} fit, and N_{data} denotes the total number of entries in the dataset.

For subsequent momentum bins, the fit is initialized using the parameter values obtained from the previous bin. This procedure ensures a smooth bin-to-bin evolution of the fitted parameters and improves the numerical stability of the two-dimensional fits.

An example of the two-dimensional dataset together with the corresponding fitted model is shown in Fig. 6.6, for both positive and negative charges.

Figures 6.7 and 6.8 show the shapes of the four model components for positive and negative charges, respectively.

6.2.4 Definition of the signal region

The signal fraction is enhanced by evaluating the ratio of the prompt-(anti)proton yield to the total yield within a restricted two-dimensional region that contains the majority of the signal. The boundaries of this signal region are determined according to the following procedure:

- Project the *signal* component of the two-dimensional PDF onto the dr axis.
- Determine the 99% quantile of the projected distribution, which defines the upper boundary of the signal region in the dr dimension. The lower boundary is fixed to $dr = 0$.
- Project the signal component onto the m_{TOF} axis.
- Measure the mode m of the distribution and its full width at half maximum (FWHM), and define the signal region in m_{TOF} as $(m - 2 \times \text{FWHM}, m + 2.5 \times \text{FWHM})$.

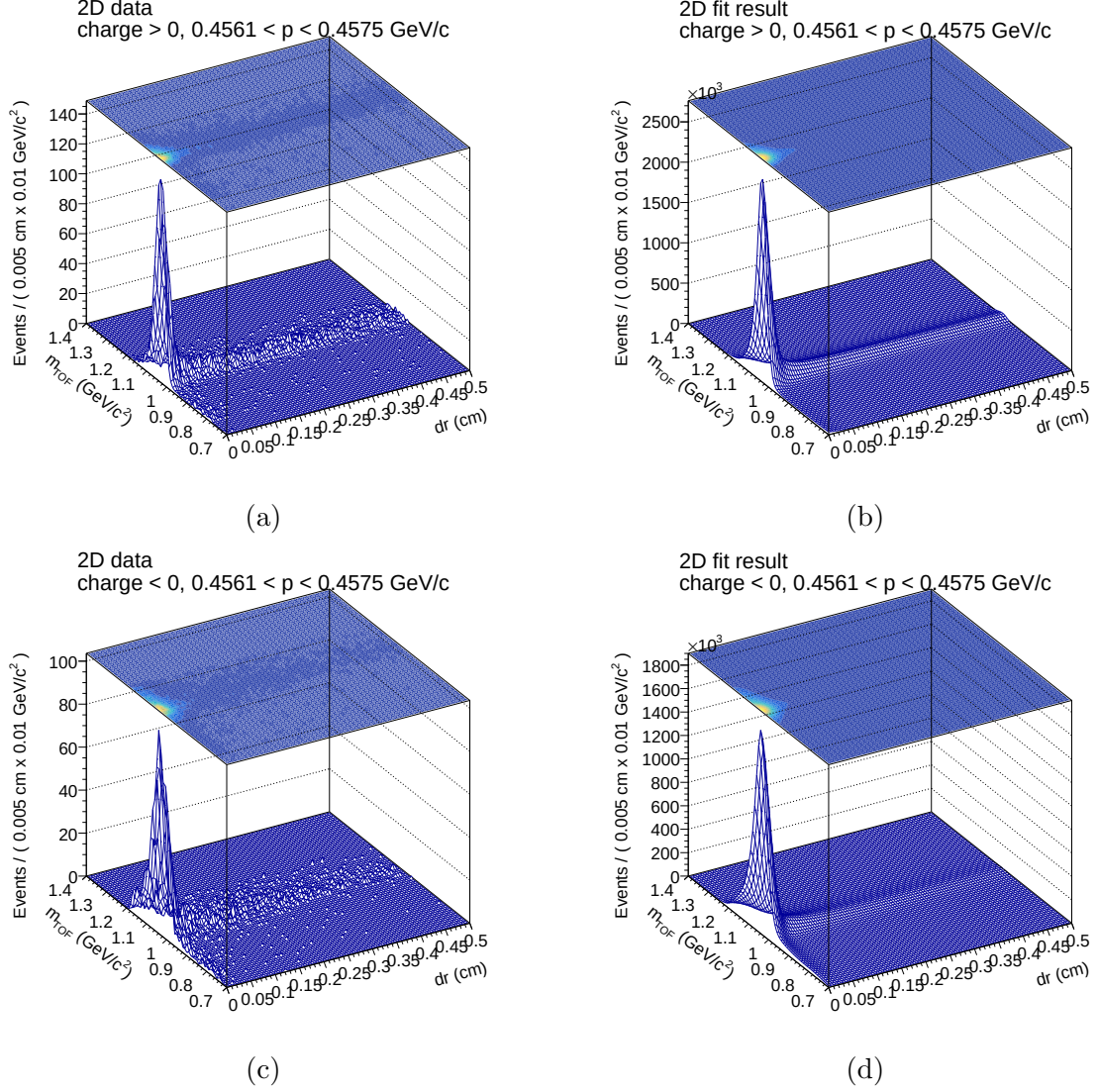


Figure 6.6: Example comparison between the 2D datasets and the fitted model for positive (a, b) and negative (c, d) tracks.

For each component i of the two-dimensional model, the effective yield within the signal region $\mathcal{R}$ is defined as

$$Y_{\text{eff},i} = Y_i \int_{\mathcal{R}} f_i(dr, m_{\text{TOF}}) ddr dm_{\text{TOF}}, \quad (6.14)$$

where Y_i is the fitted yield of component i and $f_i(dr, m_{\text{TOF}})$ is the corresponding normalized two-dimensional probability density function.

The effective prompt-(anti)proton fraction is then defined as

$$f_p^{\text{prompt}} = \frac{Y_{\text{eff},p}^{\text{prompt}}}{Y_{\text{eff},p}^{\text{prompt}} + Y_{\text{eff},\text{non-}p}^{\text{prompt}} + Y_{\text{eff},p}^{\text{non-prompt}} + Y_{\text{eff},\text{non-}p}^{\text{non-prompt}}}. \quad (6.15)$$

Figure 6.9 shows an example of the projections of the two-dimensional model onto the dr and m_{TOF} axes, with vertical lines indicating the boundaries of the signal region. As a consequence, an additional momentum-dependent selection is applied to the dr and m_{TOF} observables. The selection bands are shown in Fig. 6.10.

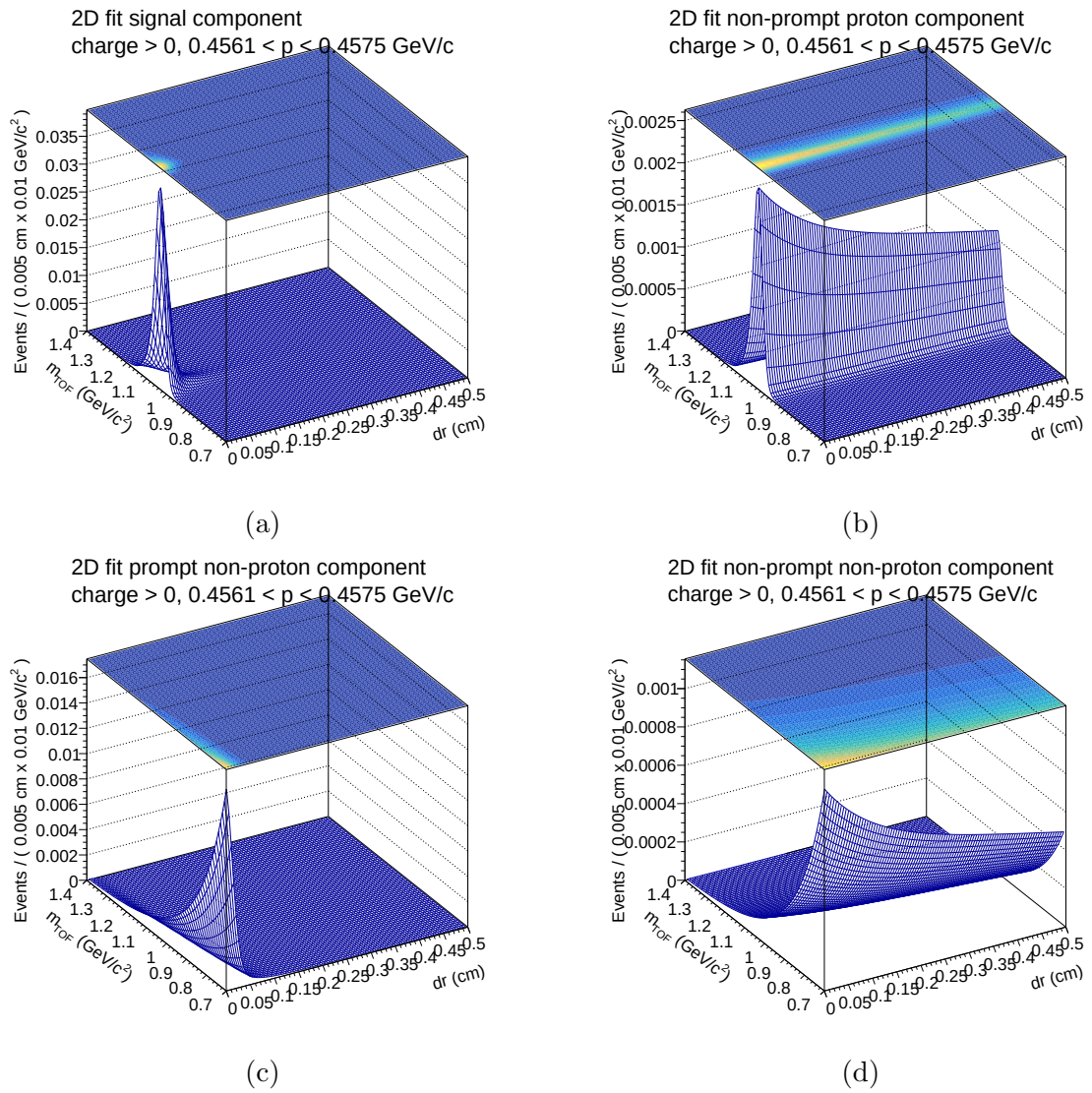


Figure 6.7: Components of the 2D model for positively charged tracks: prompt protons (a), non-prompt protons (b), prompt non-protons (c) and non-prompt non-protons (d).

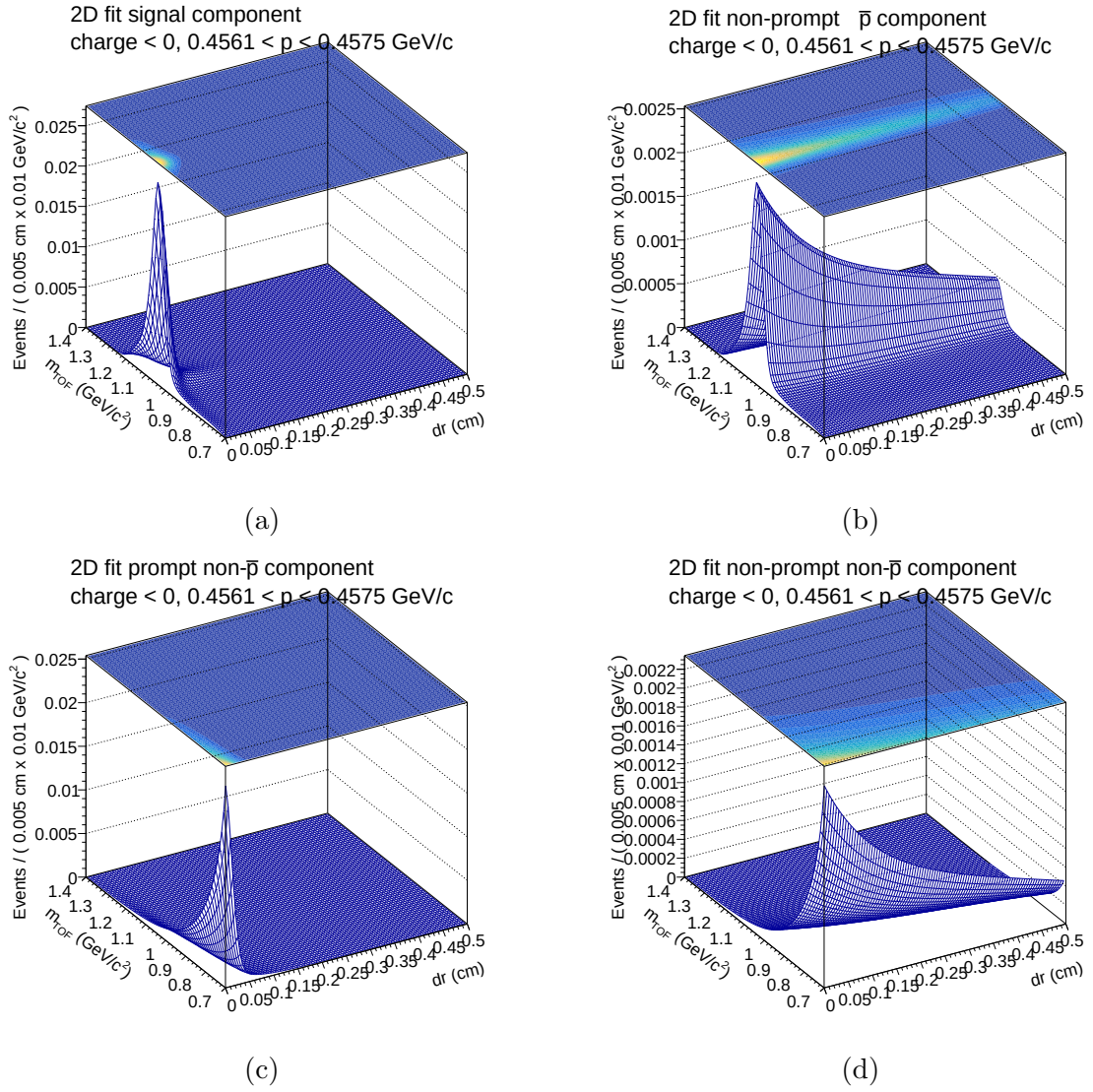


Figure 6.8: Components of the 2D model for negatively charged tracks: prompt protons (a), non-prompt protons (b), prompt non-protons (c) and non-prompt non-protons (d).

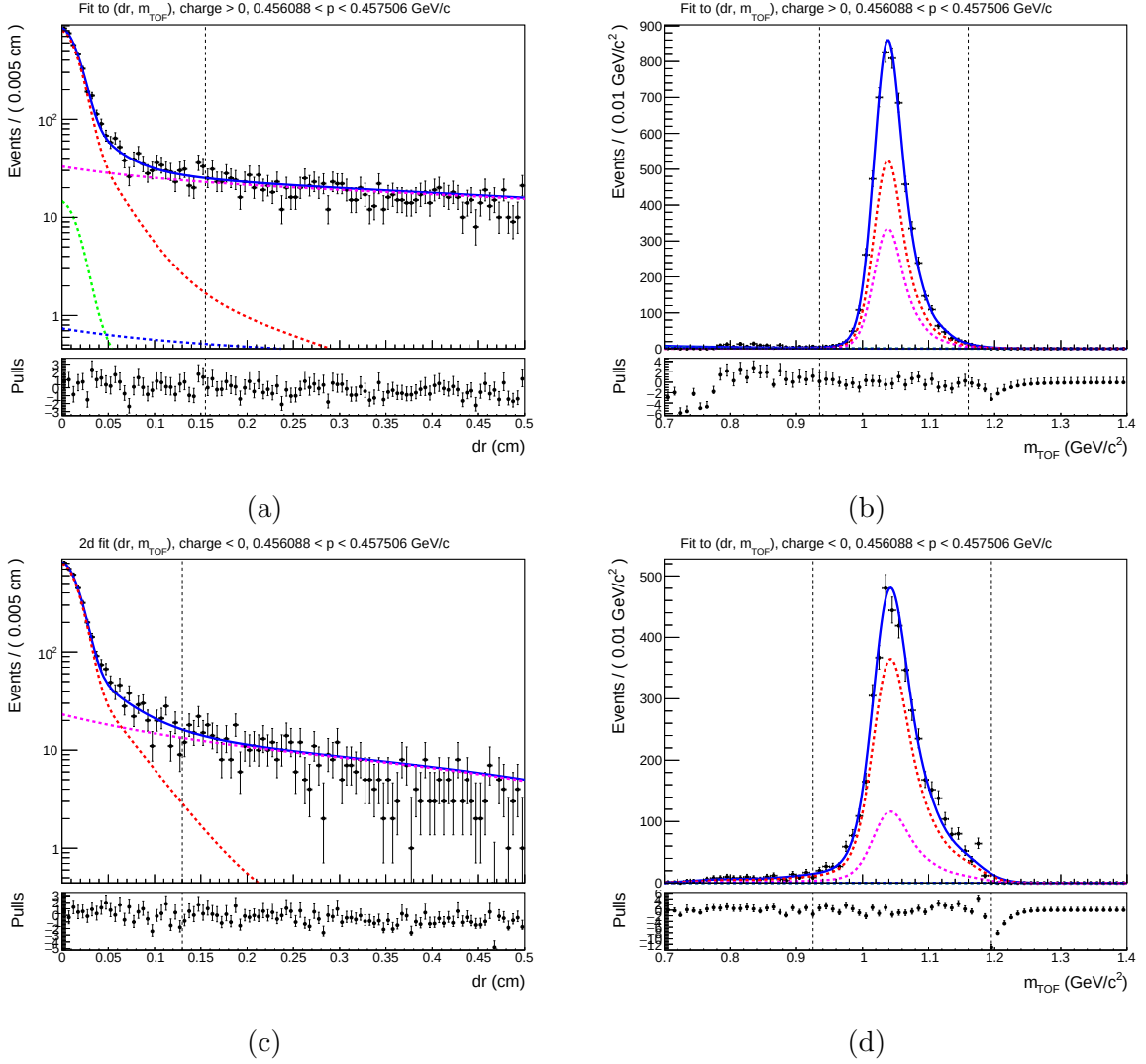


Figure 6.9: Example projections of the fitted 2D model for positive (a, b) and negative (c, d) tracks. The red dashed line represents the signal (prompt proton) component, while the magenta dashed line represents the non-prompt proton component. The green dashed line corresponds to prompt non-protons, and the blue dashed line to non-prompt non-protons. The black dashed lines enclose the signal region, where the prompt (anti)proton fraction is measured.

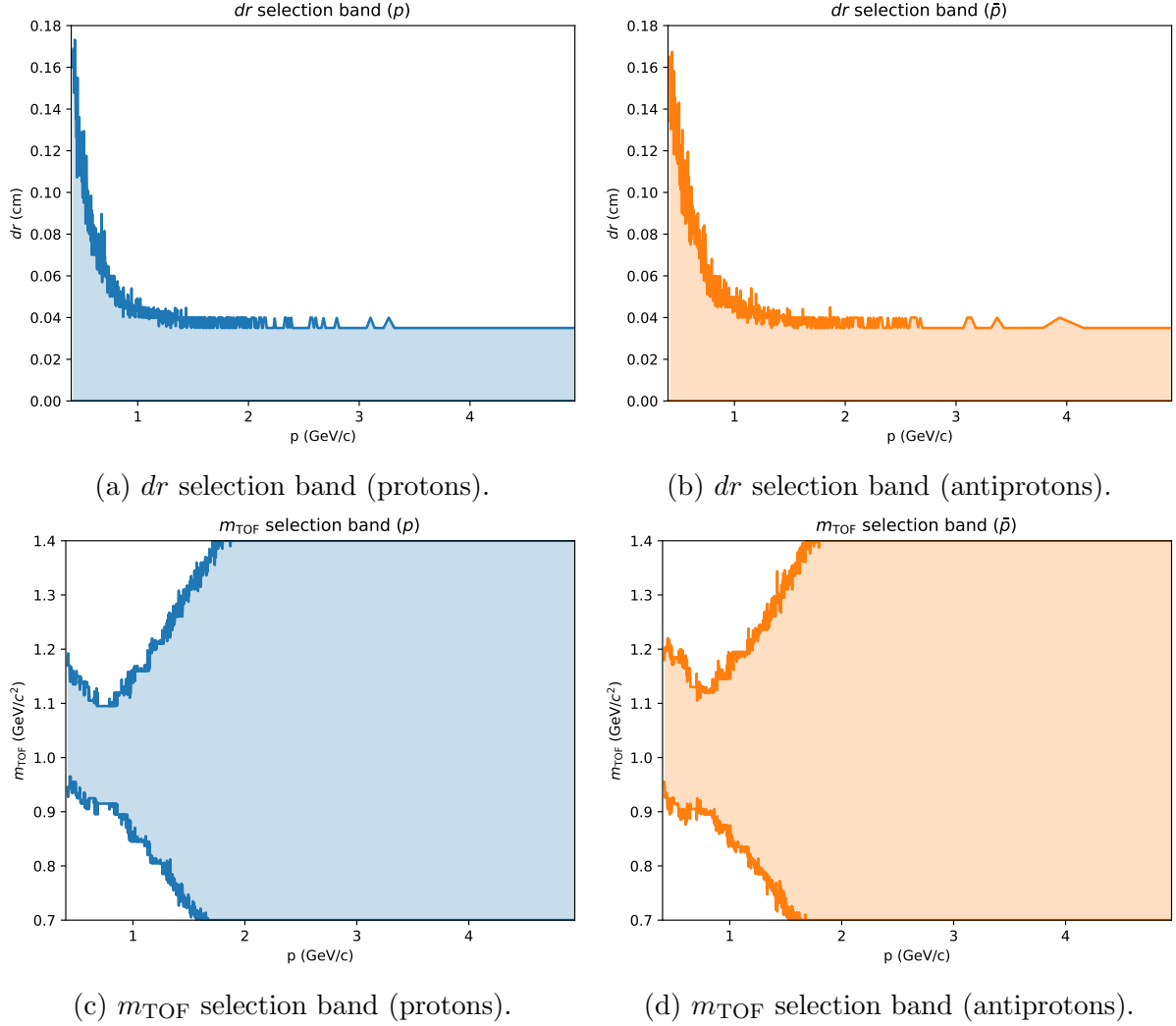


Figure 6.10: Momentum-dependent selections derived from the signal regions defined by the two-dimensional fits, shown for experimental $\Upsilon(2S)$ data. The top row shows the dr selection bands for protons (left) and antiprotons (right), while the bottom row shows the corresponding m_{TOF} selection bands.

6.2.5 Fit quality

The goodness of the fits is evaluated using the reduced chi-square statistic, $\chi_r^2 = \chi^2/\text{ndf}$. Fits are required to have a good convergence status and a positive-definite covariance matrix. These requirements retain 99.9% of the candidate antiproton fits and 100% of the candidate proton fits. For the retained fits, the scatter plot of χ_r^2 as a function of momentum is shown in Fig. 6.11. Most of the poor-quality fits ($\chi^2/\text{ndf} > 2.5$) are

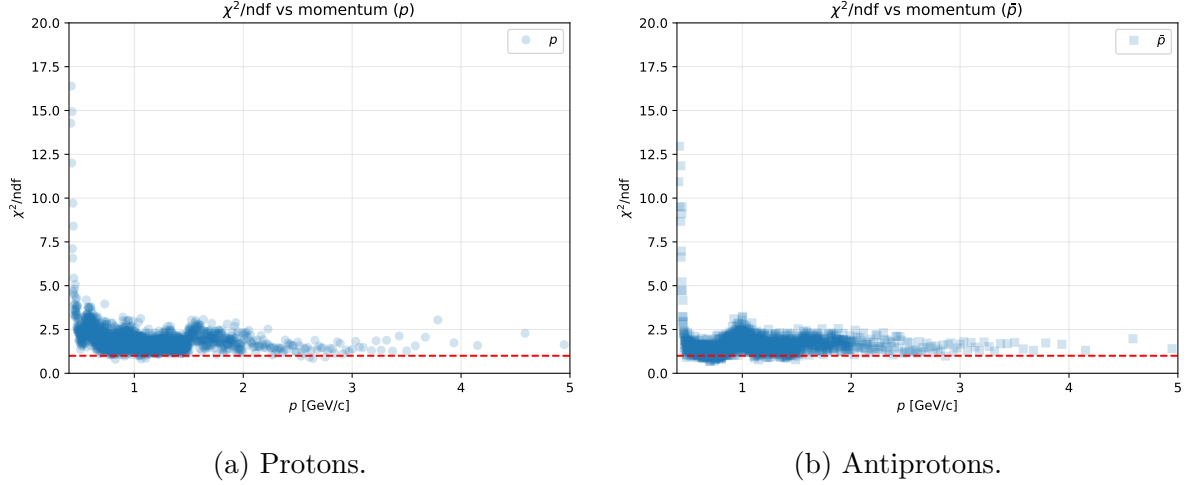


Figure 6.11: Reduced chi-square distributions of the two-dimensional fits as a function of momentum. The red dashed line corresponds to $\chi^2/\text{ndf} = 1$.

concentrated in the region $p < 450 \text{ MeV}/c$, where a sizable discrepancy between the fit function and the data is observed in the upper tail of the m_{TOF} projection. An example of this behavior is shown in Fig. 6.9d. This effect originates from the β_{TOF} selection described in Section 5.3.3. While this selection is required to improve the overall data quality, it truncates the upper tail of the m_{TOF} distribution at low momenta.

Nonetheless, since the agreement between data and model remains good over the selected region of the m_{TOF} distribution, these fits are kept. An additional requirement is imposed on the relative uncertainty of the prompt-proton fraction, requiring it to be smaller than 25% in order to ensure a reliable determination of the signal fraction. This criterion excludes only six fits for positive charges and eight for negative charges.

6.3 Prompt proton fraction

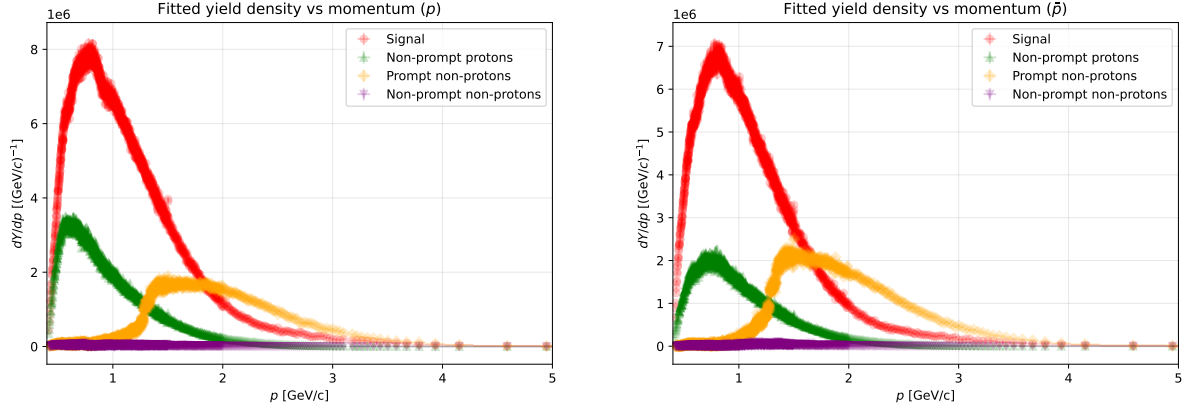
For each accepted momentum bin, the prompt (anti)proton fraction is extracted from the fitted yields and represents the probability for a reconstructed proton candidate to originate from the primary interaction vertex. The resulting fractions, obtained at discrete momentum points, exhibit smooth and slowly varying trends as a function of momentum, reflecting the combined effects of detector resolution, particle identification performance, and secondary production mechanisms.

6.3.1 Momentum dependence of fitted yields and fractions

The two-dimensional fits described in the previous sections provide discrete measurements of the fitted yields and particle fractions in bins of single-particle momentum. In

this section, the momentum dependence of these quantities is examined to assess the stability of the fit results and to highlight relevant charge-dependent features.

Figure 6.12 shows the yield density as a function of momentum for protons (left) and antiprotons (right). A smooth behavior is observed for both charges over the full momentum range, indicating that the chosen binning does not introduce artificial structures and that the fit results are stable.



(a) Proton yield density as a function of momentum.

(b) Antiproton yield density as a function of momentum.

Figure 6.12: Yield density obtained from the two-dimensional fits as a function of the single-particle momentum for protons (left) and antiprotons (right).

The prompt-track fractions extracted from the fits are shown in Fig. 6.13. No smoothing is applied at this stage, and the points correspond directly to the fit results in each momentum bin. At low momentum ($p < 1$ GeV), the prompt fraction is observed to be larger for negatively charged tracks than for positively charged ones. This behavior is consistent with the combined effect of charge-dependent interactions with the detector material and reconstruction effects at low momentum. In particular, secondary protons produced in hadronic interactions are more likely to populate the non-prompt component, while low-momentum antiprotons are suppressed by annihilation in the detector material, leading to a comparatively larger prompt fraction for negative tracks. With increasing momentum, these effects become less relevant and the prompt fractions for the two charges converge.

The proton and antiproton fractions as a function of momentum are shown in Fig. 6.14. The two distributions are nearly identical over most of the momentum range. For $p \gtrsim 1.5$ GeV, the proton fraction is observed to be slightly larger than the antiproton one. At low momentum, the proton fraction approaches unity, reflecting the high effectiveness of the momentum-dependent dE/dx selection in this region. Above $p \simeq 1.2$ GeV, the performance of this selection degrades and the fractions decrease accordingly.

6.3.2 Kernel smoothing of the prompt-proton fraction

For subsequent stages of the analysis, a continuous estimate of the *prompt-proton* fraction as a function of momentum is required in order to weight individual particles when constructing the correlation function. For this purpose only, the discrete prompt-proton fraction is smoothed using a kernel-based regression technique.

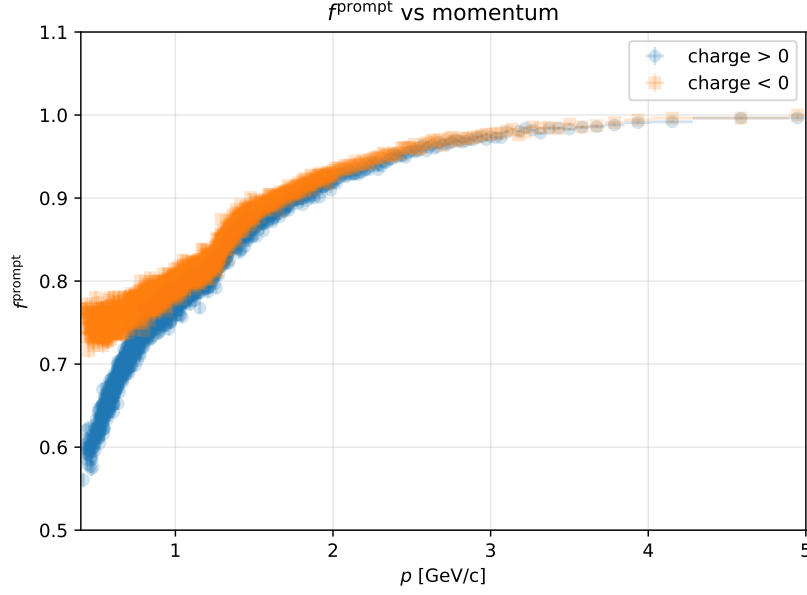


Figure 6.13: Prompt fraction as a function of momentum for protons and antiprotons. The points correspond to the discrete results of the two-dimensional fits.

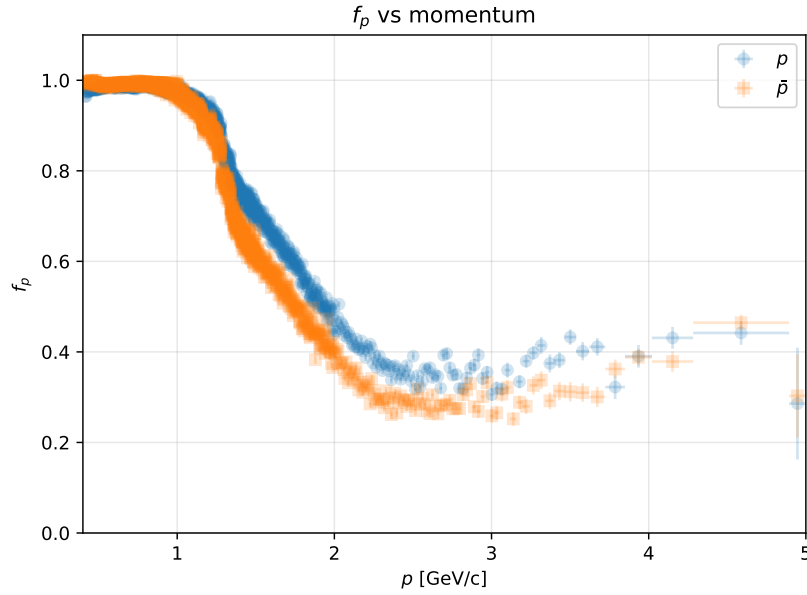


Figure 6.14: Fraction of protons and antiprotons as a function of momentum. The points correspond to the discrete fit results.

The smoothing procedure is applied to suppress local statistical fluctuations while preserving the physically expected smooth momentum dependence. The kernel width is chosen to balance statistical precision against bias from over-smoothing.

The kernel-smoothed prompt-proton fraction, together with its associated $\pm 1\sigma$ uncertainty band derived from the fit uncertainties, is shown in Fig. 6.15. This smoothed distribution is used in the subsequent analysis as a momentum-dependent weight for proton candidates in data.

The uncertainty band associated with the kernel-smoothed prompt-proton fraction is

evaluated locally at each momentum value p_0 by propagating the statistical uncertainties of the input measurements. For a given evaluation point p_0 , each measurement (p_i, f_i, σ_i) contributes with a weight

$$w_i(p_0) = K\left(\frac{p_i - p_0}{h}\right) \frac{1}{\sigma_i^2}, \quad (6.16)$$

where $K(u) = \exp(-u^2/2)$ is a Gaussian kernel and h denotes the kernel bandwidth. The smoothed value is obtained using a Nadaraya–Watson estimator,

$$\hat{f}(p_0) = \frac{\sum_i w_i(p_0) f_i}{\sum_i w_i(p_0)}. \quad (6.17)$$

The associated statistical uncertainty is approximated by treating the estimator as a locally weighted mean of independent measurements,

$$\sigma_{\hat{f}}(p_0) = \sqrt{\frac{1}{\sum_i w_i(p_0)}}. \quad (6.18)$$

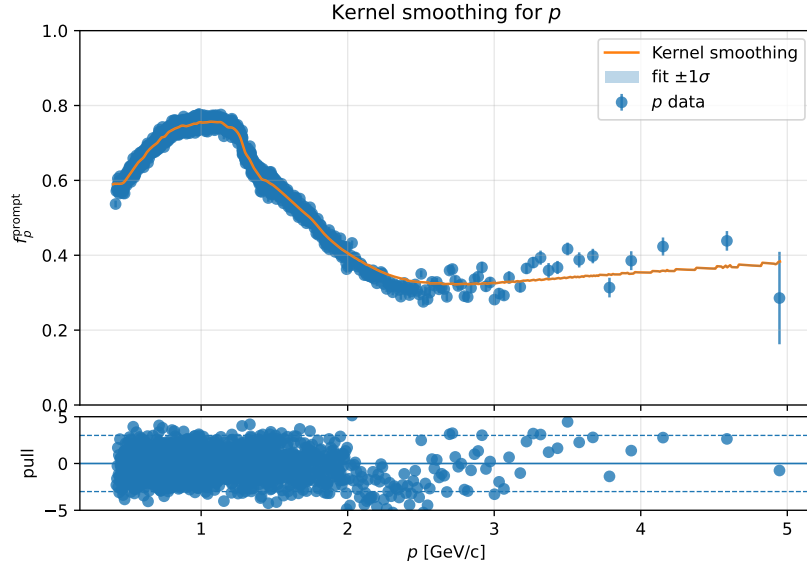


Figure 6.15: Kernel-smoothed prompt-proton fraction as a function of momentum. The shaded band indicates the $\pm 1\sigma$ uncertainty obtained from the smoothing procedure.

6.3.3 Use in correlation-function construction

The kernel-smoothed prompt-proton fraction obtained in Sec. 6.3.2 is used as a weight in the construction of the two-particle correlation function $C(k^*)$ measured in data. For each proton-proton (or antiproton-antiproton) pair, the weight is defined as the product of the single-particle prompt fractions evaluated at the momenta of the two particles. This procedure enhances the contribution from primary-primary pairs while suppressing residual contamination from pairs involving secondary protons.

The uncertainty associated with the prompt-fraction determination is propagated to the correlation function by repeating the weighting procedure using the upper and lower edges of the $\pm 1\sigma$ uncertainty band of the smoothed prompt-proton fraction. The

resulting variation of $C(k^*)$ is assigned as a systematic uncertainty related to the prompt-fraction weighting. This approach ensures a consistent treatment of the statistical limitations of the two-dimensional fits and their impact on the final femtoscopic observables.

Chapter 7

Building the correlation function

This chapter presents the construction and analysis of the proton–proton correlation function as implemented in this thesis. The methodology relies on the event mixing technique, which enables the separation of genuine two-particle correlations from trivial combinatorial effects while preserving the single-particle kinematics of the events. The procedure for constructing same-event and mixed-event pair distributions, the definition of the correlation function, and the associated technical choices are described in detail.

To validate the method, it is first applied to Monte Carlo samples at both generator level and after full detector reconstruction. The comparison between these stages highlights the impact of detector acceptance, momentum resolution, and offline selection on the observable correlation structure. A dedicated discussion is devoted to the long-range correlations arising from energy–momentum conservation, known as EMCIC effects. The parametrization used to model these contributions is introduced, its performance on simulated samples is assessed, and the implementation of the corresponding corrections is described.

7.1 Same-event and mixed-event pair construction

The raw (unnormalized) correlation function is defined as the ratio between the distribution of particle pairs originating from the same event and an uncorrelated reference built from pairs taken from different events. The construction of these two samples—referred to as same-event (SE) and mixed-event (ME) pairs—is central to the analysis and determines the sensitivity to femtoscopic effects as well as the control of non-femtoscopic backgrounds.

7.1.1 Same-event pairs

Same-event pairs are formed directly during the `basf2` reconstruction workflow. For each event, all proton candidates satisfying the selection criteria are combined in pairs, and the single-particle momentum in the pair’s rest frame (k^*) is computed from their four-momenta. Since both particles originate from the same collision, the SE distribution measured in data contains all genuine physical correlations: quantum statistics (QS), final-state interactions (FSI), and long-range effects arising from global event properties such as energy–momentum conservation (EMCIC). In the simulated sample instead, neither QS nor FSI are present. This distribution constitutes the numerator of the correlation function. The pair reconstruction is performed within the `basf2` software as if they

are produced in the decay of a virtual photon. This ensures that the SE sample remains consistent with all detector-level constraints—including tracking efficiency, acceptance, and vertex selection—and reflects exactly the environment in which the protons were reconstructed.

The outcome of the same-event pair construction is illustrated in figure 7.1, which shows the SE k^* distributions for pp and $\bar{p}\bar{p}$ pairs in reconstructed signal MC ($\Upsilon(2S)$ decays) after all selection criteria. Figure 7.2 shows the same-event relative momentum

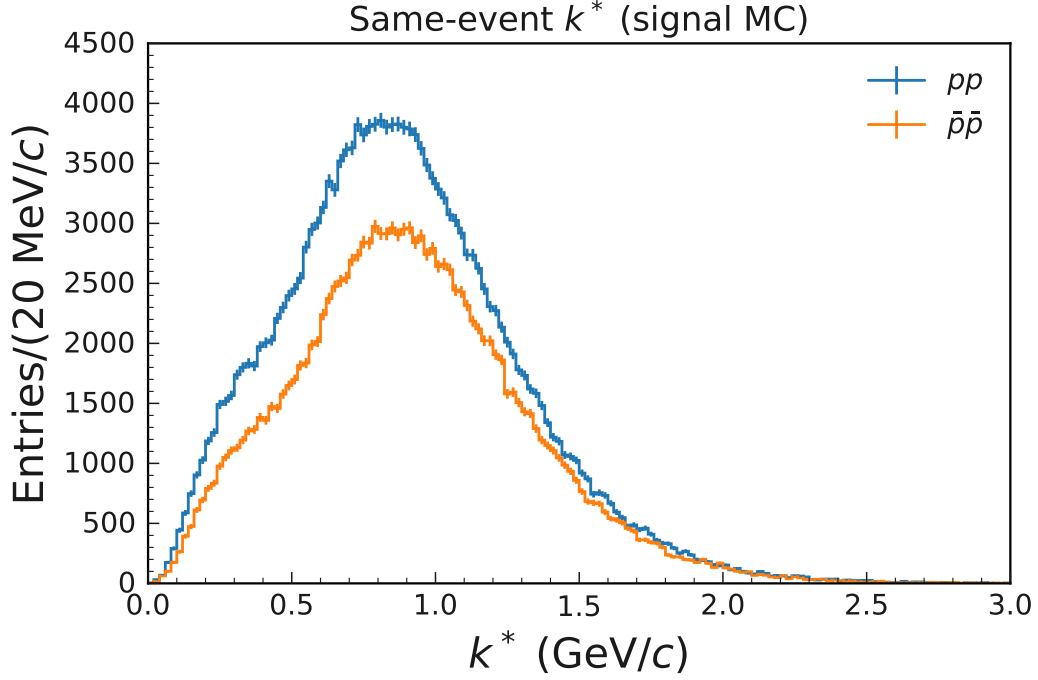


Figure 7.1: k^* distribution in $\Upsilon(2S)$ signal MC for pp (blue) and $\bar{p}\bar{p}$ (orange) pairs.

distribution k^* for pp pairs, comparing $\Upsilon(2S)$ data to signal MC. The real data is weighted using the result of the fit described in chapter 6, by associating to each track with a given reconstructed momentum p the corresponding prompt proton fraction $f_p^{\text{prompt}}(p)$. The distributions are normalized to unit integral.

7.1.2 Mixed-event pairs

To remove trivial phase-space and detector effects, an uncorrelated reference is built from events with similar global characteristics. Mixed-event pairs are constructed from single-particles ROOT ntuples by combining proton candidates taken from different events that fall within the same bin of event-shape and multiplicity. Specifically, the sample is divided into two-dimensional bins in the Fox–Wolfram moment ratio R_2 and the number of detected charged tracks N_{tracks} . The quantity R_2 is defined as

$$R_2 = \frac{H_2}{H_0}, \quad H_\ell = \sum_{i,j} |\vec{p}_i| |\vec{p}_j| P_\ell(\cos \theta_{ij}), \quad (7.1)$$

where H_ℓ is the Fox–Wolfram moment of order ℓ , the sum runs over all pairs of reconstructed particles in the event, and P_ℓ denotes the Legendre polynomial of order ℓ .

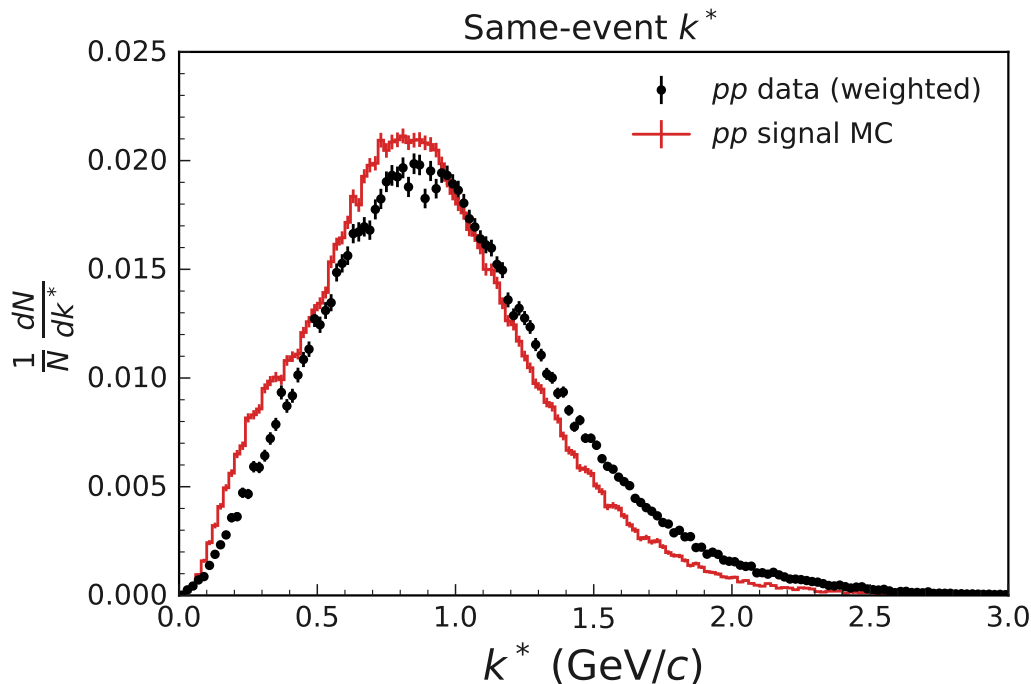


Figure 7.2: Same-event k^* distribution for pp pairs in data (black points, weighted) and $\Upsilon(2S)$ signal MC (red histogram), normalized to unit area.

The ratio R_2 provides a compact characterization of the event topology, with low values corresponding to spherical events and large values indicating jet-like configurations. Together with N_{tracks} , which encodes the event multiplicity, this choice ensures that the single-particle phase space of the mixed-event sample closely matches that of same events.

Within each (R_2, N_{tracks}) bin, every selected proton is paired with protons from 20 different events in the same bin. This fixed mixing depth results in smooth ME distributions while keeping the computational load manageable. The resulting ME pair distribution reflects the detector acceptance and the single-particle kinematics of the data, but—by construction—does not contain correlations originating from the same collision.

Figure 7.3 shows the distribution of the Fox–Wolfram ratio R_2 in $\Upsilon(2S)$ data (black points with statistical uncertainties), compared to the total generic $\Upsilon(2S)$ Monte Carlo sample (blue histogram) and the $e^+e^- \rightarrow q\bar{q}$ component (light-blue filled histogram). The data distribution exhibits a pronounced maximum at $R_2 \simeq 0.15$, followed by a smooth decrease towards larger values of R_2 . This behaviour is characteristic of hadronic $\Upsilon(2S)$ decays, which typically produce more spherical final states due to the multi-body decay topology. The peak position reflects the dominance of low-momentum, nearly isotropic events, while the tail extending to higher R_2 values corresponds to increasingly collimated, jet-like configurations.

The generic $\Upsilon(2S)$ Monte Carlo reproduces the overall shape of the distribution, but shows a moderate shift towards larger R_2 values, resulting in a slightly harder spectrum compared to data. This difference indicates that the simulated events are, on average, more jet-like than observed in data. The discrepancy is most visible in the region around the peak, where the MC underestimates the event yield at low R_2 and compensates at intermediate values.

The $q\bar{q}$ continuum component contributes predominantly at intermediate and large

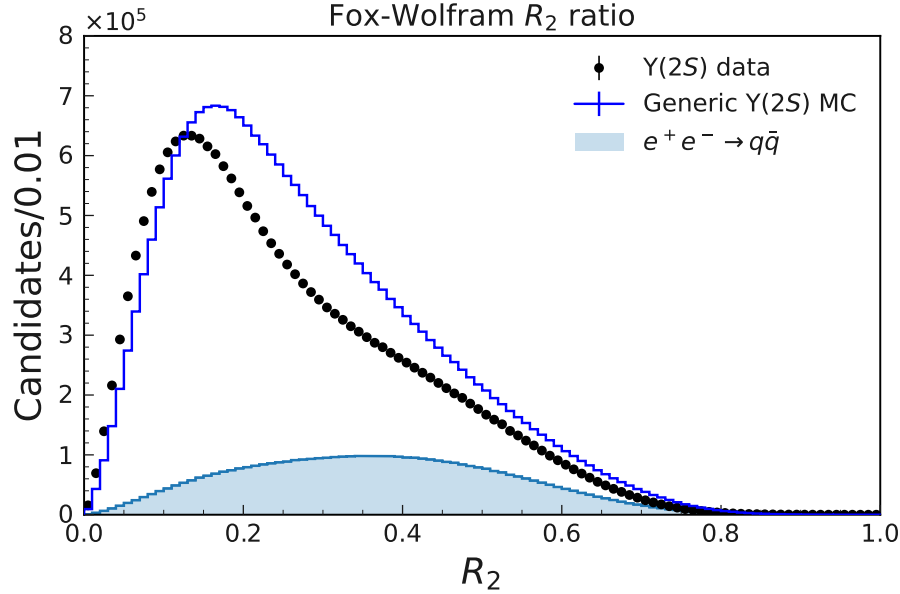


Figure 7.3: R_2 distribution in $\Upsilon(2S)$ data (black points) compared to the total $\Upsilon(2S)$ MC (blue histogram) and the $q\bar{q}$ component (light-blue filled histogram).

R_2 , as expected for two-jet-dominated topologies originating from light-quark pair production. Its relative enhancement in the high- R_2 region confirms that large values of R_2 are mainly associated with continuum-like, jetty events rather than resonant $\Upsilon(2S)$ decays. Overall, the comparison illustrates how R_2 effectively discriminates between spherical resonance decays and more collimated continuum processes, motivating its use in the definition of event-shape classes for the subsequent event-mixing procedure.

Figure 7.4 shows the charged-track multiplicity distribution N_{tracks} in $\Upsilon(2S)$ data, compared to the total generic $\Upsilon(2S)$ MC sample and the $e^+e^- \rightarrow q\bar{q}$ component. The data distribution peaks at $N_{\text{tracks}} \simeq 10$ – 12 and exhibits a long tail extending towards high multiplicities, reflecting the nature of hadronic $\Upsilon(2S)$ decays. Low-multiplicity events are strongly suppressed, while the gradual fall-off at large N_{tracks} corresponds to events with increasing particle production, including contributions from secondary interactions.

The generic $\Upsilon(2S)$ MC reproduces the overall structure of the distribution but is systematically shifted towards slightly lower multiplicities. In particular, the MC overshoots the data at low N_{tracks} and underestimates the yield in the region above the peak. This behaviour indicates that the simulated events tend to produce fewer reconstructed charged tracks on average than observed in data, resulting in a somewhat harder low-multiplicity spectrum.

The $q\bar{q}$ continuum component is concentrated at relatively small values of N_{tracks} and decreases rapidly with increasing multiplicity. This is consistent with the more jet-like topology of continuum events, which typically produce fewer charged particles than resonance decays. The comparison highlights the strong sensitivity of N_{tracks} to the underlying event structure and confirms its complementarity to R_2 in characterising the global properties of the event and in defining the event-shape classes used for event mixing.

Figure 7.5 illustrates the definition of the event-shape classes used for event mixing on the $\Upsilon(2S)$ sample. The binning is constructed using fixed intervals in R_2 and

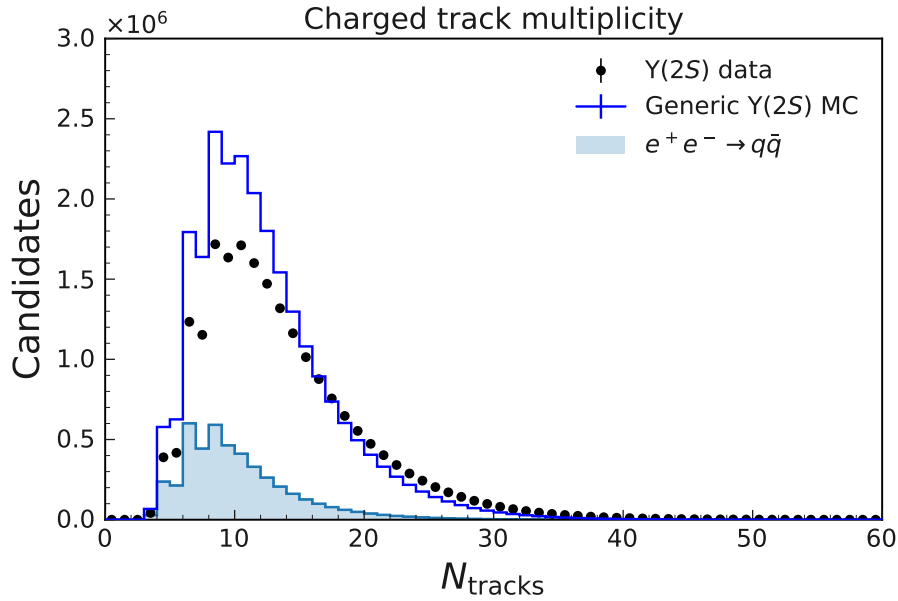


Figure 7.4: N_{tracks} distribution in $\Upsilon(2S)$ data (black points) compared to the total $\Upsilon(2S)$ MC (blue histogram) and the $q\bar{q}$ component (light-blue filled histogram).

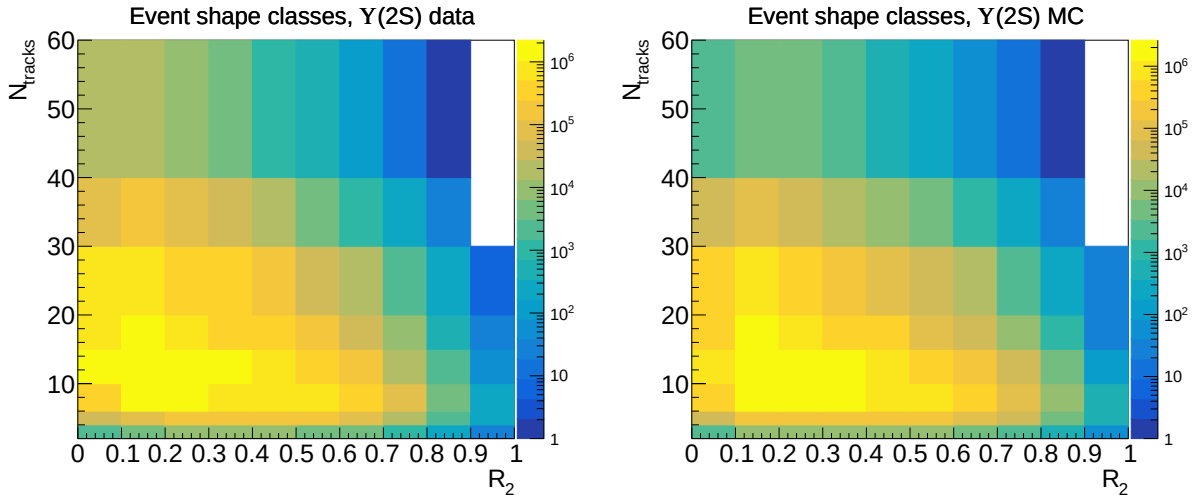


Figure 7.5: Event-shape classes in the (R_2, N_{tracks}) plane for $\Upsilon(2S)$ data (left) and $\Upsilon(2S)$ MC (right), using the same binning definition in both panels.

variable-width intervals in N_{tracks} , chosen to reflect the broad range of charged-particle multiplicities observed in hadronic events. The same classification is applied to the $\Upsilon(1S)$ sample.

The color scale represents the event density in each (R_2, N_{tracks}) bin. In data, the population is concentrated at low R_2 and intermediate-to-large N_{tracks} , corresponding to spherical, high-multiplicity events typical of $\Upsilon(2S)$ decays. As R_2 increases, the event yield decreases steadily across all multiplicity ranges, with only a limited population at large R_2 and low N_{tracks} , where more jet-like topologies dominate.

The Monte Carlo distribution shows the same global structure, but with a visible redistribution of events towards slightly larger R_2 values and lower multiplicities, consistent with the one-dimensional comparisons discussed above. Despite these differences, the

same binning scheme captures the dominant event populations in both data and simulation, ensuring that events mixed within a given class have comparable global topologies.

This two-dimensional classification allows the event mixing to preserve correlations induced by overall event shape and multiplicity, while maintaining sufficient statistical precision in each bin. The use of identical event-shape classes for data and MC ensures a consistent treatment of same-event and mixed-event samples throughout the analysis.

7.1.3 Definition of the correlation function

The two-particle correlation function is constructed from the ratio of same-event (SE) and mixed-event (ME) pair distributions as a function of the relative momentum k^* . After all kinematic, topological, and event-shape selections are applied, particle pairs are formed either within the same physical event or by combining particles from different events with similar global properties. For experimental data, the proton candidates entering the pair construction are further weighted by their prompt-proton probabilities, as described in chapter 6, in order to suppress contributions from secondary (anti)protons and the residual non-proton contamination.

The unnormalized correlation function is defined as

$$C_{\text{raw}}(k^*) = \frac{A_{\text{SE}}(k^*)}{B_{\text{ME}}(k^*)},$$

where $A_{\text{SE}}(k^*)$ is the distribution of same-event pairs, containing genuine femtoscopic correlations arising from quantum statistics and final-state interactions, while $B_{\text{ME}}(k^*)$ provides an estimate of the uncorrelated reference distribution and accounts for acceptance and phase-space effects.

Since the absolute normalization of the mixed-event distribution is in principle arbitrary, the correlation function is normalized by imposing

$$\langle C(k^*) \rangle = 1$$

in a high- k^* region where femtoscopic correlations are expected to be negligible. In practice, this is achieved by scaling B_{ME} such that the integrals of A_{SE} and B_{ME} match within the chosen normalization interval. The final, normalized correlation function is therefore given by

$$C(k^*) = \frac{A_{\text{SE}}(k^*)}{B_{\text{ME}}(k^*)} \times \mathcal{N},$$

where the normalization factor $\mathcal{N}$ is determined from the high- k^* region.

The quality of the mixed-event reference is a critical ingredient of the analysis. Residual mismatches between the SE and ME samples, for example due to imperfect matching of event topology or charged-particle multiplicity, can introduce artificial structures in the correlation function. Such effects are particularly relevant at low k^* , where the femtoscopic signal is strongest. This consideration motivates the detailed event-shape binning and event-mixing strategy described in the previous sections.

7.2 Results at generator level

To validate the pair-construction method and to isolate the physical effects expected in the absence of detector imperfections, the correlation function is first evaluated using the

$\Upsilon(2S)$ MC sample at generator level. In this configuration, the proton four-momenta correspond to the true kinematic values produced by the event generator, and the resulting distributions are free of any distortions related to tracking resolution, misidentification, or vertex selection. Only true primary protons and antiprotons are selected. This allows to identify the fundamental non-femtoscopic structures that arise purely from event kinematics.

Figure 7.6 shows the correlation functions $C(k^*)$ at generator level, separately for pp and $\bar{p}\bar{p}$ pairs, in the range $0 < k^* < 3 \text{ GeV}/c$.

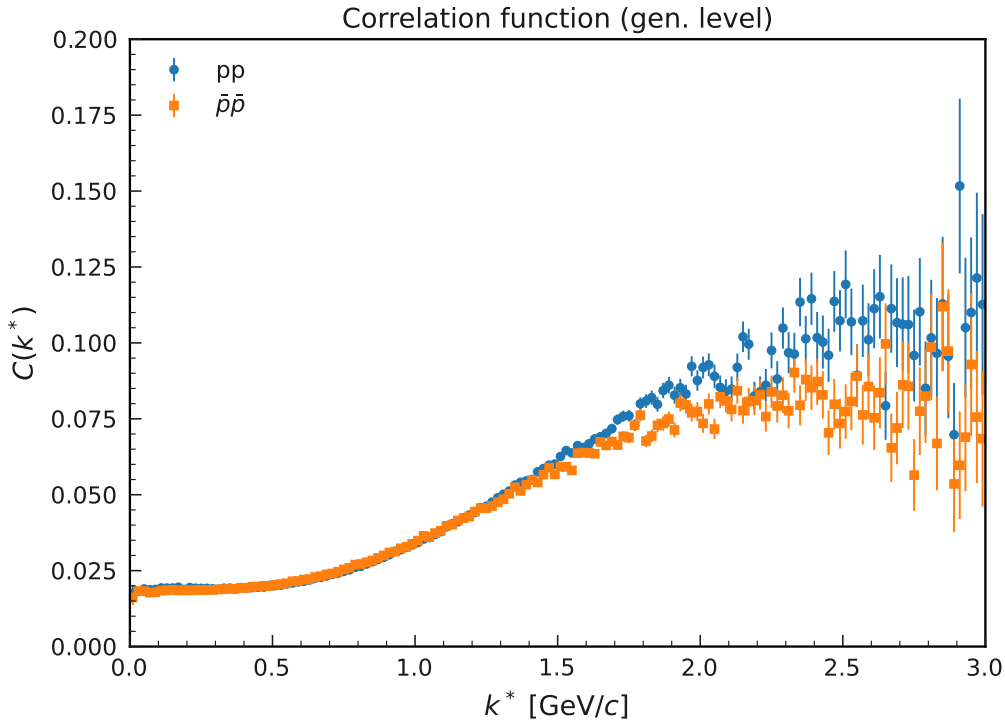
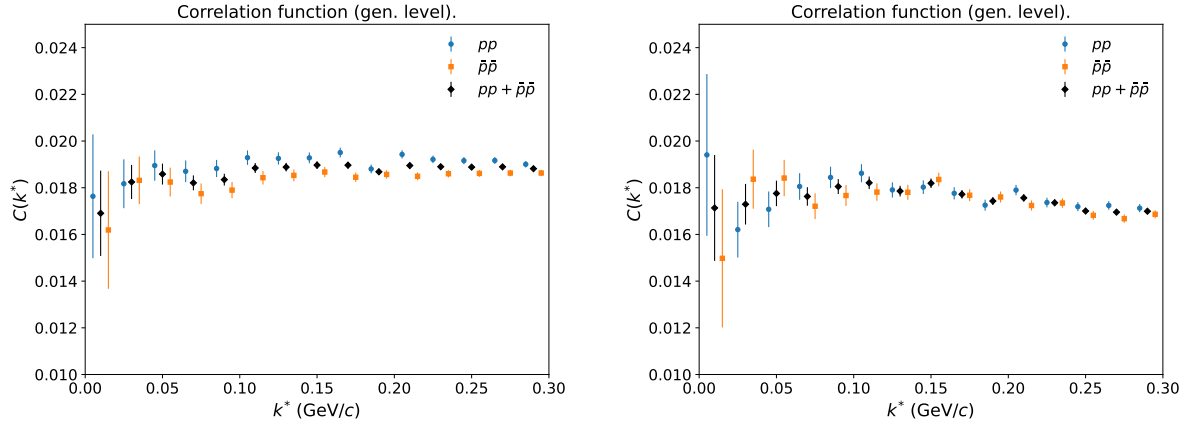


Figure 7.6: Generator-level correlation function $C(k^*)$ for pp (blue points) and $\bar{p}\bar{p}$ (orange squares) pairs in the range $0 < k^* < 3 \text{ GeV}/c$.

The correlation function obtained from the generator-level sample already exhibits a pronounced large-scale modulation, visible as a broad enhancement for $k^* > 500 \text{ MeV}/c$. This effect originates from global energy–momentum conservation: in an e^+e^- collision at fixed center-of-mass energy, the momentum of each particle is constrained by the momenta of all other particles in the event. Such kinematic correlations manifest themselves even in the absence of quantum-statistical or strong-interaction effects and are therefore present in the SE distribution but absent from the ME reference. As a consequence, the ratio $A_{\text{SE}}/B_{\text{ME}}$ acquires a characteristic long-range shape that must be accounted for in the interpretation of data. The Energy–Momentum Conservation–Induced Correlation (EMCIC) effect is described in detail in Section 7.2.2.

The correlation function is also evaluated at generator level after applying acceptance selections that emulate the detector geometrical and kinematic constraints. Specifically, tracks are required to satisfy a transverse momentum threshold of $p_{\text{T}} > 280 \text{ MeV}/c$, corresponding to the TOF acceptance, and a polar-angle selection $17^\circ < \theta < 150^\circ$, reflecting the fiducial coverage of the CDC. Fig. 7.7 shows the correlation function in the range $k^* < 300 \text{ MeV}/c$ for pp , $\bar{p}\bar{p}$, and their sum. The left panel corresponds to the

baseline selection (true primary protons and antiprotons), while the right panel shows $C(k^*)$ for pairs of candidates that survive the acceptance selection.

(a) True primary pp and $\bar{p}\bar{p}$.

(b) Including acceptance selections.

Figure 7.7: Generator-level correlation function $C(k^*)$ for pp (blue circles), $\bar{p}\bar{p}$ (orange squares), and the combined sample $pp + \bar{p}\bar{p}$ (black diamonds), shown for the nominal selection (left) and after the acceptance selection (right). For the pp and $\bar{p}\bar{p}$ samples, $C(k^*)$ is staggered by ± 5 MeV/ c to improve visualization.

Even at generator level, where tracks are not reconstructed, and in the absence of any selection beyond the requirement that both particles are true primary protons or antiprotons, a depletion of the correlation function is observed at very low relative momentum, $k^* \lesssim 50$ MeV/ c . This feature is most likely related to the hadronization model implemented in PYTHIA, in particular to diquark suppression [90] and to the so-called popcorn mechanism [91], which affect the production of nearby baryon–baryon pairs in phase space. Since this effect originates from the event generator itself, it should be regarded as a model-dependent systematic uncertainty that is not under experimental control. In practice, its impact is negligible for the present analysis, as in this low- k^* region the statistical uncertainty is expected to dominate both in real data and in reconstructed Monte Carlo samples. Moreover, a quantitative evaluation of this systematic uncertainty would require the generation of high-statistics inclusive MC samples for a grid of PYTHIA parameter settings, which would represent a disproportionate computational effort for a negligible gain in accuracy.

In the absence of any acceptance-related selection, the correlation function is observed to be reasonably flat in the region $50 < k^* < 300$ MeV/ c at generator level, as shown in Fig 7.7a. This behaviour is expected, as no genuine femtoscopic correlations are present in the simulation and no experimental effects are introduced that could distort the relative-momentum distribution. The observed flatness in this intermediate k^* range therefore provides a non-trivial validation of the correlation-function construction procedure.

When the acceptance selection is applied, however, a clear modulation of $C(k^*)$ emerges already at generator level, and the correlation function is no longer flat. This is shown in Fig 7.7b. At first sight, this behavior is unexpected, since the ratio of same-event to mixed-event distributions is designed to cancel common selection effects. The origin of this modulation is therefore not trivial and cannot be attributed to a simple acceptance bias. The following sections are devoted to a detailed investigation of this effect and to

a discussion of the mechanisms that lead to a residual structure in the generator-level correlation function once acceptance effects are taken into account.

7.2.1 Acceptance effects

The magnitude of the single particle's momentum in the pair rest frame (PRF), $k^* \equiv |\mathbf{k}^*|$, defined as half of the proton–proton relative momentum in the PRF, is a Lorentz-invariant quantity. However, $\mathbf{k}^*$ is a three-momentum vector, whose *direction* is specified by the solid angle Ω^* in the PRF.

In contrast, reconstruction and acceptance selections are defined in the laboratory frame. In this frame, a pp ($\bar{p}\bar{p}$) pair is boosted by its total momentum $\mathbf{K}$. As a consequence, although k^* itself is invariant, the *probability for a given point on the k^* shell to be accepted* depends on how the pair is boosted and oriented. In this section, it is shown that single-particle acceptance cuts can induce non-trivial structures in the measured correlation function.

The single-particle momentum density in the laboratory frame is

$$f(p) \equiv \frac{dN}{d^3\mathbf{p}}, \quad (7.2)$$

where alternative coordinate choices (e.g. p_T, η, ϕ) imply the corresponding Jacobian.

The same-event two-particle density is defined as

$$F_{\text{SE}}(p_1, p_2) \equiv \frac{d^2 N_{\text{pairs}}}{d^3\mathbf{p}_1 d^3\mathbf{p}_2}, \quad (7.3)$$

and includes all physical correlations present within a single event, such as global conservation laws, jet or string fragmentation, resonance decays, quantum statistics (QS), and final-state interactions (FSI). At generator level, QS and FSI are neglected.

The mixed-event reference is constructed by pairing particles from *different* events belonging to the same event class $\mathcal{C}$. In the absence of residual biases, its density factorizes as

$$F_{\text{ME}}(p_1, p_2 | \mathcal{C}) = f(p_1 | \mathcal{C}) f(p_2 | \mathcal{C}), \quad (7.4)$$

where $f(p | \mathcal{C})$ denotes the single-particle spectrum restricted to $\mathcal{C}$.

A single-particle acceptance selection can be expressed as

$$\mathcal{A}(p) = \mathbf{1}[\theta_{\min} \leq \theta(p) \leq \theta_{\max} \ \&\& \ p_T(p) > p_T^{\min}], \quad (7.5)$$

where $\theta_{\min} = 17^\circ$ and $\theta_{\max} = 150^\circ$ define the CDC polar-angle acceptance, and $p_T^{\min} = 280 \text{ MeV}/c$ reflects the TOF acceptance threshold. The function $\mathcal{A}(p)$ equals unity if the particle satisfies all cuts and zero otherwise.

The correlation function is defined as

$$C(k^*) = \frac{A(k^*)}{B(k^*)}, \quad (7.6)$$

with

$$A(k^*) = \int d^3\mathbf{p}_1 d^3\mathbf{p}_2 F_{\text{SE}}(p_1, p_2) \mathcal{A}(p_1) \mathcal{A}(p_2) \delta(k^* - k^*(p_1, p_2)), \quad (7.7)$$

$$B(k^*) = \int d^3\mathbf{p}_1 d^3\mathbf{p}_2 f(p_1) f(p_2) \mathcal{A}(p_1) \mathcal{A}(p_2) \delta(k^* - k^*(p_1, p_2)). \quad (7.8)$$

Although the acceptance cuts act on individual particles, they can affect pair observables through the boost between the laboratory frame and the PRF.

One defines

$$\mathbf{K} \equiv \frac{1}{2}(\mathbf{p}_1 + \mathbf{p}_2), \quad \mathbf{q} \equiv \mathbf{p}_1 - \mathbf{p}_2, \quad (7.9)$$

so that

$$\mathbf{p}_1 = \mathbf{K} + \frac{1}{2}\mathbf{q}, \quad \mathbf{p}_2 = \mathbf{K} - \frac{1}{2}\mathbf{q}. \quad (7.10)$$

Since k^* is defined in the PRF, it is convenient to rewrite Eqs. (7.7) and (7.8) in terms of pair variables. A suitable set of six independent variables is:

- the total momentum $\mathbf{P} = \mathbf{p}_1 + \mathbf{p}_2 = 2\mathbf{K}$ (3 dof),
- the Lorentz scalar $s = P^2$ (1 dof),
- the solid angle $\Omega^* = (\theta_*, \phi_*)$ defining the direction of $\mathbf{k}^*$ in the PRF (2 dof).

Given $(s, \mathbf{P})$, the boost from the PRF to the laboratory frame is characterized by

$$\beta = \frac{\mathbf{P}}{P^0}, \quad \gamma = \frac{P^0}{\sqrt{s}}, \quad P^0 = E_1 + E_2. \quad (7.11)$$

Defining $\hat{\mathbf{P}} = \mathbf{P}/|\mathbf{P}|$ and $\cos \psi \equiv \hat{\mathbf{k}}^* \cdot \hat{\mathbf{P}}$, the boosted momenta read

$$\mathbf{p}_1 = \mathbf{k}_\perp^* + \hat{\mathbf{P}} \gamma (k^* \cos \psi + \beta E_1^*), \quad (7.12)$$

$$\mathbf{p}_2 = -\mathbf{k}_\perp^* + \hat{\mathbf{P}} \gamma (-k^* \cos \psi + \beta E_2^*), \quad (7.13)$$

where $\mathbf{k}_\perp^* = k^* \sin \psi \hat{\mathbf{e}}_\perp$ is orthogonal to $\hat{\mathbf{P}}$.

The change of variables $(\mathbf{p}_1, \mathbf{p}_2) \rightarrow (\mathbf{P}, s, \Omega^*)$ introduces a smooth Jacobian J common to numerator and denominator:

$$d^3\mathbf{p}_1 d^3\mathbf{p}_2 = J(\mathbf{P}, s, \Omega^*) d^3\mathbf{P} ds d\Omega^*. \quad (7.14)$$

Since k^* depends only on s and the particle masses,

$$\delta(k^* - k^*(p_1, p_2)) = \frac{\delta(s - s_0(k^*))}{\left| \frac{dk^*}{ds} \right|_{s=s_0}}, \quad (7.15)$$

with $s_0(k^*)$ defined by $k^*(s_0) = k^*$. For equal masses,

$$s_0 = 4(m^2 + k^{*2}), \quad \frac{dk^*}{ds} = \frac{1}{8k^*}. \quad (7.16)$$

After integrating over s , both A and B reduce to

$$A(k^*) = \int d^3\mathbf{P} d\Omega^* \mathcal{W}_{\text{SE}}(\mathbf{P}, \Omega^*; k^*) \varepsilon(\mathbf{P}, \Omega^*; k^*), \quad (7.17)$$

$$B(k^*) = \int d^3\mathbf{P} d\Omega^* \mathcal{W}_{\text{ME}}(\mathbf{P}, \Omega^*; k^*) \varepsilon(\mathbf{P}, \Omega^*; k^*), \quad (7.18)$$

where $\varepsilon = \mathcal{A}(p_1)\mathcal{A}(p_2)$ is evaluated on the boosted momenta of Eq. (7.13).

The measured correlation function becomes

$$C(k^*) = \frac{\int d^3\mathbf{P} d\Omega^* \mathcal{W}_{\text{SE}}(\mathbf{P}, \Omega^*; k^*) \varepsilon}{\int d^3\mathbf{P} d\Omega^* \mathcal{W}_{\text{ME}}(\mathbf{P}, \Omega^*; k^*) \varepsilon}. \quad (7.19)$$

A fixed k^* bin corresponds to a thin spherical shell in the PRF. While all points on this shell share the same invariant relative momentum, they differ in their boost $\mathbf{P}$ and orientation Ω^* . Since acceptance cuts are applied in the laboratory frame, different regions of the shell have different probabilities to be accepted.

At fixed k^* :

- increasing $K_T = \frac{1}{2}|\mathbf{P}_T|$ tends to align the two lab-frame tracks and increase their transverse momenta;
- decreasing K_T enhances configurations where one particle is soft or emitted at large polar angle.

As a result, the acceptance weight $\varepsilon(\mathbf{P}, \Omega^*; k^*)$ is not uniform over the k^* shell.

Moreover, the conditional densities $\mathcal{W}_{\text{SE}}$ and $\mathcal{W}_{\text{ME}}$ are generally different, because same-event pairs carry correlations (e.g. momentum conservation, jets, resonance decays) that are absent in mixed events. Therefore, $A(k^*)$ and $B(k^*)$ correspond to averages of the *same* acceptance function over *different* populations of the k^* shell, leading to a smooth k^* dependence of $C(k^*)$ even in the absence of detector effects.

The acceptance-induced modulation is directly related to the relative orientation between the pair momentum $\mathbf{P}$ and the relative momentum $\mathbf{k}^*$. This can be visualized by studying the distribution of $\cos\psi = \hat{\mathbf{k}}^* \cdot \hat{\mathbf{P}}$ in different k^* regions. Fig. 7.8 shows the distributions of $\cos\psi = \hat{\mathbf{k}}^* \cdot \hat{\mathbf{P}}$ for proton–proton pairs in four representative k^* intervals, using the $\Upsilon(2S)$ generator-level MC without acceptance selection. The shown k^* bins are chosen to span from the femtoscopic region to the high- k^* region, where no femtoscopic signal is expected.

Fig. 7.9 shows the distributions of the pair transverse momentum, $K_T = \frac{1}{2}|\mathbf{P}_T|$, for SE and ME proton–proton pairs in four representative k^* intervals. The comparison highlights how the population of K_T depends on the selected relative-momentum region.

The acceptance would cancel exactly if either the selection were Lorentz invariant, or the mixing procedure enforced $\mathcal{W}_{\text{SE}} \approx \mathcal{W}_{\text{ME}}$ at fixed k^* . In practice, partial cancellation can be achieved by binning in K_T , as commonly done in heavy-ion femtoscopy. In the present analysis, however, this modulation is instead traced back to the more fundamental origin of energy–momentum conservation.

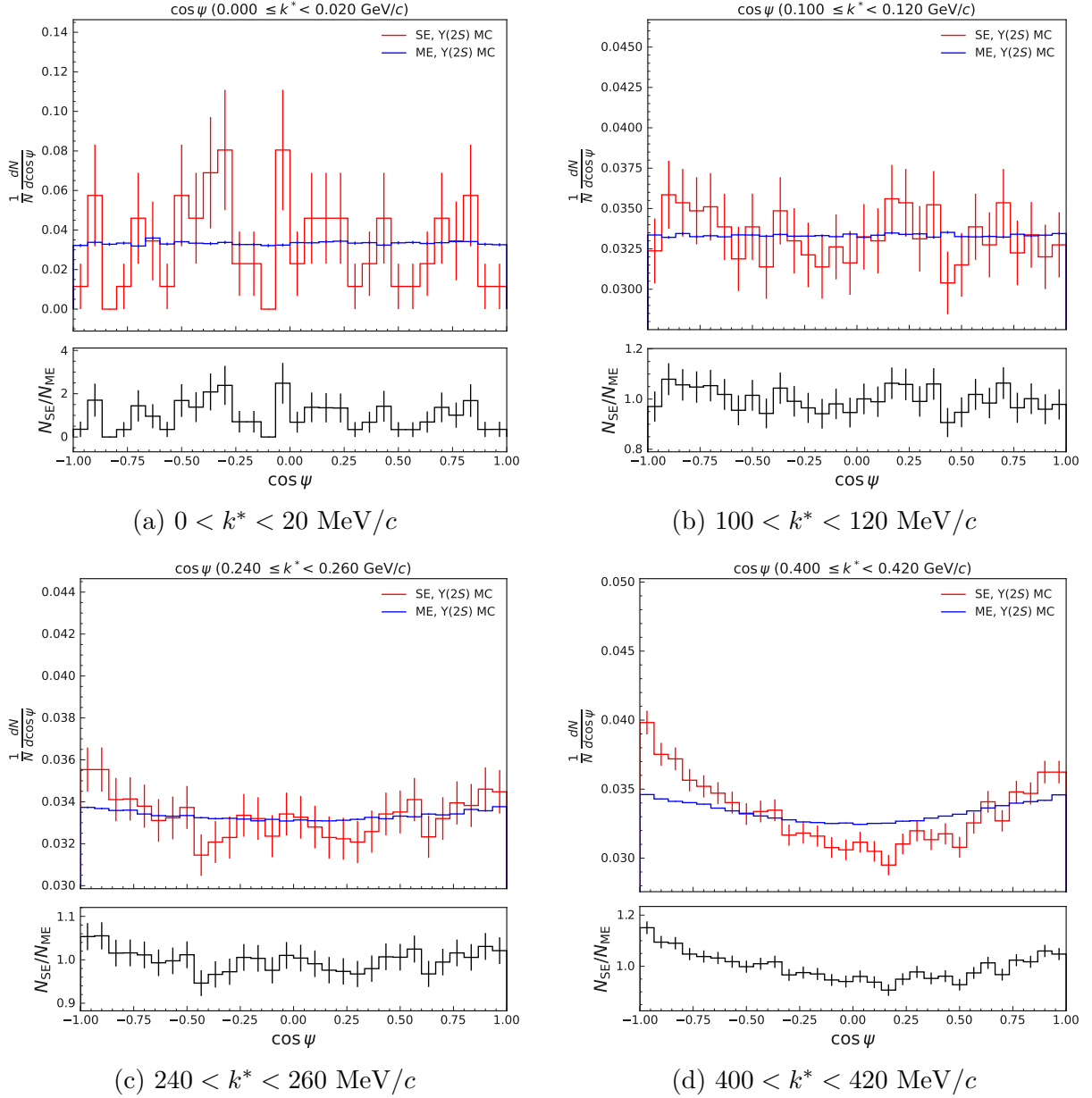


Figure 7.8: Distributions of $\cos \psi = \hat{\mathbf{k}}^* \cdot \hat{\mathbf{P}}$ in four k^* intervals: $[0, 20]$, $[100, 120]$, $[240, 260]$, and $[400, 420]$ MeV/c. These slices illustrate how the relative orientation between $\mathbf{k}^*$ and the pair momentum $\mathbf{P}$ is populated differently as a function of k^* , which in turn impacts the acceptance weight $\varepsilon(\mathbf{P}, \Omega^*; k^*)$ entering Eq. (7.19).

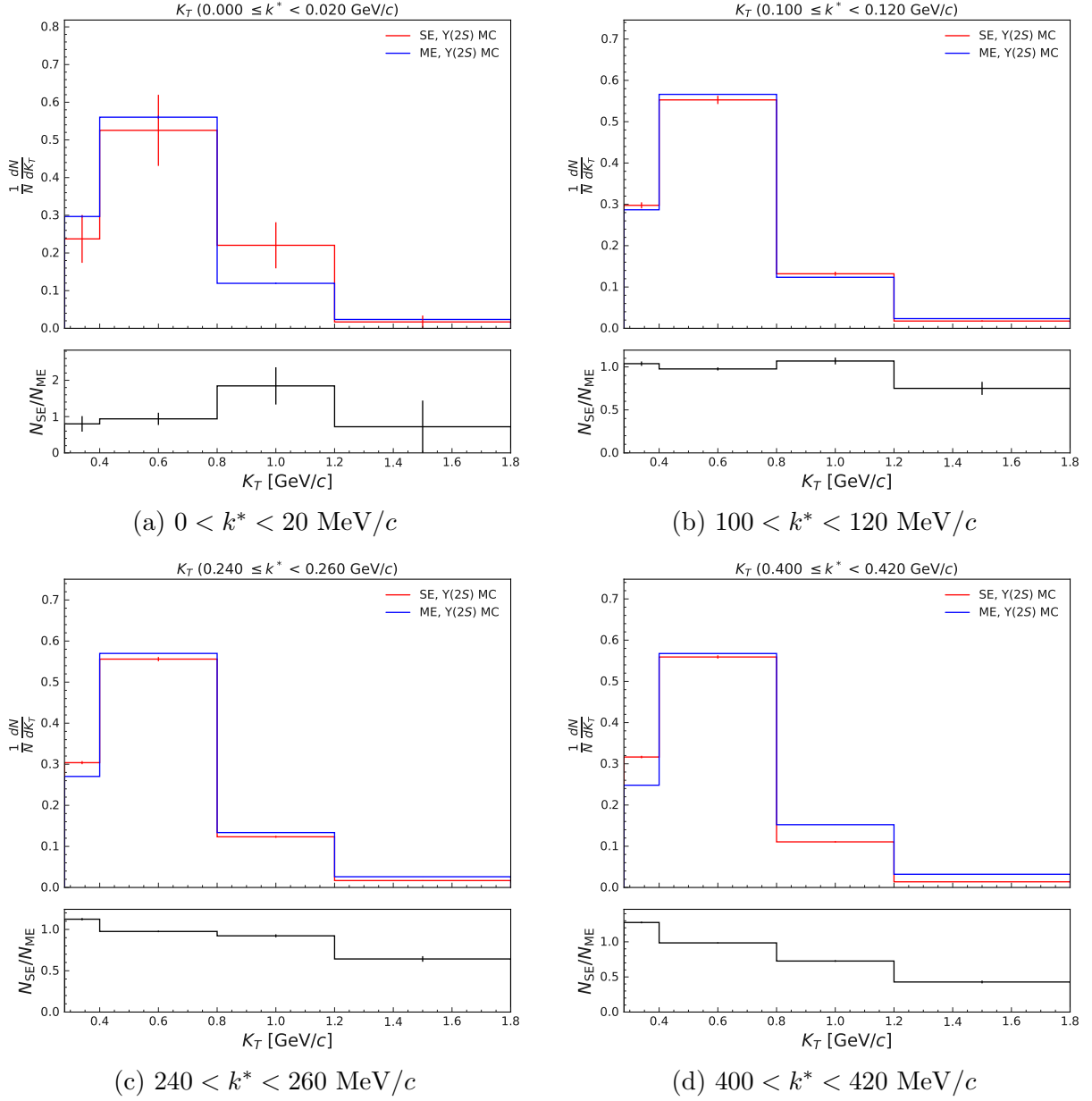


Figure 7.9: Distributions of the pair transverse momentum K_T for same-event (SE) and mixed-event (ME) proton–proton pairs in four k^* intervals: $[0, 20]$, $[100, 120]$, $[240, 260]$, and $[400, 420]$ MeV/c. The lower panels show the ratio $N_{\text{SE}}/N_{\text{ME}}$ in each k^* slice.

7.2.2 Energy–Momentum Conservation–Induced Correlation (EMCIC)

As already discussed in Sec. 7.2, a broad enhancement is observed in the proton–proton correlation function when measured over a wide relative-momentum range, $k^* \in [0.0, 3.0]$ GeV/ c , even at generator level, and in the absence of any detector acceptance effects. This structure cannot be attributed to final-state interactions and instead reflects correlations induced by global conservation laws.

Energy–Momentum Conservation–Induced Correlations (EMCIC) arise because, in a finite-multiplicity system, the momenta of the produced particles cannot fluctuate independently. The constraints imposed by global energy and momentum conservation introduce long-range, non-femtoscopic correlations among all particles in the event. As a result, even a source that is uncorrelated at the microscopic level develops a smooth, large-scale structure in the measured correlation function, typically manifesting itself as a gradual enhancement toward large values of k^* . Correctly accounting for this contribution is therefore essential in order to avoid biases in the extraction of the genuine femtoscopic signal.

A phenomenological description of EMCIC effects was proposed by Chajecski and Lisa [92]. In this approach, the non-femtoscopic baseline is expanded to leading order in terms of two-particle kinematic correlations and can be written as

$$C_{\text{EMCIC}}(k^*) \simeq 1 - 2M_1 \langle \mathbf{p}_{1T} \cdot \mathbf{p}_{2T} \rangle(k^*) - M_2 \langle p_{1z} p_{2z} \rangle(k^*) - M_3 \langle E_1 E_2 \rangle(k^*) + M_4 \langle E_1 + E_2 \rangle(k^*) - M_5(k^*), \quad (7.20)$$

where the coefficients M_i depend on the event multiplicity and on the global kinematic properties of the system. The quantities in angular brackets denote averages over particle pairs in a given k^* bin and are measured in the center-of-mass reference frame.

Despite its clear physical motivation, this parametrisation is not widely used in femtoscopic analyses. The main reason is its practical complexity: the EMCIC contribution must be reconstructed from several k^* -dependent kinematic profiles, and the coefficients M_i are strongly correlated with each other and with the femtoscopic parameters. As a consequence, many analyses prefer to model the baseline with empirical functions or to restrict the fit range to low k^* values, where EMCIC effects are less pronounced. Nevertheless, when a wide k^* interval is considered, the Chajecski–Lisa approach provides a physically well-defined framework to explicitly incorporate conservation-law effects into the modeling of the correlation function.

In the present analysis, each of the kinematic quantities entering the EMCIC ansatz is evaluated explicitly at generator level. For each proton pair, the quantities

$$\mathbf{p}_{1T} \cdot \mathbf{p}_{2T}, \quad p_{1z} p_{2z}, \quad E_1 + E_2, \quad E_1 E_2$$

are computed, and their average values are constructed as functions of k^* .

At generator level, these profiles isolate the correlations induced purely by global conservation laws, free from detector effects and final-state interactions.

Since the strength of EMCIC effects is expected to depend on the event topology, the analysis is performed not only on the fully inclusive sample but also in selected regions of the (N_{tracks}, R_2) phase space. In particular, two representative event classes are considered: a low-multiplicity, jet-like topology dominated by continuum $e^+e^- \rightarrow q\bar{q}$ events, defined by $4 < N_{\text{tracks}} < 6$ and $0.5 < R_2 < 0.6$, and a high-multiplicity, spherical

topology characteristic of $\Upsilon(nS)$ hadronic decays, defined by $15 < N_{\text{tracks}} < 20$ and $0.1 < R_2 < 0.2$. The EMCIC kinematic profiles obtained in these two event classes are shown in Fig. 7.10 and Fig. 7.11 respectively.

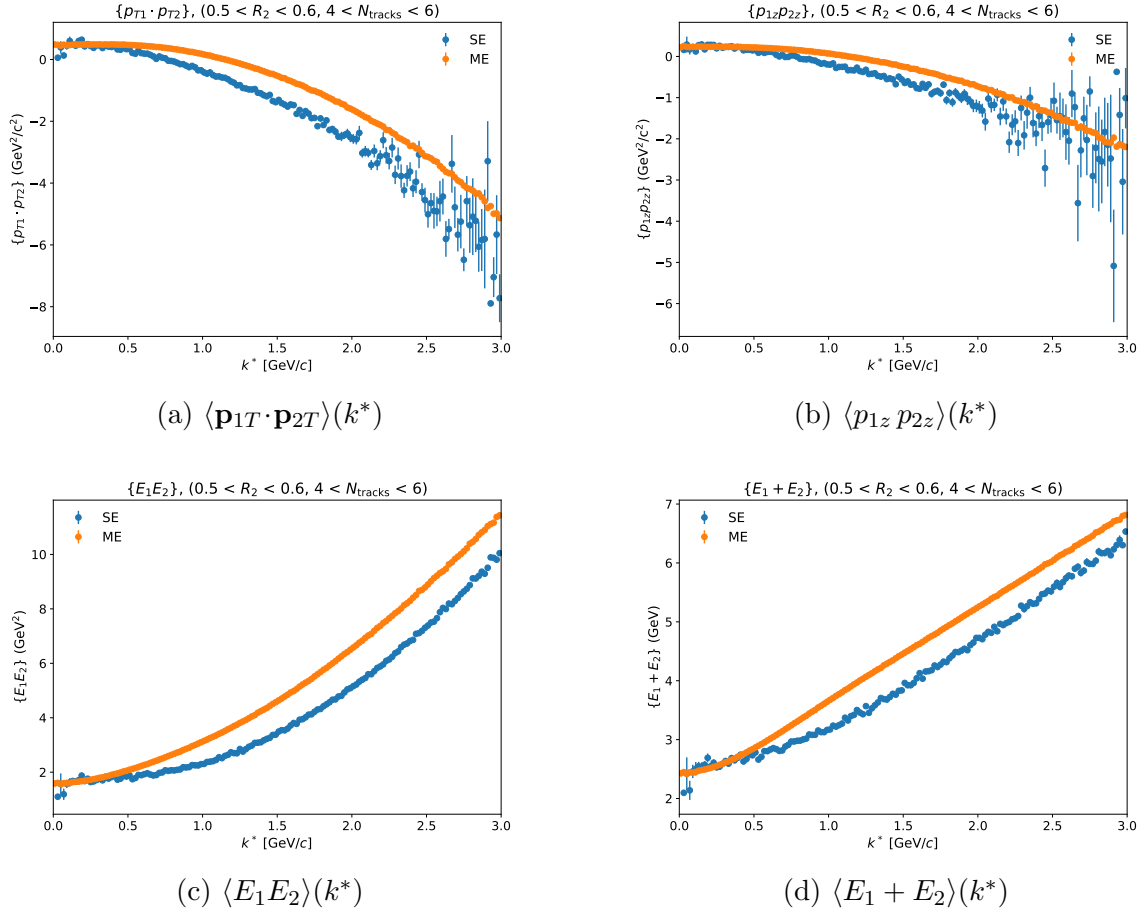


Figure 7.10: Generator-level EMCIC kinematic profiles as a function of k^* for a continuum-like, jet-like event class: $0.5 < R_2 < 0.6$ and $4 < N_{\text{tracks}} < 6$. The shown averages enter the Chajecski–Lisa parametrisation of the non-femtoscopic baseline.

In both cases, the quantities entering the Chajecski–Lisa parametrisation exhibit a smooth and monotonic dependence on k^* . The absolute amplitudes of the kinematic profiles are found to be comparable in the two example event classes. A more relevant difference emerges instead from the comparison between same-event (SE) and mixed-event (ME) distributions within each topology. In the low-multiplicity, jet-like event class, clear discrepancies between SE and ME profiles are observed over a wide range of k^* , indicating the presence of non-trivial correlations induced by global energy–momentum conservation. In contrast, in the higher-multiplicity and more spherical topology, the SE and ME profiles are nearly indistinguishable within statistical uncertainties, suggesting a strong suppression of such conservation-driven effects. This behavior is consistent with the expectation that energy–momentum conservation induces larger distortions in small, low-multiplicity systems, where individual particle momenta represent a non-negligible fraction of the total available phase space. As the event multiplicity increases, these effects become progressively diluted, leading to a reduced impact on two-particle observables and to a better approximation of factorization between single-particle kinematics and pair correlations.

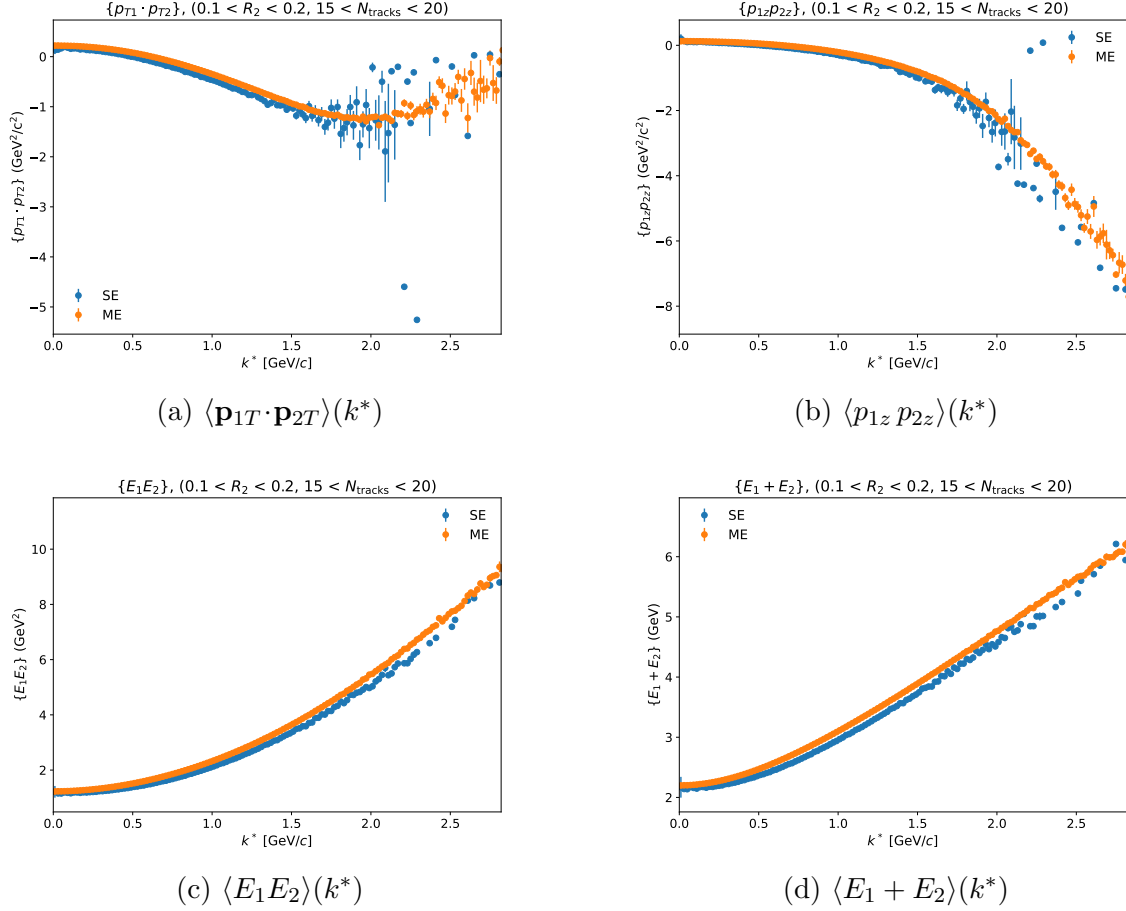


Figure 7.11: Generator-level EMCIC kinematic profiles as a function of k^* for a Υ -like, high-multiplicity and more spherical event class: $0.1 < R_2 < 0.2$ and $15 < N_{\text{tracks}} < 20$. The shown averages enter the Chajewski–Lisa parametrisation of the non-femtoscopic baseline.

The correlation functions for the full $pp + \bar{p}\bar{p}$ $\Upsilon(2S)$ sample at generator level, together with fits including the EMCIC baseline, are shown in Fig. 7.12 and Fig. 7.13. It is stressed that, at this stage, *no acceptance selection is applied*. In both topologies, the EMCIC-based model provides an adequate description of the broad, non-femtoscopic structure observed at large k^* .

In addition to the topology-selected studies discussed above, EMCIC kinematic profiles are also constructed for the fully inclusive generator-level sample, integrating over all N_{tracks} and R_2 bins. The resulting kinematic averages are shown in Fig. 7.14. As already introduced in the previous sections, the EMCIC baseline can be constructed using profiles derived either from same-event (SE) or mixed-event (ME) pairs. For the inclusive correlation function, fits employing both prescriptions are therefore performed. The comparison between the two resulting descriptions, shown in Fig. 7.15, allows a quantitative assessment of possible systematic effects associated with the choice of the EMCIC-profile construction method.

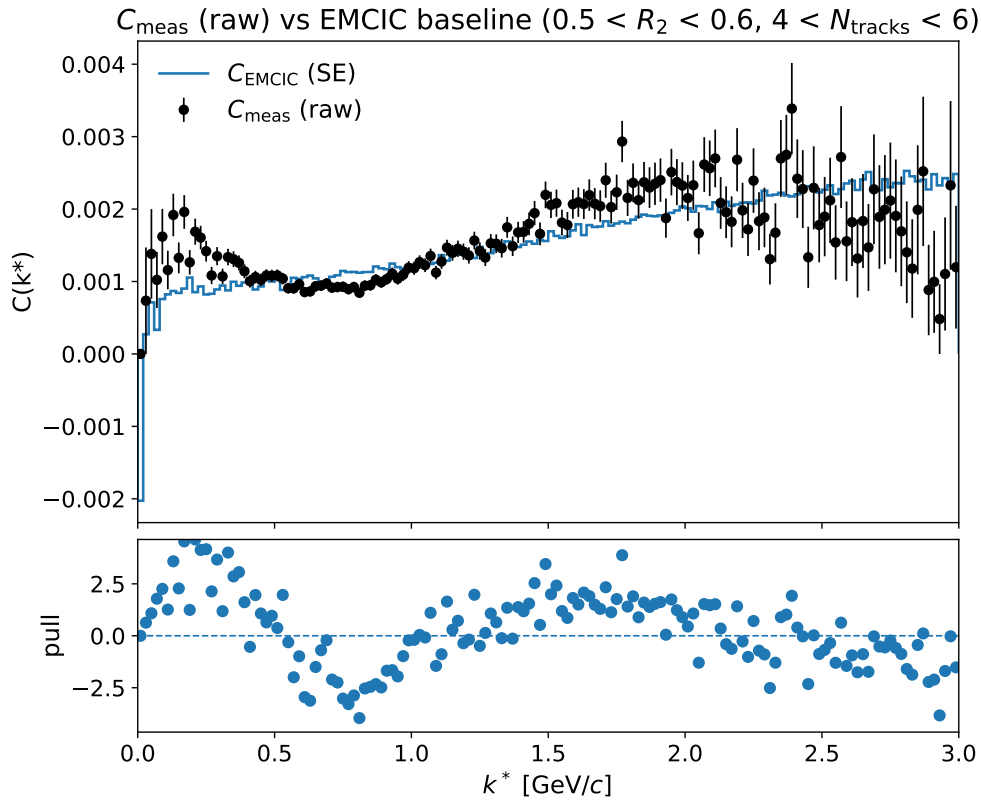


Figure 7.12: Generator-level correlation function $C(k^*)$ for the $pp + \bar{p}\bar{p}$ sample in the low-multiplicity, jet-like ($q\bar{q}$ -dominated) event class at the $\Upsilon(2S)$ resonance, together with the fits including the EMCIC baseline. No acceptance selection is applied. In the top pad, black points with error bars represent the raw measured $C(k^*)$. The blue histogram shows the fit model built from Chajeky-Lisa profiles using SE pairs. The bottom pad shows the pull between data and the fit model.

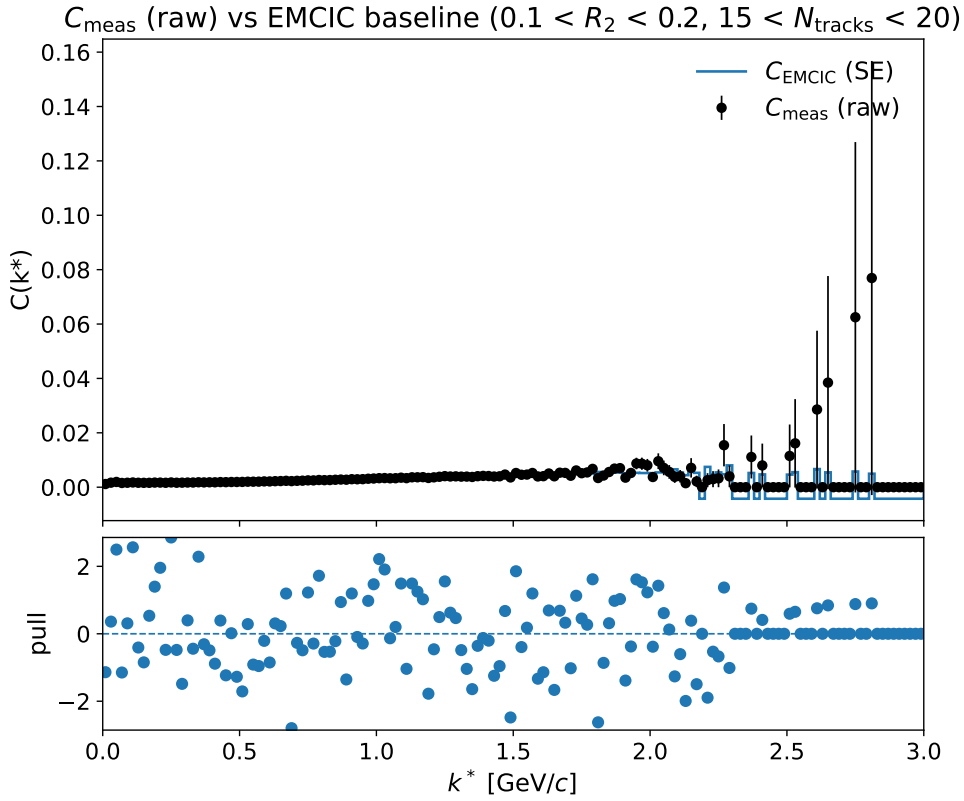


Figure 7.13: Generator-level correlation function $C(k^*)$ for the $pp + \bar{p}\bar{p}$ sample in the high-multiplicity, spherical (Υ -decay-dominated) event class at the $\Upsilon(2S)$ resonance, together with the fit including the EMCIC baseline. No acceptance selection is applied. In the top pad, black points with error bars represent the raw measured $C(k^*)$. The blue histogram shows the fit model built from Chajeky-Lisa profiles using SE pairs. The bottom pad shows the pull between data and the fit model.

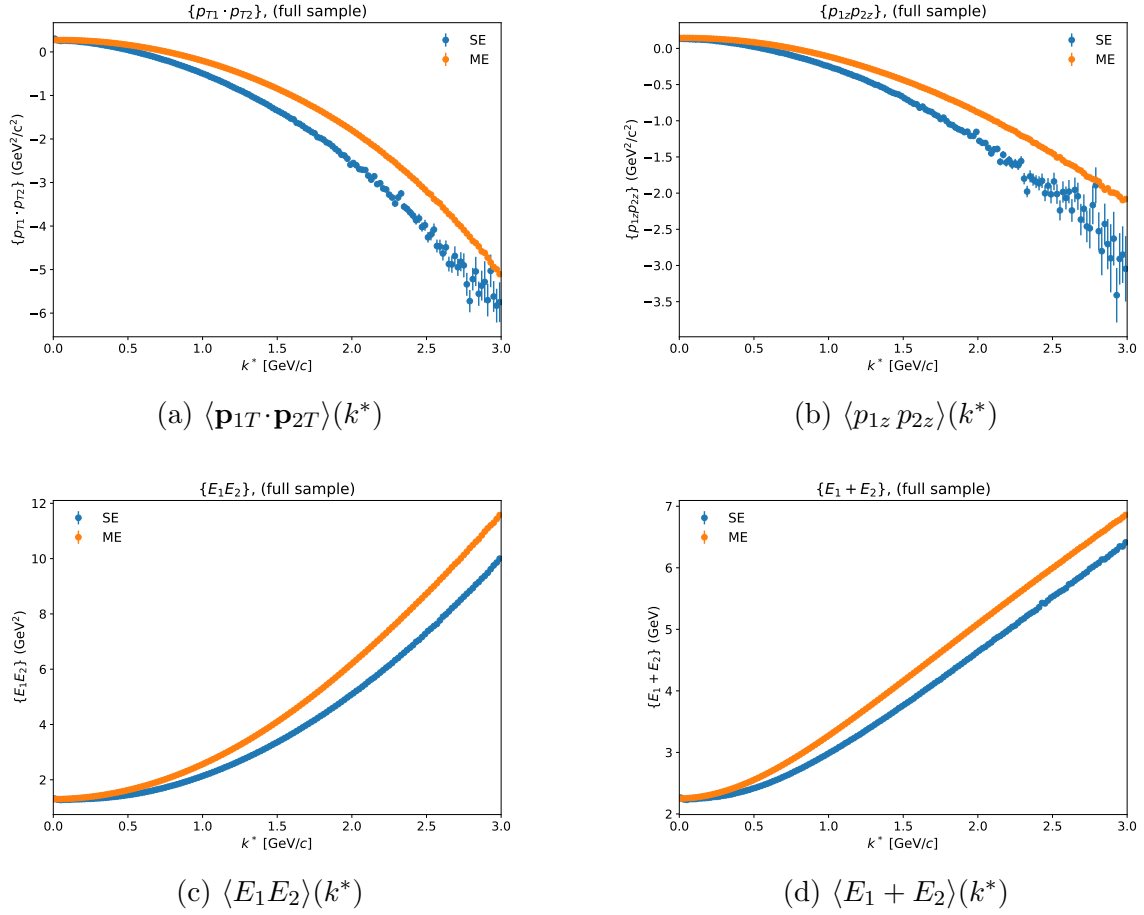


Figure 7.14: Generator-level EMCIC kinematic profiles as a function of k^* for the fully inclusive $pp + \bar{p}\bar{p}$ sample at the $\Upsilon(2S)$ resonance, integrating over all N_{tracks} and R_2 bins. The shown averages enter the Chajecski–Lisa parametrisation of the non-femtoscopic baseline. No acceptance selection is applied.

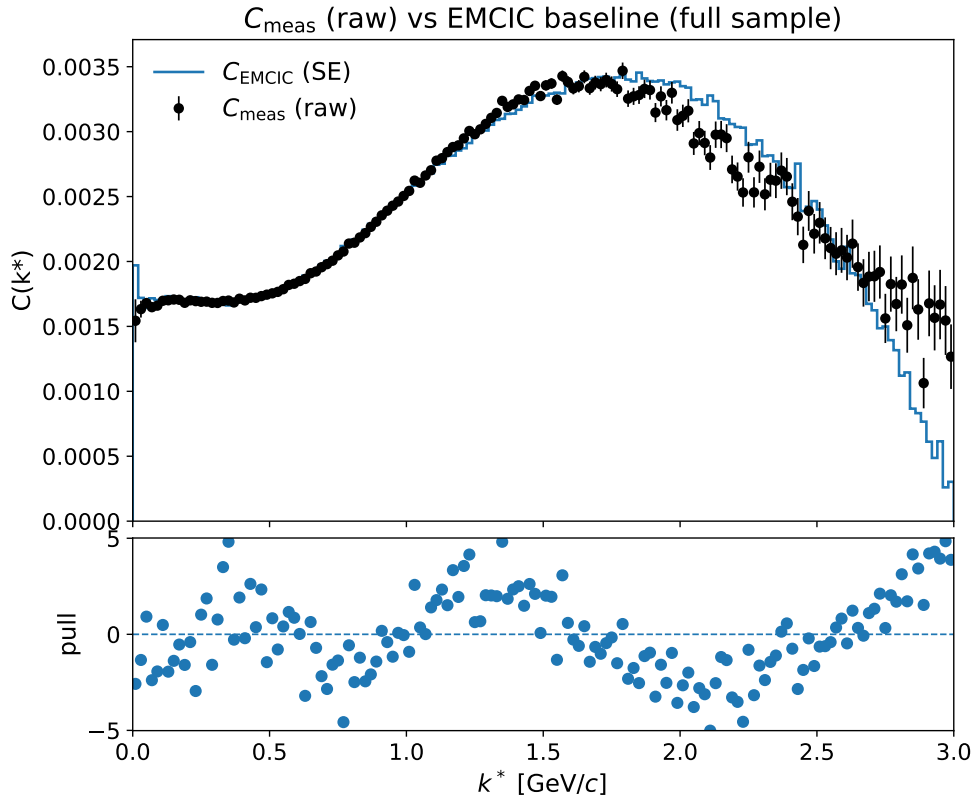


Figure 7.15: Inclusive generator-level correlation function $C(k^*)$ for the $pp + \bar{p}\bar{p}$ sample at the $\Upsilon(2S)$ resonance. The data are fitted using EMCIC baselines constructed from same-event (SE) kinematic profiles. The comparison provides an estimate of systematic effects related to the EMCIC-profile construction. No acceptance selection is applied.

7.2.3 Linear approximation of the EMCIC baseline

The EMCIC contribution to the two-particle correlation function originates from global energy-momentum conservation constraints acting on the full final state. In the Chajecski-Lisa formalism, the EMCIC-only correlation for a pair of particles with four-momenta p_1 and p_2 can be written, to leading order in the inverse event multiplicity, as

$$C_{\text{EMCIC}}(p_1, p_2) = 1 - \frac{1}{N} \left(\frac{2 \vec{p}_{1T} \cdot \vec{p}_{2T}}{\langle p_T^2 \rangle} + \frac{p_{1z} p_{2z}}{\langle p_z^2 \rangle} + \frac{(E_1 - \langle E \rangle)(E_2 - \langle E \rangle)}{\langle E^2 \rangle - \langle E \rangle^2} \right), \quad (7.21)$$

where N is the number of particles participating in the global constraint and the averages are taken over the single-particle phase space.

In femtosopic analyses, the EMCIC contribution is typically studied as a function of the pair relative momentum k^* , defined in the pair rest frame. The corresponding one-dimensional EMCIC baseline is obtained by averaging Eq. (7.21) over all pairs that populate a given k^* interval,

$$C_{\text{EMCIC}}(k^*) \equiv \langle C_{\text{EMCIC}}(p_1, p_2) \rangle_{k^*}. \quad (7.22)$$

Equation (7.21) shows that the EMCIC effect is bilinear in the particle momenta and suppressed by a factor $1/N$. After projection onto k^* , the resulting $C_{\text{EMCIC}}(k^*)$ is therefore a smooth function of the relative momentum, with no intrinsic physical scale associated with short-range structures. As demonstrated by the generator-level studies presented in the previous section, the EMCIC contribution varies slowly over the limited k^* interval relevant for femtoscopy.

Over a finite interval $k^* \in [0, k_{\text{max}}^*]$, any sufficiently smooth function can be expanded according to Taylor's theorem:

$$C_{\text{EMCIC}}(k^*) = a + b k^* + R_2(k^*), \quad (7.23)$$

where the remainder after the linear term is given by

$$R_2(k^*) = \frac{C_{\text{EMCIC}}''(\xi)}{2} (k^*)^2, \quad 0 < \xi < k^*. \quad (7.24)$$

Here $C_{\text{EMCIC}}''(k^*) \equiv d^2 C_{\text{EMCIC}}/d(k^*)^2$ represents the curvature of the EMCIC baseline as a function of k^* , and ξ is an intermediate point guaranteed by Taylor's theorem. Physically, C_{EMCIC}'' quantifies how rapidly the slope of the baseline changes with relative momentum.

The magnitude of the curvature C_{EMCIC}'' is controlled by the same physical scales that enter Eq. (7.21). Dimensional considerations imply

$$|C_{\text{EMCIC}}''(k^*)| \sim \frac{1}{N p_{\text{typ}}^2}, \quad (7.25)$$

where p_{typ} denotes a characteristic single-particle momentum scale of the sample. Using Eq. (7.24), the quadratic remainder can therefore be bounded as

$$|R_2(k_{\text{max}}^*)| \lesssim \frac{1}{2N} \left(\frac{k_{\text{max}}^*}{p_{\text{typ}}} \right)^2. \quad (7.26)$$

For the present analysis, a typical charged-particle multiplicity of $N \simeq 12$ and a conservative momentum scale $p_{\text{typ}} \simeq 0.6 \text{ GeV}/c$ can be assumed. For a femtosopic fit range $k_{\text{max}}^* = 0.3 \text{ GeV}/c$, Eq. (7.26) yields

$$|R_2| \lesssim \mathcal{O}(1\%), \quad (7.27)$$

indicating that quadratic and higher-order terms in k^* are numerically suppressed with respect to the leading linear contribution.

This estimate demonstrates that, within the k^* interval relevant for femtoscopic studies, the EMCIC contribution can be accurately approximated by a linear function of k^* . For this reason, linear or low-order polynomial baselines have been widely adopted in femtoscopic analyses as a practical way to account for long-range non-femtoscopic correlations, including EMCIC effects (e.g. Refs. [93, 94]).

7.3 Results in the reconstructed MC samples

After establishing the main features of the correlation function at generator level, the analysis is extended to fully reconstructed Monte Carlo samples. In this configuration, events are processed through the full detector simulation and reconstruction chain, and proton candidates are selected using the same criteria applied to real data. This step is essential to assess the impact of detector resolution, reconstruction efficiency, and offline selections on the measured correlation function.

To mimic the application of the prompt-(anti)proton identification weights used in data, a truth-selected reference sample is defined in Monte Carlo. Using truth-level information associated with each reconstructed track, only *true primary* protons (and, separately, true primary antiprotons) originating from the interaction point are selected. The corresponding same-event and mixed-event pair distributions are then used to build the pp and $\bar{p}\bar{p}$ correlation functions for prompt pairs. This provides a clean baseline that can be directly compared to the weighted correlation function measured in data.

Figure 7.16 shows the correlation function measured in the reconstructed $\Upsilon(2S)$ MC sample. The left panel displays the full k^* range up to 3 GeV/ c , highlighting the long-range structure induced by global energy-momentum conservation. The right panel shows a zoomed view restricted to $k^* < 300$ MeV/ c , corresponding to the momentum range used in the femtoscopic fits in data.

Over the full k^* range, the reconstructed correlation function exhibits a smooth, slowly varying modulation, qualitatively consistent with the generator-level expectation. In particular, the long-range rise of $C(k^*)$ is preserved after full detector simulation and reconstruction, confirming that the energy-momentum conservation effect is dominant against detector smearing and reconstruction effects.

Focusing on the low- k^* region relevant for femtосcopy, the correlation function is observed to be approximately flat within uncertainties for $k^* \gtrsim 40$ –50 MeV/ c . A depletion is visible in the very first k^* bins, which is less pronounced at generator level. This feature is attributed to finite tracking and momentum resolution, which smears pairs with very small true relative momentum across neighboring bins and leads to a loss of reconstructed pairs at the lowest k^* . The impact of this effect is investigated in detail using a dedicated control sample, as discussed in Section 8.4.

Apart from this low- k^* resolution effect, no additional short-range structures are observed that could mimic a femtoscopic signal. In particular, the reconstructed Monte Carlo does not exhibit spurious enhancements or suppressions in the fit region that are unrelated to known non-femtoscopic sources. This indicates that the reconstruction procedure and offline selections do not introduce artificial correlations beyond those expected from energy-momentum conservation, acceptance, and detector resolution.

The reconstruction-level study therefore confirms that the dominant non-femtoscopic

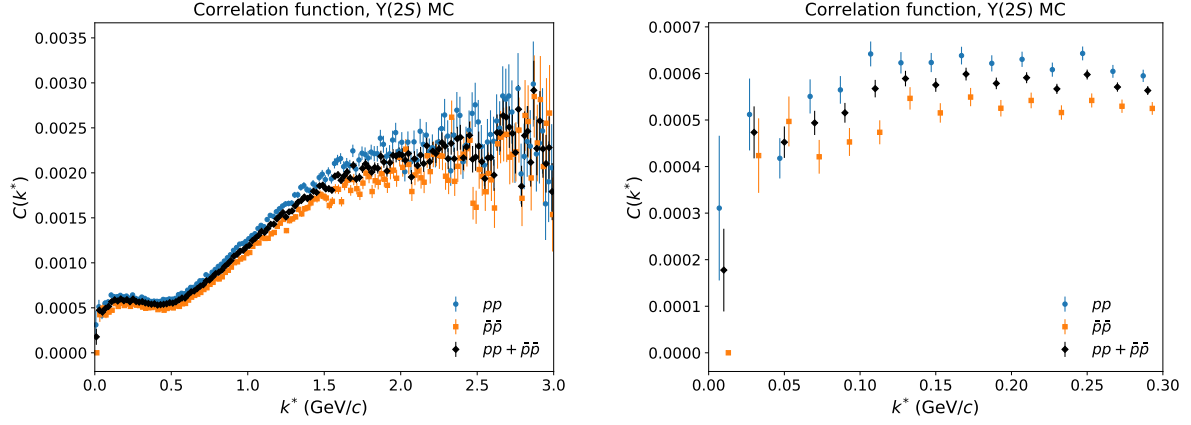
(a) Full k^* range up to 3 GeV/c.(b) Zoomed view for $k^* < 300$ MeV/c.

Figure 7.16: Correlation function measured on the reconstructed $\Upsilon(2S)$ Monte Carlo sample for true primary pp (blue dots), $\bar{p}\bar{p}$ (orange squares), and $pp + \bar{p}\bar{p}$ (black diamonds) pairs. For visualization purposes, the pp and $\bar{p}\bar{p}$ correlation functions are staggered by ± 3 MeV/c with respect to the bin centers.

contribution to the measured correlation function arises from energy–momentum conservation and acceptance effects, rather than from detector–induced artifacts. This validates the use of reconstructed Monte Carlo samples as a reliable reference for modeling the baseline in the data analysis.

Chapter 8

Systematic uncertainties

In this section, the estimation of all systematic uncertainties relevant to this analysis is described. In Sec. 8.1, the evaluation of the uncertainty associated with the choice of the N_{tracks} and R_2 binning is presented. In Secs. 8.2 and 8.3, the plan for the evaluation of uncertainties arising from the defined candidate selections are reported. In Sec. 8.4, the effect of tracking resolution on the single track's momentum and on pairs of tracks with small k^* is discussed. An additional contribution to the systematic uncertainty originates from the fit model; the strategy used to evaluate this effect is described in Sec. 8.5.

Although this chapter focuses on systematic effects, it is worth noting that the statistical uncertainties on the measured correlation function are expected to dominate over the systematic contributions in the kinematic region considered in this analysis.

8.1 Event shape binning

As discussed in Sec. 7.1.2, event mixing is performed within event classes defined on the (N_{tracks}, R_2) plane. The choice of this binning may affect the shape of the mixed-event distribution and, consequently, the resulting correlation function. To evaluate the associated systematic uncertainty, the correlation function $C(k^*)$ is recomputed using three alternative fixed (N_{tracks}, R_2) grids: (i) the **nominal** binning used throughout the analysis, (ii) a **finer** binning, and (iii) a **coarser** binning.

The nominal event-shape binning configuration is shown in Fig. 7.5. The finer and coarser configurations, used to evaluate the systematic variation, are shown in Figs. 8.1 and 8.2, respectively.

All other analysis settings are kept identical across the three configurations, including the datasets, event and track selections, k^* binning, and normalization procedure. The study is performed independently for the positively and negatively charged samples.

The effect of the different (N_{tracks}, R_2) grids on the correlation function is illustrated in Fig. 8.3, which shows the resulting $C(k^*)$ distributions obtained with the nominal, finer, and coarser binnings for positive and negative charges.

For each nominal k^* bin i , the deviations induced by the alternative binnings are defined as

$$\Delta_i^{\text{fine}} \equiv C_i^{\text{fine}} - C_i^{\text{nom}}, \quad \Delta_i^{\text{coarse}} \equiv C_i^{\text{coarse}} - C_i^{\text{nom}}.$$

The bin-wise systematic uncertainty is taken as the *envelope*

$$\delta_i^{\text{sys}} \equiv \max(|\Delta_i^{\text{fine}}|, |\Delta_i^{\text{coarse}}|), \quad (8.1)$$

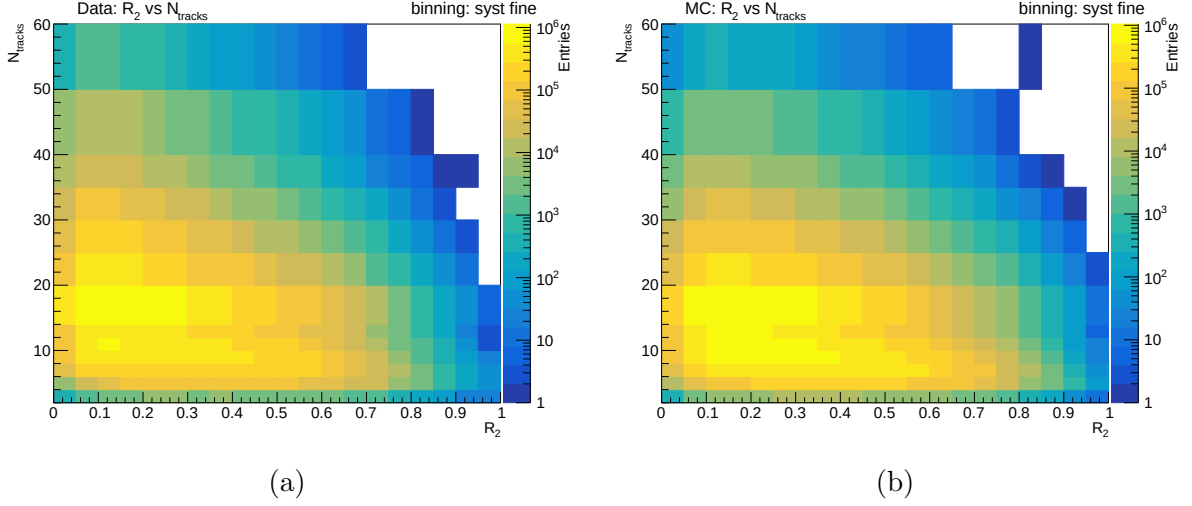


Figure 8.1: Fine binning configuration used in the evaluation of systematic uncertainties from event mixing in the $\Upsilon(2S)$ data (a) and MC (b) samples.

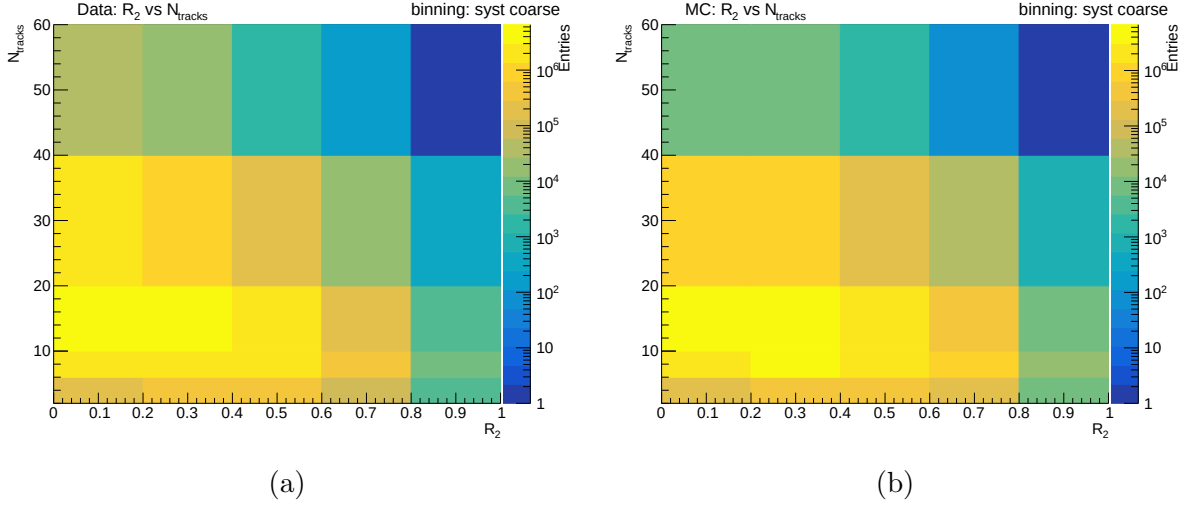


Figure 8.2: Coarse binning configuration used in the evaluation of systematic uncertainties from event mixing in the $\Upsilon(2S)$ data (a) and MC (b) samples.

and the corresponding relative uncertainty,

$$\delta_{i,\text{rel}}^{\text{sys}} \equiv 100 \times \frac{\delta_i^{\text{sys}}}{|C_i^{\text{nom}}|} \quad (\%), \quad (8.2)$$

is reported.

The resulting relative systematic uncertainty as a function of k^* is shown in Fig. 8.4, separately for positively and negatively charged samples.

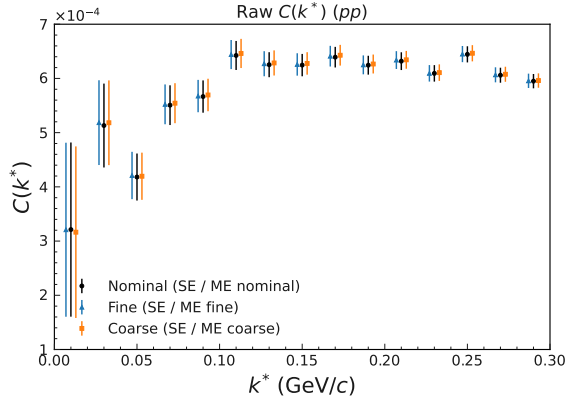
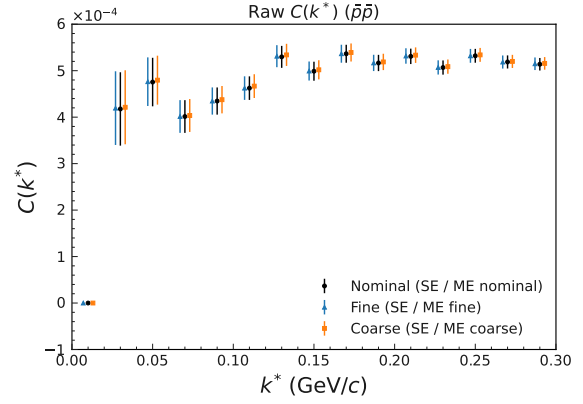
(a) $C(k^*)$ (positive charges)(b) $C(k^*)$ (negative charges)

Figure 8.3: Correlation functions obtained using the nominal, finer, and coarser (N_{tracks}, R_2) binning configurations. For visualization purposes, the $C(k^*)$ distributions corresponding to the different binnings are staggered horizontally by ± 3 MeV/ c . Error bars include statistical uncertainties only.

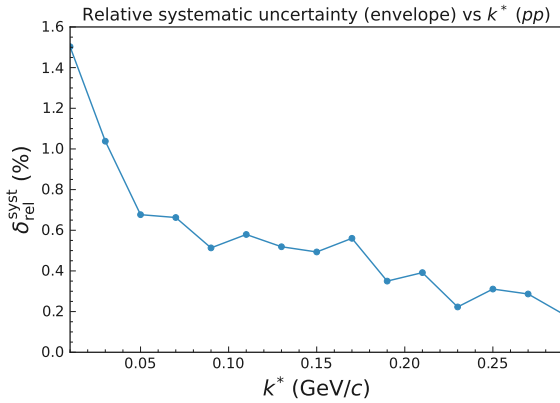
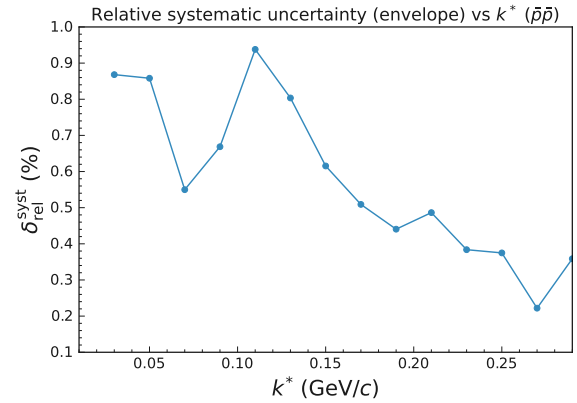
(a) Positive charges (pp)(b) Negative charges ($\bar{p}\bar{p}$)

Figure 8.4: Relative binning systematic uncertainty $\delta_{i,\text{rel}}^{\text{syst}}$ as a function of k^* , evaluated using the envelope definition in Eq. (8.1) and reported according to Eq. (8.2).

8.2 Selection on $N_{\text{hits}}^{\text{CDC}}$

The quality of charged-particle tracking is ensured by requiring a minimum number of hits in the central drift chamber (CDC). In the nominal analysis, tracks are required to satisfy $N_{\text{hits}}^{\text{CDC}} > 29$.

To evaluate the systematic uncertainty associated with this selection, the analysis is repeated on Monte Carlo samples using two alternative requirements, chosen to probe both a looser and a tighter tracking quality selection:

$$N_{\text{hits}}^{\text{CDC}} > 27 \quad \text{and} \quad N_{\text{hits}}^{\text{CDC}} > 31. \quad (8.3)$$

For each selection, the full correlation function $C(k^*)$ is reconstructed following the same procedure as in the nominal analysis. The systematic uncertainty is then estimated using the envelope method, defined bin-by-bin in k^* as the maximum deviation of the alternative correlation functions from the nominal one.

This approach captures possible effects arising from changes in tracking efficiency, momentum resolution, and track quality associated with the CDC hit requirement. The resulting envelope is assigned as the systematic uncertainty due to the $N_{\text{hits}}^{\text{CDC}}$ selection.

8.3 Momentum-dependent dE/dx selection

Prompt (anti)proton identification relies on a momentum-dependent selection based on the specific ionization energy loss, dE/dx . The selection boundaries are defined as smooth functions of the reconstructed momentum, obtained via a non-parametric smoothing procedure.

To estimate the systematic uncertainty associated with this choice, the selection limits are varied in a controlled manner by modifying the smoothing procedure. In practice, alternative sets of selection boundaries are constructed by slightly tightening and loosening the nominal limits within their statistical uncertainty. For each variation, the full analysis chain is repeated, including the construction of the correlation function.

The resulting deviations with respect to the nominal correlation function are interpreted as the systematic uncertainty associated with the dE/dx selection. This procedure accounts both for the finite statistics of the control samples used to define the selection and for possible residual modeling effects introduced by the smoothing.

8.4 Tracking momentum resolution

An additional source of systematic uncertainty arises from the finite tracking momentum resolution, which may affect the relative momentum k^* and, consequently, the shape of the correlation function at low k^* .

A dedicated study of the tracking momentum resolution has already been performed using the control decay $D^+ \rightarrow \pi^+\pi^+K^-$, as shown in detail in Sec. 8.4.2. In this channel, no significant discrepancies between data and Monte Carlo simulation were observed in the low- k^* region relevant for the present analysis.

To further validate this conclusion in topologies more closely related to the femtoscopic measurements, additional control channels are planned:

- $\Omega^- \rightarrow p\Lambda$ for the proton–proton analysis,

- $D \rightarrow \pi K_S^0$ for the $\pi\pi$ analysis.

These channels provide clean and well-understood final states, allowing a direct comparison of the reconstructed k^* distributions between data and simulation. Any observed deviation will be propagated to the correlation function and assigned as a systematic uncertainty related to the tracking momentum resolution.

If no statistically significant discrepancy is found, the associated systematic uncertainty will be considered negligible within the current precision.

8.4.1 MC study

Protons are characterised by a large specific energy loss dE/dx ; as a consequence, the reconstructed track momentum may be affected by biases introduced by the reconstruction algorithm. To assess this potential source of systematic uncertainty, momentum residuals are studied. These residuals are defined as $1 - p/p_{\text{MC}}$, where p_{MC} denotes the MC-matched momentum. Figure 8.5 shows the residuals as a function of p_{MC} . No significant reconstruction-induced bias is observed over the full momentum range.

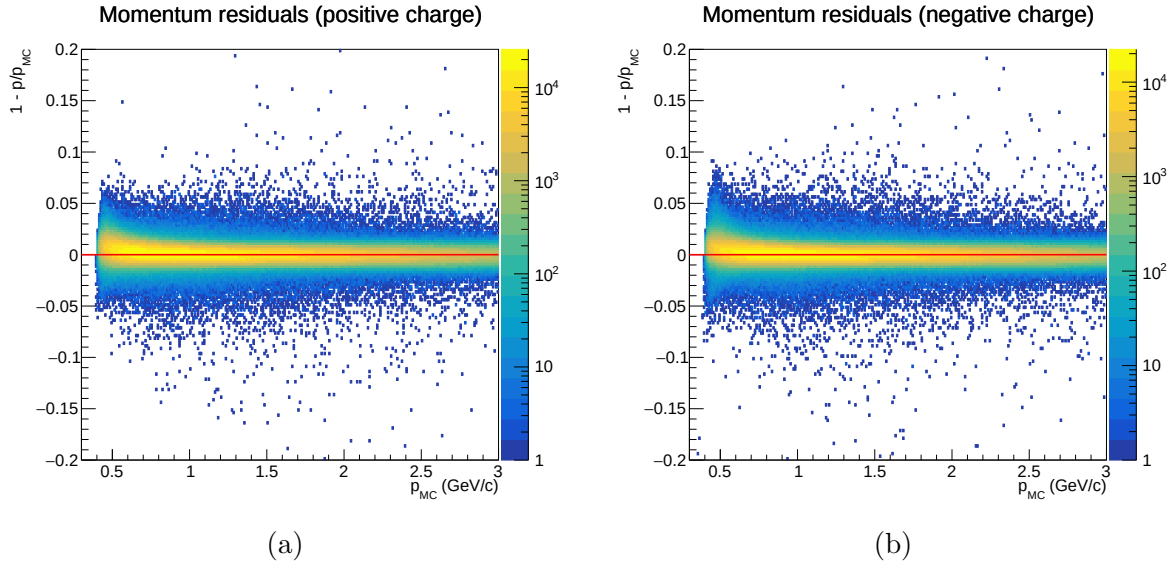


Figure 8.5: Momentum residuals for p (a) and $\bar{p}$ (b) candidates reconstructed in the $\Upsilon(2S)$ MC sample.

In addition to momentum effects at the single-track level, the reconstruction procedure introduces a smearing of the true k^* distributions, which directly enter the construction of the correlation function $C(k^*)$. This effect is quantified by comparing the MC-matched k^* to the reconstructed k^* in a two-dimensional histogram, shown in Fig. 8.6.

Figure 8.7 shows the k^* residuals integrated over the full k^* range. The full width at half maximum (FWHM) of the distributions is approximately 2 MeV/c.

A more precise determination of the k^* resolution is obtained by slicing the two-dimensional (k^*, k_{true}^*) distribution along the k_{true}^* axis and fitting the reconstructed k^* distributions with Gaussian functions. The resulting Gaussian widths, σ , are shown as a function of k_{true}^* in Fig. 8.8. The resolution is found to be approximately constant, with a value close to 2 MeV/c over the full k^* range. This is consistent with the FWHM of the k^* -integrated residual distributions.

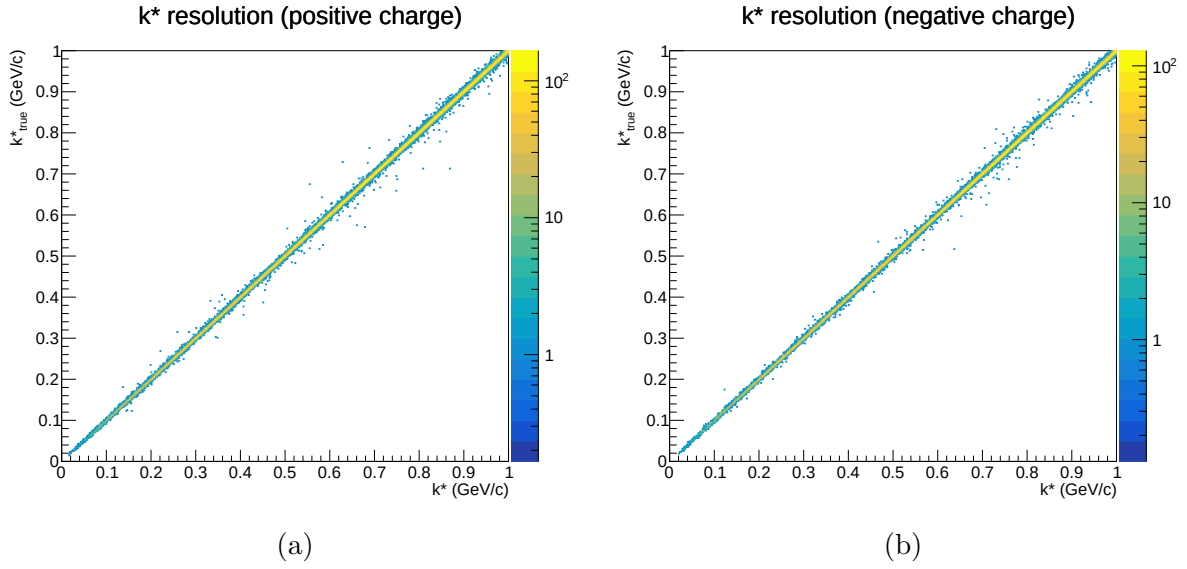


Figure 8.6: k^* resolution for pp (a) and $\bar{p}\bar{p}$ (b) candidate pairs reconstructed in the $\Upsilon(2S)$ MC sample.

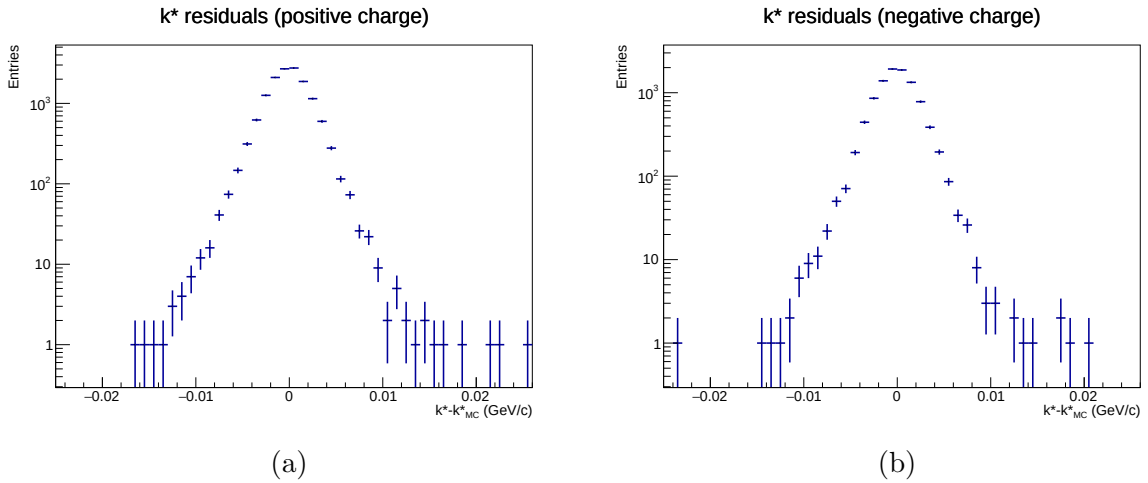


Figure 8.7: k^* residuals for pp (a) and $\bar{p}\bar{p}$ (b) candidate pairs reconstructed in the $\Upsilon(2S)$ MC sample.

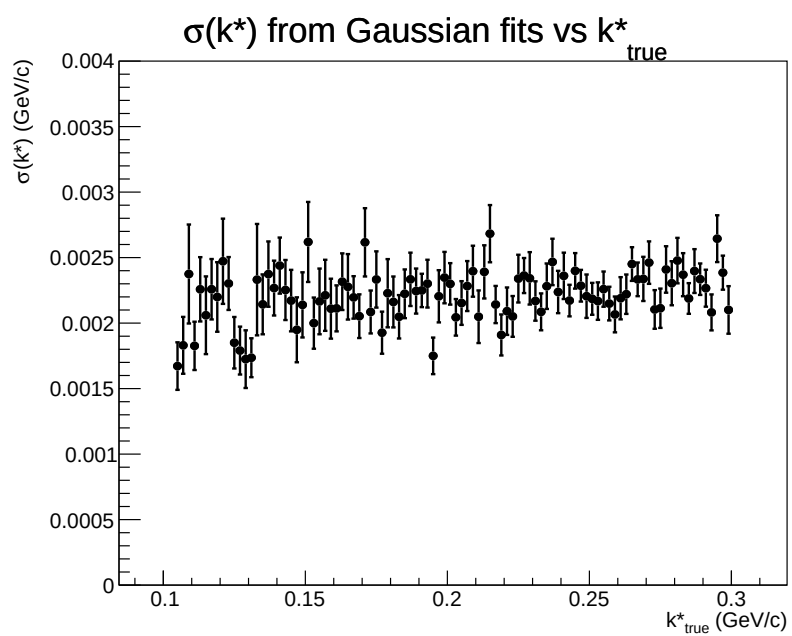


Figure 8.8: k^* resolution obtained from Gaussian fits to slices of the (k^*, k^*_{true}) distribution.

8.4.2 Momentum resolution using $D \rightarrow K\pi\pi$ decays

An attempt is made to measure momentum resolution effects in the reconstruction for pairs of tracks with the same charge and mass at low k^* . Since the particles have identical masses, this can be achieved by comparing the pair invariant-mass distribution at threshold in data and Monte Carlo. For this purpose, pairs of pions from $D^\pm \rightarrow K^\mp \pi^\pm \pi^\pm$ decays are analyzed. The candidate selection follows from the analysis "Precise measurement of the D^+ and D^0 lifetimes at Belle II" [95]. The decays $D^{*+} \rightarrow D^+ (K^- \pi^+ \pi^+) \pi^0$ are reconstructed using the following selection.

For $\pi^0 \rightarrow \gamma\gamma$ candidates, the requirements $p(\pi^0) > 200 \text{ MeV}/c$, $E(\gamma) > 30 \text{ MeV}$, and an opening angle of the diphoton $\alpha(\gamma\gamma) < 80.2^\circ$ are imposed. For charged pions, the requirements $p_T(\pi^\pm) > 300 \text{ MeV}/c$, $|dz| < 1 \text{ cm}$, $dr < 1 \text{ mm}$, $N_{\text{hits}}^{\text{SVD}} > 5$, $N_{\text{hits}}^{\text{CDC}} > 27$, and binary PID likelihoods satisfying $\text{Prob}(\pi/p) > 0.8$ and $\text{Prob}(\pi/K) > 0.8$ are applied. In addition, the two pion candidates are required to have the same charge. For kaon candidates, the requirements $p_T(K^\pm) > 30 \text{ MeV}/c$, $N_{\text{hits}}^{\text{CDC}} > 20$, $|dz| < 3 \text{ cm}$, $dr < 1 \text{ cm}$, and $\text{Prob}(K/\pi) > 0.8$ are imposed. The invariant mass of the D meson is required to satisfy $1.81 < M(D) < 1.91 \text{ GeV}/c^2$. For the D^* meson, the requirements $p_{\text{c.m.}} > 2.5 \text{ GeV}/c$, $1.9 < M(D^*) < 2.1 \text{ GeV}/c^2$, and a decay Q-value satisfying $Q < 20 \text{ MeV}$ are imposed. A vertex fit is applied to the reconstructed D mesons.

Signal events are first isolated by fitting the reconstructed D -meson mass distribution in data with an extended maximum-likelihood model composed of a Johnson S_U function for the signal component and a smooth combinatorial background described by a first-order Chebyshev polynomial. From this fit, per-event signal weights (s -weights) are extracted and used to statistically subtract background contributions in distributions uncorrelated with the fit variable. The fitted M_D distribution is shown in Fig. 8.9.

After applying the signal weights, the Dalitz distribution is constructed in data and compared with the signal Monte Carlo, as shown in Fig. 8.10.

The data-MC comparison on the Dalitz plane reveals significant shape differences, indicating that the simulated signal does not reproduce the two-dimensional structure observed in data. For this reason, no quantitative validation or tuning is attempted on the basis of this comparison. Nonetheless, candidates are selected in the region $1.6 < M_{K\pi}^2 < 1.8 \text{ GeV}^2/c^4$, and the $M_{\pi\pi}$ distribution measured on real data is compared to the one observed on MC. The comparison is shown in Fig. 8.11. For the selected data, data and MC show good agreement up to $M_{\pi\pi} \simeq 350 \text{ MeV}/c^2$. For higher masses, the s -Weighted data yields systematically fewer counts with respect to the simulated sample.

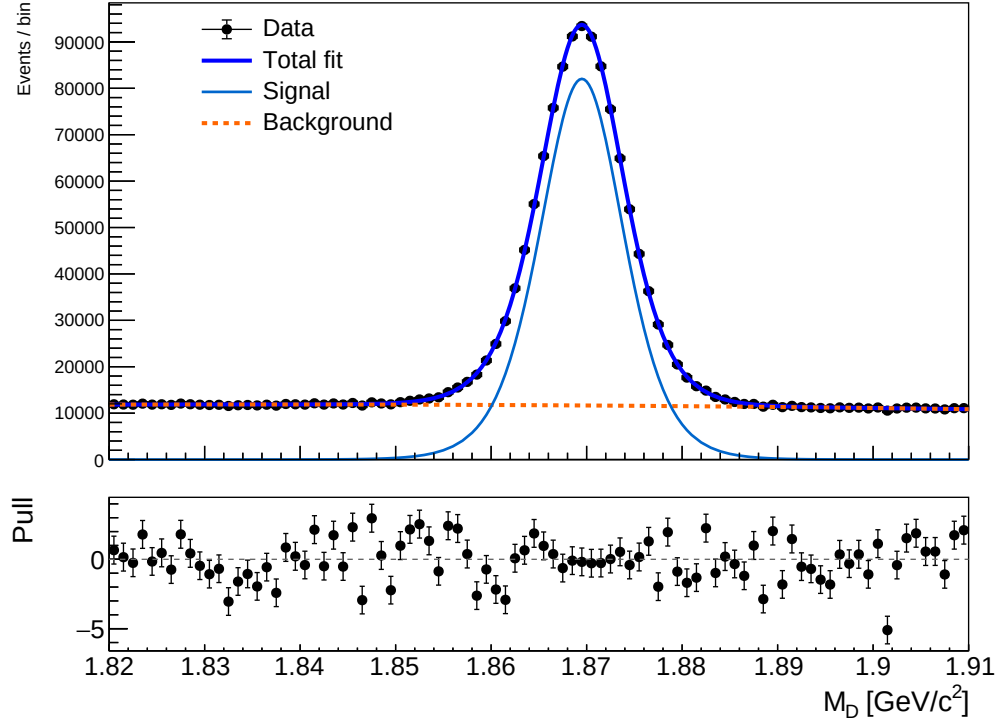


Figure 8.9: Fit to the D mass distribution in data with total model and components (signal: Johnson S_U ; background: linear). The fit provides the event-by-event signal s -weights used in the Dalitz comparison.

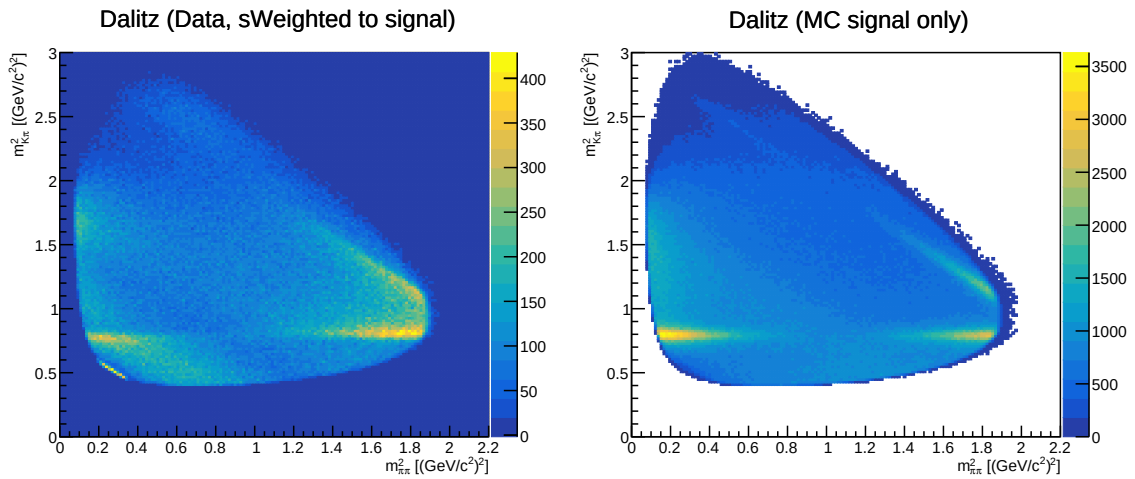


Figure 8.10: $D \rightarrow K\pi\pi$ Dalitz distributions: s -weighted data (left) versus signal-only MC (right). The observed discrepancies indicate that the Dalitz is not correctly simulated.

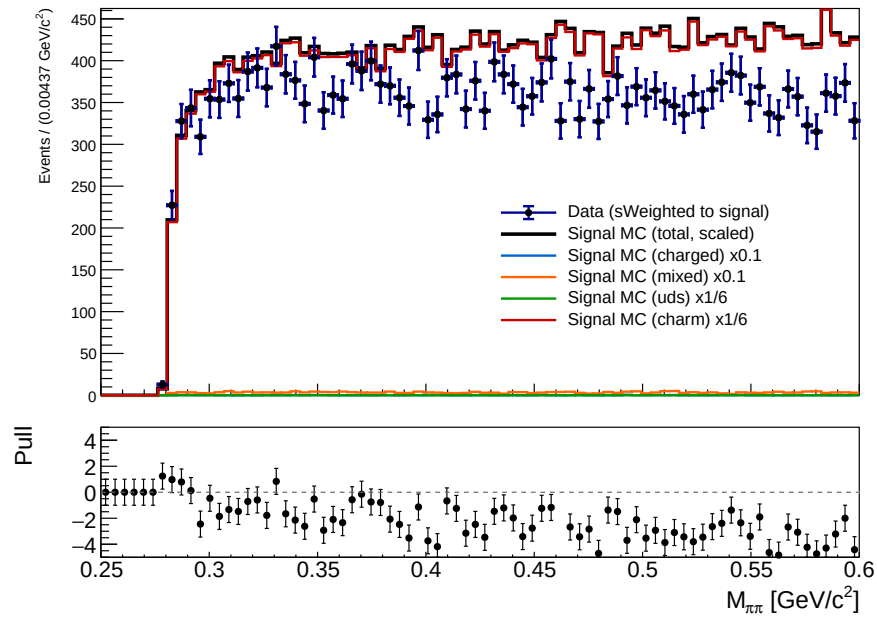


Figure 8.11: Data/MC comparison of the di-pion invariant mass distribution for selected $D \rightarrow K\pi\pi$ candidates.

8.5 Fit model

The extraction of prompt (anti)proton fractions relies on a multi-dimensional fit to the data, performed in adaptive bins of the reconstructed single-particle momentum. As such, the choice of fit model represents an intrinsic source of systematic uncertainty.

To assess the impact of the model assumptions, a series of alternative fits is performed by varying the functional forms used to describe both signal and background components. In particular, the following variations are considered:

- alternative parametrizations of the signal resolution function;
- variations of the background shapes used in the m_{TOF} and promptness observables;
- changes in the degree of smoothing applied to the non-parametric probability density functions.

For each model variation, the full fitting procedure is repeated in all momentum bins, and the resulting prompt fraction as a function of momentum is re-evaluated. The alternative prompt-fraction curves are subsequently propagated to the correlation function by reweighting the candidate pairs accordingly.

The systematic uncertainty associated with the fit model is estimated from the observed deviations of the correlation function with respect to the nominal result. Only fits satisfying strict quality criteria, based on the fit convergence status and covariance matrix quality, are retained in this evaluation.

This procedure ensures that the assigned uncertainty reflects genuine modeling ambiguities rather than statistical fluctuations or pathological fits.

Chapter 9

Results

9.1 Results on the data samples

After validating the analysis strategy on Monte Carlo samples and establishing a robust treatment of non-femtoscopic correlations, the correlation function is finally evaluated using real data. Results are presented first for the $\Upsilon(2S)$ data sample and subsequently for the $\Upsilon(1S)$ sample. The same event selection, pair construction, weighting procedure, and binning strategy adopted for Monte Carlo are applied consistently to data.

To account for the momentum dependence of the sample composition, each proton candidate is assigned a weight based on its reconstructed momentum. These weights encode the probability for a given track to correspond to a prompt proton and are propagated to the pair level in the construction of the same-event and mixed-event distributions. This ensures that the measured correlation function predominantly reflects prompt-prompt proton pairs, which are relevant for femtoscopic studies.

All correlation functions are normalized to unity in the relative-momentum interval

$$150 < k^* < 250 \text{ MeV}/c,$$

where the distributions are observed to be approximately flat and femtoscopic correlations are expected to be negligible. For a correlation function $C(k^*)$, the normalized distribution is defined as

$$C_{\text{norm}}(k^*) = \frac{C(k^*)}{\frac{1}{k_{\text{high}}^* - k_{\text{low}}^*} \int_{k_{\text{low}}^*}^{k_{\text{high}}^*} C(k^{*'}) dk^{*'}}, \quad (9.1)$$

with $k_{\text{low}}^* = 150 \text{ MeV}/c$ and $k_{\text{high}}^* = 250 \text{ MeV}/c$. This normalization is applied consistently to both data and Monte Carlo prior to any comparison or ratio.

To avoid experimenter bias, a blind analysis strategy is adopted. The interval

$$0 < k^* < 100 \text{ MeV}/c$$

is defined as the *signal region*, where final-state interaction effects are expected to produce a pronounced enhancement of the correlation function. Data points in this region are masked in all figures. Monte Carlo distributions are shown over the full range for reference. All statements in the following therefore refer exclusively to the unblinded region $k^* > 100 \text{ MeV}/c$.

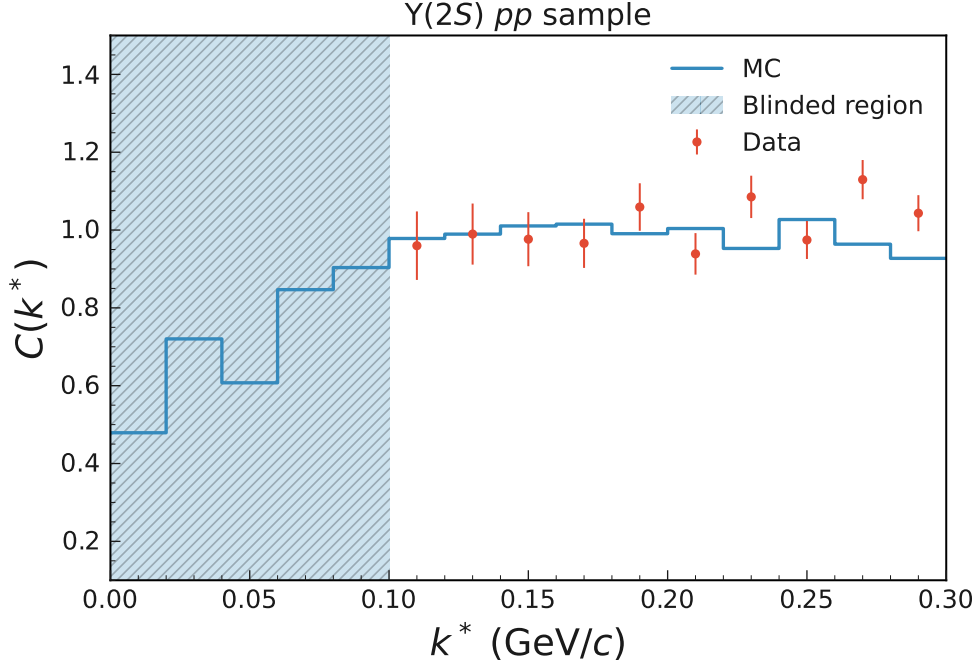


Figure 9.1: Normalized correlation function for pp pairs in the $\Upsilon(2S)$ data sample compared to fully reconstructed Monte Carlo. The region $0 < k^* < 100$ MeV/ c is blinded in data. Both distributions are normalized in $150 < k^* < 250$ MeV/ c .

Residual non-femtoscopic distortions are corrected using the double-ratio technique. The corrected correlation function is defined as

$$C_{\text{corr}}(k^*) = \frac{C_{\text{data}}(k^*)}{C_{\text{MC}}(k^*)}, \quad (9.2)$$

where both numerator and denominator are constructed with identical analysis settings.

9.1.1 Results in the $\Upsilon(2S)$ data sample

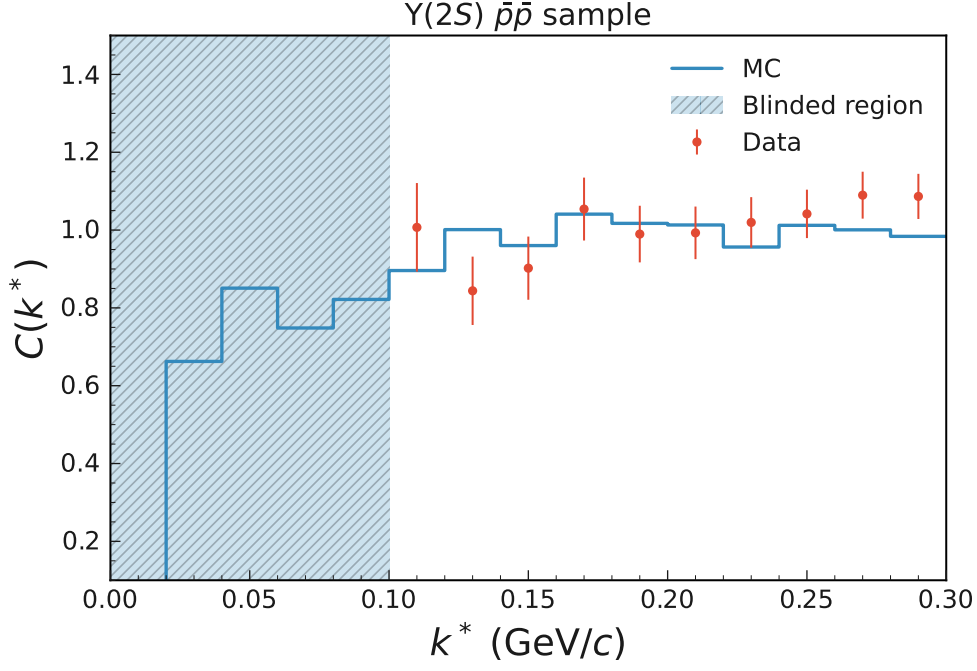
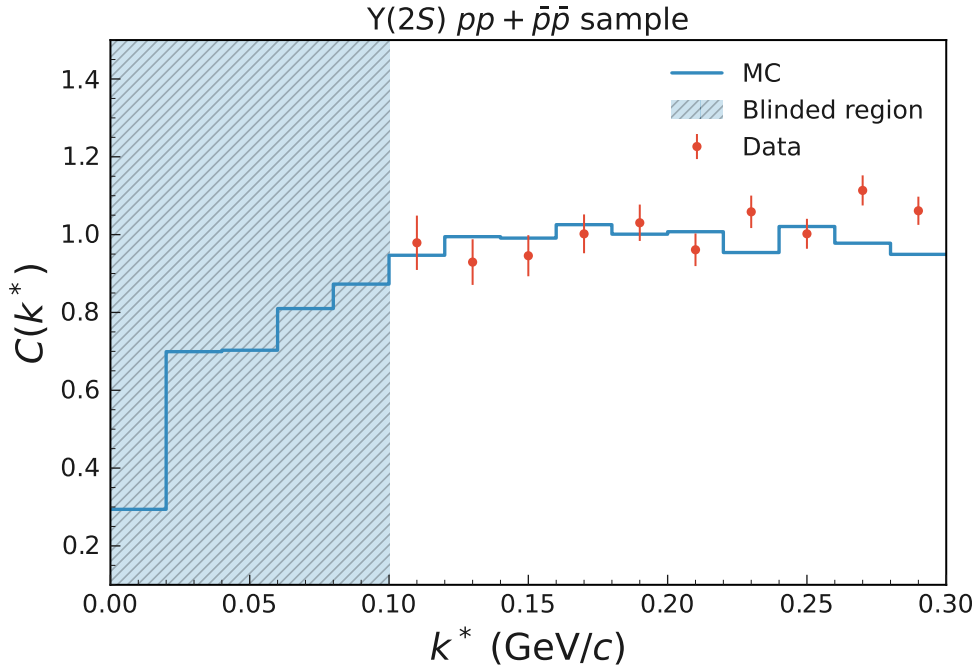
Data–Monte Carlo comparison

The comparison between data and fully reconstructed Monte Carlo for the $\Upsilon(2S)$ sample is shown separately for pp , $\bar{p}\bar{p}$, and their sum.

Figure 9.1 presents the normalized correlation function for pp pairs. In the unblinded region ($k^* > 100$ MeV/ c), the data distribution exhibits a smooth behavior and is compatible with the Monte Carlo expectation within statistical uncertainties.

The corresponding comparison for $\bar{p}\bar{p}$ pairs is shown in Fig. 9.2. The level of agreement is similar to the pp case, indicating no significant charge-dependent mismodelling.

Finally, the sum of pp and $\bar{p}\bar{p}$ pairs is displayed in Fig. 9.3. The combined sample benefits from improved statistical precision and shows no visible deviation from the Monte Carlo reference in the unblinded region.

Figure 9.2: Same as Fig. 9.1 for $\bar{p}p$ pairs.Figure 9.3: Normalized correlation function for the combined $pp + \bar{p}p$ sample in $\Upsilon(2S)$ decays.

Double ratio

The corresponding double ratios for pp , $\bar{p}p$, and their sum are shown together in Fig. 9.4. Presenting all three in a single plot provides a compact diagnostic of possible residual

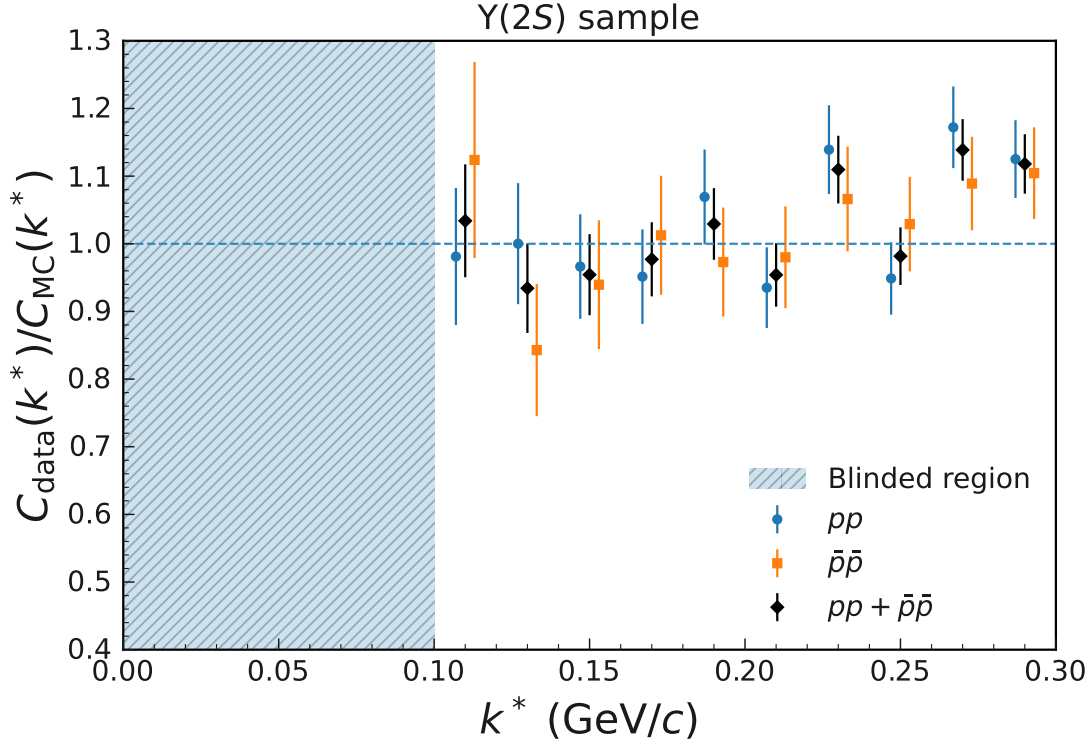


Figure 9.4: Double ratio $C_{\text{data}}(k^*)/C_{\text{MC}}(k^*)$ for the $\Upsilon(2S)$ sample. Results for pp , $\bar{p}\bar{p}$, and their sum are overlaid. The signal region is blinded.

mismodelling.

In the unblinded region, the double ratios are overall compatible with unity within statistical uncertainties. However, for $k^* \gtrsim 250 \text{ MeV}/c$ a slight excess with respect to unity is observed in data for all pair categories (pp , $\bar{p}\bar{p}$, and their sum). Although the effect remains small compared to the statistical uncertainties, its systematic presence across samples suggests a residual long-range distortion.

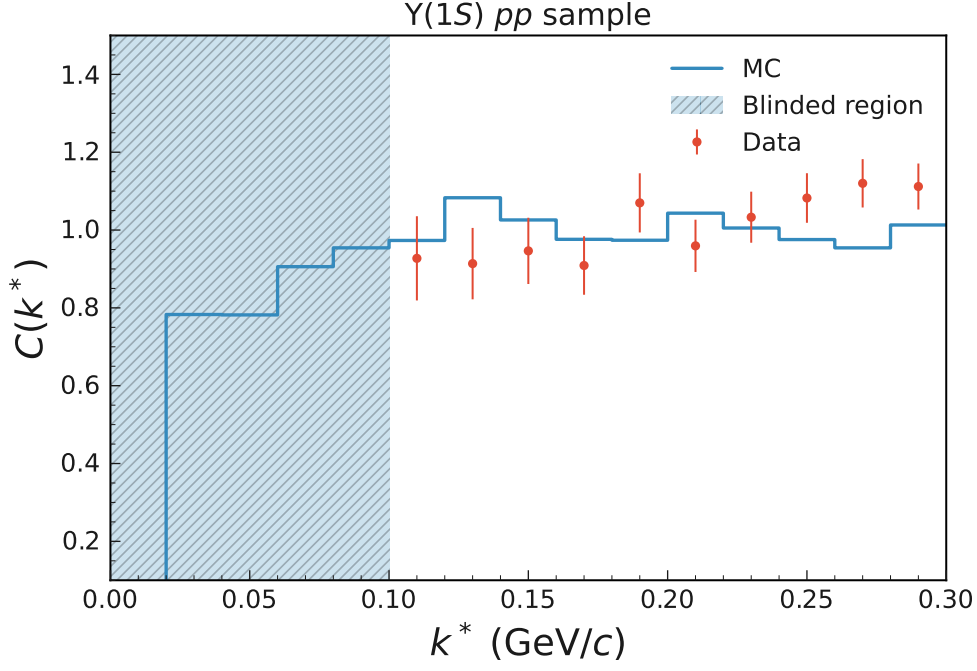
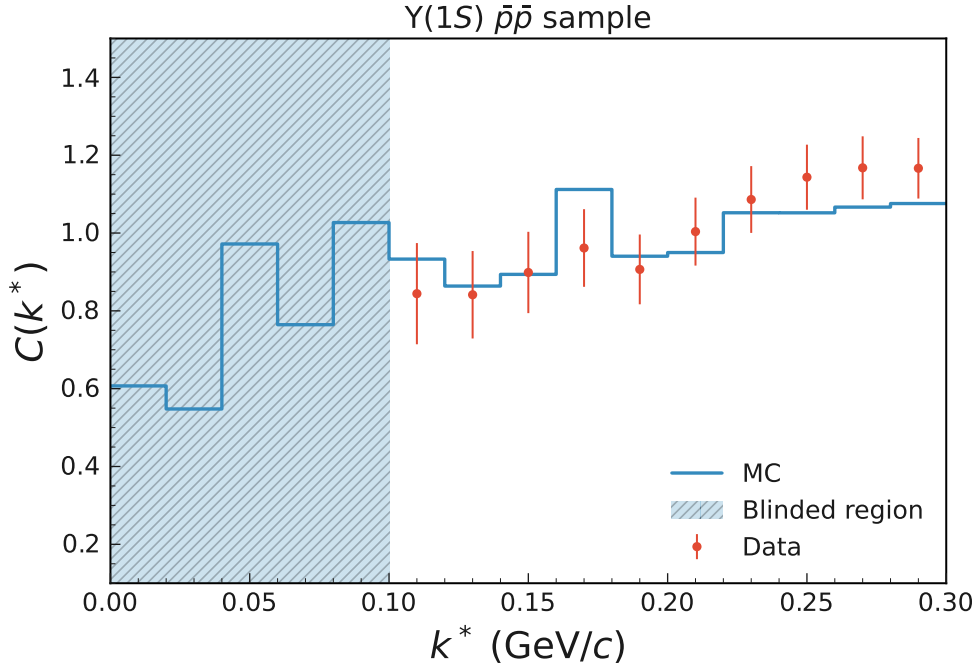
A plausible interpretation is an imperfect modelling of energy-momentum conservation induced correlations (EMCIC) in the Monte Carlo simulation. As discussed in Sec. 7.2.3, such residual effects can be effectively absorbed by introducing a linear baseline term multiplying the femtoscopic correlation function. The inclusion of this baseline component will therefore be considered in the quantitative fits after unblinding.

9.1.2 Results in the $\Upsilon(1S)$ data sample

The same analysis procedure is applied to the $\Upsilon(1S)$ data sample.

Data-Monte Carlo comparison

Figures 9.5–9.7 show the comparison between data and Monte Carlo for pp , $\bar{p}\bar{p}$, and their sum, respectively. The qualitative behavior closely resembles that observed in the $\Upsilon(2S)$ sample.

Figure 9.5: Normalized pp correlation function in the $\Upsilon(1S)$ data sample.Figure 9.6: Normalized $\bar{p}\bar{p}$ correlation function in the $\Upsilon(1S)$ data sample.

Double ratio

The double ratios for the $\Upsilon(1S)$ sample are shown in Fig. 9.8, where the results for pp , $\bar{p}\bar{p}$, and their sum are overlaid.

In the unblinded region ($k^* > 100$ MeV/c), the distributions are overall compatible

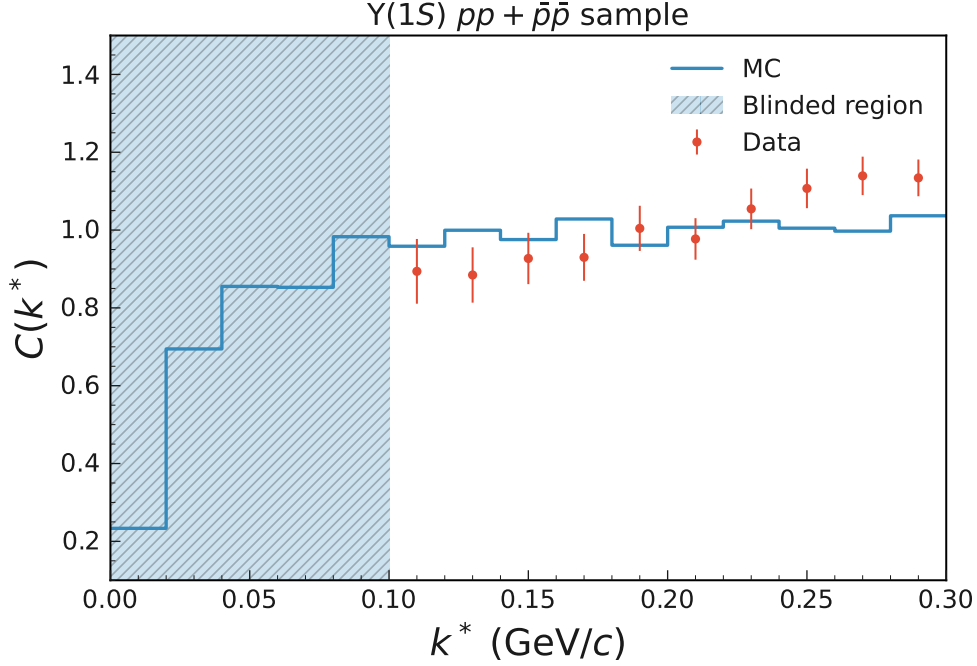


Figure 9.7: Normalized $pp + \bar{p}\bar{p}$ correlation function in the $\Upsilon(1S)$ data sample.

with unity within statistical uncertainties. Similarly to the $\Upsilon(2S)$ case, a mild upward trend is observed for $k^* \gtrsim 250$ MeV/ c in all pair categories. Although statistically limited, the consistency of this feature across charge combinations and across both data samples suggests a small residual long-range effect.

Such a behavior is compatible with a slight mismodelling of energy-momentum conservation induced correlations (EMCIC) in the simulation. As discussed in Sec. 7.2.3, these residual distortions can be accounted for by multiplying the femtoscopic correlation function by a linear baseline term. This component will therefore be included in the quantitative fits performed after unblinding.

Also in this case, the double ratios remain compatible with unity for $k^* > 100$ MeV/ c .

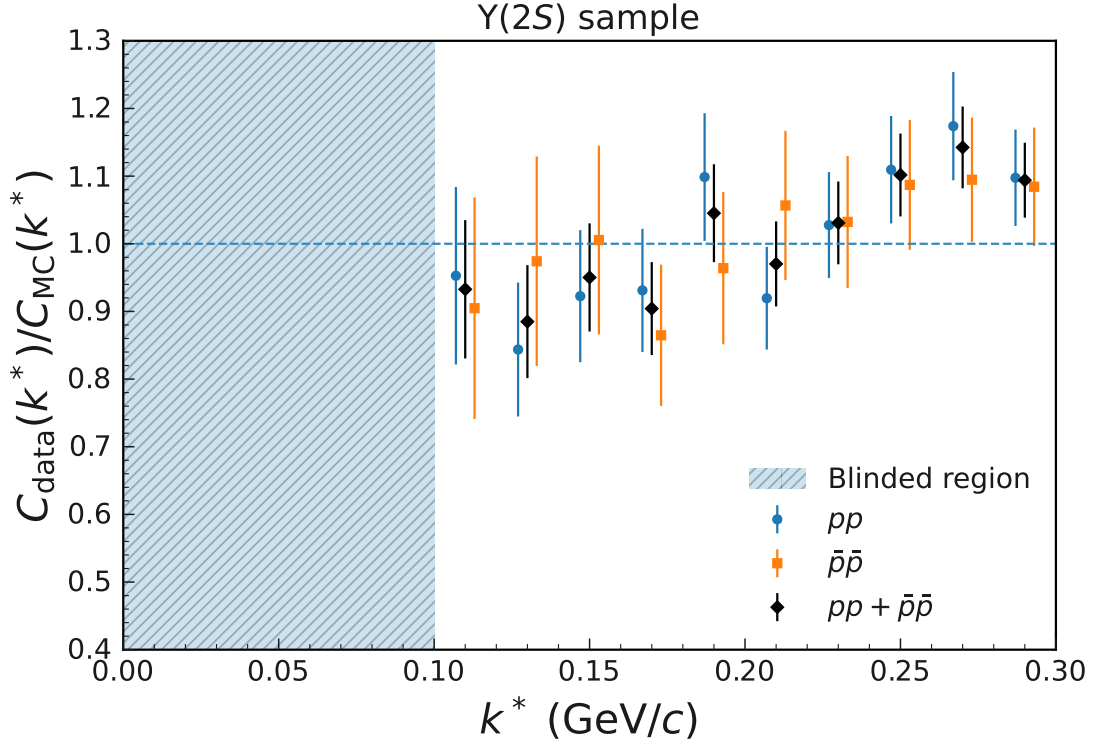


Figure 9.8: Double ratio $C_{\text{data}}(k^*)/C_{\text{MC}}(k^*)$ for the $\Upsilon(1S)$ sample. The signal region is blinded.

9.2 Discussion of the results

The discussion refers exclusively to the unblinded region $k^* > 100 \text{ MeV}/c$.

For both $\Upsilon(2S)$ and $\Upsilon(1S)$ data samples, the normalized correlation functions are overall consistent with Monte Carlo expectations within statistical uncertainties. No significant charge dependence is observed between pp and $\bar{p}\bar{p}$ pairs.

The double-ratio correction substantially suppresses long-range distortions, confirming that acceptance effects and the dominant components of energy-momentum conservation induced correlations are well reproduced by the simulation. However, a mild and systematic excess with respect to unity is visible at large relative momentum ($k^* \gtrsim 250 \text{ MeV}/c$) in both data samples and for all pair categories. Although limited in statistical significance, the persistence of this trend suggests a small residual mismodelling of non-femtoscopic correlations.

Such a behavior is compatible with an imperfect description of EMCIC effects in the Monte Carlo. As discussed in Sec. 7.2.3, these residual distortions can be effectively absorbed by introducing a linear baseline term multiplying the femtoscopic correlation function. This component will therefore be included in the final fits.

A quantitative interpretation of the femtoscopic signal in the blinded region will be performed only after completion of all systematic studies and the unblinding procedure. After unblinding, the measured correlation functions will be analyzed within the CATS framework to extract the effective size of the proton emission region in $\Upsilon(1S)$ and $\Upsilon(2S)$ decays.

9.3 The CATS framework: a toy study

Following the validation of the data–Monte Carlo agreement in the unblinded region, the quantitative extraction of femtoscopic source parameters requires a controlled and internally consistent fitting framework. For this purpose, the CATS (Correlation Analysis Tool using the Schrödinger equation) package is employed. Before applying it to real data, the full procedure is validated through a dedicated toy study.

CATS computes two–particle correlation functions from first principles. Given (i) an emission source function $S(\vec{r})$ describing the relative separation of the emitted particles and (ii) a local interaction potential $V(r)$, the stationary Schrödinger equation is solved to obtain the outgoing relative wave function $\Psi^{(+)}(\vec{r}, \vec{k}^*)$. The correlation function is then evaluated through the Koonin–Pratt integral,

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

where k^* is the relative momentum in the pair rest frame. This approach consistently incorporates quantum statistics and final–state interactions (Coulomb and strong), and allows flexible choices of source parameterizations and hadron–hadron potentials.

9.3.1 Toy generation and fit of the pp correlation function

Pseudo–data for the proton–proton correlation function $C(k^*)$ are generated using a known underlying model. The “truth” correlation includes quantum statistics, the Coulomb interaction, and the Argonne $v18$ NN strong potential, evaluated for a spherically symmetric Gaussian emission source with width

$$r_0 = 0.9 \text{ fm}.$$

On a fine k^* grid this yields a smooth theoretical spectrum $C_{\text{true}}(k^*)$.

Detector effects are incorporated through a resolution response mapping true to measured relative momentum, represented by a normalized matrix

$$R_{ij} = P(k_{\text{meas}}^* \in i \mid k_{\text{true}}^* \in j), \quad \sum_i R_{ij} = 1.$$

In practice, this is achieved via a normalized two-dimensional histogram filled with the values of reconstructed and generated k^* distributions. The smeared spectrum is obtained via

$$C_{\text{smeared}}(k_i^*) = \sum_j R_{ij} C_{\text{true}}(k_j^*).$$

To emulate finite correlation strength and normalization effects, the target spectrum used to generate pseudo–data is

$$C_{\text{tgt}}(k^*) = N \left[1 + \lambda (C_{\text{smeared}}(k^*) - 1) \right],$$

where λ represents an effective pair purity and N is a global normalization factor (unity in the nominal configuration).

Pseudo–data are generated using Poisson fluctuations in event–mixing–like counting:

$$C(k^*) = \frac{A}{B}, \quad \delta C(k^*) = C(k^*) \sqrt{\frac{1}{A} + \frac{1}{B}}.$$

The same CATS model is then fitted to the pseudo-data. For a given source size r_0 ,

$$C_{\text{true}}(k^*; r_0)$$

is computed and folded with the same response matrix to obtain

$$C_{\text{smeared}}(k_i^*; r_0) = \sum_j R_{ij} C_{\text{true}}(k_j^*; r_0).$$

The fit function is

$$C_{\text{fit}}(k^*; r_0, \lambda, N) = N \left[1 + \lambda (C_{\text{smeared}}(k^*; r_0) - 1) \right],$$

with (r_0, λ, N) as free parameters. When required, a small linear baseline term can be included, consistently with the EMCIC treatment described in Sec. 7.2.3.

The fit is performed using a standard χ^2 estimator,

$$\chi^2(r_0, \lambda, N) = \sum_i \frac{[C_i - C_{\text{fit}}(k_i^*; r_0, \lambda, N)]^2}{\delta C_i^2}.$$

Figure 9.9 shows the pseudo-data together with the best-fit curve. The extracted parameters are consistent with the input values, demonstrating closure of the full procedure and validating the resolution folding and parameter extraction strategy prior to unblinding real data.

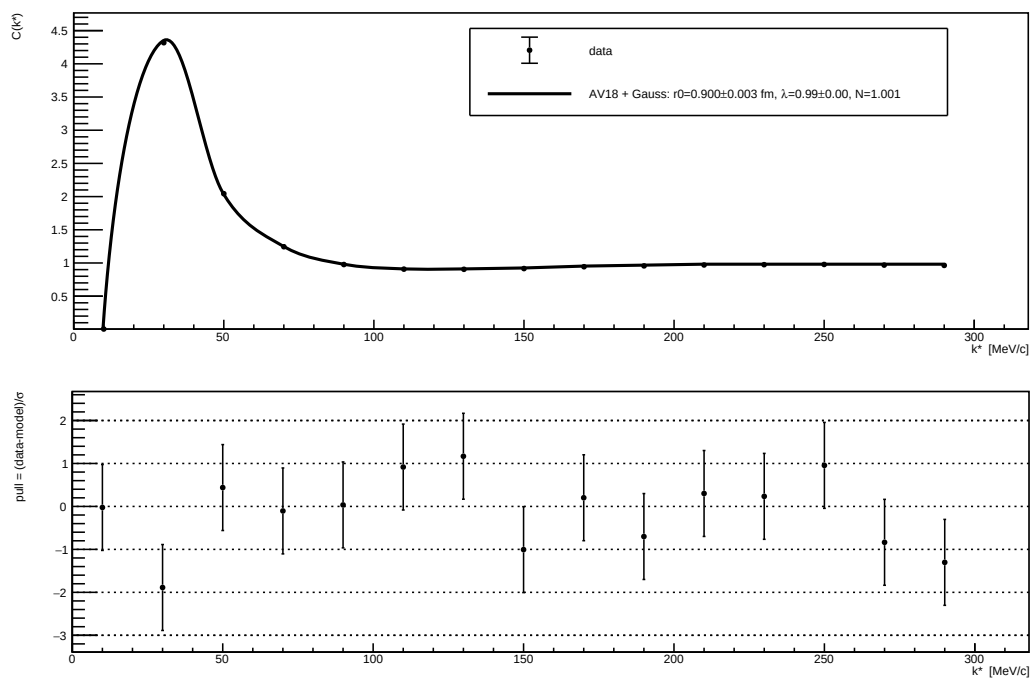


Figure 9.9: Fit of the CATS model to toy pp data. The upper panel shows the pseudo-data and the best-fit function, while the lower panel displays the fit pulls.

Chapter 10

Summary and conclusions

This work has established the first femtoscopic analysis performed at Belle, and more generally the first application of two-particle femtoscopy to hadrons produced in $\Upsilon(1S)$ and $\Upsilon(2S)$ decays. By exploiting the clean e^+e^- environment and quarkonium decays, a new laboratory has been set up to study baryon–baryon correlations at low relative momentum, complementary to measurements in hadronic and heavy-ion collisions.

A dedicated event-mixing procedure was developed in order to construct a reference distribution for the correlation function $C(k^*)$. In contrast to standard approaches based solely on multiplicity (and/or vertex constraints), mixing was performed using event-shape observables, ensuring that mixed events share similar global topology and kinematics. This choice proved essential to control the non-femtoscopic baseline in Υ decays, where global constraints and the limited final-state multiplicity can otherwise produce spurious structures.

The analysis strategy was validated extensively on Monte Carlo samples. In this stage, three initially unexpected features of the reconstructed correlation function were identified and understood in terms of detector and global-event effects: (i) a low- k^* inefficiency driven by track merging at small opening angles, (ii) a long-range modulation induced by energy–momentum conservation (EMCIC), and (iii) a residual, momentum-dependent distortion associated with detector acceptance. Among these, EMCIC provides the leading non-femtoscopic contribution over a broad k^* range, while acceptance effects are present but moderate and can be treated as a controlled systematic effect. The explicit identification of these three ingredients enabled a consistent description of the Monte Carlo correlation functions and informed the correction and fitting strategy adopted for data.

A direct comparison of $C(k^*)$ between data and simulation was then performed. The agreement is found to be good in the region 100–300 MeV/ c , where the correlation function is flat within statistical uncertainties and detector-related distortions are under control. This observation supports the reliability of the event-mixing reference, and the modeling of non-femtoscopic correlations in the simulation.

Special care is required at low k^* , where the track-merging inefficiency produces a pronounced dip. To validate the correction strategy, the control channel $D^+ \rightarrow K^- \pi^+ \pi^+$ was used (Sec. 8.4.2). This decay provides a clean and well-understood source of nearby tracks, enabling a direct probe of tracking performance and momentum-resolution effects in the regime most relevant for femtoscopy. The study demonstrates that the double-ratio correction of $C(k^*)$ can be safely employed to correct the low- k^* inefficiency without introducing artificial structures, thereby preserving the physical sensitivity of the mea-

surement.

On the interpretation side, a full fitting machinery was built based on the CATS framework, which solves the Schrödinger equation for a given emission source and interaction potential and evaluates the correlation function through the Koonin–Pratt formalism. The complete chain—including source parametrization, baseline treatment, and parameter estimation—was validated with toy-model correlation functions, demonstrating numerical stability and correct parameter recovery. This provides a solid foundation for extracting quantitative constraints on the emission source size and on the proton–proton interaction from the measured correlation functions in Υ decays.

To ensure the robustness of the analysis procedure, the signal region at low relative momentum ($k^* < 100$ MeV/ c) is kept blinded during the internal review within the Belle II collaboration; only after this step the region will be unblinded and the fit performed to extract the emission source size.

In summary, this thesis delivers the essential experimental and methodological components required to perform femtoscopy in Υ decays at Belle: a robust event-mixing strategy based on event-shape matching, a data-driven validation of the low- k^* correction using a control channel, a clear understanding of the non-femtoscopic structures driven by EMCIC and acceptance effects, and a fully tested fitting infrastructure grounded in wave-function calculations. These achievements open the path to a quantitative femtoscopic study of baryon–baryon correlations in a previously unexplored environment, and provide a blueprint for future measurements with larger data samples and extended particle-species coverage.

Bibliography

- [1] R. Hanbury Brown and R. Q. Twiss. A Test of a new type of stellar interferometer on Sirius. *Nature*, 178:1046–1048, 1956.
- [2] R. Hanbury Brown and R. Q. Twiss. Correlation between Photons in two Coherent Beams of Light. *Nature*, 177:27–29, 1956.
- [3] Gerson Goldhaber, Sulamith Goldhaber, Won-Yong Lee, and Abraham Pais. Influence of Bose-Einstein statistics on the anti-proton proton annihilation process. *Phys. Rev.*, 120:300–312, 1960.
- [4] Michael Annan Lisa, Scott Pratt, Ron Soltz, and Urs Wiedemann. Femtoscopy in relativistic heavy ion collisions: Two decades of progress. *Ann. Rev. Nucl. Part. Sci.*, 55:357–402, 2005.
- [5] R. Lednický and V.L. Lyuboshits. Final state interaction effect on pairing correlations between particles with small relative momenta. *Sov. J. Nucl. Phys.*, 35:770, 1982. *Yad. Fiz.* 35 (1982) 1316.
- [6] Shreyasi Acharya et al. p-p, p- Λ and Λ - Λ correlations studied via femtoscopy in pp reactions at $\sqrt{s} = 7$ TeV. *Phys. Rev. C*, 99(2):024001, 2019.
- [7] Shreyasi Acharya et al. Investigating the p- $\pi^\pm$ and p-p- $\pi^\pm$ dynamics with femtoscopy in pp collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. A*, 61(8):194, 2025.
- [8] E. Eichten, K. Gottfried, T. Kinoshita, K. D. Lane, and T. M. Yan. Charmonium: The Model. *Phys. Rev. D*, 17:3090, 1978.
- [9] S. Godfrey and N. Isgur. Mesons in a Relativized Quark Model with Chromodynamics. *Phys. Rev. D*, 32:189, 1985.
- [10] S. Weinberg. Phenomenological Lagrangians. *Physica A*, 96:327–340, 1979.
- [11] E. Epelbaum, H. W. Hammer, and U. G. Meissner. Modern Theory of Nuclear Forces. *Rev. Mod. Phys.*, 81:1773–1825, 2009.
- [12] J. A. Oller, E. Oset, and J. R. Pelaez. Meson-Meson Interaction in a Non-Perturbative Chiral Approach. *Phys. Rev. D*, 59:074001, 1999.
- [13] T. Hyodo and D. Jido. The nature of the $\Lambda(1405)$ resonance in chiral dynamics. *Prog. Part. Nucl. Phys.*, 67:55–98, 2012.
- [14] K. G. Wilson. Confinement of Quarks. *Phys. Rev. D*, 10:2445–2459, 1974.

- [15] S. Aoki et al. FLAG Review 2019: Flavour Lattice Averaging Group (FLAG). *Eur. Phys. J. C*, 80:113, 2020.
- [16] J. S. Ball, A. Scotti, and D. Y. Wong. One-boson-exchange model of NN and $n\bar{N}$ interaction. *Phys. Rev.*, 142:1000–1012, Feb 1966.
- [17] R. B. Wiringa, V. G. J. Stoks, and R. Schiavilla. Accurate nucleon-nucleon potential with charge-independence breaking. *Phys. Rev. C*, 51:38–51, 1995.
- [18] G. Ecker. The pion-nucleon interaction as an effective field theory. *Chin. J. Phys.*, 32:1303–1316, 1994.
- [19] E. Epelbaum et al. High-precision nuclear forces from chiral eft: State-of-the-art, challenges, and outlook. *Front. Phys.*, 2020. Review of chiral EFT applied to nuclear forces.
- [20] Sean Fleming and Thomas Mehen. Hadronic decays of the $X(3872)$ to χ_{cJ} in effective field theory. *Phys. Rev. D*, 78(094019), 2008.
- [21] J. A. Oller, E. Oset, and J. R. Peláez. Non-perturbative approach to effective chiral lagrangians and meson interactions. *Phys. Rev. Lett.*, 80:3452–3455, 1998.
- [22] E. Oset and A. Ramos. Non-perturbative chiral approach to s-wave $\bar{K}n$ interactions. *Nucl. Phys. A*, 635:99–120, 1998.
- [23] Yan-Rui Liu, Hua-Xing Chen, Wei Chen, Atsushi Hosaka, Atsushi Hosaka, et al. Pentaquark and tetraquark states. *Prog. Part. Nucl. Phys.*, 107:237–320, 2019.
- [24] Meng-Lin Du, Feng-Kun Guo, Christoph Hanhart, and Ulf-G. Meißner. Role of left-hand cuts in light-quark hadron spectroscopy. *Phys. Rev. Lett.*, 126:192001, 2021.
- [25] V. Baru, C. Hanhart, Y. S. Kalashnikova, A. E. Kudryavtsev, and A. V. Nefediev. Interplay of quark and meson degrees of freedom in a near-threshold resonance: multi-channel case. *Eur. Phys. J. A*, 44:93–103, 2010.
- [26] Kenneth G. Wilson. Confinement of quarks. *Phys. Rev. D*, 10:2445–2459, Oct 1974.
- [27] S. R. Beane, W. Detmold, K. Orginos, and M. J. Savage. Nuclear Physics from Lattice QCD. *Prog. Part. Nucl. Phys.*, 66:1–40, 2011.
- [28] Raul A. Briceño, Jozef J. Dudek, and Ross D. Young. Scattering processes and resonances from lattice QCD. *Rev. Mod. Phys.*, 90(2):025001, 2018.
- [29] Bo Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand. Parton Fragmentation and String Dynamics. *Phys. Rept.*, 97:31–145, 1983.
- [30] Bo Andersson, G. Gustafson, and B. Soderberg. A General Model for Jet Fragmentation. *Z. Phys. C*, 20:317, 1983.
- [31] Silvia Ferreres-Solé and Torbjörn Sjöstrand. The space–time structure of hadronization in the Lund model. *Eur. Phys. J. C*, 78(11):983, 2018.
- [32] Silvia Ferreres-Solé and Torbjörn Sjöstrand. The space–time structure of hadronization in the lund model. *Eur. Phys. J. C*, 78:983, 2018.

- [33] Christian Bierlich, A. Chakraborty, N. Desai, L. Gellersen, I. Helenius, P. Ilten, A. Karneyeu, L. Lönnblad, S. Mrenna, S. Prestel, C. O. Rasmussen, P. Richardson, A. Ryd, M. H. Seymour, T. Sjöstrand, P. Z. Skands, R. Verheyen, J. Webster, and L. Witzel. A comprehensive guide to the physics and usage of pythia 8.3. *SciPost Phys. Codebases*, 8, 2022.
- [34] B. R. Webber. A QCD Model for Jet Fragmentation Including Soft Gluon Interference. *Nucl. Phys. B*, 238:492–528, 1984.
- [35] Stefan Gieseke, Patrick Kirchgaßer, and Simon Plätzer. Baryon production from cluster hadronisation. *Eur. Phys. J. C*, 78(2):99, 2018.
- [36] Enrico Fermi. High Energy Nuclear Events. *Prog. Theor. Phys.*, 5(4):570–583, 1950.
- [37] R. Hagedorn. Statistical thermodynamics of strong interactions at high-energies. *Nuovo Cim. Suppl.*, 3:147–186, 1965.
- [38] F. Becattini and Ulrich W. Heinz. Thermal hadron production in p p and p anti-p collisions. *Z. Phys. C*, 76:269–286, 1997. [Erratum: *Z.Phys.C* 76, 578 (1997)].
- [39] J. Cleymans, H. B. Geyer, and F. G. Scholtz, editors. *Hadrons in dense matter and hadrosynthesis. Proceedings, 11th Chris Engelbrecht Summer School, Cape Town, South Africa, February 4-13, 1998*, volume 516, 1999.
- [40] F. Becattini and G. Passaleva. Statistical hadronization model and transverse momentum spectra of hadrons in high-energy collisions. *Eur. Phys. J. C*, 23:551–583, 2002.
- [41] V. Greco, C. M. Ko, and P. Lévai. Parton coalescence and anti-proton/pion anomaly at rhic. *Phys. Rev. Lett.*, 90:202302, 2003.
- [42] R. J. Fries, B. Müller, C. Nonaka, and S. A. Bass. Hadronization in heavy-ion collisions: Recombination and fragmentation of partons. *Phys. Rev. Lett.*, 90:202303, 2003.
- [43] J. Adams et al. Particle dependence of azimuthal anisotropy and nuclear modification of particle production at moderate transverse momenta in au+au collisions at $\sqrt{s_{NN}} = 200$ gev. *Phys. Rev. Lett.*, 92:052302, 2004.
- [44] K. Adcox et al. Formation of dense partonic matter in relativistic nucleus-nucleus collisions at rhic: Experimental evaluation by the phenix collaboration. *Nucl. Phys. A*, 757:184–283, 2005.
- [45] B. Abelev et al. Elliptic flow of identified hadrons in pb–pb collisions at $\sqrt{s_{NN}} = 2.76$ tev. *JHEP*, 06:190, 2015.
- [46] S. Navas et al. Review of particle physics – $v(1s)$ entry. *Phys. Rev. D*, 110:030001, 2024. See Particle Listings for $\Upsilon(1S)$.
- [47] S. Navas et al. Review of particle physics – $v(2s)$ entry. *Phys. Rev. D*, 110:030001, 2024. See Particle Listings for $\Upsilon(2S)$.

- [48] S. Navas et al. Review of particle physics – $\psi(3S)$ entry. *Phys. Rev. D*, 110:030001, 2024. See Particle Listings for $\Upsilon(3S)$.
- [49] D. M. Asner et al. Anti-deuteron production in $\psi(ns)$ decays and in e^+e^- annihilation at $\sqrt{s} = 10.5$ gev. *Phys. Rev. D*, 75:012009, 2007.
- [50] J. P. Lees et al. Antideuteron production in $\psi(ns)$ decays and in e^+e^- annihilation near $\sqrt{s} = 10.58$ gev. *Phys. Rev. D*, 89:111102, 2014.
- [51] Jolanta Brodzicka et al. Physics Achievements from the Belle Experiment. *PTEP*, 2012:04D001, 2012.
- [52] R. Hanbury Brown and R. Q. Twiss. A test of a new type of stellar interferometer on sirius. *Nature*, 178:1046–1048, 1956.
- [53] Gordon Baym. The physics of Hanbury Brown–Twiss intensity interferometry: from stars to nuclear collisions. *Acta Phys. Polon. B*, 29(7):1839–1855, 1998.
- [54] D. H. et al. Boal. Intensity interferometry technique in nuclear and elementary-particle collisions. *Rev. Mod. Phys.*, 62:553–602, 1990.
- [55] Urs Achim Wiedemann and Ulrich Heinz. Particle interferometry for relativistic heavy-ion collisions. *arXiv preprint*, 1999.
- [56] C. Plumberg. Hanbury Brown–Twiss interferometry and collectivity in p + p. *Phys. Rev. C*, 102(054908), 2020.
- [57] S. E. Koonin. Proton pictures of high-energy nuclear collisions. *Phys. Lett. B*, 70:43–47, 1977.
- [58] Scott Pratt. Pion interferometry of quark-gluon plasma. *Phys. Rev. Lett.*, 53:1219–1221, 1984.
- [59] Yu. M. Sinyukov. On the region of homogeneity in multiple production processes. *Nucl. Phys. A*, 498:151c–160c, 1989.
- [60] R. Lednicky. Correlation femtoscopy of multiparticle processes. In *Proceedings of the Workshop on Particle Correlations and Femtoscopy*, 2009. arXiv:nucl-th/0501065.
- [61] D. Oliinychenko, V. Vovchenko, et al. Cats: Correlation analysis tool using the schrödinger equation. *J. Phys. G*, 46:105107, 2019.
- [62] D.L. Mihaylov, E. Epple, L. Fabbietti, et al. A femtoscopic correlation analysis tool with coulomb and strong interactions (cats). *Eur. Phys. J. C*, 78:394, 2018.
- [63] Yan-Rui Liu, Hua-Xing Chen, Wei Chen, Atsushi Hosaka, Hongxia Huang, Su Hounq Lee, Bingwei Li, Ling-Yun Meng, Shuntaro Sakai, Bao-Xi Sun, Qian Wang, Wei Wang, Zhi-Gang Wang, Ju-Jun Xie, Mao Yan, Yu-Jie Yao, Yong Zhang, and Qiang Zhao. Exotic hadrons and their interactions: Theory and experiment. *arXiv preprint arXiv:2404.06399*, 2024.
- [64] STAR Collaboration. Measurement of interaction between antiprotons. *Nature*, 527:345–348, 2015.

- [65] R. Lednický and V. L. Lyuboshitz. Influence of final-state interaction on correlations of two particles with nearly equal momenta. *Soviet Journal of Nuclear Physics*, 35:770–788, 1982. *Yad. Fiz.* 35 (1982) 1316–1330.
- [66] L. Mathelitsch and B. J. VerWest. Effective range parameters in nucleon–nucleon scattering. *Physical Review C*, 29:739–746, 1984.
- [67] STAR Collaboration. Strange hadron interactions with femtosopic correlations in heavy-ion collisions at rhic. *Nature*, 588:232–238, 2020.
- [68] R. L. Jaffe. Perhaps a stable dihyperon. *Physical Review Letters*, 38:195–198, 1977.
- [69] H. Takahashi et al. Observation of a $\Lambda\Lambda$ hypernucleus. *Physical Review Letters*, 87(21):212502, 2001.
- [70] H. Nemura. Baryon–baryon interactions from lattice qcd. In *Proceedings of Science, LATTICE2018*, page 091, 2019.
- [71] D. Lonardoni, A. Lovato, S. Gandolfi, and F. Pederiva. Hyperon puzzle: Hints from quantum monte carlo calculations. *Physical Review Letters*, 114(9):092301, 2015.
- [72] K. Aamodt et al. Two-pion bose-einstein correlations in pp collisions at $\sqrt{s} = 0.9$ and 7 tev. *Phys. Rev. D*, 84:112004, 2011.
- [73] S. Schael et al. Two-particle correlations in p p, anti-p anti-p and K0(S) K0(S) pairs from hadronic Z decays. *Phys. Lett. B*, 611:66–80, 2005.
- [74] A. J. Bevan et al. The Physics of the B Factories. *Eur. Phys. J. C*, 74:3026, 2014.
- [75] A. Abashian et al. The Belle Detector. *Nucl. Instrum. Meth. A*, 479:117–232, 2002.
- [76] R. Abe et al. BELLE/SVD2 status and performance. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 535(1):379–383, 2004. Proceedings of the 10th International Vienna Conference on Instrumentation.
- [77] Belle Collaboration. The Belle Detector. *Nucl. Instrum. Methods Phys. Res. A*, 479:117–232, 2002.
- [78] H. Kichimi et al. The BELLE TOF system. *Nucl. Instrum. Meth. A*, 453:315–320, 2000.
- [79] Glenn F. Knoll. *Radiation Detection and Measurement*, 3rd ed. John Wiley and Sons, New York, 3rd edition edition, 2000.
- [80] Claus Grupen and Boris Schwartz. *Particle detectors*. Cambridge Univ. Pr., Cambridge, UK, 2008.
- [81] Collective, editor. *International Conference on Computing in High-Energy Physics (CHEP '98), 31 August–4 September 1998, Chicago, IL, USA*, Chicago, IL, 1998. n.p. CDS record 352671.

- [82] Thomas Kuhr, Christoph Pulvermacher, Michael Ritter, Thomas Hauth, and Nils Braun. The belleii core software. *Computing and Software for Big Science*, 3:1, 2019.
- [83] Belle Tracking Group. Measurement of tracking efficiency using D-meson control samples. Technical Report Belle Note BN-621, Belle Collaboration, 2005. Internal Belle note.
- [84] Anders Ryd, David Lange, Natalia Kuznetsova, Sophie Versille, Marcello Rotondo, David P. Kirkby, Frank K. Wuerthwein, and Akimasa Ishikawa. EvtGen: A Monte Carlo Generator for B-Physics, 5 2005.
- [85] Sw. Banerjee, D. Biswas, T. Przedzinski, and Z. Was. Monte Carlo Event Generator updates, for tau pair events at Belle II energies. *SciPost Phys. Proc.*, 16:003, 2025.
- [86] Piotr Golonka and Zbigniew Was. PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays. *Eur. Phys. J. C*, 45:97–107, 2006.
- [87] B. Casey. HadronB. Belle Note 390, Belle Collaboration, 2001. Internal note.
- [88] Belle Collaboration. Belle Note 684: Lambda selection at Belle. Technical report, Belle Collaboration, 2004. Internal Belle Note (not publicly published).
- [89] Kyle S. Cranmer. Kernel estimation in high-energy physics. *Comput. Phys. Commun.*, 136:198–207, 2001.
- [90] Leif Lönnblad and Harsh Shah. Baryon correlations in Pythia. *Eur. Phys. J. C*, 83(12):1105, 2023.
- [91] Bo Andersson, G. Gustafson, and T. Sjostrand. Baryon Production in Jet Fragmentation and Υ Decay. *Phys. Scripta*, 32:574, 1985.
- [92] Zbigniew Chajecki and Mike Lisa. Global Conservation Laws and Femtoscopy of Small Systems. *Phys. Rev. C*, 78:064903, 2008.
- [93] Shreyasi Acharya et al. Study of the Λ - Λ interaction with femtoscopy correlations in pp and p-Pb collisions at the LHC. *Phys. Lett. B*, 797:134822, 2019.
- [94] Shreyasi Acharya et al. Investigating the role of strangeness in baryon–antibaryon annihilation at the LHC. *Phys. Lett. B*, 829:137060, 2022.
- [95] F. Abudinén et al. Precise measurement of the D^0 and D^+ lifetimes at Belle II. *Phys. Rev. Lett.*, 127(21):211801, 2021.