PRECISION MEASUREMENT OF CHARGED PION AND KAON MULTIPLICITIES IN E^+E^- ANNIHILATION AT $Q = 10.52$ GEV

BY

MARTIN LEITGAB

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Doctoral Committee:

Professor Naomi C. Makins, Chair
Associate Professor Matthias Grosse Perdekamp, Director of Research
Professor John D. Stack
Associate Professor Brian L. DeMarco

Abstract

This thesis presents a high precision measurement of inclusive charged pion and kaon production in e^+e^- annihilation at a center-of-mass energy of 10.52 GeV. The measurements were performed with the Belle detector at the KEKB collider at KEK in Tsukuba, Japan, on a sample of 113×10^6 annihilation events. Uncertainties are kept small by applying experimental-data-driven as well as Monte Carlo-based corrections of systematic effects on measured hadron yields, such as particle misidentification, event selection and radiative corrections. This analysis represents the first precision measurement of multiplicities at low energy scales, far from the Z^0 mass energy scale of the LEP and SLC colliders where most previous precision measurements were performed. In addition, for the first time hadron multiplicities are measured for high fractional hadron energies relative to the energy of the fragmenting parton. Comparable or higher precision than existing measurements is achieved, while still maintaining high resolution in fractional hadron energy.

Measuring high precision hadron multiplicities at low center-of-mass energy from e^+e^- annihilation data will reduce uncertainties on fragmentation functions (FFs). These objects parametrize hadronization, the formation of hadrons from partons in the final state of scattering reactions with large momentum transfers. FFs cannot be calculated from first principles in the theory of Quantum Chromodynamics (QCD), which describes the interaction between color-charged particles, quarks and gluons. Thus FFs have to be extracted from experimentally measured multiplicity data from e^+e^- annihilations, lepton-nucleon scattering and proton-proton collisions in perturbative QCD (pQCD) analyses.

Reducing uncertainties on FFs not only directly enhances our understanding of the process of hadronization, which is omnipresent in any reaction with hadronic final state particles. It will also allow tests of tools and concepts of QCD which currently much of pQCD calculations rely on, such as universality and factorization. In addition, the variation of distribution functions like FFs with energy scale predicted by QCD can be tested. Finally, more precise FFs will enable us to increase our knowledge about other non-calculable quantities in QCD like the nucleon spin structure, for example.

Acknowledgments

First and foremost, I would like to express my deep gratitude to Professor Matthias Grosse Perdekamp and the Department of Physics of the University of Illinois at Urbana-Champaign. Without their help and generous support I would have not been able to write this thesis and work on fulfilling all requirements for my degree.
I would also like to thank Professor Grosse Perdekamp along with Doctor Ralf Seidl most sincerely for their excellent physics and work mentoring and patient advising. With their help I was able to complete this multi-year scientific endeavor.
I would also like to thank Doctor Anselm Vossen for many helpful comments and discussions. As well, I want to use the opportunity to thank Doctor Akio Ogawa and Doctor Francesca Giordano and the members of the Belle collaboration and our group at the UIUC for helpful discussions. Likewise, thanks also to Doctor Peter Winter, David Webber and the other members of the former UIUC MuLan/ MuSun group for computing support and discussion, as well as to Doctor John Koster and Doctor Mickey Chiu.
At last, I would like to express my gratitude to the staff members at KEK in Tsukuba, Japan, who made my stays for research and shift work most pleasant and enriching experiences.

This work was supported by funds from NSF grant NSF-PHY-12 − 05 − 671.

On the personal side, I am in deep gratitude of my parents and my family, for all their support and for equipping me with everything necessary to accomplish this important step in my life.

The last but in no way least person to thank is my lovely fiancée Toni, whom I have shared my life with for the last two years and who always was full of support and encouragement for my work.

Table of Contents

Chapter 1

Investigating Hadronization with Experiment and Quantum Chromodynamics

1.1 Introduction

The Standard Model of Particle Physics is a generally accepted theory about fundamental particles and their interactions. Among other contributions, it contains the framework of Quantum Electrodynamics (QED). QED was motivated by fundamental physics research experiments in the late 19^{th} and first half of the 20^{th} century. The theory describes the interaction of charged particles via the exchange of photons in a quantized field theoretical approach. The formulation of QED as a consistent framework to describe experimental findings but also predict future measurements was largely completed at the end of the 1940s. QED is the most stringently tested physical theory developed so far. One of its predictions, the electron magnetic moment, is experimentally verified to less than one part per trillion accuracy [1].

Through the 1960s and 1970s, new experimental facilities allowed for deep insight into nucleons, which up to then were considered point-like particles. These new particle physics experiments gave evidence for and revealed an inner structure of nucleons, composed of smaller particles (e.g. Refs. [2, 3, 4]). Following the success of the QED theory, these new experimental findings were described with similar field theoretical approaches, one of which was Quantum Chromodynamics (QCD). The theory of QCD introduces a new 'color' charge. It describes the interaction between color-charged particles, identified with quarks, q, and the mediators of the interaction, identified with gluons, g, which carry color-charge themselves. The nature of the formulation of QCD also gives rise to emergent phenomena such as parton confinement, which leads to the formation of hadrons.

Since its inception, the theory has succeeded in describing processes involving color-charged particles in a wide variety of high-energy physics experiments using perturbative techniques (perturbative QCD, pQCD). However, the coupling strength of QCD, described by the parameter α_S, exhibits a strong dependence on the energy scale of a given interaction between color-charged objects ('running coupling constant'). At low energy scales, comparable to nuclear masses ($\mathcal{O}$(1 GeV)), α_S is of the order of unity, therefore prohibiting

the application of perturbative tools. Other, non-perturbative representations of QCD (e.g. lattice QCD[1]) reach the required precision to be considered in phenomenological analyses only for few selected physical observables.

A topical area suffering from inaccessibility with QCD perturbative tools is the process of hadronization, the formation of detectable hadrons from partons in the final state of scattering reactions with large momentum transfer ('hard scattering reactions'). The only model-free consistent parameterization of hadronization is achieved by fragmentation functions (FFs). These have to be extracted in a pQCD-based approach from experimental data. FFs, however, serve as input quantities for analyses of e.g. direct tests of QCD tools and principles, the nucleon spin structure, and physics in heavy-ion collisions. Therefore, hadronization and FFs have been in the focus of large theoretical and experimental interest for several decades. Precise experimental data and extraction methods are crucial for a quantitative understanding of FFs and the process of hadronization.

The following Sections describe the process of hadronization in further detail and formally introduce FFs. The most recent extractions of FFs and the current level of understanding of hadronization will be discussed. Additionally, applications of FFs are briefly mentioned with focus on analyses of the composition of the proton spin. Finally, the last Section of this Chapter will motivate the presented measurement by the necessity to decrease current uncertainties on FFs to enhance our knowledge of hadronization and increase the precision of all FF-related analyses. A short outlook on the expected impact of the presented measurement on FF uncertainties will be given as well.

For simplicity, this thesis will treat theoretical quantities such as cross section expressions in the 'Parton Model' approximation. The Parton Model was developed largely in parallel with QCD as a model to describe hard particle scattering reactions. Characteristics of the model include the assumption that only one of the pointlike constituents of hadrons, partons, participates in a scattering reaction of a hadron with an exterior particle. No other parton in the hadron influences neither the scattering parton nor interacts with the exterior particle.

In addition, quantities described by perturbative QCD are given in leading order α_S unless otherwise noted. Similarly, spins of involved partons and hadrons are not resolved and summed over (unpolarized treatment) unless otherwise noted.

[1] In lattice QCD, or lattice regularized QCD (LQCD), Euclidean space-time is quantized on a lattice with lattice spacing a. Quark and gauge fields are placed on sites and on links between sites, respectively. Calculations can be done non-perturbatively by evaluating path integrals numerically on the lattice. The continuum theory is recovered when performing the limit of infinitesimally small lattice spacing a. For more information about LQCD refer to Ref. [5].

1.2 Describing Hadronization with Unpolarized Fragmentation Functions

1.2.1 Frameworks to Describe Hadronization

Hadronization occurs in the final state of hard scattering reactions when a QCD color-charged parton fragments into hadrons, each with zero net color charge. Hadronization is believed to always occur- experimental measurements did not produce evidence for observable final state particles with non-zero color charge, for an overview c.f. Ref. [6]. In the theory of QCD, the principle of confinement is thought to be responsible for hadronization: Two color charges are confined in vicinity of each other by a charge field, whose potential energy is assumed to increases with distance. Much in contrast to QED, the potential of QCD color-fields cannot be calculated from first principles, and neither the origins nor the quantitative mechanism of confinement are understood. Perturbative tools break down at lower energy scales where confinement occurs due to the rapidly growing QCD coupling constant α_S. Hadronization has to be made accessible to physical calculations by alternative frameworks.

Hadronization Models

In Monte Carlo-based particle physics event generators, the connection between confinement and hadronization is described by models featuring sets of additional assumptions and parameters. Common models are string fragmentation [7, 8] and cluster fragmentation[9, 10]. For illustration, the concept of string fragmentation is explained in further detail in the following.

The energy in the color field, imagined by 'color strings', between two color charges increases as the charges separate. The energy rises until the production of a quark-antiquark pair is energetically possible. Upon further separation of the color charges, the color string breaks and produces one or more quark-antiquark pairs. Over several such steps, more and more quark-antiquark pairs are produced. The hadronization process is stopped as the energy in the color fields falls below the production threshold for quark-antiquark pairs, and colorless hadrons are formed which each correspond to a small piece of color string. The string fragmentation process is illustrated in Figure 1.1.

However, hadronization models include free parameters (in the case of string fragmentation e.g. fragmentation probabilities and the width of the transverse momentum distribution of final hadrons). These parameters have to be optimized for a given physics reaction and its energy scale to achieve reasonable agreement with experimentally measured data. No first-principles description of the hadronization process is currently available.

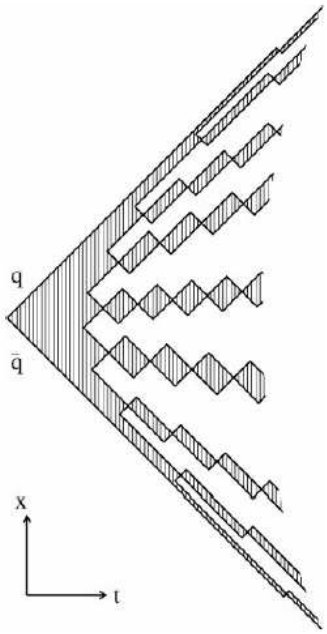


Figure 1.1: Illustrational plot for the concept of string fragmentation, taken from Ref. [11]. A quark-antiquark pair separates in space (y-axis) over time (x-axis). The color string between the partons is represented by vertical lines. The string breaks into smaller pieces if the energy in the string increases beyond the threshold to produce additional quark-antiquark pairs. If the string energy remains below this threshold, the string pieces form final state hadrons.

Description of Hadronization in the Framework of Factorization Theorems

Even though perturbative QCD calculations cannot directly access hadronization, the theory allows the formulation of a framework within which processes including hadronization can be described with perturbative tools. This framework heavily relies on the factorization theorems [12].

To illustrate the application of factorization theorems, the general process of scattering between two nucleons A and B shall be considered (e.g. proton-proton collisions at RHIC), following the description in Ref. [13]. The scattering process schematically reads

$$A + B \to C + X, \tag{1.1}$$

where C is a final state identified hadron and X an arbitrary hadronic final state. If the momentum transfer in the scattering process is sufficiently large, nucleon A will resolve the substructure of nucleon B. In such a case, the actual scattering process is performed elastically between one parton a in nucleon A and one parton b in nucleon B. The scattering process only changes the partons' momenta and is an example of a

hard scattering reaction. The elastic interaction between parton a and b can be written schematically as

$$a + b \to a' + b'. \tag{1.2}$$

The factorization theorems state that for scattering processes with sufficiently large momentum transfers (hard scattering), cross section expressions can be uniquely factorized into pQCD calculable and non-calculable parts. This is equivalent with a separation between quantities characterized by high energy scales, where pQCD is applicable, and low-energy quantities which cannot be calculated perturbatively. The factorized components of cross sections are linked with convolution integrals in their kinematic variables. Equation 1.3 illustrates the factorization of inclusive hadron production in nucleon-nucleon collisions under the factorization theorems in symbolic notation.

$$\sigma_{A+B\to C+X} \propto \sum_{\text{partons}} f_{a|A} \otimes f_{b|B} \otimes \hat{\sigma}_{ab} \otimes D_c^C. \tag{1.3}$$

The initial state of the process in Equation 1.1 is characterized by low energies. Partons a and b are bound in nucleons A and B at energies of order $\mathcal{O}(1\text{ GeV})$. Under the factorization theorems, the initial state of the process in Equation 1.1 can be described by quantities known as parton distribution functions (PDFs) $f_{a|A}$ and $f_{b|B}$. The hard scattering partonic cross section $\hat{\sigma}_{ab}$ represents the high-energy contribution in Equation 1.1 and can be calculated in pQCD.

The final state of the scattering process given in Equation 1.1 is characterized by hadronization of, e.g., the struck parton b into hadron C. Hadronization occurs at the energy level of confinement, $\mathcal{O}(1\text{ GeV})$, and therefore can only be described by non-calculable distribution functions D_c^C. These functions are identified with FFs. A schematic illustration of a nucleon-nucleon hard scattering reaction in the context of factorization is given in Figure 1.2.

To summarize, the factorization framework allows to connect the perturbatively incalculable process of hadronization with perturbative QCD via the use of FFs. The following Section introduces the kinematic dependencies of FFs and PDFs considering semi-inclusive deep inelastic scattering (SIDIS).

1.2.2 Hadronization in Semi-Inclusive Deep Inelastic Scattering in the Framework of Factorization Theorems

Relevant kinematics to describe hadronization in the factorization framework through FFs will be introduced. As an example, hadron formation in SIDIS in the picture of the parton model, in leading order α_S, is

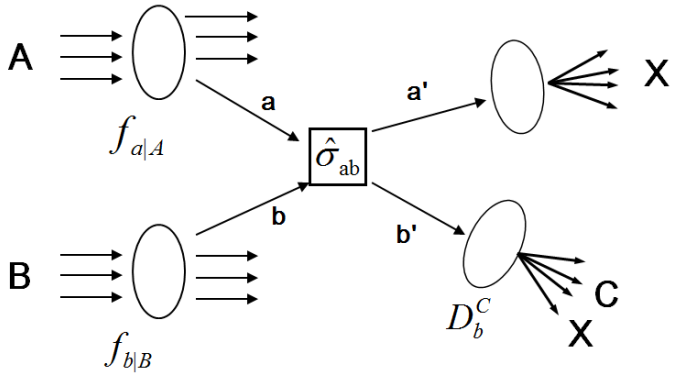


Figure 1.2: Illustration of a nucleon-nucleon hard scattering event in the framework of factorization.

considered. The description follows Refs. [13] and [14].

In a semi-inclusive deep inelastic scattering event, $l(k) + N(p) \rightarrow l^{'}(k^{'}) + h(r) + X$, an incoming lepton l with four-momentum k^{μ} is scattering off a nucleon N with momentum p^{μ}. If the scattering produces a sufficiently high momentum transfer, the incoming lepton resolves the substructure of the nucleon and directly interacts with one of the nucleon's partons i via exchange of a virtual photon γ^*. This represents another example of 'hard scattering'. The final state contains the outgoing lepton $l^{'}$ with momentum $k^{'\mu}$, an identified hadron h with momentum r^{μ} and an arbitrary hadronic state X.

Kinematic variables characterizing SIDIS events are the momentum transfer between the incoming lepton and the target nucleon, q^{μ}, the negative norm of which, $Q \equiv \sqrt{Q^2}$, measures the energy scale of the hard scattering process. The Bjorken scaling variable x gives the momentum fraction of the struck parton i in the initial nucleon, $0 \leq x \leq 1$. The variable y describes the ratio of the energy transfered to the hadron over the lepton energy in the target rest frame. The variable z represents the fractional energy of the hadron h relative to the energy of the struck parton i. All variables are defined in Equations 1.4 through 1.8.

$$q^{\mu} = k^{\mu} - \prime k^{\mu} \tag{1.4}$$

$$-(q_{\mu}q^{\mu}) = Q^2 \tag{1.5}$$

$$x = \frac{-q^2}{2p \cdot q} \tag{1.6}$$

$$y = \frac{p \cdot q}{p \cdot k} = \frac{E_k - E_{\prime k}}{E_k} \tag{1.7}$$

$$z = \frac{r \cdot p}{p \cdot q} \tag{1.8}$$

To write down the unpolarized cross section for semi-inclusive hadron production in SIDIS, the general factorized cross section expression introduced for nucleon-nucleon scattering given in Equation 1.3 can be

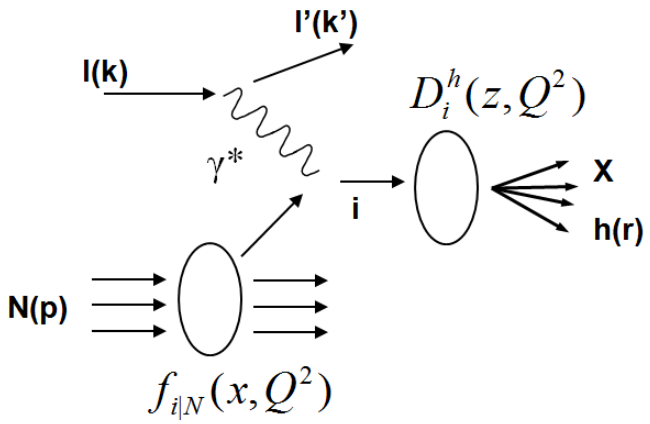


Figure 1.3: Illustration of a SIDIS event in factorization treatment.

utilized. This is possible because of the postulated property of universality of PDFs and FFs between different hard scattering processes. For SIDIS in leading order α_S, the hard scattering partonic cross sections reduce to δ-functions and the initial lepton state can be expressed in terms of kinematic variables. The cross section expression at leading order α_S thus reads

$$\frac{d^3\sigma}{dxdydz} = \frac{2\pi\alpha_{EM}^2}{Q^2}\frac{(1+(1-y)^2)}{y}\sum_{i=q,\bar{q}} e_i^2 f_{i|N}(x,Q^2)D_i^h(z,Q^2), \tag{1.9}$$

where α_{EM} represents the coupling constant of the electromagnetic interaction and e_i stands for the electric charges of the quarks. The PDFs $f_{i|N}(x,Q^2)$ describe the momentum distributions of partons i inside the nucleon N in terms of their momentum fractions x. A graphical depiction of a SIDIS event under factorization treatment is given in Figure 1.3.

The properties of FFs $D_i^h(z,Q^2)$ will be described in the following Section in further detail.

1.2.3 Properties of Fragmentation Functions

The dependence of FFs $D_i^h(z,Q^2)$ on the fractional hadron energy z cannot be calculated from first principles QCD, for instance using lattice QCD techniques. Only certain models can make predictions about the z dependence of FFs, e.g. [15]. Without relying on model assumptions, this dependence has to be extracted from experimental data. This will be described in detail in Section 1.3.

Both PDFs $f_{i|N}(x,Q^2)$ and FFs $D_i^h(z,Q^2)$ show an additional dependence on the momentum transfer squared Q^2 of the hard scattering reaction. This dependence is predicted by pQCD and can be calculated from first principles QCD by a set of coupled integro-differential equations called Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [16]. The translation of a FF from one value of Q^2 to another with the

help of DGLAP equations is referred to as 'scale evolution'. The change of FFs under changing the scale ('scaling violation') is logarithmic in Q^2. For the quark singlet FF, which is the sum of all quark and antiquark FFs, quark/antiquark and gluon FFs are convoluted in the evolution to a different scale Q due to the coupled structure of DGLAP equations. This is illustrated in the following expressions:

$$\frac{d}{d\ln Q^2} D_\Sigma^h(z,\ Q^2) \propto P_{qq}(z) \otimes D_\Sigma^h(z) + P_{gq}(z) \otimes D_g^h(z), \qquad D_\Sigma^h(z) \equiv \sum_q (D_q^h(z) + D_{\bar{q}}^h(z)). \tag{1.10}$$

The convolution integral $\otimes$ is defined as

$$f(z) \otimes g(z) \ = \ \int_z^1 \frac{f(y)}{y} \cdot g(\frac{z}{y}) dy. \tag{1.11}$$

The objects P_{ij} for $i, j = \{q, g\}$ are known as splitting functions. They are related to the probability that quarks and antiquarks radiate gluons (P_{qq}), that gluons are created from quarks or antiquarks (P_{gq}), that gluons produce a quark-antiquark pair (P_{qg}) or that a gluon radiates a gluon itself (P_{gg}). In summary, the splitting functions quantify all ways how a parton could have originated from a parent parton with higher energy. All splitting functions can be calculated in perturbative QCD. In the parton model, FFs obtain the intuitive interpretation of being related to the probability of a parton $i = \{u, \bar{u}, d, \bar{d},\ \ldots,\ g\}$ fragmenting into a hadron h with fractional energy z. It shall be noted that FFs do not give any information about the details of single hadronization events. However, they allow a fully consistent and reliable description of final state hadron distributions in parton scattering processes. Not limited to leading order treatment, FFs fulfill a sum rule resulting from energy and momentum conservation,

$$\sum_h \int_0^1 dz z D_i^h(z, Q^2) = 1. \tag{1.12}$$

Equation 1.12 can be interpreted as the requirement that a given parton fragments into a certain number of hadrons which however carry all of the parton's initial energy.

The fragmentation of partons into hadrons where the parton is one of the hadron's valence quarks is commonly described as 'favored' fragmentation, e.g. a u-quark fragmenting into a π^+ ($u\bar{d}$) meson (described by $D_u^{\pi^+}$). If this condition is not satisfied, fragmentation is considered 'unfavored', e.g. a u-quark fragmenting into a π^- ($\bar{u}d$) meson ($D_u^{\pi^-}$), or an s-quark fragmenting into π^+ ($D_s^{\pi^+}$).

Concluding the description of hadronization and hadron production in the factorization framework, the following Section describes the usage of fragmentation functions in electron-positron annihilations.

1.2.4 Describing Hadronization in Electron-Positron Annihilation with Fragmentation Functions

In electron-positron annihilation $e^+e^- \rightarrow \gamma^* \rightarrow q\bar{q}$, a virtual photon γ^* carrying the total momentum transfer q^μ is formed which decays into a back-to-back quark-antiquark pair in leading order α_S. Here the momentum transfer squared is related to the beam energy, $q_\mu q^\mu = Q^2 = 4E^2_{\text{beam}}$, in the electron-positron center-of-mass frame. The fractional hadron energy z is defined relative to the beam energy,

$$z \equiv \frac{2p^h \cdot q}{Q^2} = \frac{E_h^{cms}}{E_{\text{beam}}^{cms}}. \tag{1.13}$$

The presented analysis measures single hadron production in e^+e^- annihilation. Especially, normalized inclusive cross sections for the production of a hadron h (also referred to as 'multiplicities') are of interest. Similarly to Equations 1.3 and 1.9, Equation 1.14 shows a factorized expression for hadron multiplicities in e^+e^- annihilation, now in next-to-leading (NLO) order α_S. The notation closely follows Ref. [14].

$$\frac{1}{\sigma_{\text{tot}}^{\text{had}}(Q^2)} \frac{d\sigma(Q^2)^{(e^+e^- \rightarrow h+X)}}{dz} = \frac{\sigma_0}{\sum_q e_q^2} \Bigg[\sum_q e_q^2 (D_q^h + D_{\bar{q}}^h) + \frac{\alpha_S(Q^2)}{2\pi} \sum_q e_q^2 \big[(C_q^1 + C_q^L) \otimes (D_q^h + D_{\bar{q}}^h) + (C_g^1 + C_g^L) \otimes D_g^h \big] \Bigg]. \tag{1.14}$$

The cross section is differential in the fractional energy z and normalized to the total hadronic cross section $e^+e^- \rightarrow \textit{hadrons}$, $\sigma_{\text{tot}}^{\text{had}}(Q^2)$. The latter is obtained from the inclusive cross section by integrating over all z and summing over all hadrons, and can be written in NLO α_S as $\sigma_{\text{tot}}^{\text{had}}(Q^2) = \sum_q e_q^2 \sigma_0(Q^2) \Big[1 + \alpha_S(Q^2)/\pi \Big]$. The symbols e_q represent the electric charges of the quarks. On the right hand side of Equation 1.14, the quantity $\sigma_0(Q^2) = 4\pi\alpha_{EM}^2/Q^2$ is related to the purely electromagnetic annihilation cross section. Again following the factorization approach, the multiplicity expression in Equation 1.14 is written in terms of QCD non-calculable quantities (FFs) and the QCD-calculable coefficient functions $C_{q,g}^{1,L}$. One contribution to the multiplicity expression for single hadron production in electron-positron annihilations is illustrated in Figure 1.4.

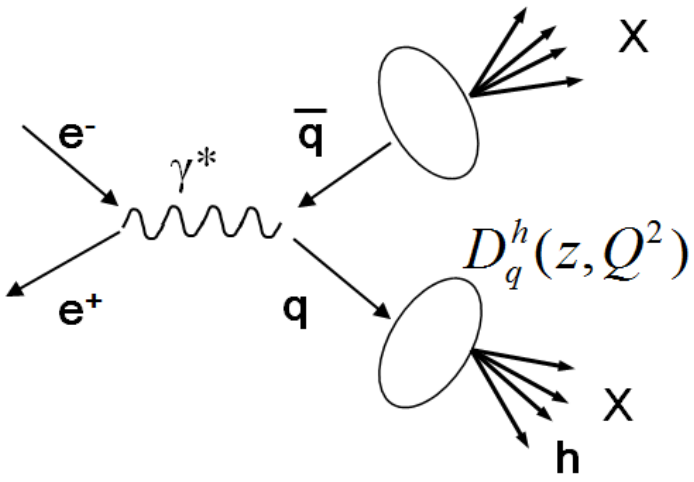


Figure 1.4: Illustration of a single hadron production event in electron-positron annihilation.

1.3 Current Extractions of Fragmentation Functions from Experimental Data

Because of their omnipresence in reactions involving hadronic final states but not being accessible to pQCD calculations, FFs have been under investigation since the inception of QCD. Especially the extraction methods for unpolarized FFs from experimentally measured data have continuously been refined. It was only in the last five years, however, that analyses which propagated experimental and theoretical uncertainties on FFs were published.

1.3.1 Fragmentation Function Extraction Schemes

In general, FFs are extracted from pQCD-based fits to experimentally measured cross sections and related observables. For best constraints on the fit, data sets from several experiments taken at different center-of-mass energies Q are analyzed simultaneously. All recent extractions of FFs follow well-established analysis procedures based on the factorization theorem. These procedures are described in the following for extractions on data from e^+e^- annihilation.

First, functional forms for FFs including freely varying parameters are assumed at a certain initial energy scale $Q_0 = \mu_0$. As illustration, the functional form chosen for FFs $D_i^h(z,\ Q^2)$ at the initial scale $\mu_0 = 1$ GeV in Ref. [14] is given by

$$D_i^h(z,\ \mu_0) = \frac{N_i z^{\alpha_i}(1-z)^{\beta_i}[1+\gamma_i(1-z)^{\delta_i}]}{B[2+\alpha_i, \beta_i+1] + \gamma_i B[2+\alpha_i, \beta_i+\delta_i+1]}, \tag{1.15}$$

where $\alpha, \beta, \gamma, \delta$ are free parameters, $B[a, b]$ represents the Euler Beta-function and N serves as normalization for $D_i^h(z,\ Q^2)$ to correctly contribute to the sum rule given in Equation 1.12.

The chosen functional forms of FFs at the initial scale are evolved with DGLAP evolution equations to the respective energy scales of the experimental datasets considered in the analysis. Expressions for hadron multiplicities are formed from the evolved FF parameterizations along the framework of factorization similar to Equation 1.14. The resulting expressions are fitted to experimental data via a χ^2 minimization. The parameter set determined by the best fit gives the set of individual FFs which is most consistent with all analyzed experimental data. Theoretical as well as experimental uncertainties are propagated through the fitting process and quoted for each individual FF. Current state-of-the-art FF extractions treat all calculable quantities with NLO precision. For inclusion of SIDIS and pp data in FF global analyses, parametrizations for PDFs have to be included to be able to form expressions in analogy to Equations 1.3 and 1.9. PDF expressions are usually adopted from other analyses which perform global fits on experimental data to extract PDFs.

The framework of FF extractions has been developed first on e^+e^- data only, because of the theoretical ease of compactly describing the QED-only initial state. The most recent of such extractions was performed by Hirai, Kumano, Nagai, and Sudoh (HKNS) [17], who also for the first time quote uncertainties on the extracted FFs, propagated from theoretical and experimental uncertainties. Additional constraints on extracted FFs not provided by e^+e^- data can be obtained under inclusion of data samples from other scattering processes, e.g. semi-inclusive deep inelastic scattering (SIDIS) or proton-proton collisions (pp), in the FF extraction analysis. Such 'global', simultaneous analyses of measurements from different processes are possible due to the postulated property of universality of distribution functions like FFs and PDFs. Global extractions of pion and kaon FFs analyzing e^+e^- and pp data (Albino, Kniehl, and Kretzer (AKK), [18]) and e^+e^-, SIDIS and pp data (de Florian, Sassot, and Stratmann (DSS), [14]) have been carried out. The latter provided a full uncertainty analysis in Ref. [19].

1.3.2 Status of Fragmentation Function

In all above analyses, the extracted singlet FF combinations agree with each other within the already small uncertainties, supporting the validity of universality and factorization. DSS FFs in Ref. [14] obtain best constraints on their fit parameters due to their fully global analysis of e^+e^-, SIDIS and pp data. Thus the status of FFs will be discussed with regards to FFs extracted from Ref. [14]. The status holds more generally for FFs from other extractions as well.

Figures 1.5 a) and b) show scaled DSS NLO FFs $zD_i^h(z,\ Q^2)$ for parton $i = \{u+\bar{u},\ \bar{u},\ s+\bar{s},\ g,\ c+\bar{c},\ b+\bar{b}\}$

fragmentation into pions and kaons from Ref. [19]. FFs are given for two energy scales, $Q = M_{Z^0}$ and $Q = 10$ GeV. Relative uncertainties of FFs are shown in the smaller panels below each FF display, calculated with a Hessian ('IH') and a Lagrange Multiplier ('LM') method.

The following general features are visible. For pions, FFs describing favored fragmentation are fairly well-determined with uncertainties below 10% for $z < 0.7$. Unfavored and gluon pion FFs show larger uncertainties up to 20% for $z < 0.7$. For high $z > 0.7$, all pion FFs carry uncertainties of more than 20%. For kaons, uncertainties are even larger due to fewer and less precise experimental data: Favored kaon FF uncertainties remain below 25% for medium z and below about 30% at high $z > 0.7$. Unfavored kaon FFs carry uncertainties larger than 50% at high z. As additional illustration, Appendix A shows pion and kaon FFs from AKK, in comparison to FFs from HKNS and DSS at $Q = M_{Z^0}$. Although the data sets of all three extractions are largely overlapping, substantial differences are visible for unfavored and gluon FFs and in general at high z. The FFs with the largest variance among the different extractions also show the largest uncertainties as described in the discussion of DSS FFs above.

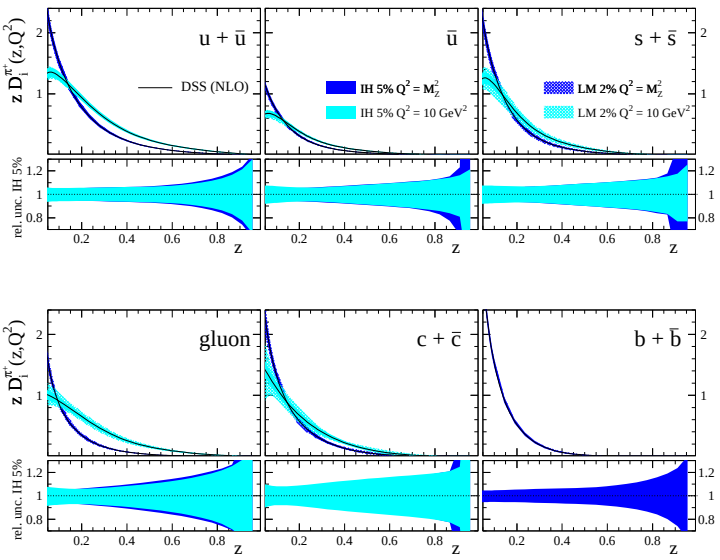
1.4 Relevance and Applications of Fragmentation Functions

1.4.1 Fragmentation Functions as Probes for Hadronization and QCD

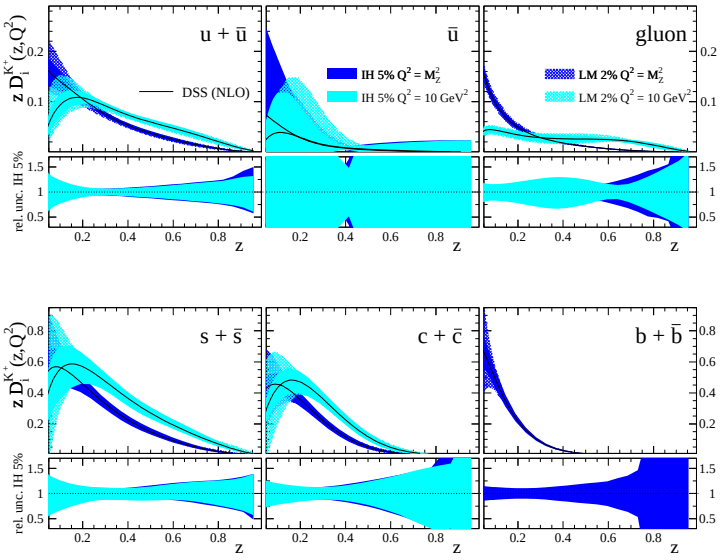
As mentioned above, FFs represent the only access to a quantitative description of hadronization due to the limitation of perturbative techniques in QCD at the low energy scales governing hadron structure. Therefore it is of inherent importance to enhance our knowledge of FFs to improve our understanding of how colorless hadrons are formed from color-charged partons.

Precise information on FFs also allows tests of the fundamental theorems of factorization and the principle of universality of distribution functions between different scattering processes. These concepts can be motivated from QCD, but not all of them have been rigorously proven yet. Therefore, they receive most of their validity and consistency confirmations from comparing results between hadronic physics analyses for different scattering processes, such as global FF extractions. In addition, precise FFs allow tests of the QCD-predicted scale evolutions of distribution functions. Therefore, increasing the theoretical and experimental precision of FF extractions will enable more stringent tests of factorization, universality and scale evolution.

On the other hand, global analyses treating data from different scattering processes simultaneously are founded on the assumed applicability of factorization and universality and the accuracy of scaling evolution predictions. The validity and relevance of results from global analyses is strongly bound to the validity and



(a)



(b)

Figure 1.5: DSS NLO FFs $zD_i^h(z,\ Q^2)$ for parton $i = \{u + \bar{u},\ \bar{u},\ s + \bar{s},\ g,\ c + \bar{c},\ b + \bar{b}\}$ fragmentation into pions (a) and kaons (b) from Ref. [14] shown in Ref. [19] at energy scales $Q = M_{Z^0}$ and $Q = 10$ GeV. Uncertainties are calculated with Lagrange Multiplier ('LM') and Hessian ('IH') uncertainty techniques.

precision of these concepts. Increasing the precision of FF extractions will therefore contribute to further strengthen the legitimacy of global analyses.

In addition, enhanced understanding of FFs especially at high z is expected to provide novel opportunities for tests of resummation concepts of QCD. Currently, the precision of FFs at high z is insufficient to perform quantitative studies in this kinematic regime. As will be described below, this can be related to a lack of experimental multiplicity data at high z. Increased precision of FFs at high z will allow tests of the relevance of all-order resummation effects of $\log(1-z)$-enhanced terms in the QCD-calculable coefficient functions of FF cross section expressions [20, 18]. Corresponding studies using preliminary data of this thesis are given in Ref. [20].

Furthermore, increasing the precision on FFs will enable more accurate QCD predictions for hadron production cross sections at RHIC or LHC. Such improvements will be necessary to facilitate the discovery of signatures of new physics beyond the Standard Model. Another application of precise FFs is to study the modifications of PDFs in heavy ion collisions [21].

Another topical area relying on hadron production measurements from RHIC and LHC and benefiting from increased precision in FFs is the determination of the nucleon spin structure. This analysis will shortly be described in the following.

1.4.2 Fragmentation Functions as Analysis Input Quantities- Investigating the Origin of Nucleon Spin

Introduction

The total nucleon spin is measured to be 1/2 in units of $\hbar$ if projected along the direction of movement of the nucleon. The spin contributions of the nucleon's constituents to the total spin are usually described by so-called 'helicity PDFs',

$$\Delta f_i(x,Q^2) \equiv f_i^+(x,Q^2) - f_i^-(x,Q^2). \tag{1.16}$$

The notation here is adopted from Ref. [22]. The quantities $f_i^{+,-}(x,Q^2)$ in Equation 1.16 describe the distribution of a parton i with positive/negative helicity in a nucleon with positive helicity. The distributions $f_i^{+,-}(x,Q^2)$ depend on the fraction of the nucleon momentum x carried by the parton, and on the hard scattering scale Q at which the parton is probed, in analogy to FFs. The first moment of each distributions, $\Delta f_i^1(Q^2) \equiv \int_0^1 \Delta f_i(x,Q^2)dx \equiv \Delta I(Q^2)$, yields the spin contribution of parton i to the nucleon spin for a given hard scattering scale Q. In analogy to FFs, one can also define a quark singlet helicity PDF, the first moment of which $\Delta\Sigma(x,Q^2)$ is the sum of all quark and antiquark spin contributions. Using the first

moment of the singlet and gluon helicity PDFs together with the sum of angular orbital momenta of all quarks and gluons $L_z(Q^2)$ allows the formulation of the nucleon spin sum rule as

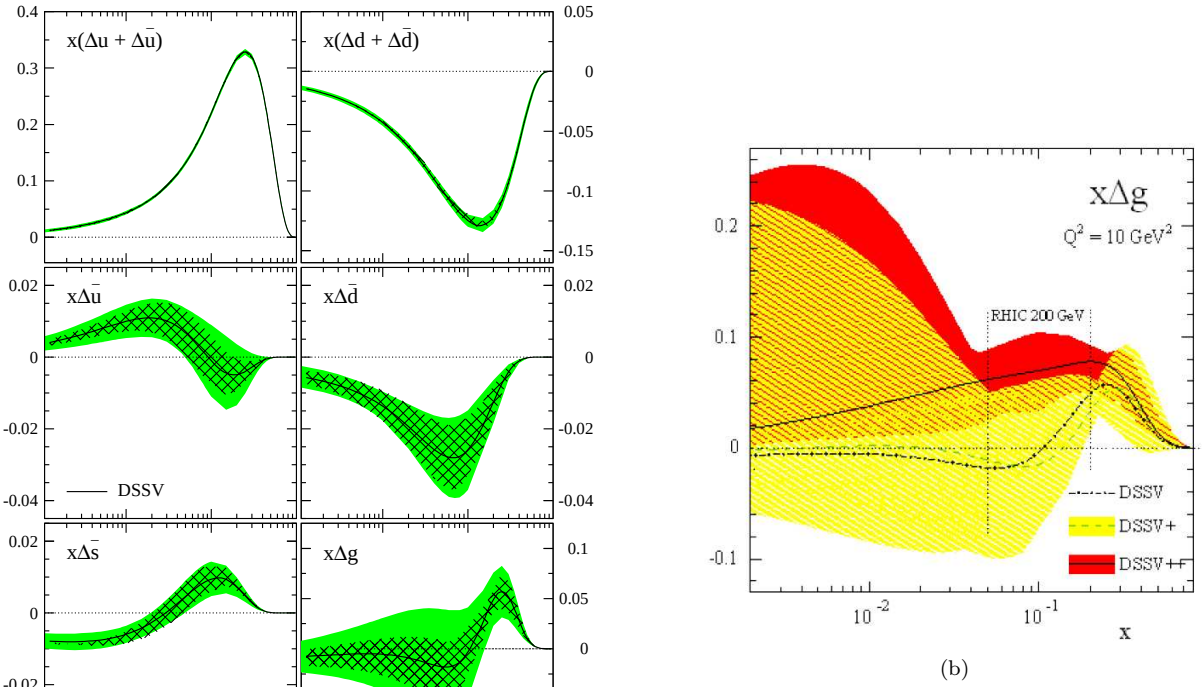
$$J_z = \frac{1}{2} = \frac{1}{2}\Delta\Sigma(Q^2) + \Delta G(Q^2) + L_z(Q^2). \tag{1.17}$$

The notation has been adopted from Ref. [23]. Due to large experimental and theoretical efforts, knowledge about the individual contributions in Equation 1.17 increased continuously over the last decades. The term 'spin crisis' has emerged in the late 1980s [24] when it became evident that only about a quarter, as opposed to the expected majority, of the nucleon's spin is contributed by quarks and antiquarks. The main focus of the field is currently set in theoretical work on determining the role of the orbital angular momentum contributions $L_z(Q^2)$ and in experimental work to increase the precision of the antiquark and gluon helicity PDFs.

Helicity PDFs provide direct insight into nucleon structure. However, their dependence on x cannot be calculated from first principles QCD and has to be extracted in QCD analyses from experimental data, analogously to FFs. FFs enter such analyses as non-perturbative input quantities in expressions for experimentally measured observables under the factorization theorem.

Recent Extractions of Helicity Parton Distribution Functions

One recent extraction has been reported in Ref. [23] on measurements from polarized lepton-nucleon inclusive deep inelastic scattering (DIS) data and polarized SIDIS data ($l + p \to l' + X$ (DIS) and $l + p \to l' + h + X$ (SIDIS)). Another recent extraction [25] additionally included data from semi-inclusive polarized proton-proton collisions ($p + p \to h + X$), assuming universality of distribution functions like helicity PDFs and FFs. Results show fairly well-determined u and d quark helicity PDFs with small uncertainties of the order of 5 to 10%. Antiquark and gluon helicity PDFs, however, carry large uncertainties up to 100%. Samples of extracted helicity PDFs from [25] and a preliminary updated result for their gluon helicity PDF [26] can be seen in Figures 1.6 a) and b), respectively. The gluon helicity PDF update shown in Figure 1.6 b) was extracted with recently released experimental data and for the first time does not feature a node, being wholly positive over the displayed range of x. Both analyses find that uncertainties are largest on the $\bar{u}, \bar{d}$ and gluon helicity PDFs which is consistent with the least precise FFs. Therefore, lower uncertainties on FFs will be necessary to properly analyze upcoming data from π^0 measurements at RHIC for a more precise determination of ΔG. In addition, increased precision of FFs will enable consistency studies with upcoming W-boson measurements at RHIC. These measurements allow access to possible flavor symmetry breaking in the light quark sea $\Delta\bar{u} \neq \Delta\bar{d}$, which is predicted to be large by model calculations (e.g. Ref. [27]). Also,



(a) (b)

Figure 1.6: a) Helicity PDFs from Ref. [25] at $Q^2 = 10$ GeV^2 with $\Delta\chi^2 = 1$ uncertainty bands calculated from Lagrange Multiplier and Hessian uncertainty techniques. b) Preliminary update of the QCD analysis from Ref. [25], released in Ref. [26], taking into account new experimental data ('DSSV+', 'DSSV++'), compared to the original result from Ref. [25] ('DSSV').

the $\bar{s}$ helicity PDFs have been found to show unexpected differences if different sets of FFs are used [23]. A more precise set of FFs, especially with a more precise strange quark to kaon FF, will enable more accurate determinations of the $\bar{s}$ helicity PDF.

In this Section, several aspects of the impact of more precise FFs in physics have been identified. The next Section will motivate the analysis presented in this thesis to achieve the desired increase in precision of FFs.

1.5 The Presented Measurement- Reducing Uncertainties on Fragmentation Function

So far, FFs have been introduced as non-calculable objects essential to the description of the process of hadronization within the limits of the framework of pQCD and factorization. The present status of FFs was discussed and sizable uncertainties at high z and for unfavored and gluon FFs were identified. In addition, the relevance of FFs for tests of QCD tools and principles, for searches for new physics beyond the standard model at the LHC and for the determination of the nucleon spin structure was described. These applications of FFs would significantly benefit from more precise sets of FFs. The next Section will discuss means to reduce FF uncertainties in the context of hadron multiplicity measurements. This will give rise to the motivation and the scope of the measurement presented in this thesis in the following Section. The Chapter will conclude with a brief outlook on the impact of the presented measurement on FF extractions.

1.5.1 Fragmentation Function Uncertainties Due to Lack of High Precision Hadron Multiplicity Data at Low Q^2 and High z

For the three most recent extractions of FFs described in Section 1.3, all groups of authors state explicitly the comparably large uncertainties of FFs at high $z > 0.7$. In all extractions, the reason for these observations is identified with a lack of precise experimental data for $z > 0.7$. In addition, authors of [14] and [17] point out high uncertainties on the gluon FFs $D_g^h(z,\ Q^2)$ for all z. Again, the low precision is attributed to the fact that no precise experimental data are available to constrain the gluon FFs. The FFs with largest uncertainties in the DSS analysis are consistent with the FFs showing the largest differences among the AKK, HKNS and DSS extractions in Appendix A. In part these differences can be motivated from slightly differing analysis concepts and assumptions as well as from the inclusion of pp datasets in AKK, and SIDS and pp datasets in DSS in addition to e^+e^- data only. On the other hand, however, differences arise from the lack of precise and high z datasets which would largely limit the variance of FFs between different extractions.

The authors of Ref.s [14] and [17] stress that experiments at e^+e^- colliders operated at Q 10 GeV (so-called 'B-factories') are highly suitable to relieve the shortcomings of current FFs. The reasons are two-fold.

First, large cross sections and luminosities of B-factory colliders allow the collection of large amounts of statistics, such that the high $z > 0.7$ dependence of hadron multiplicities can be measured even with fairly high precision.

Secondly, constraints for the gluon FFs can be obtained from the lower center-of-mass energy of B-

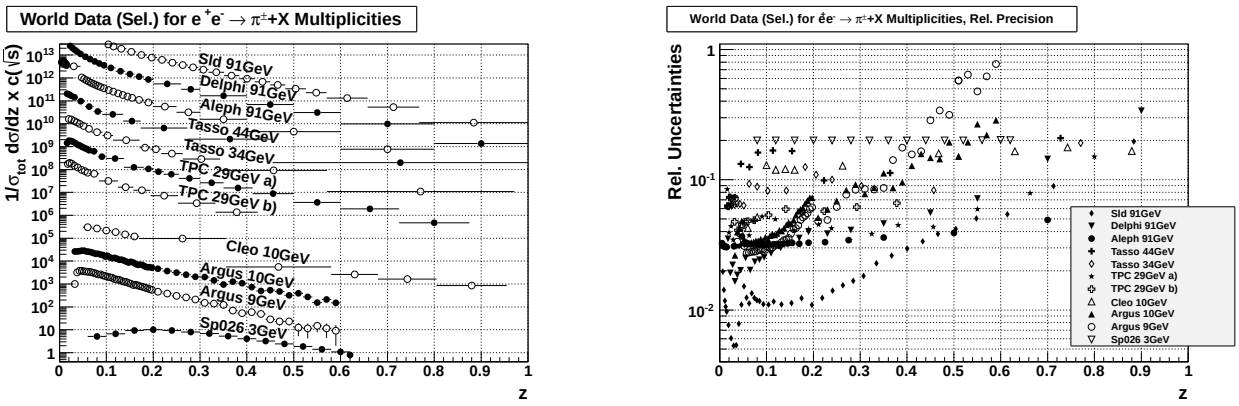
factory colliders compared to previously available precision data from LEP and SLC. In general, multiplicity measurements from e^+e^- annihilation only provide weak constraints on gluon FFs in global analyses, since the corresponding factorized multiplicity expressions only contain gluon FFs in NLO terms at $\mathcal{O}(\alpha_S)$ (cf. Equation 1.14). However, due to the nature of DGLAP evolution equations, evolving the quark singlet function between different energy scales mixes quark/antiquark and gluon FFs already at leading order α_S. Taking advantage of this property, strong constraints on FFs can be obtained by simultaneously analyzing precision datasets taken at high and low center-of-mass energies.

The measurement presented in this thesis will complement the bulk of presently available precision multiplicity data at Z^0 boson mass energy levels from LEP and SLC, providing an excellent opportunity for constraining FFs through significant scale evolution leverage. An e^+e^- multiplicity dataset not containing b quark fragmentation is also expected to add slightly different constraints on quark and antiquark FFs compared to the ones present from $Q = M_{Z^0}$ data including $b\bar{b}$ events. In addition, simultaneous analysis of precision datasets from Z^0 boson mass and Belle energy levels provides limited but possibly significant FF flavor sensitivity due to negligible electroweak contributions at Belle energies.

The limitations of currently available hadron multiplicities measured in e^+e^- annihilation experiments are illustrated in Figures 1.7 from Refs. [28, 29, 30, 31, 32, 33, 34, 35, 36] and 1.8 from Refs. [28, 30, 31, 34, 35, 36, 37], respectively. Shown are representative selections of e^+e^- World multiplicity data for identified, charge-integrated pions and kaons in Figures a), respectively, together with their relative precision in Figures b). As can be seen in the Figures, high precision multiplicities with relative uncertainties below 10% for $z > 0.3$ are only provided by LEP and SLC experiments at high $Q = M_{Z^0}$. Measurements at low Q have been performed by the ARGUS and TASSO collaborations but are not sufficiently precise to constrain the gluon FFs via QCD evolution. The z dependence of hadron multiplicities for $z > 0.7$ is virtually unmeasured for neither pions nor kaons.

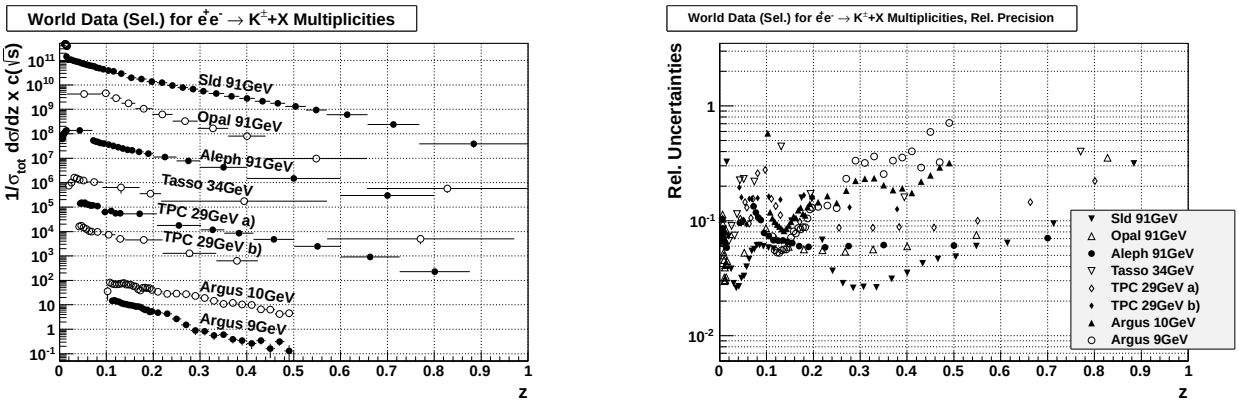
1.5.2 Impact of the New Belle Results

After discussing the requirements for hadron multiplicity measurements to reduce uncertainties on current FFs in the previous Section, the new measurement presented in this thesis is introduced. Precise measurements of multiplicities in the inclusive production of identified charged pions and kaons in electron-positron annihilation are carried out at a center-of-mass energy of $Q = 10.52$ GeV. The measurements are performed with the Belle detector at the KEKB e^+e^- collider using a data sample containing 113×10^6 selected events.



(a) (b)

Figure 1.7: a) Representative selection of available World pion multiplicity data from Refs. [28, 29, 30, 31, 32, 33, 34, 35, 36]. For enhanced visibility, multiplicities are scaled with Q-dependent constants. b) Relative uncertainties of the multiplicity datasets shown in a). All systematic and statistical uncertainties are added in quadrature.



(a) (b)

Figure 1.8: a) Representative selection of available World kaon multiplicity data from Refs. [28, 30, 31, 34, 35, 36, 37]. For enhanced visibility, multiplicities are scaled with Q-dependent constants. b) Relative uncertainties of the multiplicity datasets shown in a). All systematic and statistical uncertainties are added in quadrature.

This thesis presents the analysis and results for hadron multiplicities,

$$\frac{1}{\sigma_{\rm tot}^{\rm had}(Q^2)}\frac{d\sigma(Q^2)^{(e^+e^-\to h^\pm+X)}}{dz} = \frac{1}{N_{\rm evts}} \times \frac{dN_{h^\pm}}{dz} \tag{1.18}$$

for $h^\pm = \{\pi^\pm, K^\pm\}$ pions and kaons, from events $e^+e^- \to q\bar{q}$ where $q = \{u, d, s, c\}$ and for fractional hadron energies of $0.2 \leq z < 0.98$ $(\pi),\ 0.97$ (K), respectively. The combined statistical and systematic uncertainties for $\pi^\pm$ $(K^\pm)$ are 4% (4%) at $z = 0.6$ and 18% (52%) at $z = 0.9$.

This analysis represents the first measurement of the z-dependence of pion and kaon multiplicities for

z-values above 0.7. The reported measurement is also the first precision measurement of multiplicities from e^+e^- annihilations at a low center-of-mass energy, far from the Z^0 resonance which the LEP and SLC colliders operated on. The measurement results are expected to improve the precision of FFs and enable tests of QCD tools and assumptions with unprecedented accuracy. Moreover, QCD-based analyses employing FFs as input quantities, e.g. analyses to extract the nucleon spin structure, are expected to improve in precision as well.

Most recently, first results of FF extraction analyses including preliminary data of this measurement were presented at a topical workshop at RIKEN in Japan (c.f. Refs. [38, 39]). The measured multiplicities are found consistent with previously analyzed multiplicities from e^+e^- annihilations, SIDIS events and pp collisions. Uncertainties especially for gluon FFs and at high z are reported to decrease significantly due to the precision of the preliminary data of this measurement. Final results of updated FF extractions will be published after publication of the presented measurement.

1.6 Summary

In this Chapter, the theory of QCD was introduced as the current description of the interaction between quarks and gluons. FFs have been introduced as objects used to parametrize the non-perturbative aspects of hadronization in cross section expressions in the QCD factorization framework. Recent extractions of FFs were discussed in detail. Applications of FFs with increased precision in particle physics were pointed out. Means to increase currently low precision on FFs at high z and the gluon FFs were identified with high precision multiplicity measurements at low $Q = 10$ GeV. Finally, the analysis presented in this thesis was introduced as a precision measurement of charged pion and kaon multiplicities at the e^+e^- KEKB collider at $Q = 10.52$ GeV. In the following Chapter, the Belle detector will shortly be discussed in the context of the performed measurement. The third Chapter represents the main part of this thesis and provides detailed descriptions about the analysis and associated correction steps on the way from raw experimental pion and kaon yields to the final multiplicities. Chapter four shows the final multiplicity results and contains discussions of the results in the context of the current knowledge of hadronization and FFs.

Chapter 2

The Belle Experiment

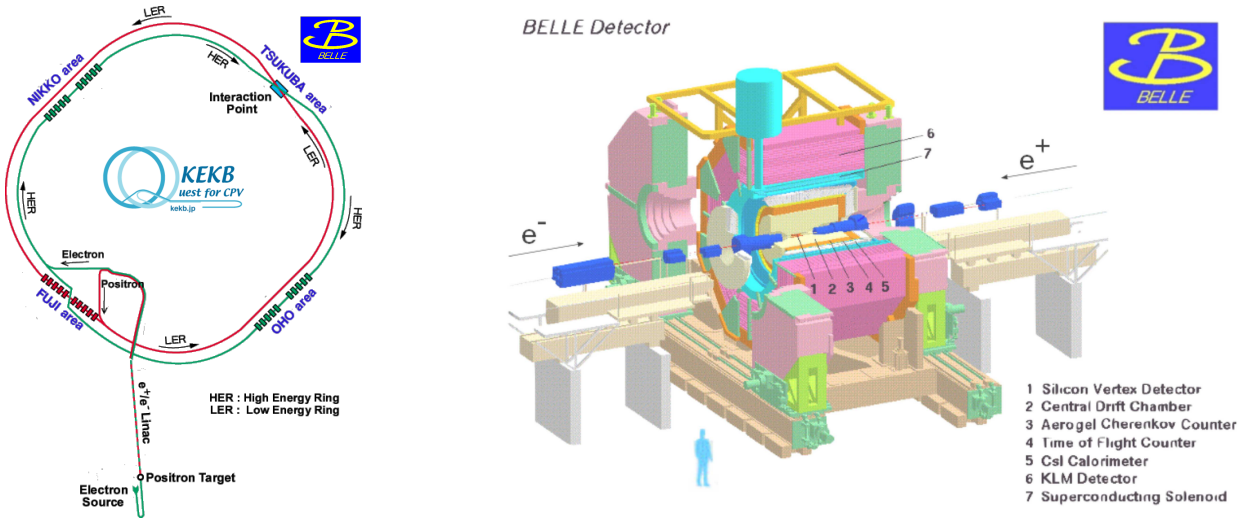
2.1 Introduction to the KEKB Accelerator

In 1964, CP-symmetry violation in the kaon-system was observed at Princeton/ Brookhaven [40]. Analogously, CP-violation was assumed to occur in the B-meson-system. To investigate the B-meson system in detail, two so-called B-factories have been constructed, BABAR/PEPII at SLAC in the USA and KEKB at KEK in Japan. Both of them collided electrons and positrons from circular storage rings. The KEKB accelerator in Japan provided high luminosity particle beams for the Belle-detector which was installed at one of the interaction points of KEKB. Figures 2.1 a) and b) give sketches of the KEKB accelerator rings and the Belle detector.

The Belle experiment intended to measure rare B-meson decay modes and to study CP-violation parameters. The KEKB accelerator was therefore operated at the center-of-mass energy of the $\Upsilon(4S)$ resonance, which is situated at $Q = 10.58$ GeV. The $\Upsilon(4S)$ decays with a branching fraction of greater than 96% into a B-meson pair [5]. The $\Upsilon(4S)$ resonance energy is labeled as the 'on-resonance' energy $Q_{\text{on-res}}$. To obtain the ability to analyze background contributions of processes at the resonance, the KEKB accelerator was also run at the slightly lower 'continuum' center-of-mass energy $Q_{\text{cont}} = 10.52$ GeV. The difference to the on-resonance energy is about 5 half-widths of the $\Upsilon(4S)$ meson [5]. Therefore, only quark-antiquark pairs of flavors $\{u,\ d,\ s,\ c\}$ and lepton-antilepton pairs are produced at the continuum level.

In order to facilitate the study of the B-mesons in the $\Upsilon(4S)$ decay, the beam momenta were chosen to be asymmetric. The momentum of the electron beam was set to 8 GeV/c, whereas the positron beam was kept at a momentum of 3.5 GeV/c. The rest frames of all produced particles in Belle e^+e^- annihilations are therefore boosted along the direction of the e^- beam in the laboratory frame. This circumstance extends the lifetimes of the B-mesons in the $\Upsilon(4S)$ decay and thus allows to resolve their properties more accurately. In order to account for expected higher particle momenta in the 'forward' direction, i.e. in the direction of the electron beam, the Belle detector was designed asymmetrically with respect to the interaction point.

The detector was operated from early 1999 through mid-2010. CP-violation in the B-meson-system was



(a) (b)

Figure 2.1: Illustrations of the KEKB accelerator at Tsukuba, Japan (a) and of the Belle detector at the Tsukuba interaction area (b). Both plots are adapted from Ref. [43].

reported at Belle in 2001 [41]. In addition to the main physics goals, the considerable detector performance and the high statistics available make it possible to carry out precision measurements in hadron spectroscopy and QCD fragmentation physics. A world-record peak luminosity of $2.11 \cdot 10^{34}$ cm^{-2}s^{-1} was achieved in 2009 [42].

2.2 The Belle Detector

In the following Sections, qualitative surveys of the most important tracking and particle identification detector subsystems will be given. The Belle Silicon Vertex Detector (SVD) will be described by information taken from Refs. [44] and [45]. The Sections about the Belle Central Drift Chambers (sCDC/ CDC), the Aerogel Cherenkov Counter (ACC) and the Time-of-Flight (ToF) Detectors will closely follow Refs. [46] and [47] unless otherwise noted. Similarly, the descriptions of the Electromagnetic Calorimeter (ECL) and the K_L and Muon Detection System (KLM) also base on Ref. [46]. For these two Sections, additional information is used from Refs. [48] and [49], respectively.

The Cartesian detector coordinate system is centered at the nominal electron positron interaction point and additionally specified by setting the positive z-direction along the electron beam. The polar angle θ is defined with respect the positive z-axis as well. This direction is also referred to as the 'forward' direction, as opposed to the 'backward' direction along the negative z-axis. The polar angle region from about 34° to 120° is labeled as the 'barrel' region. Five charged particle species are considered stable in the dimensions

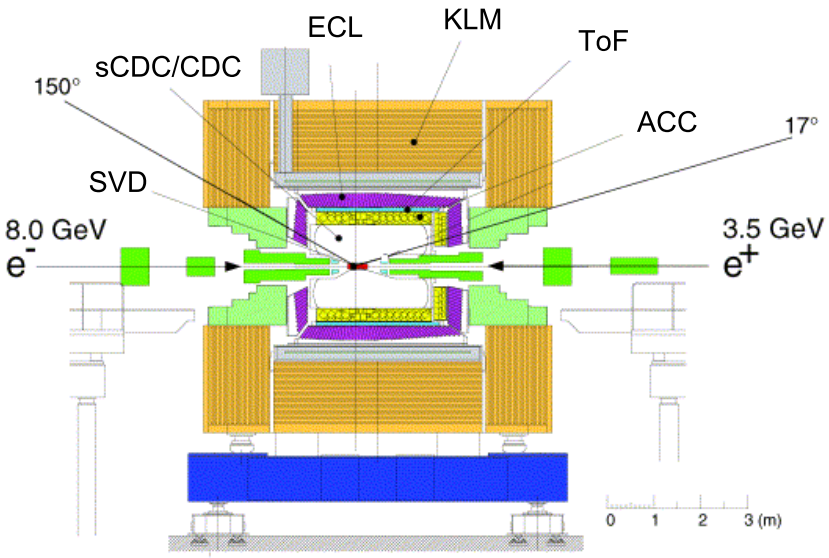


Figure 2.2: Schematic view of the Belle detector perpendicular to the beam axis. The tracking and particle identification subdetectors which are described in this Section are indicated, and the acceptance of the Belle tracking is given. The plot has been adapted from Ref. [46].

of the Belle detector, which are $\{e,\ \mu,\ \pi,\ K,\ p\}$. An illustration of the presented subdetector systems is given in Figure 2.2.

2.2.1 Silicon Vertex Detector (SVD)

In order to satisfactorily resolve B-meson-decays at KEKB, a vertex reconstruction resolution of less than 100 μm is to be achieved by the track sensitive parts of the detector. For the Belle experiment, the concept of a strip Silicon Vertex Detector (SVD) was chosen to achieve this resolution. The SVD was installed outside a beam pipe of 15 mm radius and can be seen in Figure 2.2 as a cylinder colored in red at the center of the detector on the beam axis.

After the last major upgrade in 2003, the Belle SVD featured 4 layers at radii of 20, 43.5, 70 and 88 mm, respectively. The four layers consist of Double-sided Silicon Strip Detector modules (DSSDs) in a barrel-only design [44]. The beam pipe and the configuration of the different SVD layers are illustrated in Figure 2.3.

Roughly speaking, the passage of an ionizing particle creates electron-hole pairs in the bulk of the semiconducting Silicon. These charges are collected by electric fields and produce a signal on electrodes on each side of the DSSDs. The signals created along the way of the through-going particle allow reconstruction of the particle's track and its production vertex.

A characteristic quantity to describe the performance of vertex detectors is the impact parameter reso-

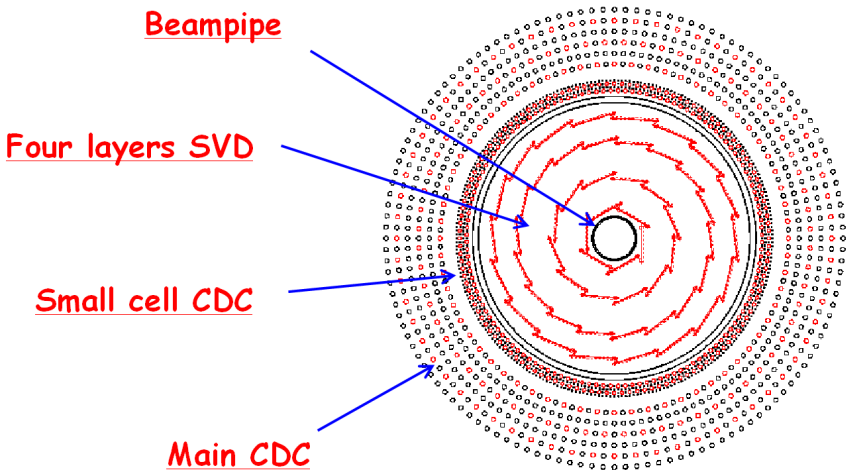


Figure 2.3: Shown is a projecting sketch, view along the z-axis, of the Belle beam pipe, the SVD and the sCDC/CDC. The plot has been adapted from Ref. [50].

lution. This observable can be used, e.g., to quantify the uncertainty in the measurement of the distance between the primary collision vertex and a secondary decay vertex. In Ref. [45], the impact parameter resolution of the Belle SVD detector perpendicular and along the beam axis, respectively, is reported as

$$\sigma_{dr\phi}^{ip} = 21.9 \oplus \frac{35.5}{p\ \beta\ \sin^{3/2}\theta}\ [\mu\text{m}], \tag{2.1}$$

$$\sigma_{dz}^{ip} = 27.8 \oplus \frac{31.9}{p\ \beta\ \sin^{5/2}\theta}\ [\mu\text{m}], \tag{2.2}$$

where p represents the track momentum, θ its polar angle and $\oplus$ indicates that the different terms are to be summed in quadrature.

2.2.2 Central Drift Chamber (CDC)

The SVD is radially surrounded by a multi-wire small cell Central Drift Chamber (sCDC) and the 'main' CDC, as illustrated in Figure 2.3. The sCDC as well as the CDC consists of cylindrical layers of wires. The sCDC is represented by the first four wire layers which are mounted with a reduced wire spacing and share the same gas volume with the main CDC. [51] This volume is filled with a 50%-50% mixture of ethane and helium. Ref. [46] also reports that the SVD and the drift chambers are sensitive to particle interactions in the acceptance region of $17° < \theta < 150°$. The innermost layer of a total of 33728 wires is mounted at a radius of 108.5 mm, the outermost at 863 mm. [52]

Throughout the sCDC and the CDC, 126 μm diameter Al field-wires are used for the purpose of providing

a homogeneous electric field. If a charged particle passes through the chambers, the gas molecules are ionized along its way. The created ions and electrons drift through the electric field and the latter are collected by 30 μm Au-plated tungsten anode sense wires. The signals from gas amplification at the sense wires allow reconstruction of particle trajectories and measurements of particle energy loss in the drift chamber gas.

The SVD and the drift chambers measure tracks and momenta of charged particles inside a 1.5 T magnetic field provided by the Belle Superconducting Solenoid, which surrounds most of the Belle subdetectors. Due to the magnetic field, the tracks of charged particles are curved by the Lorentz force to helices. In the track reconstruction process, track measurements are collected and fitted with helices to determine the trajectories of the measured particles. To improve vertex reconstruction, the measurements from SVD and the drift chambers are combined to fit a total trajectory in the inner detectors. The curvature of the tracks in the magnetic field allows to determine the momenta of the measured particles.

The drift chambers also provide important measurements for particle identification. The magnitude of the signals in the drift chambers associated with a reconstructed helix determine the energy loss dE/dx of the respective particle along its way through the gas. The resolution of the energy loss measurement in the Belle sCDC and CDC is given as 7% for minimum ionizing particles.

On the other hand, the energy loss for different particle species $\{e,\ \mu,\ \pi,\ K,\ p\}$ can be calculated as a function of the laboratory frame momentum. The energy loss of a measured track can be compared to the values which have been calculated for all particles of species j assuming the momentum of the measured track. The difference in dE/dx allows to extract a likelihood $L_j^{\mathrm{sCDC/CDC}}$ for the measured track to be of particle species j.

In Figure 2.4, values of dE/dx measured with the Belle drift chambers are given in dependence of the logarithm of their momenta. The calculated values for the energy loss of $e,\ \pi,\ K$, and p are indicated by red lines.

2.2.3 Aerogel Cherenkov Counter (ACC)

Radially following the CDC, the Aerogel Cherenkov Counter is providing additional information for the assignment of particle species to tracks. This subdetector takes advantage of the Cherenkov effect.

If a particle passes through a medium faster than the speed at which photons can propagate in the material, the particle emits Cherenkov photons. The speed of light in the medium is related to the refractive index of the material, n_{mat}, via $c_{\mathrm{mat}} = c/n_{\mathrm{mat}}$. On the other hand, the velocity of a particle depends on its mass for a given momentum. Thus the refractive index of the medium can be set, e.g., such that fast pions emit Cherenkov light but kaons and protons/antiprotons of the same momentum do not.

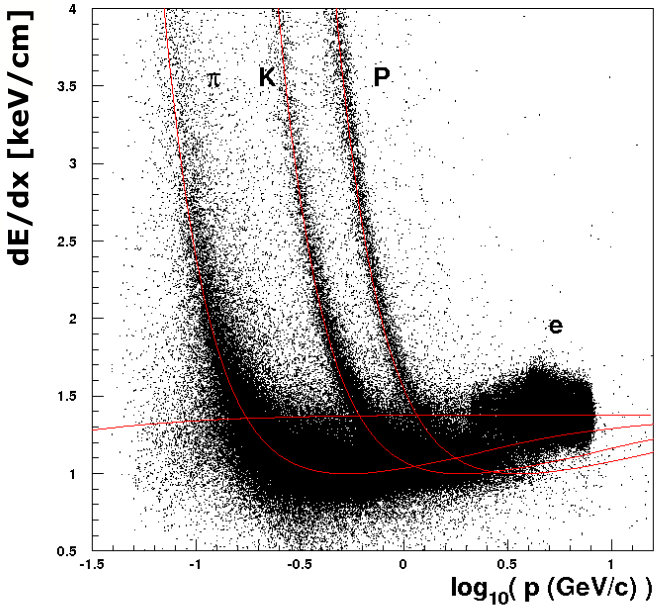


Figure 2.4: Measured values of dE/dx for a sample of tracks reconstructed in the Belle experiment against the logarithm of their momenta. The calculated values for e, π, K, and p are indicated by red lines. The measured values for electrons are truncated at a logarithm of about 0.4 in the plot to enhance the visibility of the measured values of energy loss for π, K and p. The plot has been adapted from Ref. [53].

For the ACC, the medium consists of silica aerogel, a colloidal form of SiO_2. The Cherenkov photons are collected with Fine-Mesh Photo Multiplier Tubes (FM-PMTs). As an effect of the asymmetric beam momenta at the KEKB accelerator, the momenta of B-meson decay products show a strong dependence on their polar angle θ. Thus the ACC modules in the barrel part are divided into 5 sections with differing aerogel refraction indices, depending on the polar angle θ. In total, the ACC detector consists of 1188 modules. The barrel ACC covers an angular region of $34° < \theta < 127°$, which is extended by the forward ACC down to 17°.

For a given particle passing through the ACC, a characteristic number of electrons is created by Cherenkov photons in the FM-PMTs. These numbers can also be calculated from simulations using Monte Carlo methods, taking into account the refractive indices of the aerogel modules. The measured numbers are compared to the calculated values for different particle species j. Similarly as for the drift chambers, likelihood values L_j^{ACC} for a measured track to be of particle species j can be extracted from these comparisons.

The Cherenkov threshold momenta can be calculated from particle momenta and particle masses along with the refractive index of the aerogels used in the ACC, n_{aer}, via the relations $c_{\mathrm{aer}} = c/n_{\mathrm{aer}}$ and $p = m\gamma v$. The symbol γ refers to the relativistic factor $\gamma = 1/\sqrt{1-\frac{v^2}{c^2}}$. The Cherenkov threshold momentum for a particle species i in the ACC is equal to the momentum of a particle of species i which propagates at a velocity $v = c_{\mathrm{aer}}$. Thus the Cherenkov threshold momenta can eventually be calculated according as

$$p_{\mathrm{thresh\ ACC}} = \frac{m \cdot c}{\sqrt{n_{\mathrm{aer}}^2 - 1}}. \tag{2.3}$$

As an illustration, Table 2.1 gives the refraction indices in the barrel part and the resulting threshold laboratory frame momenta for different particle species $j = \{e,\ \mu,\ \pi,\ K,\ p\}$. A detailed graphical illustration of the ACC components is given in Figure 2.5.

Refractive Index	Cherenkov Threshold Momentum for Particle Species i				
	e [MeV/c]	μ [GeV/c]	π [GeV/c]	K [GeV/c]	p [GeV/c]
1.010	3.6	0.75	0.98	3.48	6.62
1.013	3.2	0.65	0.86	3.05	5.80
1.015	2.9	0.61	0.80	2.84	5.40
1.020	2.54	0.53	0.69	2.46	4.67
1.028	2.14	0.44	0.59	2.07	3.94

Table 2.1: ACC threshold momenta corresponding to the choice of different refractive indices for the charged particles considered stable at Belle. The values for the refractive indices are taken from Ref. [46].

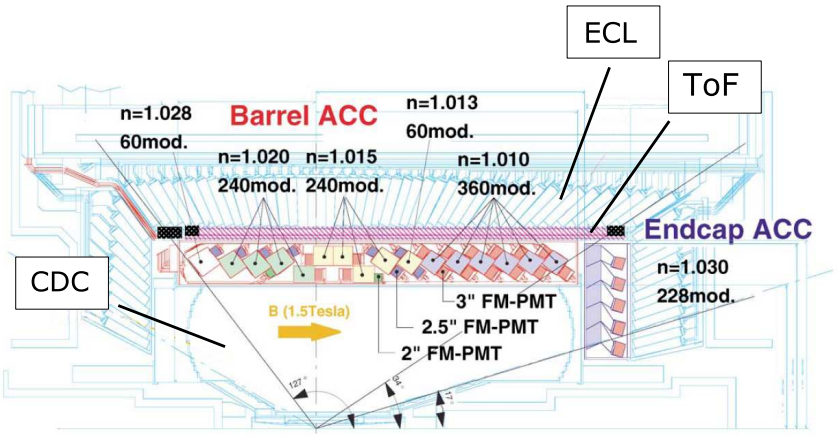


Figure 2.5: Detailed view of the CDC, ACC, ToF and ECL subdetectors in Belle perpendicular to the beam axis. As additional information, the sizes of the Fine-Mesh Photo Multiplier Tubes (FM-PMTs) used for the ACC are given in inches. The plot has been adapted from Ref. [46].

2.2.4 Time-of-Flight Detector (ToF)

Proceeding radially to the outside in the Belle detector, the next subdetector is the Time-of-Flight (ToF) detector. This detector measures the time it takes a particle to propagate the distance of 1.2 m from the interaction point to the ToF detector. The measurement is started by the KEKB radio-frequency clock which is precisely synchronized with the beam collisions. Roughly speaking, the stop signal for the time measurement is provided by one of 128 Polyvinyl-toluene plastic scintillators in the ToF detector. With a charged particle passing through a plastic scintillator, photons are created. These photons are collected in FM-PMTs and give the time and location of the particle passage through the detector.

The ToF detector achieves a time resolution of 100 ps and provides effective kaon-pion discrimination momenta below roughly 1.2 GeV/c. The majority of all B-meson decay products are found below this momentum in the laboratory frame. The coverage of the ToF detector ranges from 34° to 120° in the polar angle θ. The ToF modules are indicated in Figure 2.5 as a narrow band outside the barrel ACC subdetector.

In order to extract PID information from the ToF measurements, the measured times of flight are again compared to calculated values for different particle species j. From this comparison, again a likelihood L_j^{ToF} can be extracted for the measured particle to be of species j.

As an illustration, Table 2.2 gives approximate values for the times-of-flight of all 5 particle species which are considered stable within the dimensions of the Belle detector. These values have been computed from their momenta and masses using the relations $v = s/t$ and $p = m\gamma\, v$. The path-length s is assumed to be 1.2 m which is the radial distance from the nominal interaction point to the ToF detector. The symbol γ again refers to the relativistic factor $\gamma = 1/\sqrt{1-\frac{v^2}{c^2}}$.

The approximate time-of-flight for particles from the interaction point to the ToF detector can eventually be given as

$$t_{\mathrm{ToF}} \;=\; s \cdot \sqrt{\frac{1}{c^2} + \frac{m^2}{p^2}}. \tag{2.4}$$

In Equation 2.4, the particle tracks are assumed to be straight lines, neglecting the magnetic field of 1.5 T in the Belle detector. Calculated values for t_{ToF} assuming particle momenta of 1, 2 and 3 GeV/c are given in Table 2.2.

Track Momentum [GeV/c]	Times-of-Flight for Particle Species i [ns]				
	e	μ	π	K	p
1.0	4.004	4.209	4.274	4.892	5.572
2.0	4.003	4.008	4.013	4.123	4.421
3.0	4.003	4.005	4.007	4.057	4.194

Table 2.2: Times-of-flight for all particle species considered stable in the Belle detector for a path-length of 1.2 m from the interaction point to the ToF detector. Straight trajectories are assumed. The time resolution of the ToF detector is given as 0.1 ns.

2.2.5 Electromagnetic Calorimeter (ECL)

As the last detector inside the Belle Superconducting Solenoid, the Electromagnetic Calorimeter (ECL) is primarily designed to detect $e^{\pm}$ and photons from B-meson and π^0 decays. The ECL is composed of arrays of 8736 CsI(Tl) crystals.

Charged particles and photons passing through the crystal interact with the atoms and produce electromagnetic showers of secondary particles. These particles are collected and read out with silicon photo-diodes. The length of each crystal is chosen to be 30 cm. The ECL features a forward, barrel and backward part and covers an overall range from 17° to 150° in the polar angle θ. The energy resolution achieved in the barrel ECL is 1.7%.

The ECL plays an important role for $e^{\pm}$ particle identification. The energy deposit and the shower shape of $e^{\pm}$ show characteristic differences as compared to hadrons. Electrons and positrons are expected to deposit almost all of their energy in the ECL. Based on these differences, several likelihood values can be computed which give an estimate if the observed track measurements have been created by an $e^{\pm}$ particle.

A first likelihood value is extracted from a χ^2 expression which measures the position matching between an extrapolated charged track from the tracking detectors to a detected energy deposit in the ECL. Another likelihood is calculated from the ratio of deposited energy in the ECL and the measured momentum in the CDC for each track. Due to the low mass of the electrons and positrons and the typically low energy deposit of hadrons in the ECL, this likelihood offers a good discrimination power. A third likelihood is extracted

from the transverse shower shape of ECL signals, taking advantage of different characteristic shower shapes between $e^{\pm}$ and hadrons.

The ECL subdetector can be seen in Figures 2.2 and 2.5 and surrounds the drift chambers, the ACC and the ToF detector.

2.2.6 K_L and Muon Detection System (KLM)

The outermost detector of the Belle experiment is the K_L and Muon Detection System (KLM). This detector is integrated in the magnetic flux return yoke of the Belle Superconducting Solenoid and was installed to detect K_L and muons. The KLM features a total acceptance region in θ of 20° to 155° and consists of about 15 alternating layers of glass-electrode Resistive Plate Chambers (RPCs) and iron absorption plates. The KLM surrounds all other detectors, as indicated in Figure 2.2.

The RPCs are designed to detect charged particles and consist of two parallel plate electrodes under high voltage which represent the boarders of a gas-filled gap in between. The gas used in the KLM is a mixture of HFC/R-134a, argon, n-butane and iso-butane. A particle which passes through the gaps of the RPC modules ionizes the gas atoms which in turn produces a local discharge on the plates in the high electric field. The discharge allows to reconstruct the track of the original particle in the KLM.

The KLM is the main subdetector used for muon identification. The minimum momentum threshold for muon detection is given as 0.6 GeV/c in the laboratory frame. Each track which has been reconstructed in the SVD and the drift chambers is extrapolated to the KLM. All KLM signals from RPC measurements within 25 cm or 5σ of the extrapolated track in the respective RPC plane are associated with the track.

From particle interaction properties, the expected range of charged hadrons and muons through the material in the absorption layers and the RPCs can be calculated. The difference between the measured and calculated range ΔR allows to extract likelihoods $L_k^{\mathrm{KLM},\ \Delta R}$ for a track to be of species k, where $k\ =\ \{\mu,\ \pi,\ K\}$.

Additionally, all RPC measurements associated with a track are fitted to a trajectory themselves. The reduced χ^2 of the fitted KLM measurements is used to calculate another likelihood $L_k^{\mathrm{KLM},\ \chi^2}$ describing the measured track to be a muon, pion or kaon. These two likelihoods are eventually combined to yield the total likelihood for a track to be of a species k,

$$L_k^{\mathrm{KLM}}\ \doteqdot\ L_k^{\mathrm{KLM},\ \Delta R}\ \cdot\ L_k^{\mathrm{KLM},\ \chi^2}. \tag{2.5}$$

2.3 Monte Carlo Simulation of the Belle Experiment

In general, a variety of aspects, e.g. background contributions to analyzed processes or the influence of detector acceptance effects on collected data, can significantly modify the results of physics analyses. Statistical concepts are widely used in physics to learn about and correct for the effects of data corrupting processes which are expected to occur in a given experiment. These processes are most often studied as follows.

In the first step, usually 'event generators' create particles according to known and expected physics processes for the respective experiment. These generators most commonly use Monte Carlo (MC) methods to create particles according to expected yields and momentum distributions. Secondly, the generated particles are propagated through a simulation of the detector used in the experiment. The interaction of the generated particles with the detector material is generally simulated by Monte Carlo algorithms as well. In order to obtain simulation results which can be compared with data taken with the real detector, the virtual detector description and its real counterpart have to match to a high level of precision.

At the Belle experiment, the MC event generator EvtGen [54] is used. This generator has been developed initially by the CLEO collaboration and is primarily designed to simulate physics processes of B-meson decays. More relevant for this thesis, particle generation below the B-meson production threshold is performed using the PYTHIA 6.205 [55] event generation package.

The Belle detector is simulated by 'GSIM', a GEANT [56]-based detector simulation. The data delivered from the GSIM package are stored in the same data structures as used for real data, see Section 2.4. Additionally, particle generator information is stored for every track. By analyzing the data from the detector simulation and the particle generator, underlying processes and their effect on real data analyses can be studied.

However, the Monte Carlo approach is usually not able to model the response of real particle identification subdetectors to a high level of accuracy. This represents an important aspect for the objective of performing a precision measurement. This thesis will study particle identification relying on a combination of data-driven as well as MC methods.

2.4 Belle Data Handling

2.4.1 Data Organization

At Belle, all data are segmented into units called 'Experiments'. These units refer to time periods with different detector setups or different physics goals being analyzed. In general, the changes in the properties

of the detector setup during the same Experiment are assumed to not significantly affect extracted analysis quantities. All experiment numbers are uneven; Experiment 29 does not exist. The numbers of the Experiments considered in this analysis are represented by the interval $[7; 55]$. The data of these Experiments has been taken from roughly the beginning of the detector operation in late 1999 up to late 2006 [57]. The total integrated luminosities of Experiments $[7; 55]$, resolved for KEKB operation energies, are given in Table 2.3.

Experiment Range	On Resonance	Continuum	Other Energies	Total
$[7; 55]$	605 fb^{-1}	68 fb^{-1}	27 fb^{-1}	700 fb^{-1}

Table 2.3: Total integrated luminosities of the datasets used in this analysis for different KEKB operation energies as defined in Section 2.1. The luminosity values have been taken from Ref. [58] and are rounded to fb^{-1}.

2.4.2 Data File Structure and Data Processing

All data collected from Belle subdetectors are processed in several analysis steps. Properties of successfully reconstructed tracks are stored in Data Summary Tapes (DSTs). The logical structure of the data files is defined via the 'PANTHER' event and I/O management package [59] which has been developed by the Belle collaboration. [46]

The amount of data in the DSTs is vast, and physics analyses are most often analyzing only small subsamples of these data. In order to make physics analyses more efficient, reduced datasets - 'skims' - have been introduced. All events and tracks matching the respective skim data selection criteria are collected into skim Micro DSTs (MDSTs). Also event shape variables of the selected events are stored. Data which have been taken via the MC simulation are treated analogously and only feature additional entries with the generator information for each generated particle.

In the MDSTs, the properties of reconstructed tracks which have been selected by the respective skim are stored. All MDST data are stored in the format of MDST 'Panther Tables'. These tables contain most information which is used in physics analyses, e.g. kinematic and PID variables of each track. To process the data files for physics analyses, the Belle Analysis Framework (BASF) [60] has been developed by the collaboration. Roughly speaking, this framework allows to link user analysis codes to data files at run time. The data transfer between the user codes and the data files is performed by the PANTHER package. [46]

For all analyses of the output data of BASF processes, the object-oriented software environment ROOT [61] is used. For all data fitting tasks, the ROOT class 'TMinuit' is applied. All fits are carried out via the default χ^2 minimization algorithm in TMinuit. A fit precision of 10^{-10} was set manually to suppress dependencies

of the fit results on the used version of ROOT and computing hardware parameters. All uncertainties of fit parameters given and used in the extraction of probabilities have been calculated by the 'MINOS' package implemented in ROOT.

2.4.3 Example for Data Skimming: The Belle HadronB Skim

As already mentioned above, skimmed data subsamples are created to enhance processing convenience for physics analyses in the Belle experiment. Ref. [62] describes the main hadronic analysis skim 'HadronB', which shall shortly be sketched in this Section.

The HadronB skim was designed to select 100% of all $B\overline{B}$ events and as many 'continuum' events (production of quark-antiquark pairs $e^+e^- \rightarrow q\overline{q}$) as possible. All non-hadronic events should be rejected. To achieve this result, cuts are applied to track properties as well as event variables taken from the Belle DST Panther Tables. As an illustration, three event selection cuts shall be discussed in the following.

The event selection cuts are used to reject background events. The tracks of background events produce signals in the detectors but do not originate from an e^+e^- annihilation. Possible sources for such events are cosmic rays and beam related background, for instance beam particles hitting residual gas molecules in the beam pipe or parts of the beam pipe. The first HadronB event selection cut discussed here requires at least three tracks of an event with transverse momenta, p_t, higher than 0.1 GeV/c.

Two other cuts examine distances of closest approach of reconstructed tracks to the nominal e^+e^- interaction point (impact parameters). The first cut rejects events which primarily consist of particles produced in the wall of the beam pipe. It requires track impact parameters in the radial direction of less than 2 cm (1.5 cm), which is equal to the beam pipe radius [63] for Experiments < 31 (≥ 31, after the installation of the updated SVD). Another cut requires track impact parameters along the z axis of less than 4 cm to reject misreconstructed tracks or tracks from cosmic ray particles.

Additional cuts in the HadronB skim impose requirements on so-called event shape variables determined from reconstruction. An example of such variables is the sum of all reconstructed particle masses in the jet with the most reconstructed mass ('Heavy Jet Mass', HJM). The event shape variable cuts are included, e.g., to reject leptonic events $e^+e^- \rightarrow l^+l^-$. For explicit formulations of the HadronB skim cuts on track and also event shape variables see Ref. [62].

A cut commonly used for physics analyses is the skim 'HadronBJ'. This skim is created by combining events in the HadronB skim with events selected in the J/ψ skim. [64] Events containing two oppositely charged tracks are included in the J/ψ skim if their momenta are above 0.8 GeV/c and the invariant mass of the two tracks is in the range $[2.5; 4]$ GeV/c^2 [64]. Events passing the HadronBJ event selection are used

as the measurement sample for the analysis of pion and kaon multiplicities presented in this thesis.

Chapter 3

Experimental Data Sample for Measurement

3.1 Measurement Data Sample

The presented measurement is performed on a subset of the available data for the given Experiments. To enter the analysis sample, each event is required to fulfill the 'HadronBJ' data skim criteria. This skim was introduced in the previous Chapter in Section 2.4.3. Only data from the 'continuum' energy level, about 60 MeV/c^2 below the mass resonance of the $\Upsilon(4S)$, are considered, containing events $e^+e^- \to q\bar{q}$ with $q = \{u, d, s, c\}$. Technically, also 'on-resonance' data could be used for the present analysis, as long as all tracks are required to show a fractional energy of $z >\sim 0.5$. This is motivated by the fact that almost all b quarks are produced as B and $\bar{B}$ mesons on-resonance, which decay more or less at rest in the center of mass frame. Thus they and also all of their decay products can carry a maximum z of $<\sim 0.5$. Since the precision of this analysis is limited by systematic uncertainties, no inclusion of on-resonance data has been performed.

3.1.1 Additional Event and Track Selection Cuts

In addition to passing the HadronBJ selection criteria, each event is required to also show a 'visible energy' larger than 7.0 GeV to enter the analysis sample. The visible energy is defined as the sum of 'good' track momenta and 'good' photon energies. This requirement should reject $\tau\tau$ events where neutrinos can reduce the reconstructed energy in the detector.

Other cuts are imposed on parameters of reconstructed tracks: Only tracks within the Belle Time-of-Flight detector acceptance and with reconstructed laboratory frame momenta larger than 500 MeV/c are entering the analysis data sample. These cuts will be motivated in the following Section 4.1. Furthermore, all tracks are required to be associated to at least 3 Silicon Vertex Detector (SVD) hits, and the closest approach of each track helix to the run-by-run interaction point position has to be closer than 1.3 cm radially and 4 cm in beam direction to reject cosmics and ghost tracks. Also all tracks are rejected for which reconstruction suggests a fractional energy $z > 1.0$. All event and track selection cuts are summarized in Table 3.1. The

species of reconstructed tracks are extracted via cuts on particle identification (PID) likelihood values, which will be discussed in Section 4.1.

Event Selection

Cuts	**Criterion**
HadronBJ	pass
$E_{\rm vis}$	> 7.0 GeV

Track Selection

Cuts	**Criterion**
p_{lab}	>= 500 MeV/c
$cos\theta_{lab}$	$[-0.511; 0.842)$
$N^{SVD}_{r\phi+z}$	>= 3
IP_dr	<= 1.3 cm
IP_dz	<= 4.0 cm
z	<= 1.0

Table 3.1: Selection criteria for events/tracks to enter the present analysis. All quantities are explained in the text.

3.1.2 Statistics of Measurement Sample

After all event selection cuts, about $113 * 10^6$ events were analyzed. With all track selection cuts considered, about $105 * 10^6$ events contribute at least one track to the analysis. The integrated luminosity for the data sample is about 68.03 fb^{-1} [57]. Figure 3.1 shows plots of raw experimental yields for negatively charged electrons, muons, pions, kaons and antiprotons versus reconstructed fractional energy z with Poisson uncertainties. The assigned track species is inferred from the applied PID likelihood cuts. The spectra are plotted starting from $z = 0.2$ for leptons, pions and kaons and from $z = 0.3$ for protons with a constant bin width of 0.01 in z. In the muon plot, the presence of lepton pair production and 2γ events in HadronBJ skim data is clearly visible at high z, which will be further examined in Section 4.1.5.

3.1.3 Minimum Value of z in Analysis

Since this analysis only considers tracks with laboratory frame momenta higher or equal to 0.5 GeV/c, an acceptance turn-on will occur in the corresponding spectra for low values of the normalized center-of-mass energy z. Additionally, data at low z ($<\sim 0.1$) are most commonly excluded from fragmentation function extractions due to singularities in the DGLAP equations and non-negligible mass effects [14]. To be sufficiently far from mass effect regions for pions and kaons and to avoid the low z turn-on from the laboratory frame momentum cut, this analysis extracts pion and kaon spectra for values of z no lower than 0.2.

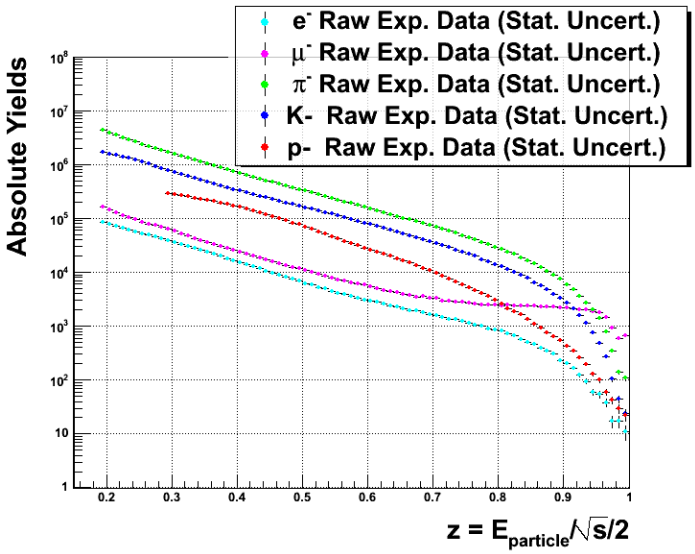


Figure 3.1: Plot of raw experimental data yields analyzed in the present measurement. Statistical uncertainties are shown.

3.2 Performed Corrections

The 'raw' yields given in Figure 3.1 are corrected for several effects to finally obtain charged pion and kaon multiplicities versus fractional energy z, with a binning of 0.01 in z. These effects include particle misidentification, sample impurities, kinematic smearing, decay-in-flight/reconstruction, event and track selections, acceptance effects and initial/final state radiation (ISR/FSR) effects. Finally, the yields are normalized by number of events to arrive at multiplicities. Equation 3.1 summarizes schematically the construction of multiplicities for particle species i in the k^{th} z bin from the Belle raw yields. The particle misidentification and kinematic smearing corrections are performed as matrix corrections; matching indices are implicitly summed over.

$$N^{\text{mult.; } i}_{(z_k)} = \frac{1}{n_{\text{evts}}} \epsilon^{\text{ISR/FSR; } i}_{(z_k)} \epsilon^{\text{accept.; } i}_{(z_k)} \epsilon^{\text{evt. sel.; } i}_{(z_k)} \epsilon^{\text{DIF/recon.; } i}_{(z_k)} \mathcal{P}^{-1;\ \text{smear.; } i}_{z_k z_l} \epsilon^{\text{impur.; } i}_{(z_l)} \mathcal{P}^{-1;\ PID}_{ij;\ (p_{\text{lab}}, cos\theta_{\text{lab}})} N^{\text{raw;}j}_{(p_{\text{lab}}, cos\theta_{\text{lab}}; z_l)}. \tag{3.1}$$

3.3 Uncertainty Treatment

In the present measurement, uncertainties are propagated as bin-by-bin statistical and systematic uncertainties throughout all corrections. Bin-by-bin correlations will only be considered in the particle misidentification and momentum smearing corrections.

All uncertainties occurring in all corrections for systematic effects are propagated as described below.

Systematic Corrections with Bin-by-Bin Correction Factors

In case a systematic correction can be performed with bin-by-bin correction factors, the statistical uncertainties of the corrected yields $N_{(z)}^{\text{corr.}}$ are given by the statistical uncertainties of the uncorrected yields $N_{(z)}^{\text{uncorr.}}$ scaled by the correction factor ϵ,

$$\sigma^{\text{stat.}}_{N^{\text{corr.}}_{(z)}} \;=\; \epsilon \; \sigma^{\text{stat.}}_{N^{\text{uncorr.}}_{(z)}} . \tag{3.2}$$

The systematic uncertainties of the corrected yields contain the contributions from the uncertainty of the correction factor and the scaled systematic uncertainties of the uncorrected yields,

$$\sigma^{2,\text{ syst.}}_{N^{\text{corr.}}_{(z)}} \;=\; N^{2,\text{ uncorr.}}_{(z)} \; \sigma^2_\epsilon \;+\; \epsilon \; \sigma^{2,\text{ syst.}}_{N^{\text{uncorr.}}_{(z)}} . \tag{3.3}$$

If the correction factor carries statistical and systematic uncertainties, the uncertainty of the correction factor is given by the square sum of those,

$$\sigma^2_\epsilon \;=\; \sigma^{2,\text{ stat.}}_\epsilon \;+\; \sigma^{2,\text{ syst.}}_\epsilon . \tag{3.4}$$

Systematic Corrections with Correction Matrices

The correction for particle misidentification and for momentum smearing both are performed by using correction matrices. In these cases, the statistical uncertainty of the corrected yields are equal to the sum of all contributing products of matrix elements and uncorrected yields,

$$\sigma^{2,\text{ stat.}}_{N^{\text{corr.}}_{m}} \;=\; \Sigma_n A^2_{mn} \sigma^{2,\text{ stat.}}_{N^{\text{uncorr.}}_{n}} . \tag{3.5}$$

In analogy to Equation 3.3, the systematic uncertainties of the corrected yields contain all contributions from matrix element systematic uncertainties and scaled systematic uncertainties from uncorrected yields,

$$\sigma^{2,\text{ syst.}}_{N^{\text{corr.}}_{m}} \;=\; \Sigma_n (N^{2,\text{ uncorr.}}_n \; \sigma^2_{A_{mn}} \;+\; A^2_{mn} \; \sigma^{2,\text{ syst.}}_{N^{\text{uncorr.}}_{n}}) . \tag{3.6}$$

The goal of the measurement is to keep relative systematic and statistical uncertainties on the percent-level for most of the range in z. At high values of z, uncertainties are expected to increase since production cross sections and thus statistics will fade out.

Chapter 4

Corrections for Systematic Effects

4.1 Correction for Particle Misidentification

For a measurement of charged hadron multiplicities, hadrons need to be identified and their energy determined. This information is distorted by the non-deterministic measurement process in particle physics detectors, data skimming and analysis cuts and thus needs to be corrected for these aspects.

4.1.1 Framework for PID Correction

In order to select hadrons of a certain species j from the Belle dataset, cuts on particle identification (PID) likelihood values associated with each track can be specified. These likelihood values are the result of combining PID information from the dE/dx measurement in the drift chamber, the amount of Cherenkov light produced in the ACC, the time-of-flight measurement in the ToF, the energy deposit and the shower shape in the ECL and the track range and track continuity in the KLM. For details on the calculation of likelihood values from the Belle detector measurements see Appendix B.1. The set of likelihood cuts used throughout the entire PID correction in this analysis can be found in this Appendix as well.

Assigning a species hypothesis j to a track by cutting on its likelihood values can be incorrect due to the probabilistic nature of the PID likelihoods, i.e. the track might actually have been created in the detector by a particle of species i. Several studies in the Belle collaboration have already been performed to determine the PID performance of the Belle detector.[1] These studies found the rates for particle mis-identifications to be $O(10\%)$ at Belle, depending on the kinematic properties of particles and the chosen set of likelihood cuts. In the context of the presented analysis being a precision measurement, particle misidentification needs to be corrected for.

At Belle, five charged particle species are considered 'stable' over the dimensions of the detector, $\{e,\ \mu,\ \pi,\ K,\ p\}$. If one considers a vector of particle yields with likelihood-selected species $j\ =\ \{e,\ \mu,\ \pi,\ K,\ p\}$, this vector

[1] Some of these studies have been taken as starting points for the PID analysis in the presented measurement. They will be referenced in later sections of this thesis.

is assumed to be connected to a second vector containing particle yields of 'true/physical' species i by a matrix $\mathcal{P}$,

$$\overrightarrow{N}^{\text{likelihood-cut ID}} \;=\; \mathcal{P}\overrightarrow{N}^{\text{real ID}}. \tag{4.1}$$

$$[\mathcal{P}]_{ji} \;\doteq\; \begin{pmatrix} \mathrm{p}_{e\,\to\,e} & \mathrm{p}_{\mu\,\to\,e} & \mathrm{p}_{\pi\,\to\,e} & \mathrm{p}_{K\,\to\,e} & \mathrm{p}_{p\,\to\,e} \\ \mathrm{p}_{e\,\to\,\mu} & \mathrm{p}_{\mu\,\to\,\mu} & \mathrm{p}_{\pi\,\to\,\mu} & \mathrm{p}_{K\,\to\,\mu} & \mathrm{p}_{p\,\to\,\mu} \\ \mathrm{p}_{e\,\to\,\pi} & \mathrm{p}_{\mu\,\to\,\pi} & \mathrm{p}_{\pi\,\to\,\pi} & \mathrm{p}_{K\,\to\,\pi} & \mathrm{p}_{p\,\to\,\pi} \\ \mathrm{p}_{e\,\to\,K} & \mathrm{p}_{\mu\,\to\,K} & \mathrm{p}_{\pi\,\to\,K} & \mathrm{p}_{K\,\to\,K} & \mathrm{p}_{p\,\to\,K} \\ \mathrm{p}_{e\,\to\,p} & \mathrm{p}_{\mu\,\to\,p} & \mathrm{p}_{\pi\,\to\,p} & \mathrm{p}_{K\,\to\,p} & \mathrm{p}_{p\,\to\,p} \end{pmatrix} \tag{4.2}$$

The matrix $\mathcal{P}$ contains information about all possible particle identifications and mis-identifications. The matrix elements $[\mathcal{P}]_{ji}$ are so-called PID probabilities $\mathrm{p}_{i\,\to\,j}$, which describe the probability that a track created in the detector by a particle of 'true/physical' species i is assigned a 'likelihood-cut' species j, based on the PID likelihood values stored in the datasets. [2] To solve Equation 4.1 for the desired yields $\overrightarrow{N}^{\text{real ID}}$, both sides have to be multiplied by the inverse of the matrix $\mathcal{P}$ from the left. To perform the inversion, all PID probability matrix elements have to be known.

Technically, all of the matrix elements can be extracted from an existing GEANT [56]-based detector simulation for the Belle detector. However, this simulation is expected to show deficiencies in accurately modeling the response of the PID-related detectors. This would introduce hard-to-estimate systematic uncertainties in the PID correction analysis and thus prompts to attempt the extraction of all PID probabilities from experimental data, which is further described in Section 4.1.1.

Binning and Analysis Acceptance for PID Correction The PID performance of the Belle detector is expected to vary with the polar angle and the momentum of particle tracks. Therefore a binning was introduced in the laboratory frame variables p_{lab} and θ_{lab}, following Belle Note 779, Ref. [65]. The chosen binning in p_{lab} and $\cos\theta_{\text{lab}}$ in the present analysis is given explicitly in Table 4.1.

The barrel Time-of-Flight (ToF) detector represents the detector with the most limited scattering angle acceptance. To maximize the quality of PID information for each track entering the analysis data set, all accepted tracks are required to pass through the barrel ToF detector, roughly equal to an interval in the cosine of the laboratory frame scattering angle of about $[-0.511; 0.842)$. In order to ensure that all tracks

[2] The off-diagonal elements of $\mathcal{P}$ are commonly referred to as PID 'fake-rates', which represent the relative number of tracks being selected by likelihood cuts as being of species $j \neq i$. Relative numbers of tracks which are assigned the same species $j = i$ as the original particle are usually called PID 'efficiencies' (diagonal elements of $\mathcal{P}$).

reach at least the iron yoke of the superconducting magnet and so can also deposit PID information in the ECL and KLM detectors, a minimum laboratory momentum cut of 0.5 GeV/c is imposed on each track as well.

Figure 4.1 shows the laboratory frame binning superimposed on a distribution of likelihood-selected tracks from hadronic events on the ($p_{\rm lab}$, cos$\theta_{\rm lab}$) plane. To perform the PID correction, complete PID probability matrices have to be extracted for every bin.

Appendix 2.2 gives details on how PID likelihoods are extracted from detector measurements at Belle. Also the used set of PID likelihood cuts and information about Belle PID subdetector performance can be found there.

(a)

Bin Number	Range in $p_{\rm lab}$ [GeV/c]
0	$[0.5; 0.65)$
1	$[0.65; 0.8)$
2	$[0.8; 1.0)$
...	widths of 0.2 GeV/c
12	$[2.8; 3.0)$
13	$[3.0; 3.5)$
14	$[3.5; 4.0)$
15	$[4.0; 5.0)$
16	$[5.0; 8.0)$

(b)

Bin Number	Range in cos$\theta_{\rm lab}$
0	$[-0.511; -0.300)$
1	$[-0.300; -0.152)$
2	$[-0.152; 0.017)$
3	$[0.017; 0.209)$
4	$[0.209; 0.355)$
5	$[0.355; 0.435)$
6	$[0.435; 0.542)$
7	$[0.542; 0.692)$
8	$[0.692; 0.842)$

Table 4.1: Binning for the particle misidentification correction in $p_{\rm lab}$ (a) and cos$\theta_{\rm lab}$ (b) of the reconstructed particles. The binning in cos$\theta_{\rm lab}$ has been adopted from Ref. [65].

Extraction of PID Probabilities from Experimental Data PID probabilities $\mathrm{p}_{i \,\to\, j}$ can be extracted from track samples where both the 'real'/physical and the likelihood-cut species, i and j, are known. Such track samples can be obtained from experimental data by reconstructing unstable particles. The procedure is commonly used and as such also was employed in the existing PID studies in Belle.

As an example, PID probabilities $\mathrm{p}_{\pi,\, K \,\to\, j}$ have been extracted from analyzing $D^* \to \pi^+_{slow} + D^0 \to \pi^+_{slow} + (K^-\pi^+_{fast})$ decays. To reconstruct D^* particles, ordered track triplets with charge combination $-,+,+$ are selected from the datasets. To distinguish the D^* decay products, the kaon can be 'tagged' by its opposite charge compared to the pions. The two pions can be distinguished via their kinematics- because of the small mass difference between the D^* and the D^0, only little phase space is open to the 'slow' pion, whereas the energy of the 'fast' pion contains a good fraction of the D^0 mass energy.

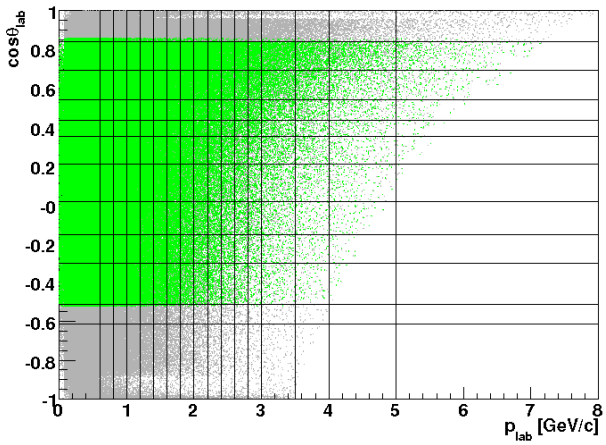


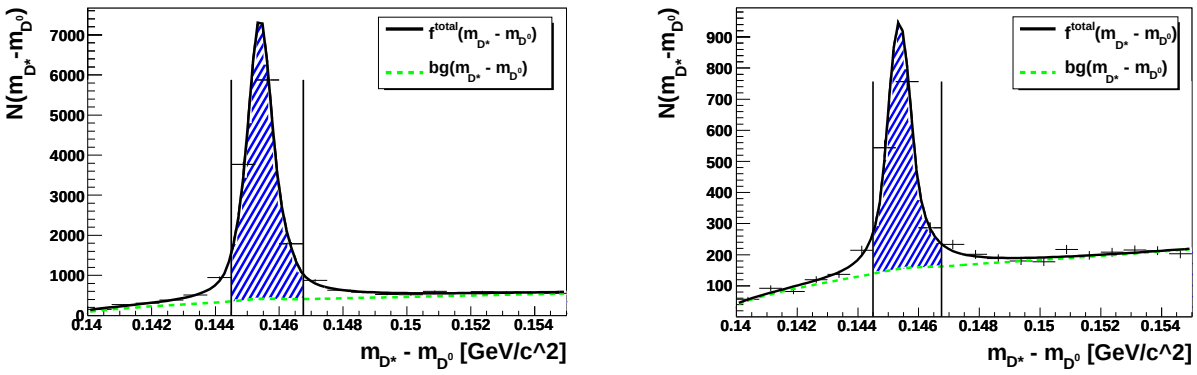
Figure 4.1: Representative sample of tracks likelihood-selected as $\{e,\ \mu,\ \pi,\ K$ or $p\}$ from $0.4 \cdot 10^6$ hadronic events, plotted on the (p_{lab}, cosθ_{lab}) plane. The plot also contains an illustration for the binning in (p_{lab}, cosθ_{lab}) chosen in this analysis. The kinematic area considered in this analysis is highlighted in green. The highlighted range in the polar angle roughly corresponds to the geometrical acceptance of the ToF detector.

After assigning masses m_K, m_π, m_π to the ordered triplet tracks, the invariant mass difference $m_{D^*} - m_{D^0}$ can be formed. The distribution of this variable is sharply peaked around 145 MeV/c^2. Fitting the peak with the sum of a signal and a background function and integrating the signal part over a defined interval around the peak allows to extract an approximation for the number of 'true' reconstructed D^* particles. This number is equal to the number of 'true' kaons, slow and fast pions in the sample, yielding numbers N_i with 'real'/physical species $i\ =\ \pi,\ K$. For illustration, Figure 4.2 a) shows the distribution for $m_{D^*} - m_{D^0}$ where the kaon tracks fulfill p_{lab} in $[1.4; 1.6)$ GeV/c and $cos\theta_{\mathrm{lab}}$ in $[0.209; 0.355)$.

The procedure described above is repeated to form a second distribution, but now additionally imposing a likelihood-cut on the PID likelihood variables of the, e.g., kaon tracks. If the likelihood-cut is constructed to select e.g. pions, the integral of the signal function over the same interval as before around the peak of the distribution yields the number $N_{K\ \to\ \pi}$, the number of 'true' kaons being likelihood-selected as pions. As a result, the PID probability $\mathrm{p}_{p\ \to\ K}$ can be calculated by the ratio of $N_{K\ \to\ \pi}$ over N_K, or more generally, by

$$\mathrm{p}_{i\ \to\ j}\ \doteqdot\ \frac{N_{i\ \to\ j}}{N_i}. \tag{4.3}$$

. The probability uncertainty is determined by propagating the uncertainties on N_K from fit parameters and on $N_{K\ \to\ \pi}$ from a binomial approach through the yield ratio. Figure 4.2 b) shows the distribution of Figure a) with an additional cut on the kaon track likelihoods to select pions.



(a) (b)

Figure 4.2: Fits of experimental data invariant mass distributions from a sample of decay products of $D^* \to \pi(D^0 \to K\pi)$. Figures a) and b) show distributions of track triplets containing a negatively charged track with assigned kaon mass, without and with an additional cut on the track PID likelihood values to select the species $j = \pi$. Extracting the PID probability $p_{(K^- \to \pi^-)}$ from the ratio of the hatched areas yields 0.111 ± 0.004 for negatively charged kaons with laboratory frame momentum in $[1.4; 1.6]$ GeV/c and laboratory azimuthal angle θ in $[77.9; 89.0]$ degrees.

By choosing different likelihood cuts to form the second distribution, all PID probabilities $\mathrm{p}_{K \to j}$ with $j = \{e,\ \mu,\ \pi,\ K,\ p\}$ can be extracted. To extract probabilities $\mathrm{p}_{\pi \to j}$ instead, the likelihood cuts are applied to the 'fast' pion tracks. The same analysis is performed in each kinematic bin where statistics allow to still fit the distributions with reasonable precision, such that PID matrices can be filled for as many kinematic bins as possible,

$$\mathrm{p}_{i \to j\ (\mathrm{charge},\ p_{\mathrm{lab}},\ \cos\theta_{\mathrm{lab}})} \doteq \frac{N_{i \to j\ (\mathrm{charge},\ p_{\mathrm{lab}},\ \cos\theta_{\mathrm{lab}})}}{N_{i\ (\mathrm{charge},\ p_{\mathrm{lab}},\ \cos\theta_{\mathrm{lab}})}}. \tag{4.4}$$

Necessity of Re-Extracting PID Probabilities Despite Existing Belle PID Studies The existing PID studies at Belle have extracted PID probabilities for all diagonal and the largest off-diagonal elements of the matrix $\mathcal{P}$. However, the binning in these analyses is for the most part quite different to the one presented in this analysis, and usually track/particle charges are not resolved in the probability extractions. Also, since the extracted probabilities have a unique relation to the used likelihood-cuts, the results of existing PID studies with likelihood cuts different to the ones used here are not directly usable in the present analysis. Finally, for the inversion of the matrix $\mathcal{P}$, all possible fake rates have to be known as well with sufficiently high precision, only few of which have been extracted in existing Belle PID studies.

In the following, detailed descriptions about the extraction procedures of all PID probabilities will be given. All fits have been performed using the class TMinuit from the software analysis environment ROOT. Throughout, the fit iterations have been increased from the default value of 5000 to 10000. The fit precision

has been kept at the default value, 10^{-6}.

4.1.2 Extraction of $p_{\pi,\ K\ \to\ j}$ PID Probabilities from D^* Decays

The internal Belle analysis notes 'Belle Notes' (BN) 321 [66] and 779 [65] have performed pion and kaon PID probability extractions on samples of D^* decay products. Especially BN 779 is taken as a reference for this part of the present analysis.

Selection of D^* Decay Product Triplets

The procedure of extracting track candidate triplets for the decay $D^{*+}(2010) \rightarrow \pi^+_{slow} + D^0(1865) \rightarrow \pi^+_{slow} + (K^-\pi^+_{fast}),\ +\ c.c.$ has already been introduced in Section 4.1.1. To make the candidate extraction from the vast Belle data sets more efficient, sets of preselected events already exist at Belle in the form of the kaon identification (KID) skim data files.

KID Skim An event from the general Belle hadronic event skim (HadronBJ) is recorded in the KID skim files if the criteria given in Table 4.2 are met by at least one track triplet. The invariant mass of a kaon and a pion track pair should not differ from the nominal D^0 mass by more than 0.1 GeV/c^2, the center-of-mass (cms) momentum of D^* candidates scaled to the quark energy has to be larger than 0.35, and the invariant mass difference between the D^* and the D^0 candidates, Δm, is required to be smaller than 170 MeV/c^2.

Observable	Criterion
$m_{D^0} - m_{K\pi_{\mathrm{fast}}}$	$(-0.1; 0.1)$ GeV/c^2
$\frac{\lvert\overrightarrow{p}^{cms}_{\pi_{\mathrm{slow}}K\pi_{\mathrm{fast}}}\rvert}{\sqrt{s}/2}$	$(0.35; 1.0)$
Δm	< 0.17 GeV/c^2

Table 4.2: Selection criteria of the Belle KID skim for D^* candidates, taken from Ref. [67]. All quantities are explained in the text.

The amount of statistics used to extract $p_{\pi,\ K\ \to\ j}$ PID probabilities comprises all KID skim on-resonance and continuum data from experiments $7 - 55$, corresponding to about $43 * 10^6$ events.

Analysis Selection Cuts All events are required to pass the HadronBJ hadronic event selection cuts. Additional cuts are applied to track triplets found from KID skim events to reject combinatorial background. These cuts are derived from BN 779 and are given explicitly in Table 4.3.

Selections are imposed on track kinematics to ensure proper tracking: Tracks are only considered for forming triplets if their laboratory frame scattering angle θ falls within the CDC tracking volume of $[17; 150)$ deg. For reliable tracking, a minimum laboratory frame momentum of 100 MeV/c is required- a compromise between the momentum necessary to traverse the barrel CDC (about 200 MeV/c) and the minimum momentum to reach the CDC, about 70 MeV/c. The slow pion momentum distribution peaks at about 190

MeV/c.

Additional requirements are applied to track triplets. The second cut in Table 4.3 refers to a rejection of D^0 decays producing a pion and a kaon with close momenta, where the mass assignment cannot be unambiguously decided anymore by the track charge. Thus a cut is imposed requiring the invariant mass of a pion and kaon track pair with swapped masses to lie outside the D^0 mass window [66]. Additionally, a fit of the kaon and pion candidate tracks to a common vertex is required to be successful.

Another cut to prevent ambiguous track triplets from entering the analysis is made on the angle between the momentum of the kaon candidate in the D^0 rest frame and the center-of-mass momentum of the D^0 candidate. This cut rejects each D^0 candidate the momentum of which is parallel or oriented back-to-back to its kaon decay product [66]. In such decays, the kaon takes over most of the energy available and the fast pion can hardly be distinguished from the slow pion from the D^* decay. The kaon candidate is boosted into the D^0 rest frame to give a better resolution of the angle $\theta_{K\ D^0}$.

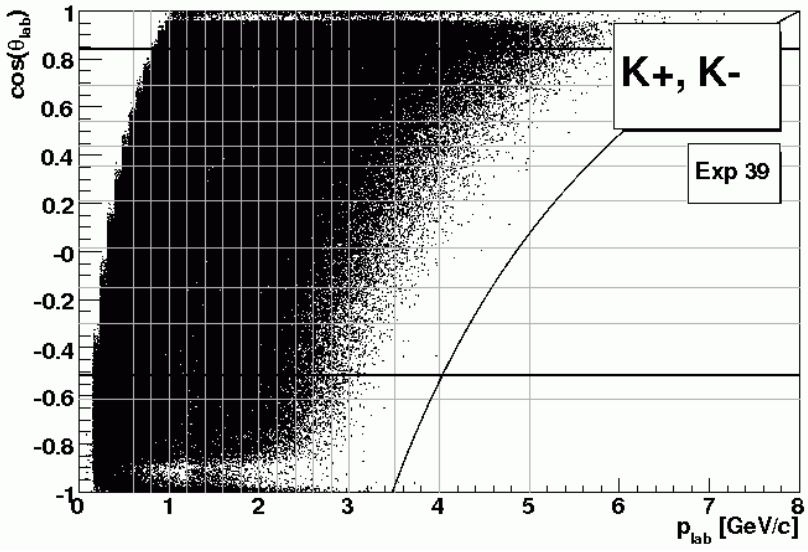
Finally, a lower limit on the D^* momentum in the center-of-mass frame also ensures that the D^0 decay products are sufficiently asymmetric in kinematics to be unambiguously taggable by their charges [65]. The high limit corresponds to the momentum of a D^* particle with the maximum center-of-mass energy achievable, $\sqrt{s}_{\text{on res}}/2$,

$$|\overrightarrow{p}^{\text{max cms}}_{\pi_{\text{slow}}K\pi_{\text{fast}}}| \cdot c = \sqrt{(\sqrt{s}_{\text{on res}}/2)^2 - m^2_{D^*} \cdot c^4}, \tag{4.5}$$

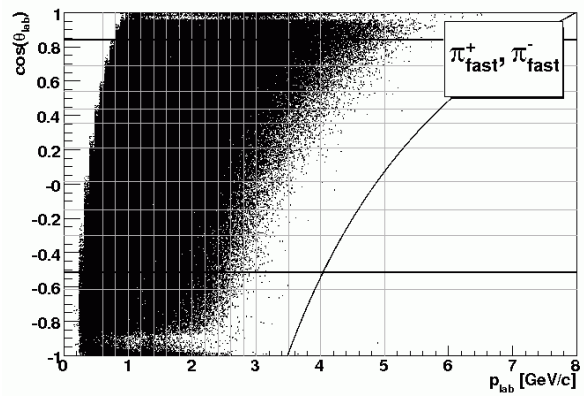
to reject 'ghost tracks'. [3] In order to keep all valid D^* candidate momenta rather than cutting out all background contributions, $\sqrt{s}_{\text{on res}}$ instead of $\sqrt{s}_{\text{cont}}$ is used here. Eventually, all selected candidate triplet Δm values are filled in histograms ranging from 140 to 155 MeV/c^2.

As an illustration for the available statistics in the kinematic bins analyzed in the present analysis, Figures 4.3 show all selected D^* decay particles K, π_{fast} and π_{slow} from all Experiment 39 KID skim files, at the on-resonance as well as the continuum energy level. As expected, the slow pion decay products only cover a very confined kinematic region and therefore will not be considered in the extraction of pion PID probabilities $\text{p}_{\pi \to j}$.

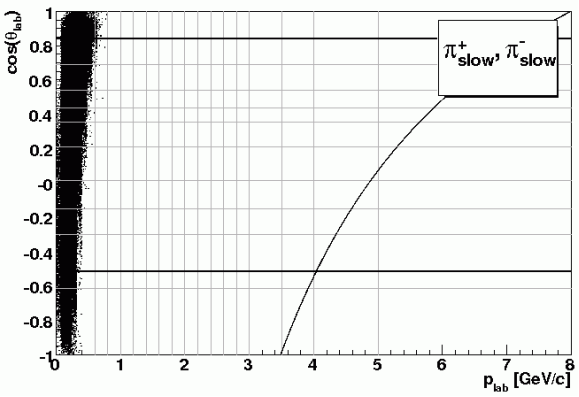
[3] Generally, tracks are reconstructed from detector measurements which are produced by one particular particle. Sometimes independent measurements from different particles are accidentally combined in the reconstruction process. These measurements form so-called 'ghost' or 'fake' tracks, as they do not represent the trajectories of actual particles. Ghost tracks can feature unphysical values of track observables, e.g. normalized hadron energies $z > 1.0$.



(a)



(b)



(c)

Figure 4.3: Illustrations of the kinematic coverage of the D^* decay products K^+, K^- (a), $\pi^+_{\rm fast}$, $\pi^-_{\rm fast}$ (b) and $\pi^+_{\rm slow}$, $\pi^-_{\rm slow}$ (c) selected from Experiment 39 KID skim data. The two accentuated horizontal lines denote the geometrical acceptance limits of the ToF subdetector, also representing the acceptance limit for the present analysis. The curved line on the right hand side of each Figure shows the physics limit $z = 1.0$ for the presence of kaons or pions, respectively, in the laboratory frame.

Observable	Criterion
$m_{K\pi_{\rm fast}}$	$(1.835; 1.895)$ GeV/c^2
$m^{rev}_{K\pi_{\rm fast}}$	$\notin$ $[1.835; 1.895]$ GeV/c^2
θ_{KD^0}	$(-0.8; 0.8)$
$\lvert\overrightarrow{p}^{cms}_{\pi_{\rm slow}K\pi_{\rm fast}}\rvert$	$(2.0; 4.89)$ GeV/c
Δm	$(0.0; 0.155)$ GeV/c^2

General Track Selection for K, π_{fast}, π_{slow}

Observable	Criterion
p_{lab}	$>$ 100 MeV/c
θ_{lab}	$[17; 150)$ deg

Table 4.3: Set of criteria to select D^* candidate track triplets from the Belle KID skim files as applied in the present study. All cuts are taken or adapted from Ref. [65]. All quantities are explained in the text.

PID Probability Extraction

In order to still be able to derive a PID statement also in kinematic areas with very low statistics, all histograms are segmented into only 20 bins and are fitted if they contain more than or equal to 180 entries. The signal and background shapes of the invariant mass distributions are quite complex and require many-parameter fit functions. Extensive Monte Carlo studies have been performed to find a flexible but yet low-statistics-stable functional description of the signal and combinatorial background shapes.

To obtain a well converging fit despite low statistics and only 20 data points, the following procedure has been applied for each kinematic bin and PID selection.

Fit Procedure In order to obtain a representative description of the combinatorial background by the background fit function, at first an auxiliary histogram is created by replacing the signal peak with a linear interpolation between the closest bins to the interval $[0.1425; 0.1515)$ GeV/c^2. This interval roughly describes the invariant mass range where signal tails are still significant, which was estimated in Monte Carlo studies. The uncertainty of the bin contents in the interpolation interval is doubled in order to allow for background fits to potentially behave non-linearly under the signal peak.

To fit the combinatorial background of the Δm invariant mass distribution, the threshold function used in BN 779 was employed,

$$bg(\Delta m) \;\div\; par^{\rm bg}_{scale}\;(\Delta m \;-\; 0.1395)^{par^{bg}_0}\;e^{\;par^{bg}_1\;(\Delta m\;-\;0.1395)}. \tag{4.6}$$

Separate starting parameter values are assigned from Monte Carlo studies for each histogram with different PID selections, and parameter limits are imposed on the background function parameters $par^{bg}_{0,1}$ of $(-10; 1000)$ and $(-1000; 1000)$, respectively, to prevent errors when using the $'I'$ integral fit option in

ROOT's TMinuit.

In some kinematic areas and for certain PID cuts on the tracks filling the histogram, no significant D^* signal peak is present in the distribution. In order to obtain a means to quantitatively check if such a situation is given for a histogram under study, the full data histogram including the signal peak is fitted with the background function $bg(\Delta m)$ alone ('pure background fit'). As initial conditions for this fit, the fit parameter results and bounds from the fit to the auxiliary histogram from the previous step are assigned. If the parameter results of the pure background fit stay below their bound limits, the probability from integrating the χ^2 distribution from the obtained value of χ^2 is kept in memory for later comparison to the fit including the signal function.

Next, the data histogram is fitted with a combined fit function of signal and background. During initial Monte Carlo studies, a physical background contribution in the invariant mass distributions was found, originating from a 3-prong decay of the D^0 passing through all selection cuts. This contribution is accounted for by an additional fit function consisting of one or two single Gauss functions, depending on which functional choice described the 3-prong peak in Monte Carlo better. All parameters for these Gaussian functions have been optimized for each histogram with and without PID selection in Monte Carlo studies on integrated kinematics. For the fitting procedure on experimental data, the amplitude, mean and width parameters of these Gaussians are fixed relative to the corresponding parameters of the actual signal fit function.

The signal peak description in the invariant mass distributions is performed by a modified 'Crystal Ball' function. The modification achieves a function which consists of power-law tails for invariant mass regions far from the mean,

$$sig_{\text{tail}}(\Delta m) \;\doteqdot\; par^{\text{sig}}_{scale} \; \frac{(\frac{par^{\text{sig tail}}_{0a}}{abs(par^{\text{sig tail}}_{1a})})^{par^{\text{sig tail}}_{0a}} \; e^{-0.5 \; abs(par^{\text{sig tail}}_{1a})}}{(\frac{par^{\text{sig tail}}_{0a}}{abs(par^{\text{sig tail}}_{1a})} - par^{\text{sig tail}}_{1a} - \frac{\Delta m - par^{sigGauss}_{0}}{par^{sigGauss}_{1a}})^{par^{\text{sig tail}}_{0a}}} \tag{4.7}$$

and a single Gaussian function for the description of the peak area,

$$sig_{\text{Gauss}}(\Delta m) \;\doteqdot\; par^{\text{sig}}_{scale} \; e^{-0.5 \; (\frac{\Delta m - par^{sigGauss}_{0}}{par^{sigGauss}_{1a}})^2}, \tag{4.8}$$

which are independent on each side of the mean. The normalization parameter par^{sig}_{scale} is the same for both function parts. The function is designed to be fitted asymmetrically, such that both sides around the mean can have different power-law and Gaussian parameters $par^{\text{sig tail}}_{0a,b}$, $par^{\text{sig tail}}_{1a,b}$ and $par^{sigGauss}_{1a,b}$, respectively. Continuity with respect to the mean is ensured by proper normalization of the single function parts. Apart from this asymmetric Crystal Ball function, in some PID channels adding an extra single Gaussian

function to the signal function improved the fit of the signal peak. The single Gaussian then shares the mean and sigma parameters with the Gaussian part of the Crystal Ball signal function and is fixed relative in size compared to the Crystal Ball part.

To fit the data histogram including the signal peak with a combined signal and background function, the background function is assigned the fit results from the auxiliary histogram fit. Starting parameters for all signal function parts are again assigned from extensive Monte Carlo studies for each histogram with and without PID on integrated kinematics. To reduce parameters to be fitted, the widths of all involved Gaussians in the signal function are assigned fixed relations to one single width fit parameter. For the same purpose, only the scaling parameter $par^{\mathrm{sig}}_{scale}$, the mean parameter $par_0^{sigGauss}$ and the joint width parameter of the signal functions are fitted. All other signal parameters are fixed for the fit. Bounds are again assigned to the background parameters $par^{bg}_{0,1}$. Better fit results were achieved by not fixing the background parameters at this step due to the long tails in the D^* signal which interfere with the background fit function.

If the fit of the full data histogram is successful, the histogram is refitted with the same fit function, taking the results of the previous fit as starting values for all fit parameters. All background parameters now are fixed as well, only the three signal function parameters (scale factor, mean position and joint width) are fitted, with 17 degrees of freedom. The results of this fit ('full fit function fit'), if successful, are used to extract PID probabilities. This fit procedure is repeated for each data histogram.

The cuts imposed on PID likelihood variables to select $\{e, \mu, \pi, K, p\}$ are not complete and thus leave some tracks unselected. Those tracks are collected in histograms labeled $unsel$. In all cases throughout the PID correction, the effect of those unselected tracks on the PID correction has been found to be below about 1%. Joint e/μ probabilities have been extracted from all PID studies to match the impossibility to distinguish between the electron and muon decay channels of the J/ψ particle with purely kinematical means, cf. Section 4.1.4. This will be revisited in future improvements.

Extraction of PID Probability Ratios The successfulness of fits are determined by comparing the probabilities from integrated χ^2 distributions for both the fits with and without signal part. Several possible cases are considered, a few of which are mentioned in the following: If the full fit function fit of a histogram without PID selection, e.g. filled with pion track kinematics, fails or the pure background function fit obtains a higher fit probability, no PID probability $\mathrm{p}_{\pi \,\to\, j}$ can be calculated and no PID statement can be extracted from the respective kinematic bin. If the pure background fit obtains a higher probability than the full fit function fit for a histogram with additional PID selection, so-called 'zero-PID-probabilities' are calculated, cf. Section 4.1.2. If the full fit function fit obtains a higher probability than the pure background fit, the

significance of the peak in the histogram is tested. This should avoid cases where the full fit function fits a statistical fluctuation in the absence of an actual signal and thus achieves a higher probability than the pure background fit.

In case both fits of the histogram without additional PID selection for species i and with additional PID selection $i \rightarrow j$ have been successful in a given kinematic bin, the corresponding PID probability can be extracted. As introduced in Section 4.1.1, the signal parts (not including the potential contribution from the D^0 3-prong decay) of the fit results for histograms with and without PID selection are integrated to obtain yields $N_{\pi,\ K \rightarrow j}$ and $N_{\pi,\ K}$, respectively. In order to be able to compare yields $N_{\pi,\ K \rightarrow j}$ to yields $N_{\pi,\ K}$, the integral interval has been fixed to $[0.1445; 0.14675]$ GeV/c^2, corresponding to the limits of 3 histogram bins around the signal peak. A simple ratio $\frac{N_{\pi,\ K \rightarrow j}}{N_{\pi,\ K}}$ yields the PID probability.

The uncertainties on the extracted PID probabilities are the uncertainties of the ratio components propagated through the ratio $\frac{N_{\pi,\ K \rightarrow j}}{N_{\pi,\ K}}$. Since the histograms $N_{\pi,\ K \rightarrow j}$ and $N_{\pi,\ K}$ are highly correlated (in fact, they are exactly the same except for an additional PID cut to select species j on all contributing tracks of species i), all yields $N_{\pi,\ K \rightarrow j}$ are considered to obey binomial statistics rather than Gaussian/Poisson statistics. Thus only the fit parameter uncertainties in the integrals forming the yields $N_{\pi,\ K}$ are propagated through the ratio. This is achieved by following Gaussian uncertainty propagation, so applying partial derivatives by each fitted parameter to the signal function and integrating each term in the integration interval. All terms are multiplied by their respective fit parameter uncertainty from TMinuit and summed up. The binomial uncertainties of yields $N_{\pi,\ K \rightarrow j}$ are also propagated through the ratio.

Some ratios yielded PID probabilities larger than 1.0. Most of them remain consistent with values < 1.0 if the probability uncertainties are taken into consideration. However, since binomial uncertainties are only defined for probabilities < 1.0, all ratios amounting to values $>= 1.0$ are set to 0.99 by default.

Extracting PID Statements without Successful Full Fit Function Fit- 'Zero Probabilities'
'Zero-probabilities' are calculated from a histogram in three cases: The pure background function fit is successful for a histogram with PID selection and the full fit function fit is not; the full fit function fit achieves only a lower probability than the pure background fit; the full fit function fit fails the signal peak significance test. The PID probability value for this kinematic bin and PID channel is set to 0.0, since no signal peak can be analyzed.

However, one can estimate the sensitivity of the 0.0-statement by extracting the largest possible signal which could 'hide' behind the statistical fluctuations found in the histogram with additional PID selection. The largest net (positive) fluctuation in the histogram can be calculated by examining the deviations of bin

contents of any bin-triplet (the same width as the integration interval) with respect to the result of the pure background fit. Then the signal equal to the largest fluctuation, divided by the integrated signal yield of the fit on the histogram without PID selection, gives the largest possible PID probability value extractable from the histogram. This value is assigned as the uncertainty of the zero-probability statement.

Fit Statistics Summing up kinematic bins (153), resolved charges (2), number of decay products from D^* decays (3) and histograms without and with PID selection $(1 + 5)$, the total number of data histograms to be fitted amounts to 5508. However, statistics were not sufficient for fitting in 2857 bins- notably, almost all histograms for π_{slow} and most of the histograms with the *unsel* PID selection contained less than the required 180 entries. So only 2633 histograms were fitted. Of these, 2239 full fit function fits were successful and 128 zero probabilities were extracted based on successful pure background fits. A detailed manual examination of each fit has been performed, and 77 fits have been rejected due to anomalies in the fit results.

Sample fits of data histograms from different kinematic bins and PID selections are shown in Figure 4.4. The code in brackets given in each pad refers to the respective $(p_{\mathrm{lab}},\ \cos\theta_{\mathrm{lab}})$ bin and the PID selection on the tracks filling the histogram, where applicable. The sample distributions are taken from bins in the ToF backward, barrel and forward region with either low, medium or high laboratory frame momentum, with bin numbers as defined in Table 4.1.

As additional information, the PID probability extracted using data from the fit are shown as well, where applicable. The global integration interval is indicated by two vertical black lines. The two green lines indicate the window within which the data distribution is approximated by a linear interpolation to create the auxiliary histogram (which is fitted to obtain initial fit parameters for the final background function). The black curve indicates the full fit function, the green curve the background function including the D^0 3-prong contribution, visible as slight bumps under the signal peak. The rightmost pad in the top row shows a data histogram where the full fit function fit was successful, but the signal peak did not achieve enough statistical significance compared to statistical fluctuations. Thus a zero-probability was extracted. The pure background fit is shown in black, the full fit function fit in red. The first pad in the second row shows a fit which has been rejected in the manual fit examination process.

In Appendix B.3, cumulative χ^2 plots are shown, separated by charge, different physical species i (fast pion and kaon), and different additional PID cuts, but integrated over all kinematic bins. All histograms are fitted with the corresponding theoretical χ^2 distribution for 17 degrees of freedom, the fit is performed by scaling the theoretical distribution to the χ^2 histogram. The scale parameter fit result is mostly close to 1.0, and overall the distributions show good agreement with the fit χ^2 values. We conclude that the

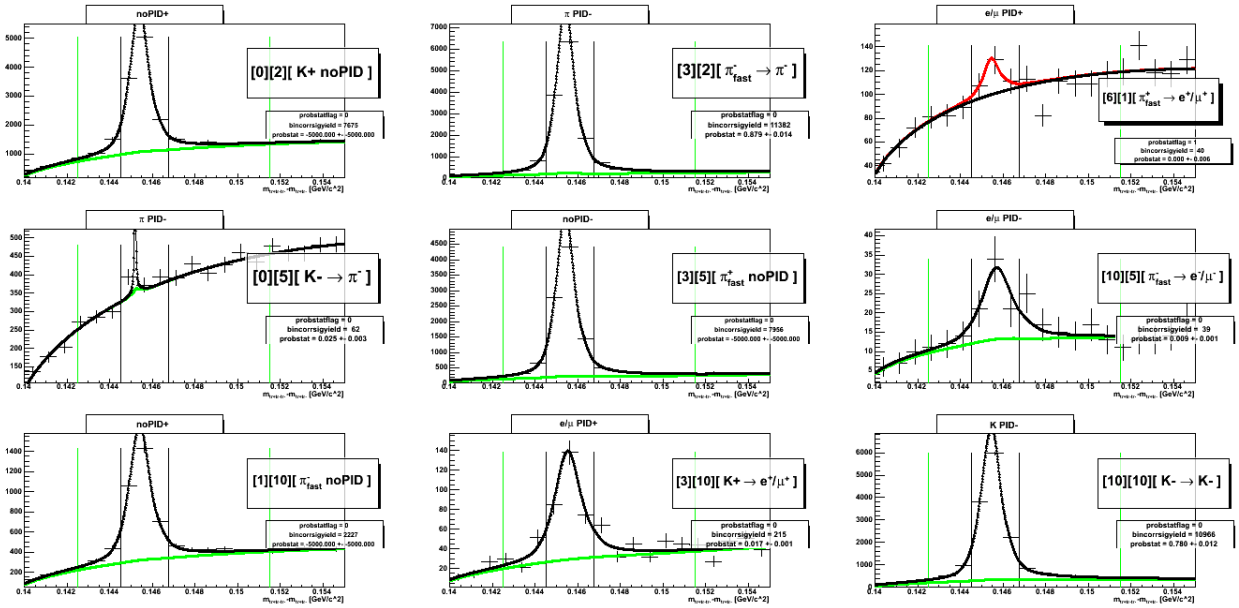


Figure 4.4: Nine analyzed sample distributions from the extraction of PID probabilities on the set of D^* candidates. Further descriptions of the plots can be found in the text.

parameterization chosen is valid and describes the data distributions well.

Also shown in the Appendix are 2-dimensional plots of all kinematic bins, on a plane in the laboratory frame momentum and the cosine of the laboratory frame scattering angle θ. The color coding is explained in the legends of each plot; white kinematic bins indicate that the fit has been rejected by hand because of anomalies. White lines describe kinematic areas with equal normalized *cms* hadron energy z, the rightmost line indicating $z = 1$. Judging from these plots, the 'kinematic coverage', so the area where PID statements can be extracted from D^* decays, is fairly large and covers most of the low and medium z, even high z for PID efficiencies $\mathrm{p}_{\pi,K \to i}$. Missing probabilities will have to be calculated from extrapolation of available probabilities, which will be described in Section 4.1.5.

Samples of Extracted PID Probabilities $\mathrm{p}_{\pi,\ K \to j}$ from D^* Decays

Figures 4.5 and 4.6 show samples of extracted PID probabilities from D^* decays. The different subfigures give plots of probabilities $\mathrm{p}_{K^+ \to j^+}$ and $\mathrm{p}_{\pi^- \to j^-}$, respectively, for all likelihood-cut selected species j. The probabilities are plotted against p_{lab} in different scattering angle intervals (top 5 pads) and against $\cos\theta_{\mathrm{lab}}$ for different p_{lab} bins (bottom 3 pads) at the center of each bin. Kinematic areas where no PID statement could be extracted from the data histograms are indicated by markers plotted below 0.

The top 5 pads show probabilities for tracks travelling through the backward, backward/barrel, barrel, forward/barrel and forward regions in $\cos\theta_{\mathrm{lab}}$, respectively. Analogously, the lowest three plots show the

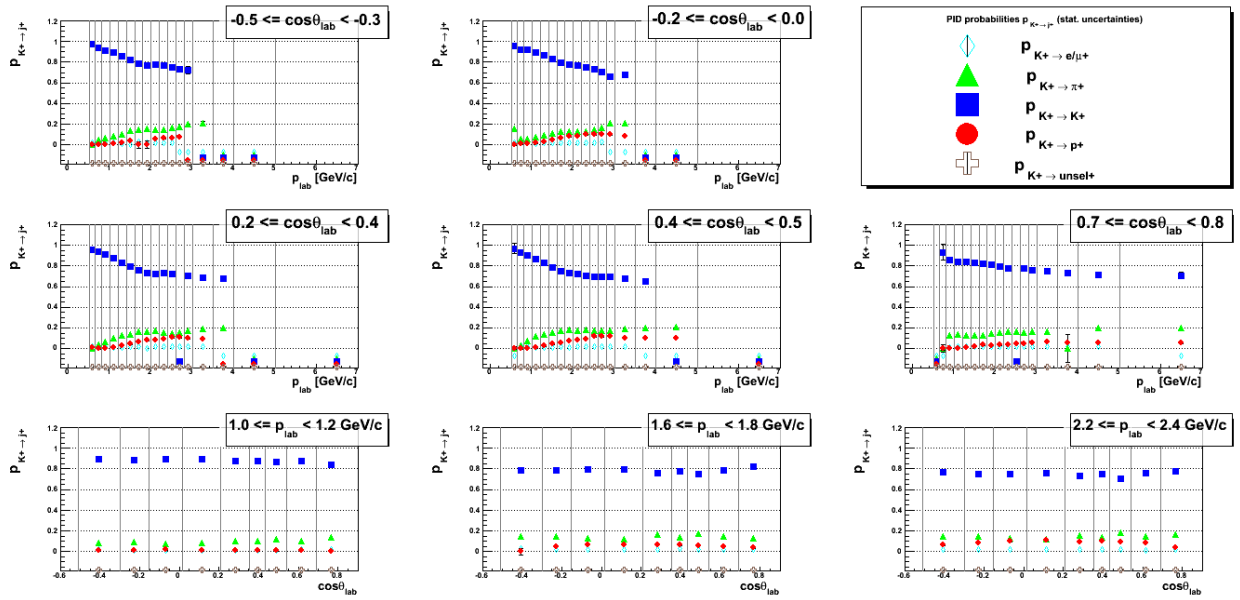


Figure 4.5: Samples of extracted PID probabilities $p_{K^+ \to j^+}$ from D^* decays. The different subfigures give plots of probabilities for all likelihood-cut selected species j against p_{lab} in different scattering angle intervals (top 5 pads) and against $\cos\theta_{\mathrm{lab}}$ for different p_{lab} bins (bottom 3 pads) at the center of each bin.

respective probabilities against $\cos\theta_{\mathrm{lab}}$ for tracks with low, medium and high momentum. Uncertainties are given by error bars for all probabilities.

Consistency Check for Extraction of Fit Probabilities Values for the integrated signal yields $N_{\pi,\,K\,\to\,j}$ and $N_{\pi,\,K}$ can also be obtained by an alternative analysis to check the quality of the fit for the signal peak in each data histogram. First, the data histogram bin contents of bins in the integration window specified above are summed up. Then the integral of the full fit function fit background part (retaining the D^0 3-prong contribution) is performed over the same integration interval and is subtracted from the summed histogram bin contents. The result represents a comparison value for the integrated signal yield from the full fit function fit. This procedure is performed for both histograms with and without PID selection.

With the so obtained alternative values for all integrated signal yields, one can extract alternative PID probabilities to test the effect of possible deviations between fits and data shapes on the extracted probabilities. The uncertainties of such alternative PID probabilities can be calculated again by applying a binomial approach for the numerator, and for the denominator by propagating the Poisson uncertainty from the bin content sum and also propagating all background fit parameter fit uncertainties through the ratio. Ratios larger than 1.0 are again set to 0.99 to respect the binomial interpretation of the uncertainty on the yields

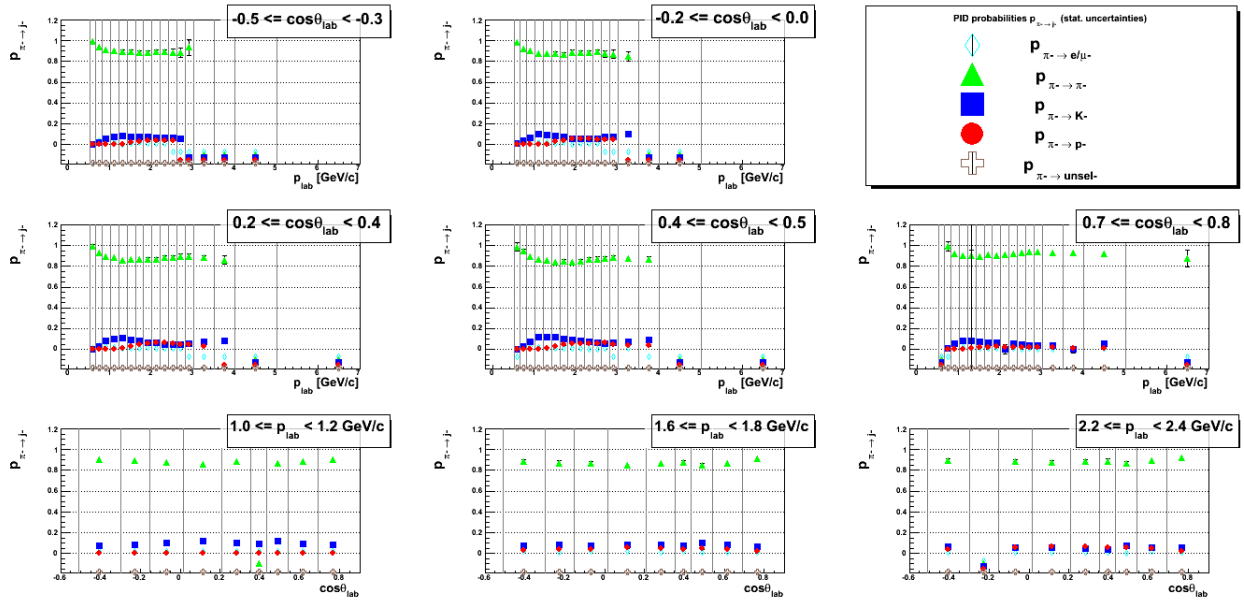


Figure 4.6: Samples of extracted PID probabilities $\mathrm{p}_{\pi^- \to j^-}$ from D^* decays. The different subfigures give plots of probabilities for all likelihood-cut selected species j against p_{lab} in different scattering angle intervals (top 5 pads) and against $\cos\theta_{\mathrm{lab}}$ for different p_{lab} bins (bottom 3 pads) at the center of each bin.

extracted from histograms with additional PID selection.

Distributions of differences to the probabilities extracted from integrals of the fitted signal function parts are given in Appendix B.3. Most absolute differences remain below 1% and are centered around 0. Under consideration of the uncertainties of both sets of probabilities, differences are found to stay within 1σ of the difference. No significant deviation was found. The signal fits and the extracted probabilities are thus assumed to be reliable.

Discussion of Extracted Pion and Kaon PID Probabilities

The probabilities shown in Figures 4.5 and 4.6 feature an obvious dependence on the laboratory frame momentum, whereas they only slightly vary in different regions of the polar angle θ_{lab}. The polar angle dependence of the extracted probabilities is mostly induced by the ACC detector, as it represents the only PID subsystem in Belle which was intentionally designed differently for different polar angle regions. In the following, an attempt to motivate the dependence of the shown probabilities on the momentum will be given discussing the interplay of the subdetectors used to extract PID likelihoods.

Discussion of Kaon PID Probabilities $\mathbf{p}_{K \to j}$ Hadron PID likelihood values are extracted by combining measurements from the drift chambers, the ACC and the ToF detectors, cf. Appendix 2.2. Especially

for kaons, the measurements in the drift chambers and the ToF detector provide good discrimination power versus pion and proton/antiproton misidentification, which is reflected by the high kaon efficiencies and low fake-rates in the first momentum bins. However, the high kaon efficiency at low momenta deteriorates with increasing momentum. To explain this development, the lines of expected characteristic energy loss for pions and kaons in the drift chambers shall be considered. These lines approach each other with increasing momentum and finally intersect at a momentum of roughly 1 GeV/c, cf. Figure 2.4 in Appendix 2.2. This causes the discrimination power from the drift chamber measurements to vanish at higher momenta.

Additionally, the measured times-of-flight for pions and kaons in the ToF detector also become comparable as illustrated in Table 2.2 in Appendix 2.2. Table 2.2 also suggests that the difference in times-of-flights for pions and kaons is of the order of the time resolution given for the ToF detector at momenta around 2 GeV/c. This causes the discrimination power of the ToF measurements for pion and kaon separation to decrease at higher momentum. However, the descent of the kaon efficiency and the respective ascent in the kaon-pion fake-rate is found to be rather modest. This can be attributed to the strong pion-kaon discrimination provided from the ACC measurements for track momenta between about 0.8 and 3 GeV/c depending on the polar angle of the tracks, as given in Table 2.1 in Appendix 2.2. The ACC discrimination at least modestly counterbalances the losses of discrimination power in the drift chambers and the ToF detector.

The kaon-proton/antiproton misidentification rates are slightly lower than the kaon-pion fake-rate. This can be related to the slightly better discrimination power provided by the drift chamber and especially the ToF measurements for kaon and proton/antiproton separation than for kaon and pion separation, cf. Figure 2.4 and Table 2.2. The decrease of the proton/antiproton fake-rate at momenta around 3 GeV/c and in turn the slight increase of the kaon efficiency is most likely caused by the onset of additional discrimination power from ACC measurements. The threshold momenta given in Table 2.1 suggest good kaon and proton/antiproton separation for momenta above about 2.5 GeV/c, depending on the polar region of the track.

All kaon-lepton fake-rates remain at the order of 1% or below. For electrons/positrons, this low misidentification rate is most probably caused by the significantly different shape of the lines of expected energy loss for kaons and electrons/positrons in the drift chambers. Furthermore, the ACC gives good discrimination between kaons and electrons/positrons as the Cherenkov threshold for the latter is situated at momenta of a few MeV/c, in contrast to threshold momenta for kaons of about 2.8 GeV/c. Finally, the difference between hadronic and leptonic shower shapes in the ECL detector most probably largely contributes to the low misidentification rate. Muon likelihoods are solely extracted from measurements in the KLM detector.

The difference in interaction properties of muons and all other particles presumably allows to very well separate kaons from muons in any momentum and polar angle region.

Discussion of Pion PID Probabilities $\mathbf{p}_{\pi \to j}$ As all hadron likelihoods are extracted from the drift chambers, the ACC and the ToF measurements, most of the reasoning given for kaon PID probabilities can be applied to pion probabilities too. Again decreasing from high values at low momentum, the pion efficiency only slightly drops because of the loss of discrimination power in the drift chambers at momenta of 1 GeV/c as explained above. The pion efficiency reaches a minimum due to the loss of the ToF measurement significance against kaons around 1.3 GeV/c. However, additional discrimination power is provided from the ACC measurements above roughly 0.8 GeV/c, as at particle momenta higher than the pion Cherenkov threshold, pions can well be separated from kaons and protons/antiprotons. This seems to cause the pion efficiency even to slightly increase at momenta above 1.5 GeV/c. Once the discrimination power versus kaons provided by the ACC measurements is lost at the kaon Cherenkov threshold of roughly 2.8 GeV/c, the pion efficiency slightly drops for higher momenta.

The pion-kaon fake-rate reaches a local maximum at momenta of about 1 GeV/c. This is assumed to be caused by the intersection of the lines of energy loss for pions and kaons in Figure 2.4, which in turn does not allow to use drift chamber measurements for pion-kaon separation anymore. The fake-rate then decreases again for higher momenta as the ACC provides good pion-kaon discrimination for momenta above the pion Cherenkov threshold. A second maximum is reached by the pion-kaon fake-rate at momenta above the kaon Cherenkov threshold, which implies that the ACC measurements cannot be used anymore for pion-kaon separation for higher momenta.

The misidentification of pions as protons/antiprotons is low in general. Judging from Table 2.2, the pion-proton separation gained from time-of-flight measurements gradually deteriorates with increasing momentum. Nevertheless, the ToF measurements allow to separate pions from protons/antiprotons well up to momenta of 2 GeV/c, considering the time resolution of the ToF detector of 100 ps, cf. Table 2.2. The ACC measurements are expected to continuously give good discrimination for virtually all momenta above the pion Cherenkov threshold at roughly 0.8 GeV/c. At momenta below about 1 GeV/c, the drift chambers provide good discrimination along with the ToF measurements to separate pions from protons/antiprotons. However, the lines of expected energy loss for pion and proton/antiproton tracks in the drift chambers intersect at momenta slightly below 2 GeV/c. This causes the discrimination power of the dE/dx measurements to vanish, as already described above. The loss of discrimination power from the drift chambers could explain the increase of the pion-proton/antiproton fake-rate to a maximum at momenta of about 2 GeV/c. The

lines of expected energy loss separate again for higher momenta, which might explain the following descent of the pion-proton/antiproton fake-rate.

Analogously as mentioned in the discussion of the kaon PID probabilities, all lepton fake-rates remain below 5%. The same reasons as given in the kaon probability discussion can again be applied to explain the low pion-lepton fake-rates.

Summary of the Pion and Kaon PID Probability Extraction from D^* Decays

Summing up the extraction of kaon and pion PID probabilities from D^* decays, all available KID skim data files between Experiment 7 and 55 have been analyzed, yielding $43 * 10^6$ candidate D^* events. The kinematic coverage of extracted probabilities reaches close to the kinematic limits of the Belle experiment but will have to be extended to perform a correction for particle misidentification on the entire laboratory frame kinematic plane. The statistical precision of the extracted probabilities is high. A consistency check has been applied to ensure tight correspondence between fits and data histograms, and no significant deviation was found.

The extracted pion and kaon PID probabilities vary significantly depending on which kinematic bin is analyzed. In the present study, kaons are found to be misidentified as pions at the order of 20% at medium and high momentum. The fake-rate of kaons misidentified as protons is also sizeable and reaches values of about 15% also at medium and high momenta. Pions are found to be misidentified as kaons at probabilities of about 10% and misidentified as protons at probabilities of about 5%.

For the purpose of entirely populating PID probability matrices $\mathcal{P}$, the following Section will discuss proton PID probability extractions by analyzing hadron decay products of the Λ baryon.

4.1.3 Extraction of Proton/Antiproton PID Probabilities $p_{p \to j}$ from Λ Decays

Similarly to the previous section, Belle studies about the extraction of proton/antiproton PID probabilities $p_{p \to j}$ employing the decay of the Λ baryon $\Lambda^0(1116) \to p^+ \pi^- + c.c.$, exist. In principle, the decay channel would allow to extract pion probabilities as well, however the phase space available to the pions is severely restricted due to the very close mass of Λs and protons. Thus only proton probabilities are extracted in this study. As a reference for Λ candidate selections, internal Belle analysis notes BN 670 [68], BN 684 [69] and BN 1126 [70] were consulted.

Selection of Λ Decay Product Pairs

The kinematic reconstruction of $\{p, \pi\}$ track pairs is performed in the Belle raw data processing. From all available tracks, combinatorial track pairs are formed and assigned the proton and pion mass, respectively. If, among several other conditions, the two helices of the tracks approach each other sufficiently closely, the two tracks are fitted to a common vertex. If this fit is successful and the invariant mass of the track pair lies inside a 30 MeV/c^2 window around the nominal Λ mass, the two tracks are stored in the Belle data table 'MDST_Vee2', which is available for all hadronic events.[4] All HadronBJ skim on-resonance and continuum data from Experiments $7 - 55$ have been analyzed, corresponding to about $2.7 * 10^9$ events.

Analogously to the procedures described in Section 4.1.2, the proton/antiproton PID probabilities will be extracted by fitting $p\pi$ invariant mass spectra without and with additional PID selection cuts on the proton candidates. In order to reduce the quite sizable combinatorial background, BN 670 introduces quality cuts to reduce background contributions in the analyzed mass spectra. The cuts are performed using observables which have initially been defined in the analysis of the decay $K_s \to \pi^+\pi^-$ to select K_s candidates [72]. However, these observables can also be applied to the analysis of the Λ decay channel under study here. Definitions of the observables used in Ref. [68] are given in Table 4.4. For the present analysis, cut values from Ref. [69] have been used. Those cut values are summarized in Table 4.5.

All cuts given in Table 4.5 have been applied for the present study without alteration. Further background removal is achieved by requiring the fit of two decay product candidate tracks to a common vertex to be successful. Candidates are not considered if their vertex lies outside or in the material of the beampipe. Additionally, pion-proton/antiproton candidate track pairs with an invariant mass within 30 MeV/c^2 of the nominal Λ mass are required to produce an invariant mass outside of this window if the mass assignment is reversed. This should prevent track pairs from entering the analysis which have ambiguous mass assignments.

[4] For further information about this track selection process, see Ref. [71].

Cut Variable	Definition
fl	distance between Λ decay vertex and event-dependent Interaction Point (IP) in the xy plane.
$d\phi$	angle in the xy plane between the vector from the event-dependent IP to the Λ decay vertex and the Λ momentum vector.
Δz	distance in the z direction between the decay product helices at the Λ decay vertex.
Δr_{min}	minimum of the impact parameter of the two decay product helices after propagation to the event-dependent IP.

Table 4.4: Kinematic observables to remove background processes in the $p\pi$ invariant mass spectra created from the Belle data table MDST_Vee2. All quantities are evaluated in the laboratory frame. All definitions are taken from Ref. [73], except for the definition of Δz which can be found in Ref. [74].

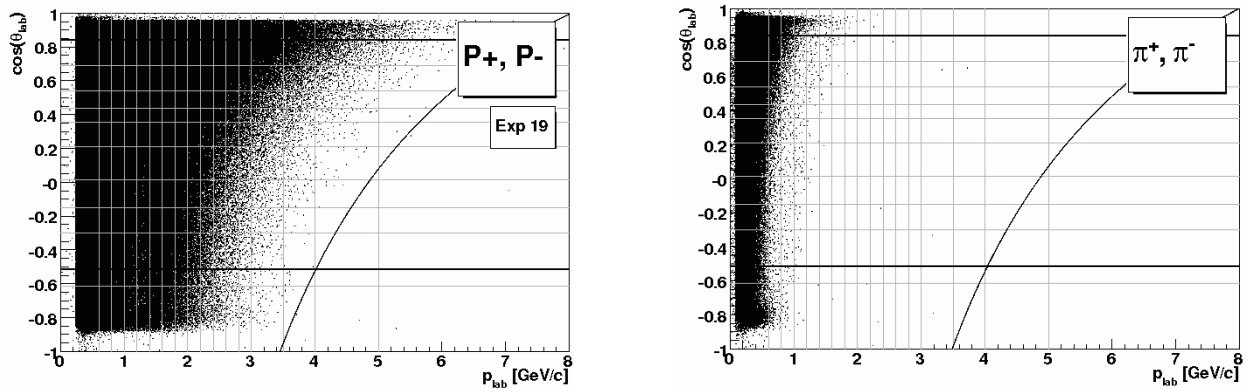
$p_{\rm lab}^{p\pi}$ [GeV/c]	fl [cm]	$d\phi$ [rad]	Δz [cm]	Δr_{min} [cm]
$<$ 0.50	(0.17; 15.0)	$<$ 0.60	$<$ 1.9	$>$ 0.059
(0.50; 1.50]	(0.24; 15.0)	$<$ 0.10	$<$ 2.1	$>$ 0.033
$>$ 1.50	(0.35; 15.0)	$<$ 0.07	$<$ 7.7	$>$ 0.018

Table 4.5: Criteria for selecting Λ candidates from the Belle data table 'MDST_Vee2' in the present analysis. All cuts given are adopted from BN 684, with the additional requirement that all candidate pair vertices lie within 15 cm from the interaction point.

Additional cuts to reject physics backgrounds have been adopted from Ref. [70]. Contributions come from the decay $K_s \rightarrow \pi^+\pi^-$, where one pion is given the proton mass in the reconstruction. To remove this contribution, Ref. [70] suggests to form the invariant mass of both candidate tracks after assigning pion masses and then discarding all pairs which fall within 12 MeV/c^2 of the nominal K_s mass. This procedure was applied in the present analysis as well. A second contribution is given by γ conversion electrons and positrons which are given pion and proton masses in the reconstruction. The suggested procedure is to give the two candidate tracks the electron mass and discard the pair if the pair invariant mass falls below 50 MeV/c^2. This procedure has also been implemented in the present study.

Similar single track cuts as in the D^* analysis have been employed in the Λ analysis as well. Both tracks are required to pass through the Belle detector within the CDC acceptance and carry a momentum in the laboratory frame of $>$ 100 MeV/c. All kinematic data for the proton/antiproton and pion candidate tracks are extracted from the helices stored in the Belle data table 'MDST_Vee2_Daughters'. Eventually, the invariant mass values of all selected candidate pairs are filled in histograms spanning a 30 MeV/c^2 invariant mass interval around the nominal Λ mass, ranging from 1.101 to 1.131 GeV/c^2.

As an illustration, Figures 4.7 give the available charge-integrated Λ decay product statistics on the ($p_{\rm lab}$, cos$\theta_{\rm lab}$) bin grid. The distributions of all Λ decay products selected from all Experiment 19 MDST_Vee2 tables are shown. As can be seen from the pion plot, only little phase space is available for the pion in the Λ decay, thus pion probabilities will not be extracted in this part of the present study.



(a) (b)

Figure 4.7: Kinematic Coverage Illustrations of the Λ decay products p^+, p^- (a) and π^+, π^- (b) selected from all Experiment 19 data. The two accentuated horizontal lines denote the geometrical acceptance limits of the ToF subdetector, also representing the acceptance limit for the present analysis. The curved line on the right hand side of each Figure shows the physics limit $z = 1.0$ for the presence of protons, respectively, in the laboratory frame.

PID Probability Extraction

PID probabilities were extracted using the same procedures as described in Section 4.1.2 for analyzing D^* decay samples. All filled data histograms were segmented into 20 bins and are fitted if the number of entries is larger than 180.

After applying all candidate selection cuts given in the previous section, the invariant mass distributions of most data histograms show turn-on and turn-off behaviors in the first and last bins. In order to still be able to fit the background with a linear function as suggested in previous Belle PID studies, the first and last 2 bins on each side of the histogram were excluded from the fit, leaving 16 bins to be fitted in each histogram. Another effect possibly related to the applied cuts are non-linearities in the combinatorial background distributions, especially for the proton and pion PID selections. However, in the case of the proton PID selection histograms, the background levels are very low compared to the signal height since the proton PID efficiency is usually high. The deviations of the backgrounds from the linear fit under the signal peak has been found to be of the order of the statistical uncertainty of the histogram bins in the signal peak area. Thus the background non-linearities are neglected and a linear background function is applied. For the pion PID selection histograms, background non-linearities are most significant for low track momentum. The proton fake-rates into pions are of the order of a few per cent in this kinematic area. The background non-linearities are estimated to change the signal yields in the fits by about 10%, amounting to changes in

the PID probabilities of about one per mille. These changes have been neglected.

For fitting the Λ signal peaks, functions containing the modified asymmetric Crystal Ball functions introduced in the D^* fit descriptions in Section 4.1.2 and single Gaussians have been used. Again extensive Monte Carlo studies have been performed to extract fast-converging starting parameters for each fit function from histograms for each PID selection cut on integrated kinematics.

Fit Procedure Again for obtaining start parameters for the background fit, an auxiliary histogram is created for each data histogram where the signal peak is removed and instead a linear interpolation between the bin contents at 1.105 and 1.125 GeV/c^2 is filled. For the fit, no limits on the parameters have been applied. To quantitatively separate insignificant from significant signal peaks, the data histograms are again fitted with the background function alone with start parameters assigned from the results of the auxiliary histogram fit. The probabilities from integrating the χ^2 distribution to the obtained value of χ^2 of this pure background fit are kept in memory for later comparison to the fit including the signal function.

Next, the data histogram is fitted with signal and background functions combined. The background function parameters are assigned the fit results from the auxiliary histogram fit. Contrary to the D^* case, it has been found that the full fit would converge even though the background parameters were not fixed, so they were left free. For the signal part, in analogy to the D^* case, only the signal scaling parameter $par^{\mathrm{sig}}_{scale}$, the mean $par_0^{sigGauss}$ and joint width parameter of the signal function are fitted. All other signal parameters are fixed for the fit. The results of this fit, if successful, are used to extract PID probabilities. This fit procedure is repeated for each data histogram.

Since the first two and last two bins of the histograms were left out due to turn-on and turn-off behavior, the fits were performed with 11 degrees of freedom.

Extraction of PID Probability Ratios For determination of signal significance, extraction of PID statements from fit results and uncertainty treatment the same procedures as in the D^* analysis have been applied. The integration interval for the signal function has again been chosen to be 3 histogram bins wide, ranging from 1.113 to 1.1175 GeV/c^2. Similarly to the D^* extraction, probabilities from the *unsel* PID selection cut were found to mostly remain below 1%. Again, joint e/μ probabilities have been extracted in analogy to the D^* analysis. This will be revisited in future improvements. Extracted probabilities with values larger than 1.0 were set to 0.99 to be consistent with the binomial uncertainty interpretation of the yields $N_{p \to j}$.

Fit Statistics For the Λ analysis, 3672 histograms were filled, and 1589 of them showed sufficient statistics to fit the data distributions. Again, the pion phase space is severely restricted to low laboratory momenta due to similar masses of the Λ and the proton. Of the 1589 fitted histograms, 1296 produced successful full fit function fits, from 87 histograms zero probabilities were extracted based on successful pure background function fits. After examining each histogram by hand, 90 fits have been rejected due to fit anomalies.

Similarly to the D^* analysis, Figure 4.8 shows sample fits of data histograms from different kinematic bins and PID selections. The code in brackets shown in each pad refers to the respective (p_{lab}, $\cos\theta_{\mathrm{lab}}$) bin and the PID selection on the tracks filling the histogram, where applicable. The sample distributions are again taken from bins in the ToF backward, barrel and forward region with either low, medium or high laboratory frame momentum, with bin numbers as defined in Table 4.1. All specifications given in the description of the D^* analysis apply.

The second plot from the left in the top row of Figure 4.8 shows a fit of a data histogram for antiprotons filled with additional pion PID cuts. The background is not optimally described by the linear fit function- however, as reasoned in the previous section, the pion fake-rate is about 1.7 per cent. If one estimates that the background non-linearity changes the signal yield $N_{p^- \to \pi^-}$ by 10%, the PID probability would change by 1.7 permille, which is neglected. The third plot from the left in the top row shows a data histogram where the fit with the full fit function failed because of very low signal to background ratio. A zero-probability is extracted, the pure background fit is shown in black, the full fit function in red. The third plot from the left in the second row shows a data histogram which has not been fitted by the fitting algorithm because of insufficient statistics.

Cumulative χ^2 distributions are shown in Appendix B.4 for all data histogram fits, separated by charge and different additional PID cuts, but integrated over all kinematic bins. All χ^2 histograms are fitted with the corresponding theoretical χ^2 distribution for 11 degrees of freedom, the fit is performed by scaling the whole distribution to the χ^2 histogram. Compared to the D^* analysis, the χ^2 distributions from fits to the data histograms have longer tails to high χ^2. The most pronounced tails can be found for fits on histograms filled without PID selection, and for histograms with additional pion and proton PID selection cuts. The tails for the latter two groups of histograms can probably be motivated by the occurring background non-linearities discussed in the beginning of this section. The high-χ^2 tails in the histograms without PID selection is probably an effect of the high statistics in the histograms and thus very low uncertainties on the bin contents. Any small deviation by the fit function from the data distribution is punished by a large χ^2 contribution. The reliability of the fits to determine the signal yield used for the probability extraction will be revisited in Section 4.1.2.

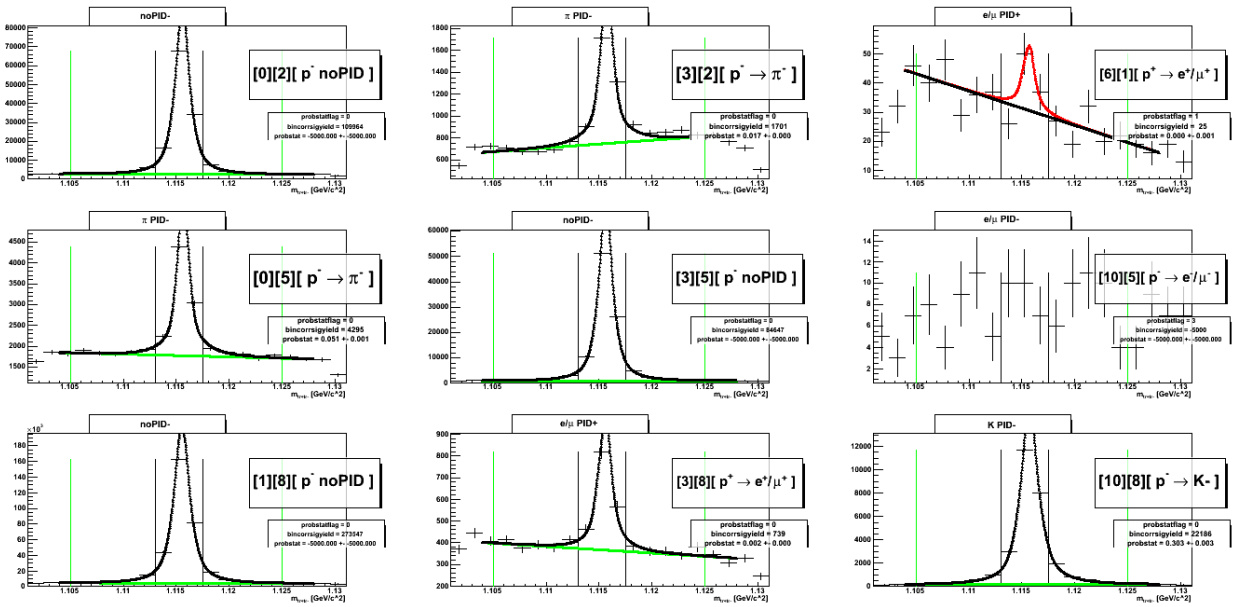


Figure 4.8: Nine analyzed sample distributions from the extraction of PID probabilities on the set of Λ candidates. Further descriptions of the plots can be found in the text.

The Appendix also contains 2-dimensional plots of all kinematic bins to display the result of PID statement extraction for each bin, similarly to the D^* analysis. The color coding is explained in the legends of each plot; white kinematic bins indicate that the fit has been rejected by hand because of anomalies. The kinematic coverage is again vast and covers most of the z range for the hadronic PID probabilities $\mathrm{p}_{p \to \pi,\, K,\, p}$. It can be seen especially from the e/μ PID plots that fake-rates of protons misidentified as leptons are low, such that PID statements could only be extracted for low to medium z where sufficient statistics in the data histograms is available. Similarly as in the D^* analysis, missing probabilities will be calculated from extrapolations of available PID statements, c.f. Section 4.1.5.

Samples of Extracted PID Probabilities $\mathrm{p}_{p \to j}$ from Λ Decays

Sample PID probabilities from Λ decays are shown in Figure 4.9 in analogy to the D^* analysis. The single plots show probabilities $\mathrm{p}_{p^- \to j^-}$ for all likelihood-cut species j, plotted against p_{lab} in different scattering angle intervals (top 5 pads) and against $\cos\theta_{\mathrm{lab}}$ (bottom 3 pads) for different p_{lab} bins. Kinematic areas where no PID statement could be extracted from the data histograms are again identified by markers plotted below 0.

The top 5 pads show probabilities for tracks travelling through the backward, backward/barrel, barrel, forward/barrel and forward regions in $\cos\theta_{\mathrm{lab}}$, respectively. Analogously, the lowest three plots show the respective probabilities against $\cos\theta_{\mathrm{lab}}$ for tracks with low, medium and high momentum. Uncertainties are

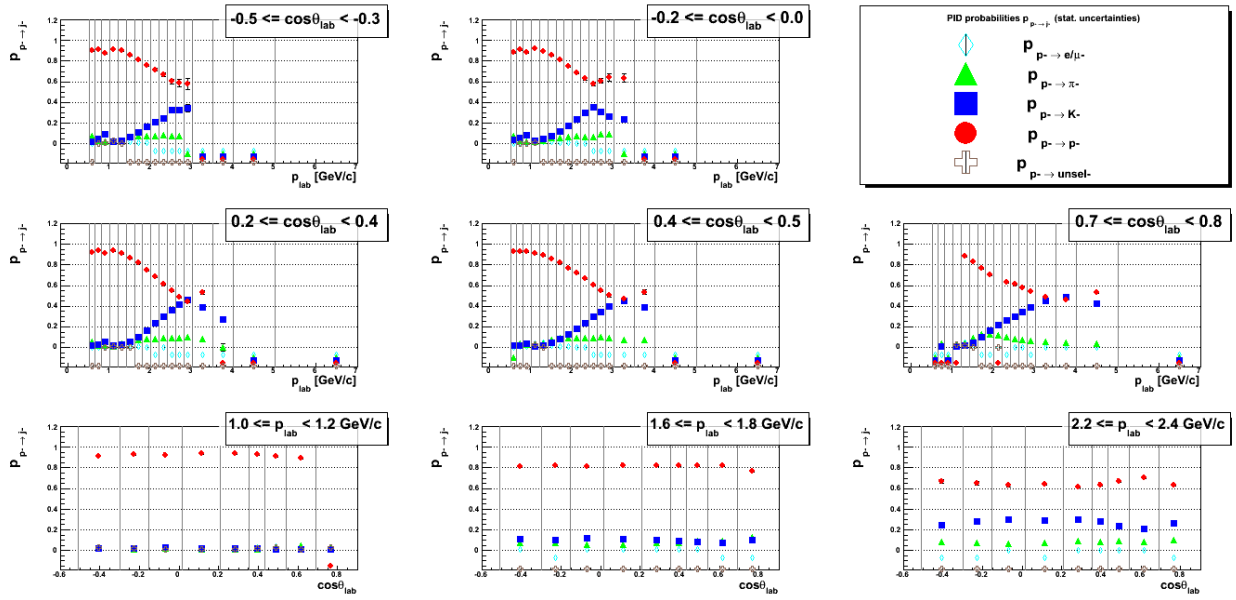


Figure 4.9: Samples of extracted PID probabilities $\mathrm{p}_{p^- \to j^-}$ from Λ decays. The different subfigures give plots of probabilities for all likelihood-cut species j against p_{lab} in different scattering angle intervals (top 5 pads) and against $\cos\theta_{\mathrm{lab}}$ (bottom 3 pads)for different p_{lab} bins at the center of each bin.

given by error bars for all probabilities.

Consistency Check for Extraction of Fit Probabilities To check the fit quality of the signal peak in each data histogram, alternative values for the integrated signal yields $N_{p \to j}$ and N_p have been extracted by applying a procedure in analogy to the D^* analysis. As a reminder, the alternative signal yield is calculated by subtracting the fitted background from the bin contents of the data histogram in the signal yield integration window. This way, possible deviations from the signal yield fit from the data distributions and their effect on the extracted PID probabilities can be monitored.

Plots with all calculated differences between the bin content-based probabilities and the signal function integral probabilities are given in Appendix B.4. Despite the imperfect χ^2 distributions discussed in an earlier section, almost all absolute differences remain below 1% and are centered around 0. If one considers also the uncertainties on both probabilities, the differences stay within 1σ of the joint difference uncertainty. We thus conclude that even though the χ^2 distributions are not optimal, adequate description of the signal yields in the data histograms is achieved and the extracted PID probabilities are assumed to be reliable.

Discussion of Extracted Proton PID Probabilities $\mathbf{p}_{p \to j}$

As already observed in the D^* analysis, the extracted PID probabilities feature an obvious dependence on the laboratory frame momentum but only slightly vary in different regions of the polar angle θ_{lab}. Similarly to kaon and pion efficiencies, the proton/antiproton efficiency is found above 90% at low momenta. The drift chambers and the ToF detector provide good discrimination power via the measurement of the energy loss dE/dx and the time-of-flight. At a track momentum of roughly 1 GeV/c, the lines of expected energy loss in the drift chambers for kaons and protons/antiprotons intersect, cf. Appendix 2.2, Figure 2.4, so no discrimination power can be provided from these measurements. This is reflected by a dip in the proton/antiproton efficiency and a gain in the kaon fake-rate which is confined to a narrow momentum region around 1 GeV/c. Since the energy loss curves for protons/antiprotons and kaons are very close for higher momenta as well, the kaon-proton/antiproton discrimination for the most part relies on ToF detector measurements for momenta between about 1 and 3 GeV/c. As illustrated in Table 2.2, the differences between times-of-flight of protons/antiprotons and kaons decrease with increasing particle momenta, providing a plausible reason for the proton/antiproton efficiency to almost decrease linearly between about 1 and 3 GeV/c. At a momentum of roughly 3 GeV/c, the proton/antiproton efficiency reaches a minimum and is even slightly surpassed by the kaon fake-rate. This can be interpreted as the effect of an almost vanishing discrimination power from the ToF measurements, since the differences in times-of-flight for protons/antiprotons and kaons approach the order of the time resolution of the ToF detector at momenta above 3 GeV/c, cf. Table 2.2. A slight enhancement of the proton/antiproton efficiency and a decrease of the kaon fake-rate at even higher momenta can be found. This might be attributed to the onset of discrimination power in ACC measurements, as all kaons with momenta above about 3 GeV/c produce Cherenkov photons in the ACC aerogels, whereas protons/antiprotons by design barely ever produce an ACC signal in the Belle experiment, cf. Table 2.1.

As to the proton/antiproton-pion fake-rates, the measurements from the drift chambers and the good time-of-flight separation keep this misidentification low at low momenta. The lines of expected energy loss for pions and protons/antiprotons intersect at momenta around 1 GeV/c, cf. Figure 2.4. However, all pions above roughly 0.8 GeV/c produce Cherenkov photons in the ACC, whereas protons/antiprotons are expected to 'fire' the ACC at momenta no lower than at least 4 GeV/c, cf. Table 2.1. Thus the ACC measurements offer high discrimination power for track momenta above about 1 GeV/c, compensating for the loss of the drift chamber discrimination. Because of the dominance of the ACC measurements, the loss of ToF discrimination power for increasing track momenta is only very lightly reflected by a slight rise in the pion fake-rate. The notable drop of the pion fake-rate at momenta higher than about 3 GeV/c might be attributed to a recovery of discrimination power from drift chamber measurements. The lines of expected

energy loss for protons/antiprotons and pions are considerably separated again for momenta higher than about 2.5 GeV/c.

All proton/antiproton-lepton fake-rates remain below 5%. The difference between electrons and protons/antiprotons in the characteristics of ECL shower shapes and the different energy losses in the drift chambers provide good discrimination power. Especially the ACC measurements largely prevent misidentification over almost the entire momentum range as electron Cherenkov threshold momenta lie at the order of MeV/c, whereas protons/antiprotons produce ACC signals only at momenta close to the momentum limit in the Belle experiment.

Summary of the Proton PID Probability Extraction from Λ Decays

For the extraction of proton PID probabilities, candidate pairs for Λ decay products have been selected from the Belle data table 'MDST_Vee2' for all hadronic skim HadronBJ data from Experiments $7 - 55$, corresponding to about $2.7 * 10^9$ events. The kinematic coverage of all probabilities extracted is again fair but will need to be extended to perform a correction for particle misidentification on the entire laboratory frame kinematic plane. As in the D^* case, the statistical precision of the extracted probabilities is high. The same consistency check as in the D^* analysis has been applied to test a good correspondence between fits and data histograms, and no significant deviations were found.

Similarly to the extracted pion and kaon PID probabilities, proton efficiencies and fake-rates show significant variations depending on track kinematics. The proton/antiproton-pion fake-rate reaches values at the order of 10 to 15% at medium to high momenta in the Belle experiment. The proton/antiproton-kaon fake-rate even rises above 50% at high momenta.

After the extractions of all available hadron probabilities $\mathrm{p}_{\pi,\ K,\ p\ \rightarrow\ j}$, the next section will discuss the extraction of lepton PID probabilities $\mathrm{p}_{e/\mu\ \rightarrow\ j}$ from J/ψ decays.

4.1.4 Lepton PID Probabilities $p_{e/\mu \to j}$ from J/ψ Decays

In order to fully populate PID probability matrices, finally lepton PID probabilities need to be extracted. Several studies investigating electron and muon PID probabilities exist in Belle, notably Belle analysis notes BN 777 [75] and BN 954 [76], together with two NIM papers, Ref. [48] and Ref. [49]. Most of those analyses use events $e^+e^- \to e^+e^-l^+l^-$ where $l = \{e, \mu\}$ to extract their primary probabilities. Then the impact of hadronic event environments on the probabilities is checked by embedding single electron or muon tracks into hadronic events or analyzing reconstructed samples of decays $J/\psi \to l^+l^-$.

In event samples containing reactions $e^+e^- \to e^+e^-l^+l^-$, the impurity level is usually estimated to be a few per cent, but cannot be quantitatively determined for experimental data. In our study, we would like to extract lepton probabilities with high precision, so we decided to study decays $J/\psi \to l^+l^-$ in a similar fashion to what was described in the D^* and Λ analyses so far.

Selection of J/ψ Decay Product Pairs

In an attempt to extract lepton PID probabilities from a sample which is completely unbiased as to PID selections, we select J/ψ candidates from all hadronic HadronBJ events.[5] Similarly to the probability extractions described so far, the goal is to reconstruct decayed particles with purely kinematical means. However, in the case of the J/ψ decay, low signal-to-background ratios are found. To reduce the background, studies were performed to kinematically reconstruct B-meson channels producing J/ψ particles, but these additional kinematical conditions were found to add even more combinatorial background. Another way to increase signal-to-background ratios is to employ so-called 'tagging' cuts, where PID cuts are applied to one particle in the candidate pair and the other unbiased track is analyzed for the extraction of PID probabilities. However, this additional feature will be examined in future improvements of the study but is not implemented in the presented results. Since without PID tagging it is not possible to distinguish the decay channels $J/\psi \to e^+e^-$ from $J/\psi \to \mu^+\mu^-$, only joint PID probabilities $p_{e/\mu \to j}$ are extracted.

From all HadronBJ events, J/ψ candidate track pairs are selected by requiring opposite track charges and a minimum laboratory frame momentum of 500 MeV/c. Additionally, all tracks must pass through the ToF detector scattering angle acceptance in order to enter the analysis. In Monte Carlo studies differences in signal yields between J/ψ invariant masses calculated from electron or muon mass assignments have been found to be below 1%. Thus all tracks are given the electron mass. A vertex fit of the two candidate tracks to the interaction point of each run is required to be successful.

[5] The existing skim enhanced with events containing J/ψ particles (psiskim [77]) makes use of PID likelihood cuts in the skimming procedure and thus would represent a biased sample.

Additional cuts are imposed which help reduce the combinatorial background by half. These cuts have been optimized in Monte Carlo studies to not reject more than 1% of all true J/ψ decay products and are imposed on the observables given in Table 4.6, the cut values are displayed in Table 4.7. The cuts on R_2 and the thrust variable select events with a more spherical topology, since J/ψ are fairly abundant decay products of B-mesons decaying virtually at rest in the Belle center-of-mass frame.[6] The energy asymmetry cut exploits the fact that the J/ψ decays into two identical decay products with similar energies only distorted by the (slow) motion of the J/ψ. Candidate pairs with highly asymmetric energy distributions are rejected by this cut. Finally, the cut on the distance of the track helices at the fitted candidate J/ψ vertex rejects pairs which do not approach each other closely enough at the vertex.

Cut Variable	Definition
R_2	second order Fox-Wolfram moment
thrust T	maximum of expression $\frac{\sum_h \lvert\vec{P}_h^{cms}\times\hat{n}\rvert}{\sum_h \lvert\vec{P}_h^{cms}\rvert}$ under variation of axis $\hat{n}$
Δz	distance in the z direction between the decay product helices at the J/ψ decay vertex.
A_E	energy asymmetry between center-of-mass energies of positively and negatively charged track candidate.

Table 4.6: Kinematic observables to remove background processes in the invariant mass spectra reconstruction the bi-leptonic decay of J/ψ particles.

R_2	T	Δz [cm]	A_E
$<$ 0.51	$<$ 0.883	$<$ 0.43	$<$ 0.43

Table 4.7: Criteria for selecting J/ψ candidates in the present analysis.

Eventually, the candidate invariant mass is filled in a 20 bin histogram ranging from 3.0 to 3.18 GeV/c^2 for each kinematic bin on the ($p_{\rm lab}$, $\cos\theta_{\rm lab}$) plane. All HadronBJ skim on-resonance and continuum data from Experiments 7$-$55 have been analyzed, corresponding to about $2.7{*}10^9$ events. Similarly to the sections described above, lepton PID probabilities are extracted by fitting J/ψ invariant mass spectra without and with additional PID selection cuts on the PID likelihood values of the lepton candidate tracks.

As mentioned earlier, most of the J/ψ particles selected in this analysis are decay products of B-mesons. At Belle, these mesons are produced more or less at rest in the center-of-mass frame. Thus the J/ψ particles themselves can only carry a maximum fractional energy z of around 0.5, which then also represents the maximum possible energy for its decay products. As a consequence, PID probabilities can only be extracted on a confined kinematical region from J/ψ decays. As an illustration, Figure 4.10 gives charge-integrated J/ψ decay product candidates selected by the cuts specified above from all HadronBJ data in Experiments 7 through 55 on the ($p_{\rm lab}$, $\cos\theta_{\rm lab}$) bin grid.

[6] A similar cut on the value of R_2 has already been used in BN 954.

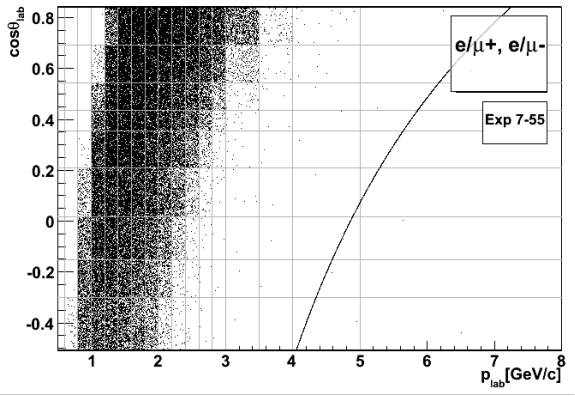


Figure 4.10: Kinematic Coverage Illustrations of selected J/ψ decay product candidates e/μ^+, e/μ^- selected from all Experiment 7 through 55 data. The y range is limited to the geometrical acceptance limits of the ToF subdetector, also representing the acceptance limit for the present analysis. The curved line on the right hand side of each Figure shows the physics limit $z = 1.0$ for the presence of muons in the laboratory frame.

PID Probability Extraction

PID probabilities are extracted using again the same procedures as in the previously described analysis sections. All filled data histograms are fitted if the number of entries is larger than 180.

Fit Procedure Background shapes are fitted with second or third order polynomials, where starting parameters again have been determined from Monte Carlo studies on histograms integrating over kinematics but resolving PID selection cuts. Fit functions for signal shapes were also extracted from extensive Monte Carlo studies. In these studies, the signal contributions from channels $J/\psi \rightarrow e^+e^-$ and $J/\psi \rightarrow \mu^+\mu^-$ were treated separately and fitted with independent functions, among which are modified asymmetric Crystal Ball functions introduced in the D^* fit descriptions in Section 4.1.2 and single Gaussians. In the data histogram fits, the J/ψ signal peak is fitted with the sum of the optimized electron and muon signal function, only leaving a global scale parameter and joint mean and width parameters free for fitting.

The fit procedure itself remains the same as previously described in the D^* and Λ analyses. First an auxiliary histogram is created by removing the signal part and interpolating a linear function between the data histogram bin contents at 3.04 and 3.12 GeV/c^2. The parameter results of this fit are used as starting parameters for the pure background fit function. The integrated χ^2 distribution probability of this fit is retained for later comparison to the full fit function results to test the significance of the signal peak. Finally, the data histogram is fitted with the sum of the background and signal function, leaving all background parameters and the three mentioned signal function parameters free for the fit. No parameter limits are applied. The probability of the fit is compared to the value achieved by the pure background

function fit. In case the total function probability is larger, the significance of the peak is tested. The results of this fit, if the peak is found significantly larger than statistical fluctuations in the histogram, are used to extract PID probabilities. This fit procedure is repeated for each data histogram.

All histogram bins are considered in the full fit function fit. Since the numbers of parameters in the full fit function are 6 or 7 (depending on which order polynomial is used), the number of degrees of freedom is 14 or 13, respectively.

Extraction of PID Probability Ratios For the J/ψ analysis, the integration window for the signal function was chosen to be $[3.072; 3.108]$ GeV/c, corresponding to four bins in the data histograms around the signal peak. Analysis techniques analogous to the D^* and Λ analyses were applied to extract PID statements from the data histograms. Because of the large combinatorial background, significant signal peaks were only found for the histograms without PID selection and the 'diagonal' PID selection e/μ, rarely for the pion selection. Thus only zero probabilities were extracted for the kaon, proton and unselected PID histograms. Also, the signal peak significance fades out at higher laboratory frame momenta, which does not allow to extract any PID probabilities anymore. Extracted probabilities with values larger than 1.0 were again set to 0.99 to be consistent with binomial uncertainty interpretation of the yields $N_{e/\mu \to j}$.

Fit Statistics Out of the 1836 data histograms of all possible kinematics and charge bins and PID selections, 1277 showed enough entries to be fitted. However, since most entries are combinatorial background and J/ψ signal peaks are present on a limited kinematic range only, 277 histograms were fitted successfully with the full fit function. However, zero probabilities from successful pure background function fits were extracted from 413 histograms. After the final manual examination of each histogram, 39 fits have been rejected due to fit anomalies.

Analogously to the D^* and Λ analyses, Figure 4.11 shows sample fits of data histograms from different kinematic bins and PID selections. For PID selections other than for e/μ, only zero probabilities could be extracted. In each pad, the code in brackets refers to the respective $(p_{\mathrm{lab}},\ \cos\theta_{\mathrm{lab}})$ bin and the PID selection on the tracks filling the histogram, where applicable. From top to bottom, the sample distributions are taken from bins in the ToF backward, barrel and forward region with, from left to right, low, medium or high laboratory frame momentum. Bin numbers are defined in Table 4.1. All specifications given in the description of the D^* and Λ analyses apply.

In Appendix B.5 cumulative χ^2 histograms are shown for all data histogram fits, separated by charge and different additional PID cuts, but integrated over all kinematic bins. All χ^2 histograms are fitted with the corresponding theoretical χ^2 distribution for 13 or 14 degrees of freedom, depending on the chosen order

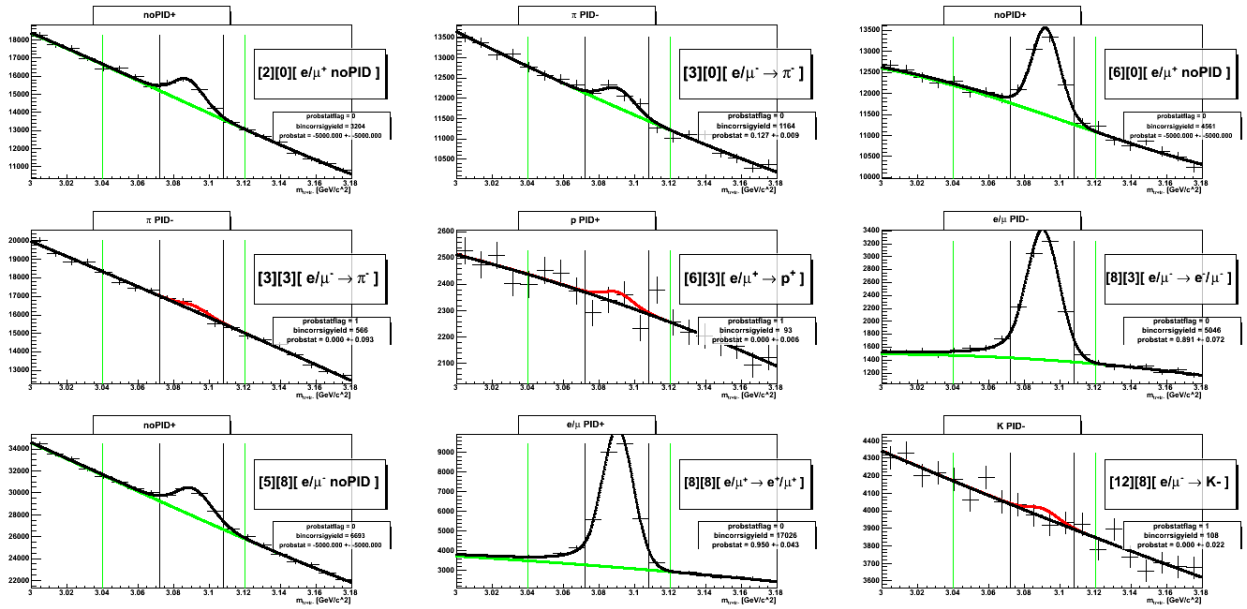


Figure 4.11: Nine analyzed sample distributions from the extraction of PID probabilities on the set of J/ψ candidates. Further descriptions of the plots can be found in the text.

of the polynomial background function. The fit is performed by scaling the theoretical distribution to the χ^2 histogram. The fit results indicate that very good descriptions of the signal and background data shapes have been achieved. Thus it is concluded that all fits used for PID probability extractions are reliable and properly describe the data distributions.

Also shown in the Appendix are 2-dimensional plots of all kinematic bins to display the result of PID statement extraction for each bin. The legend in each plot explains the color coding of each bin; white kinematic bins indicate that the fit has been rejected by hand because of anomalies. As stated above, the kinematic coverage in the J/ψ decay sample is limited due to confined kinematical abundance of the parent particle. The blue bins in the plots for pion, kaon and proton PID histograms indicate that only zero probability PID statements could be extracted from the data histograms. This is caused by combinatorial background diluting signal significance and in addition PID fake-rates from leptons into hadrons being mostly small. Extrapolating probabilities will be important in order to achieve fully populated PID probability matrices, c.f. Section 4.1.5.

Samples of Extracted PID Probabilities $\mathrm{p}_{e/\mu \,\to\, j}$ from J/ψ Decays

Figure 4.12 shows sample PID probabilities from the analysis of J/ψ decays. Similarly to the D^* and Λ sections before, the single plots show probabilities $\mathrm{p}_{e/\mu^- \,\to\, j^-}$ for all likelihood-cut species j, plotted against p_{lab} in different scattering angle intervals (top 5 pads) and against $\cos\theta_{\mathrm{lab}}$ (bottom 3 pads) for different p_{lab}

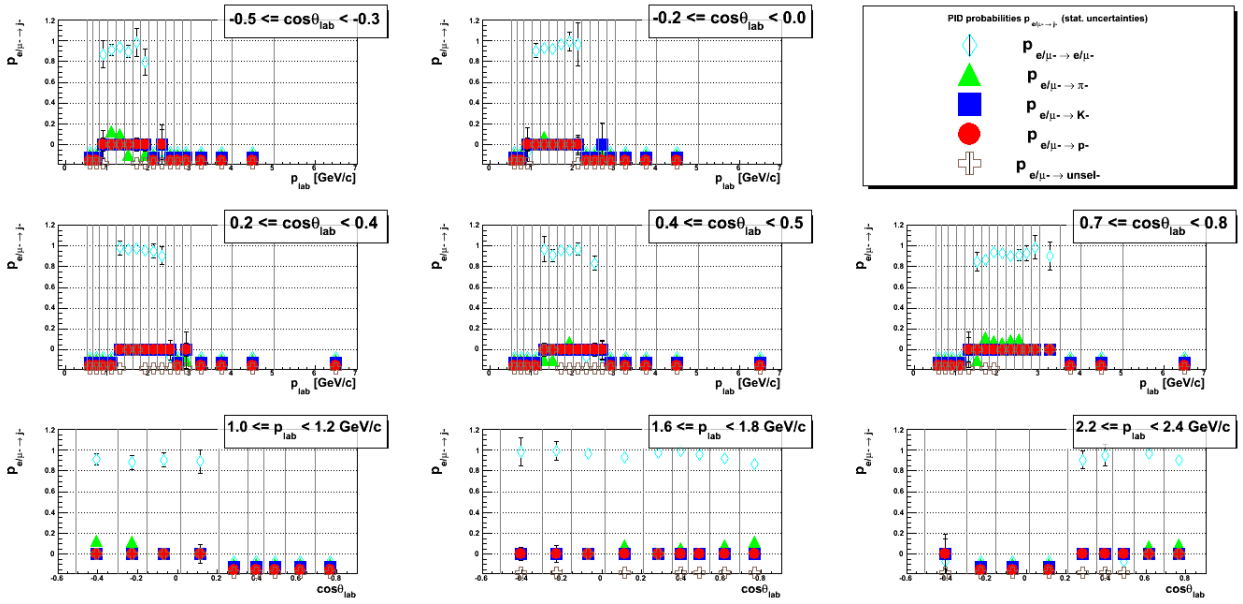


Figure 4.12: Samples of extracted PID probabilities $p_{emu^- \to j^-}$ from J/ψ decays. The different subfigures give plots of probabilities for all likelihood-cut species j against p_{lab} in different scattering angle intervals (top 5 pads) and against $\cos\theta_{\mathrm{lab}}$ (bottom 3 pads) for different p_{lab} bins at the center of each bin.

bins. Marker symbols plotted below 0 indicate kinematic areas where no PID statement could be extracted from the data histograms.

The top 5 pads show probabilities for tracks travelling through the backward, backward/barrel, barrel, forward/barrel and forward regions in $\cos\theta_{\mathrm{lab}}$, respectively. Analogously, the lowest three plots show the respective probabilities against $\cos\theta_{\mathrm{lab}}$ for tracks with low, medium and high momentum. Uncertainties are given by error bars for all probabilities.

Consistency Check for Extraction of Fit Probabilities Analogously to the D^* and Λ analyses, the fit quality of the signal peak was checked in each fitted data histogram by extracting alternative values for the integrated signal yields $N_{e/\mu \to j}$ and $N_{e/\mu}$ as described in the D^* analysis applying the same procedure. The alternative signal yield is calculated by subtracting the fitted background from the bin contents of the data histogram in the signal yield integration interval. This makes it possible to monitor potential deviations of the signal yield fit from the data distributions and their effect on the extracted PID probabilities.

Distributions of calculated differences between the bin content-based probabilities and the signal function integral probabilities are given in Appendix B.5. Similarly to the analyses described earlier in this thesis, also in this part of the study most absolute differences remain below 1% and are centered around 0. Additionally considering the uncertainties on both probability expressions, the differences stay within 1σ of the joint

difference uncertainty. Therefore we state that the signal yields are representatively well described by our fit functions and that the extracted PID probabilities are assumed to be reliable.

Discussion of Extracted Lepton PID Probabilities

Due to the kinematically limited availability of J/ψ particles as described above, PID probabilities from J/ψ decays are extracted on a smaller kinematic range than the hadron probabilities calculated in earlier sections. In the accessible kinematic regions, lepton PID probabilities are found to not vary significantly with the laboratory frame momentum or the polar angle θ_{lab}.

Discussion of Lepton PID Probabilities $\mathbf{p}_{e/\mu \to j}$ Similar to the hadron efficiencies, also lepton efficiencies stay above or around 90% for most kinematic areas. The most significant fake-rates are found to be the lepton-pion probabilities. These reach values as high as 13%, cf. Figures 4.12. Since only joint e/μ probabilities have been extracted, their values and kinematic dependencies represent convolutions of electron/positron and muon probabilities scaled by their relative abundancy. Thus the extracted lepton probabilities will be discussed in the context of the properties of PID subdetectors by analyzing the electron/positron and muon contributions separately.

For electron/positron efficiencies, electron/positron PID likelihoods are extracted from measurements in the drift chambers, the ACC and the ECL detectors. The lines of expected energy loss in the drift chambers for electrons/positrons and pions intersect at track momenta of about 150 MeV/c and stay close to each other for higher momenta, cf. Figure 2.4. For kaons and protons, discrimination is still provided for higher momenta as well. Judging from the threshold momenta in the ACC, cf. Table 2.1, the ACC should provide good discrimination between electrons/positrons and hadrons- for pions up to momenta of about 0.8 GeV/c, for kaons up to about 2.5 GeV/c and for protons up to about 4 GeV/c, depending on the scattering angle of the tracks. Even though at higher momenta the CDC and the ACC lose discrimination power, the high lepton efficiency found in the J/ψ decay studies suggest that the ECL detector provides most of the discrimination power between electrons/positrons and other particles for momenta higher than about 1 GeV/c. The high rejection of hadrons in the ECL electrons/positron identification is stated in Ref. [46].

Non-zero fake-rates of electrons/positrons into hadrons, especially pions, can be mostly motivated by the gradual loss of discrimination power in the CDC and ACC for momenta around about 1 GeV/c. Also the ToF measurements discrimination power decreases above momenta of 1 GeV/c, cf. Table 2.2. Additionally to CDC, ToF and ACC likelihood cuts, the hadron PID selection cuts in this analysis also impose requirements on the *eid* likelihoods mostly consisting of ECL measurements. For track kinematics above momenta of 1 GeV/c, it seems that the only discrimination power between hadrons and electrons/positrons can emerge

from ECL measurements. Thus fake-rates of electrons/positrons into pions and other hadrons for higher momenta are expected to be of the size of misidentifications based on ECL measurements alone.

For muons, the muon PID likelihood is solely extracted based on measurements from the K-long and Muon detector. The efficiency is expected to be high, since misidentifications are assumed to occur only because of punch-through hadrons which reach the KLM detector. It is estimated that the efficiency would decrease slightly with higher momenta as more and more hadrons reach the KLM detector. As to fake-rates into hadrons, muons show electromagnetic material interaction behavior very similar to the one of pions, in terms of their energy loss in the drift chambers, the Cherenkov threshold momenta and their times-of-flight. So slight non-zero fake-rates similar to the pion-kaon and pion-proton fake-rates would be expected. However, hadron PID cuts also require low muon likelihood. This might cause the muon-hadron fake-rates to be mostly similar to the size of KLM muon identification inefficiencies.

In summary, lepton efficiencies are expected to be high, whereas fake-rates into hadrons are predicted to be low for both electrons/positrons and muons. The extracted joint lepton PID probabilities from J/ψ decay samples are found to show consistency with these expectations.

Summary of the Lepton PID Probability Extraction from J/ψ Decays

In order to complete extractions of experimental data PID probabilities, decay samples of J/ψ particles have been analyzed for all hadronic skim HadronBJ data from Experiments $7-55$, corresponding to about $2.7 * 10^9$ events. Due to the limited kinematic availability of the parent particles, lepton PID probabilities are also extracted only in a limited kinematic range. However, all calculated lepton PID probabilities have been checked for consistency with analogous procedures as described in the D^* and Λ analyses and have been found reliable and representative for the experimental PID performance of the Belle detector. Lepton efficiencies $\mathrm{p}_{e,\ \mu\ \rightarrow\ e/\mu}$ are found to be around 90%, hadron fake-rates mostly below 10%.

4.1.5 Extrapolation of PID Probabilities $p_{i \to j}$

In previous sections, PID probabilities $p_{i \to j}$ have been extracted from experimental data for all accessible kinematic regions in the analyzed decay samples. However, in some kinematical areas probabilities could not be calculated due to lack of statistics. In order to be able to perform a correction for PID misidentification of raw data for all kinematic bins within the ToF acceptance and above a minimum momentum of 500 MeV/c, available probabilities have to be extrapolated into kinematic bins with incomplete PID probability matrices.

Lepton Probability Deconvolution

In the analysis of J/ψ decays in Section 4.1.4, the two leptonic decay channels cannot be separated by purely kinematic means. Thus joint e/μ PID probabilities have been extracted. The value of joint probabilities depends on the relative abundance of electrons/positrons and muons in the analyzed sample. Down to an experimental accuracy of one percent [5], the two leptonic decay channels occur with the same branching fraction. Thus a $1:1$ mixing ratio of electrons/positrons and muons in J/ψ decay samples can be assumed. However, all extracted lepton probabilities are uniquely tied to this mixing ratio, since otherwise either the electron/positron or the muon PID performance would give the dominating contribution to a joint e/μ PID probabilities. If the probabilities are to be used for correcting a data sample for particle misidentification and the correction should produce correct results, the same $1:1$ mixing ratio has to be present in the data sample. Since the mixing ratio from raw experimental data cannot be determined, it was decided to perform a lepton PID deconvolution before the extrapolation step.

For this purpose, all extracted joint lepton PID probabilities $p_{e/\mu \to j}$ with $j = \{e/\mu, \pi, K, p\}$ from J/ψ decays and all hadron-lepton fake-rates $p_{i \to e/\mu}$ with $i = \{\pi, K, p\}$ are deconvoluted into separate probabilities $p_{e,\, \mu \to j}$ with $j = \{e, \mu, \pi, K, p\}$. This deconvolution is performed by using probabilities extracted from the 'Monte Carlo HadronBJ' data sample, which is described in further detail in Section 4.1.5. Since the MCHadronBJ mixing ratio of electrons/positrons and muons are different to the ones in the J/ψ sample as well, the abundances of electrons/positrons and muons in the MCHadronBJ sample are taken into account.

At first, the deconvolution of hadron-lepton fake-rates $p_{had \to e/\mu}$ into separate probabilities will be described, where $had = \{\pi, K, p\}$. Here the relative abundances of electrons/positrons and muons do not enter since only hadron yields without and with PID selections are involved. Recalling that probabilities are calculated from yield ratios, combined hadron-lepton fake-rates can simply be written as the sum of

individual hadron-electron/positron and hadron-muon fake-rates,

$$\mathrm{p}_{had \,\to\, e/\mu} \;=\; \mathrm{p}_{had \,\to\, e} \;+\; \mathrm{p}_{had \,\to\, \mu}. \tag{4.9}$$

The basic assumption is that the ratio of individual fake-rates $\mathrm{p}_{had \,\to\, e,\mu}$ over the joint fake-rates $\mathrm{p}_{had \,\to\, e/\mu}$ is the same in the raw experimental data sample as in the MCHadronBJ sample. With this assumption, the deconvolution can be performed by

$$\mathrm{p}^{\text{deconv. } J/\psi}_{had \,\to\, lep} \;=\; \frac{\mathrm{p}^{\mathrm{MC}}_{had \,\to\, lep}}{\mathrm{p}^{\mathrm{MC}}_{had \,\to\, e/\mu}} \; \mathrm{p}^{J/\psi}_{had \,\to\, e/\mu}, \tag{4.10}$$

where $lep = \{e, \mu\}$. Uncertainties on all associated quantities are propagated through the deconvolution expressions. As a conservative estimate, the largest probability uncertainty occur

Next, lepton-hadron fake-rates $\mathrm{p}_{e/\mu \,\to\, had}$ are deconvoluted into probabilities $\mathrm{p}_{e,\mu \,\to\, had}$. Again starting from the definition of probabilities as ratios of yields, joint probabilities $\mathrm{p}_{e/\mu \,\to\, had}$ can be rewritten as

$$\mathrm{p}_{e/\mu \,\to\, had} \;=\; \frac{\mathrm{p}_{e \,\to\, had} \; N_e \;+\; \mathrm{p}_{\mu \,\to\, had} \; N_\mu}{N_e \;+\; N_\mu}. \tag{4.11}$$

Applying this relation to the J/ψ sample with mixing ratio $N_e : N_\mu \;=\; 1:1$ yields

$$\mathrm{p}^{J/\psi}_{e/\mu \,\to\, had} \;=\; \frac{\mathrm{p}_{e \,\to\, had} \;+\; \mathrm{p}_{\mu \,\to\, had}}{2}. \tag{4.12}$$

Again, PID probabilities from the MCHadronBJ sample are used to perform the deconvolution. For this deconvolution a general assumption is made similar to the the one leading to Equation 4.10: The ratio of individual lepton-hadron fake-rates $\mathrm{p}_{lep \,\to\, had}$, now weighted with the corresponding yield N_{lep}, $lep = \{e, \mu\}$, over joint fake-rates $\mathrm{p}_{e/\mu \,\to\, had}$ is the same in the raw experimental data sample as in the MCHadronBJ sample,

$$\frac{\mathrm{p}^{\text{exp. data}}_{lep \,\to\, had} \; N^{\text{exp. data}}_{lep}}{N^{\text{exp. data}}_{e} \;+\; N^{\text{exp. data}}_{\mu}} \; \frac{1}{\mathrm{p}^{\text{exp. data}}_{e/\mu \,\to\, had}} \;=\; \frac{\mathrm{p}^{\mathrm{MCHadronBJ}}_{lep \,\to\, had} \; N^{\mathrm{MCHadronBJ}}_{lep}}{N^{\mathrm{MCHadronBJ}}_{e} \;+\; N^{\mathrm{MCHadronBJ}}_{\mu}} \; \frac{1}{\mathrm{p}^{\mathrm{MCHadronBJ}}_{e/\mu \,\to\, had}}. \tag{4.13}$$

With this assumption, the deconvolution can be performed with yields $N^{\mathrm{MCHadronBJ}}_{lep}$ available from the

MCHadronBJ probability extraction and with $N_e^{J/\psi} = N_\mu^{J/\psi}$ via

$$\mathrm{p}_{lep \to had}^{\text{deconv. } J/\psi} = 2\, \mathrm{p}_{e/\mu \to had}^{J/\psi} \frac{\mathrm{p}_{lep \to had}^{\text{MCHadronBJ}} N_{lep}^{\text{MCHadronBJ}}}{\mathrm{p}_{e \to had}^{\text{MCHadronBJ}} N_e^{\text{MCHadronBJ}} + \mathrm{p}_{\mu \to had}^{\text{MCHadronBJ}} N_\mu^{\text{MCHadronBJ}}} \tag{4.14}$$

.

Again all uncertainties are propagated through the deconvolution expression. Also as before, the largest uncertainty of all involved J/ψ and MCHadronBJ uncertainties and the propagated deconvoluted probability uncertainty is assigned to the deconvoluted probability. The deconvolution is not performed if any of the involved probabilities are not extracted or zero-probabilities.

Finally, PID probabilities $\mathrm{p}_{e/\mu \to e/\mu}$ also have to be separated into probabilities $\mathrm{p}_{lep_1 \to lep_j}$, with $lep_{1,2} = \{e, \mu\}$. This process can be seen as just a combination of the deconvolution steps described above, splitting first the likelihood-selected species $j = e/\mu$ of the probability, and eventually the real/physical species $i = e/\mu$.

In the first step, probabilities $\mathrm{p}_{e/\mu \to e/\mu}$ are split into probabilities $\mathrm{p}_{e/\mu \to lep_j}$ in analogy to Equation 4.10,

$$\mathrm{p}_{e/\mu \to lep_j}^{\text{deconv. } J/\psi} = \frac{\mathrm{p}_{e/\mu \to lep_j}^{\text{MC}}}{\mathrm{p}_{e/\mu \to e/\mu}^{\text{MC}}} \mathrm{p}_{e/\mu \to e/\mu}^{J/\psi}. \tag{4.15}$$

, only assuming that the ratio of individual fake-rates $\mathrm{p}_{e/\mu \to lep_j}$ over the joint probabilities $\mathrm{p}_{e/\mu \to e/\mu}$ in the raw experimental data sample is the same as in the MCHadronBJ sample. Then, the rest of the deconvolution can be performed following the reasoning around Equations 4.13 and 4.14 to obtain probabilities $\mathrm{p}_{lep_i \to lep_j}$, where $lep_{i,j} = \{e, \mu\}$,

$$\mathrm{p}_{lep_i \to lep_j}^{\text{deconv. } J/\psi} = 2\, \mathrm{p}_{e/\mu \to lep_j}^{J/\psi} \frac{\mathrm{p}_{lep_i \to lep_j}^{\text{MCHadronBJ}} N_{lep_i}^{\text{MCHadronBJ}}}{\mathrm{p}_{e \to lep_j}^{\text{MCHadronBJ}} N_e^{\text{MCHadronBJ}} + \mathrm{p}_{\mu \to lep_j}^{\text{MCHadronBJ}} N_\mu^{\text{MCHadronBJ}}} \tag{4.16}$$

. Under substitution of Equation 4.15 for $\mathrm{p}_{e/\mu \to lep_j}^{J/\psi}$, one obtains the final expression as

$$\mathrm{p}_{lep_i \to lep_j}^{\text{deconv. } J/\psi} = 2\, \mathrm{p}_{e/\mu \to lep_j}^{MC} \frac{\mathrm{p}_{e/\mu \to e/\mu}^{J/\psi}}{\mathrm{p}_{e/\mu \to e/\mu}^{\text{MC}}} \frac{\mathrm{p}_{lep_i \to lep_j}^{\text{MCHadronBJ}} N_{lep_i}^{\text{MCHadronBJ}}}{\mathrm{p}_{e \to lep_j}^{\text{MCHadronBJ}} N_e^{\text{MCHadronBJ}} + \mathrm{p}_{\mu \to lep_j}^{\text{MCHadronBJ}} N_\mu^{\text{MCHadronBJ}}} \tag{4.17}$$

.

Again all uncertainties are propagated through the expressions for the deconvoluted probability, and only the largest involved uncertainty is eventually assigned to the deconvoluted probability. The deconvolution is not performed if any involved probability is not extracted or a zero-probability.

Samples of deconvoluted lepton probabilities can be seen in Section 4.1.5. The deconvoluted probabilities are reasonably close to their Monte Carlo counterparts under consideration of their uncertainties. As mentioned throughout this Section, these uncertainties have been assigned conservatively to account for the 'indirect' calculation of lepton probabilities based on the joint probabilities extracted from J/ψ experimental data samples.

It should be noted explicitly that once the joint probabilities are separated, they can be applied to data samples with arbitrary mixing ratios $N_e : N_\mu$, so also to the raw experimental data sets analyzed in this measurement. Additionally, the deconvolution will not be necessary anymore once PID-tagged J/ψ probabilities are available.

Auxiliary Monte Carlo PID Probabilities from the 'Monte Carlo HadronBJ' Sample

PID Probabilities from Monte Carlo Data In the probability extrapolation described below and also in the lepton probability deconvolution explained in the previous Section, PID probabilities extracted from Monte Carlo (MC) simulation data are used. In general, for particle physics experiments usually Monte Carlo particle generators are employed to create particles according to the physics processes in the experiment under study. The generated particles are then propagated through a virtual model of the detector of the given experiment. At Belle, the detector simulation is designed to produce the same data output structures as the real detector and thus creates analogous datasets with reconstructed tracks and associated likelihood values. The benefit of extracting PID probabilities from MC is the fact that the original generated particle species i is stored in the datasets for each reconstructed track, so no background separation has to be performed.

In order to obtain PID probabilities from a data sample as close as possible to the raw experimental data samples, quality cuts of the latter have been mimicked in the MCHadronBJ event selection as well, cf. Section 3.1.1. Histograms with and without PID requirement are filled in dependence of reconstructed track kinematics, in direct analogy to the PID studies described in earlier Sections. The real/physical species i of all reconstructed tracks are inferred from the *get_hepevt* function which relates a reconstructed track to its generated particle in Monte Carlo data. PID probabilities are subsequently extracted similarly to the probability extractions described earlier- since no combinatorial background is present and for each single particle its real/physical species i and likelihood-cut selected species j are known, no fitting procedure is needed. Thus the yields N_i^{MC} and $N_{i \to j}^{\mathrm{MC}}$ allow for direct extraction of all PID probabilities $\mathrm{p}_{i \to j}$ from the

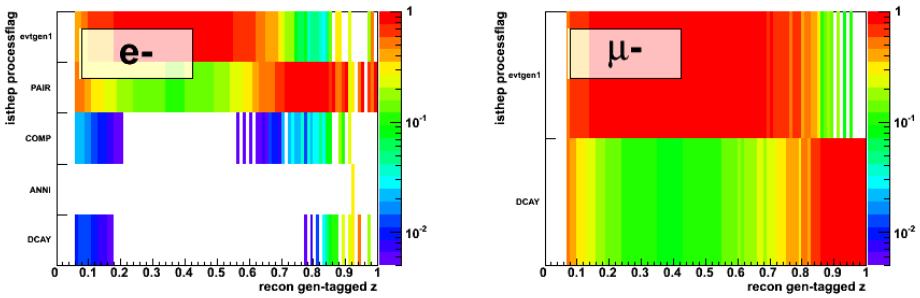


Figure 4.13: Plot of flag for generating process for electrons and muons, from Experiment $7-55$ for 3 streams of Belle 'uds' and 'charm' Monte Carlo data for the 'continuum' energy level (about $607*10^6$ events). The fractional energy z is determined with the mass of the real/physical species i but with reconstructed momentum. The flag 'evtgen1' refers to the particle being generated in the physical reaction $e^+e^- \rightarrow q\bar{q}$. The flag 'PAIR' refers to electrons created in processes $\gamma \rightarrow e^+e^-$, the flag 'DCAY' refers to particles created from heavier particles in decay-in-flight processes.

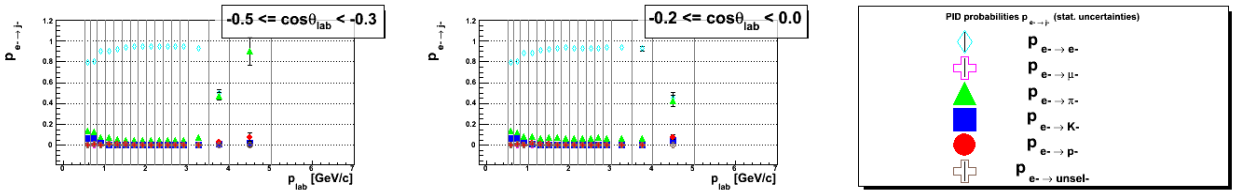
familiar yield ratio.

Necessity to Analyze 'Monte Carlo HadronBJ' Sample for MC Probabilities The standard Belle Monte Carlo contains reactions $e^+e^- \rightarrow q\bar{q}$. In experimental reality however, also other processes occur in electron-positron collisions, pass the hadronic skim selection criteria and then are present in the experimental raw data used for this study, as e.g. $\tau\tau$ or 2γ events. A test was performed by extracting PID probabilities from the standard Belle Monte Carlo data alone. It turns out that for $z \geq 0.75$, leptons from events $e^+e^- \rightarrow q\bar{q}$ mainly originate from pair production (electrons) and decay-in-flight (muons), cf. Figure 4.13.

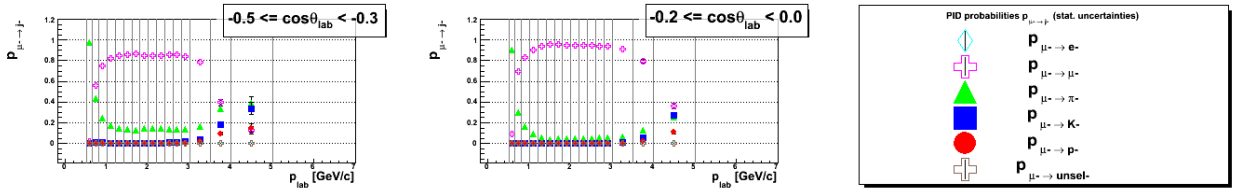
These leptons show substantially higher PID fake-rates into hadrons than leptons produced in quark fragmentation and QED events since they are produced in the detector after the decay of an initial particle, cf. Figure 4.14. This makes it difficult for the detector tracking to properly pick up the kinematic parameters of the particle. Also for that reason, PID measurements often do not match the expected detector interaction behavior of leptons. This leads to PID likelihood values which make the particles prone to fail the lepton likelihood cuts but in turn pass the hadron selection cuts, and so contribute to lepton-hadron fake-rates.

In the PID correction (described in a later Section), high lepton-hadron fake-rates result in large yield gains for pions, and large yield drops for leptons and kaons arising from the respective inverted probability matrices. These matrices create unphysical PID-corrected spectra at high z, cf. Figure 4.15.

After further study, it was found that including other processes in the extraction of MC probabilities, especially from $\tau\tau$ and 2γ events, adds many more leptons at high z, cf. Figure 4.16 and Appendix B.6. Those leptons achieve high PID efficiencies at high momenta (especially for muons, cf. Figures B.20 in



(a)



(b)

Figure 4.14: Electron (a) and muon (b) PID probabilities extracted from Experiment $7-55$ for 5 streams of Belle 'uds' and 'charm' Monte Carlo data for the 'on-resonance' energy level (about $5.6*10^9$ events). All specifications given for corresponding earlier plots hold, e.g. Figure 4.5.

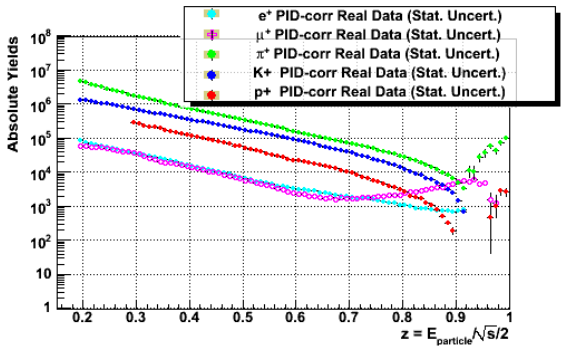


Figure 4.15: Plot of PID-corrected experimental data yields, using only generic Belle Monte Carlo PID probabilities to extrapolate probabilities from experimental data at high track momenta, as introduced in Section 4.1.5. Only statistical uncertainties are shown. Data points below zero are not plotted.

the Appendix), which then produce physically consistent PID-corrected spectra. Another way of describing the reasoning is to state that the PID probability matrix used to correct for particle misidentification has to match the matrix which acts on the real/physical yields in Equation 4.1. The attempt to correct raw experimental yields for PID misidentification with different PID information than what they were 'encoded with' is bound to fail, because the multiplication of $[\mathcal{P}]_{ji}$ and $[\mathcal{P}]^{-1}_{ji}$ will not achieve unity in the correction step, cf. Equation 4.1. Since the attempt was made to create a collection of data samples in Monte Carlo which closely resembles the experimental data HadronBJ skim, the Monte Carlo data samples were collectively labeled as 'Monte Carlo HadronBJ'.

The MCHadronBJ Data Samples With the support of our collaborator Inami-san, Monte Carlo samples for different processes occurring in e^+e^- collisions have been analyzed at Nagoya University. A list of all analyzed processes and their cross sections at Belle on-resonance energy level are given in Table 4.8, together with the corresponding cross sections of generic Belle MC quark-antiquark production. No data files were available for the 2-photon production of τ pairs- since the production cross section is the lowest of all processes (0.018 nb, half of the otherwise smallest 2γ charm quark-antiquark production), the omission of this sample is not considered to have significant impact on the extracted PID probabilities. In lieu of a likewise missing $'eedd'$ sample, $'eeuu'$ was used as a substitute, properly scaled by the cross-section ratio, cf. Table 4.8. Also the effect of weights $\neq 1.0$ used in the generation for the $Bhabha$, $eecc$ and $eess$ samples at Nagoya are taken into account in the analysis. The effect of the energy difference between on-resonance and continuum energy levels (60 MeV)on the cross sections has been neglected.

Process $e^+e^- \rightarrow$	Cross Section [nb]
$q\bar{q},\ q = \{u, d, s, c\}$	3.3
$B^(+)B^-, B^0B^0$	1.1
$Bhabha$	123.5($\pm$0.2)
$\mu\mu$	1.005($\pm$0.001)
$\tau\tau$	0.9187($\pm$0.0003)
$eeee$	40.9
$ee\mu\mu$	18.9
$eeuu$	11.7
$eedd$	0.798
$eess$	0.227
$eecc$	0.030

Table 4.8: Component processes of the MCHadronBJ data sample and production cross sections in e^+e^- collisions at Belle on-resonance energy level. The first two cross sections are quoted from Ref. [62], the remaining ones from Ref. [78].

Overall, from two MC production streams 557 (572) runs of on-resonance data and 63 (64) runs of continuum data for Experiment 41 conditions have been analyzed for all processes. Including standard

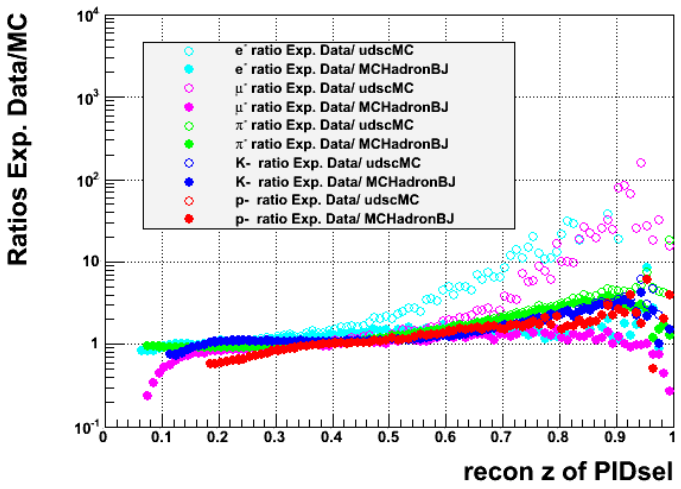


Figure 4.16: Ratios of experimental data over standard Belle $uds + charm$ MC (empty circles) and over MCHadronBJ data (full circles) for 63 runs of Experiment 41 at continuum energy, versus reconstructed z. All shown species are assigned to negatively-charged tracks based on their PID likelihood values. The same selection cuts are applied in the analyses of all data samples. No uncertainties are shown.

Belle MC $q\bar{q}$ data, the analyzed data sample consists of in total about $430 * 10^6$ events. To illustrate the improvement of correspondence between experimental data and the Monte Carlo PID data sample under consideration of additional processes other than $e^+e^- \rightarrow q\bar{q}$, Figure 4.16 shows ratios of experimental data over standard Belle $uds + charm$ MC and over MCHadronBJ data for 63 runs of Experiment 41 at continuum energy. The x axis shows the reconstructed fractional energy z for likelihood-selected species j. All samples are extracted with the exact same analysis module. As can be seen from the plot, discrepancies to experimental data (which reach up to a factor of 100 in the muon spectra at high z) can be greatly reduced by considering the additional physics processes given in Table 4.8. The remaining discrepancies, especially at high z, are most probably due to non-optimized generator tuning to properly describe high momentum products of quark fragmentation processes. The focus of the Belle Monte Carlo lies on reproducing kinematics of B meson decays with high accuracy, which only involve particles with fractional energy $z <\sim 0.5$.

In Appendix B.6, plots are given which compare the contributions of all MCHadronBJ processes to PID-selected yields for species j versus fractional energy z. Also for comparison, experimental data yield distributions are included in the plots. Significant contributions to lepton spectra at high z come from processes $eeuu$, $ee\mu\mu$ and $\tau\tau$ events for likelihood-selected electrons/positrons and from processes $\mu\mu$, $\tau\tau$ and $ee\mu\mu$ for muons, respectively.

MCHadronBJ PID Probabilities PID probabilities have been extracted from the MCHadronBJ sample following the procedure outlined in Section 4.1.5. In case the histogram without PID selection is empty for a

given generated species i, no PID statement can be extracted for the respective kinematic bin. This was only encountered for the highest momentum and most backward kinematic bin for $i = p^7$. PID information for that bin is copied from the next-lower laboratory frame momentum bin. The error made by this substitution is estimated to be negligible due to fair continuity of probabilities against the laboratory frame momentum and large uncertainties because of low statistics.

For all histogram with PID selection with no entries, zero-probabilities are calculated. The uncertainty on the probability here is set equal to one over the yield in the histogram without PID selection. This can be interpreted as an upper bound for a possible non-zero probability and reflects the probability for a single entry in the histogram with PID selection.

Uncertainties for the extracted yield ratio probabilities are again calculated assuming binomial uncertainties for yields from histograms with PID selection, and Poisson uncertainties for yields from histograms without PID selection.

Appendix B.6 shows samples of PID probabilities extracted from the MCHadronBJ sample, in analogy to similar plots in the extraction of probabilities from experimental data in earlier Sections. All descriptions to motivate the kinematic dependence of the extracted probabilities given in earlier Sections apply. Prominent in the plots are the slow turn-on for lepton efficiencies and high lepton-hadron fake-rates at low momenta. For high momenta, lepton efficiencies stay large, as expected and in contrast to probability extractions from 'uds' and 'charm' Belle Monte Carlo alone, compare Figure 4.14.

Probability Extrapolation

Extrapolation Algorithm In order to perform the PID correction on the entire kinematic bin grid in the laboratory frame with resolved track charges, 306 5×5 PID matrices have to be fully populated. As already mentioned in the previous Section, the three most backward bins at highest laboratory frame momentum are excluded from the analysis since all of them would be populated by particles with $z > 1.0$. Thus 7500 PID probabilities (in addition to 1500 *unsel* probabilities used for putting further constraints on the extrapolation) have to be known. From the PID probability extractions studies described in the previous Sections, 4699 probabilities are available from experimental data. About half of the 4301 missing probabilities are lepton probabilities due to the limited kinematical availability of J/ψ particles.

An extrapolation algorithm has been developed to extrapolate available PID probabilities to kinematical areas where no PID statement could be extracted due to insufficient statistics or rejected fits. The intention is to find extrapolation techniques which are driven by experimental data and only additionally use information

[7] The most backward bin for highest momenta actually is the fourth bin, since all lower bins are outside the physical range of $z <= 1.0$.

from Monte Carlo. As initially mentioned, Monte Carlo descriptions of PID subdetectors are expected to not model the detector response sufficiently well to extract reliable PID information at a high level of precision.

Three methods for deriving unextracted PID probabilities from extracted experimental and Monte Carlo probabilities have been developed. First, the 'crossfade' method uses PID probabilities available in immediately neighboring bins on the 2-dimensional kinematic plane of ($p_{\mathrm{lab}},\ cos\theta_{\mathrm{lab}}$). The value for the missing probability is extrapolated as the average of the neighboring probabilities, with all uncertainties propagated. If only one neighboring probability in the p_{lab} or the $cos\theta_{\mathrm{lab}}$ direction is available, this probability is directly taken as output of the crossfade method.

The second method ('unitarity' method) takes advantage of the probabilistic character of PID probabilities. When grouped into a matrix $[\mathcal{P}]_{ji}$, all columns have to sum up to 1.0 within the probability uncertainties. In other words, by design of the likelihood cuts in this analysis, a track in the Belle data tables has to be likelihood-selected as one of species $j = \{e,\ \mu,\ \pi,\ K,\ p\}$ with a probability of 100%[8]. The unitarity method is applied for extrapolation whenever all elements in a matrix column are known except for one element. The extrapolation result of this method is then equal to 1 minus the sum of all available probabilities minus the square root of their added uncerainties[9]. All uncertainties are propagated as well.

The last extrapolation method uses a similar approach as the crossfade method. The two main differences are however, that only experimental probabilities neighboring in p_{lab} direction are considered, and that not the experimental probabilities directly but their differences to MCHadronBJ probabilities enter this extrapolation method ('MC crossfade' method). At least one non-zero-probability extracted from experimental data is required for this method to provide output. For the extrapolation, the average difference of the neighboring experimental and Monte Carlo probabilities is calculated. The value for the missing probability is then extracted by combining the average difference with the value of the Monte Carlo probability at the position of the unavailable experimental data probability. All uncertainties are again propagated.

Extrapolating the missing probability by considering the charge-conjugate probability if available did not prove to be a viable concept since charge differences between probabilities are physically expected and also have been found in experimental data.

Finally, the values from all extrapolation methods for a missing probability are averaged to yield the final extrapolated probability, and all uncertainties are propagated. Since Monte Carlo data are used at some points in the extrapolation of experimental probabilities, uncertainties are assigned conservatively: Every

[8]The in fact rare occurence of likelihood-unselected tracks is taken into account by the calculated PID probabilities $\mathrm{p}_{i \to unsel}$. These probabilities are extrapolated where available with the methods described in analogy to all other probabilities, and eventually added in quadrature to the uncertainty of every probability for higher accuracy.

[9]The last term is considered since this method tends to overestimate the missing probabilities without considering uncertainties.

extrapolated probability is assigned the largest uncertainty from all experimental probabilities involved in the extrapolation step.

All experimental and MC data input zero-probabilities are set to one per mille to enhance stability of the extrapolation and the matrix inversion lateron. The algorithm is started at the first laboratory frame momentum bin and probes all $cos\theta_{\rm lab}$ bins from backward to forward direction for missing probabilities. The latter are calculated with the methods given above.

Dual Probability Extrapolations Attention has to be paid to make sure the algorithm only extrapolates existing probability tendencies, but does not create tendencies by itself which are unphysical or do not agree with experimental data trends. This is made difficult by the fact that only few probabilities are extracted from experimental data at highest laboratory frame momenta. The algorithm is only very lightly constrained in these kinematic regions since only few 'pivot' points are available. In order to impose tighter constraints on the extrapolation results, it has been decided to extract two sets of extrapolated PID probabilities.

In order to construct the first set of extrapolated probabilities, all extrapolation methods described in the Section above are employed. To obtain the second set, the same extrapolation methods as for the first set are used, if both neighbors of the missing probability in momentum direction have been extracted from experimental data. If one or both of the neighboring probabilities are missing as well[10], the value for the probability under consideration is directly copied from the corresponding MCHadronBJ probability, and the MCHadronBJ uncertainty is assigned. Thus the second set of extrapolated probabilities is dominated by Monte Carlo input, especially at high momenta.

Since the MCHadronBJ probabilities mostly differ from the probabilities extracted from experimental data, unitarity for columns of the extrapolated PID matrix in the second set might not be given anymore. Thus the unitarity extrapolation method is slightly altered for the extraction of the second set of PID probabilities: If the last missing probability of a matrix column is filled from a MCHadronBJ probability, its uncertainty is set equal to either the value of one minus the sum of all other probabilities in the same matrix column or to the uncertainty of the MCHadronBJ probability, whichever is larger.

Eventually, the PID correction of the raw experimental measurement sample is performed once for each set of probabilities. The final PID-corrected yields are then calculated as the average of the two different corrections, which is described in the following Chapter.

Extrapolation Results In this Section, results of the performed PID probability extrapolations are given. Shown are, against the laboratory frame momentum for different scattering angle bins, probabilities which

[10]This can also be the case due to the fact that the missing probability is situated at the end or start point of the momentum fiducial area of $p_{\rm lab} >= 500\ MeV/c$ and $z <= 1.0$.

have been extracted from experimental and MCHadronBJ data or which have been calculated by the extrapolation algorithm. Also results from the lepton deconvolution procedure described in Section 4.1.5 are shown.

In each plot, inverted triangles given below 0.0 indicate the extrapolation method used to arrive at the extrapolated probability. Red triangles stand for the usage of the crossfade method, green triangles for the unitarity method and blue triangles for the MC difference method. Experimental probabilities are given as symbols with color indicating the likelihood-selected species j, where $j = e$ (cyan), μ (magenta), π (green), K (blue), p (red). Gray symbols describe MCHadronBJ probabilities, black symbols indicate final extrapolated probabilities for the first set of probabilities. All probabilities from the second extrapolation set can be inferred from the probabilities shown: The second set contains all colored experimental data probabilities and black extrapolated probabilities, unless the missing probability lacks at least one colored neighbor versus the laboratory frame momentum. In that case, the probabilities of the second extrapolated set are equal to the MCHadronBJ probabilities in gray. Additional information is given for each plot by the text box.

Figure 4.17 shows probabilities for all PID efficiencies $\mathrm{p}_{i \to i}$. Figure 4.18 shows extrapolation results for the largest fake-rates $\mathrm{p}_{i \to j}$ for each species i. In general, it can be seen that hadron probabilities are available from experimental data analyses over most of the kinematic range, and only few have to be inferred from extrapolation. Lepton probabilities are relying on more significantly. Another noteworthy observation is that probabilities extracted from experimental data are mostly close to probabilities from MCHadronBJ data, but yet often significantly different. Also it should be noted again that the extrapolation algorithm assigns very conservative uncertainties, which are usually larger than the uncertainties for probabilities extracted from experimental data, to account for the algorithm being an 'indirect' way of determining PID information in kinematical bins.

Similarly, the lepton PID deconvolution produces probabilities with conservative uncertainty estimates. Colored markers in all lepton plots show joint probabilities before deconvolution. Black markers show deconvoluted probabilities in bins where colored markers are available, and describe extrapolation results otherwise.

Figure 4.19 shows an illustrational plot of half of all 7500 probabilities, from the second extrapolation set.

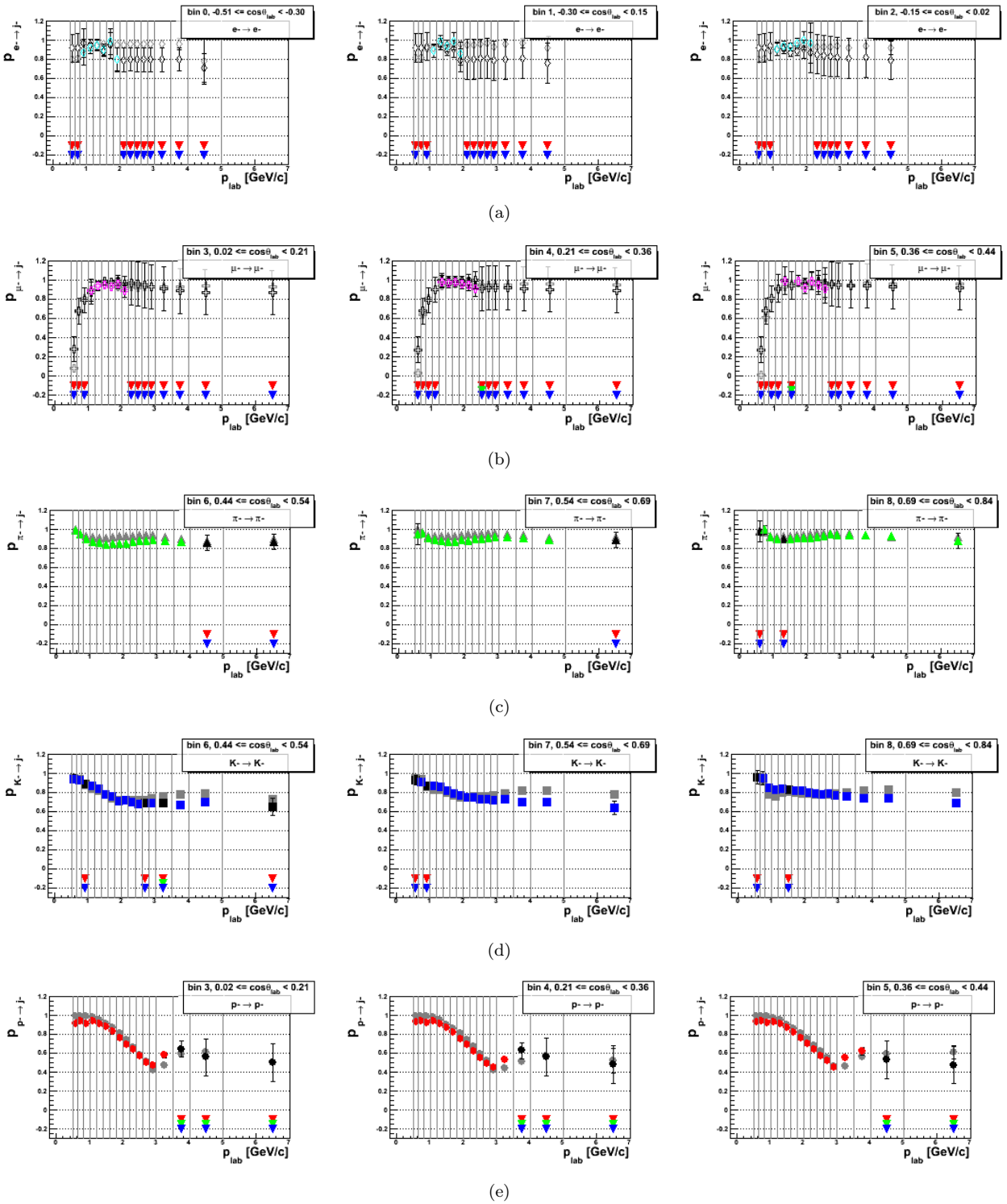
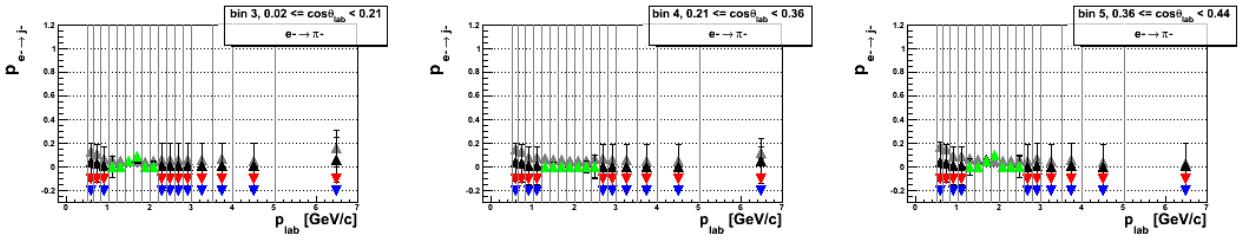
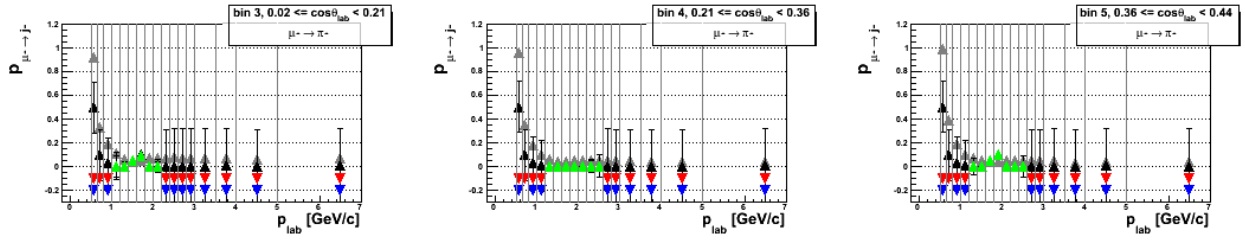


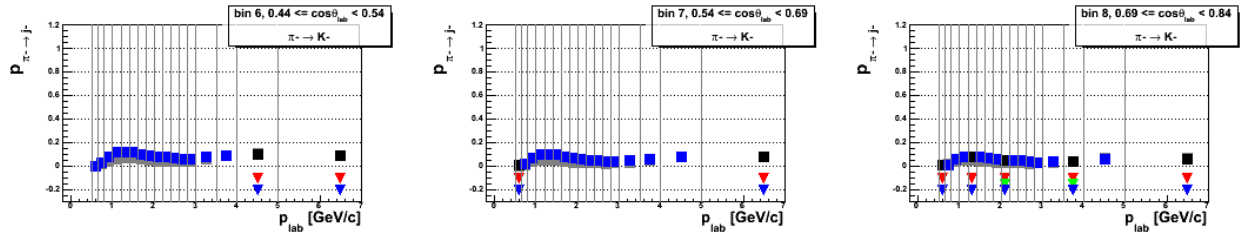
Figure 4.17: First set of extrapolated PID efficiencies $p_{i \to i}$ for negatively charged electrons (a), muons (b), pions (c), kaons (d) and protons (e). All plot specifications are given in the text.



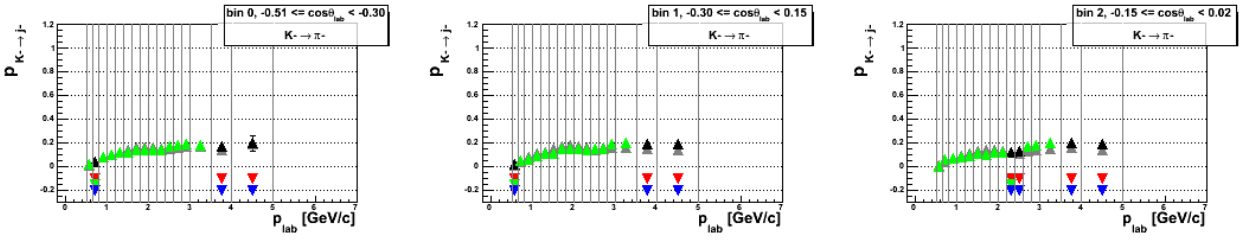
(a)



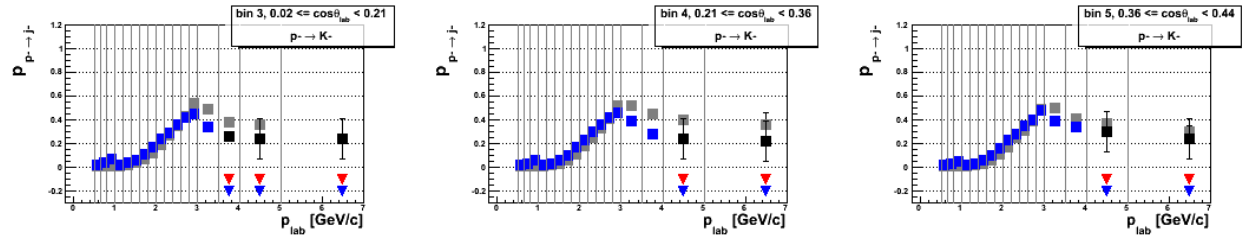
(b)



(c)



(d)



(e)

Figure 4.18: First set of extrapolated largest PID fake-rates $p_{i \to j}$ for negatively charged electrons (a), muons (b), pions (c), kaons (d) and protons (e). All plot specifications are given in the text.

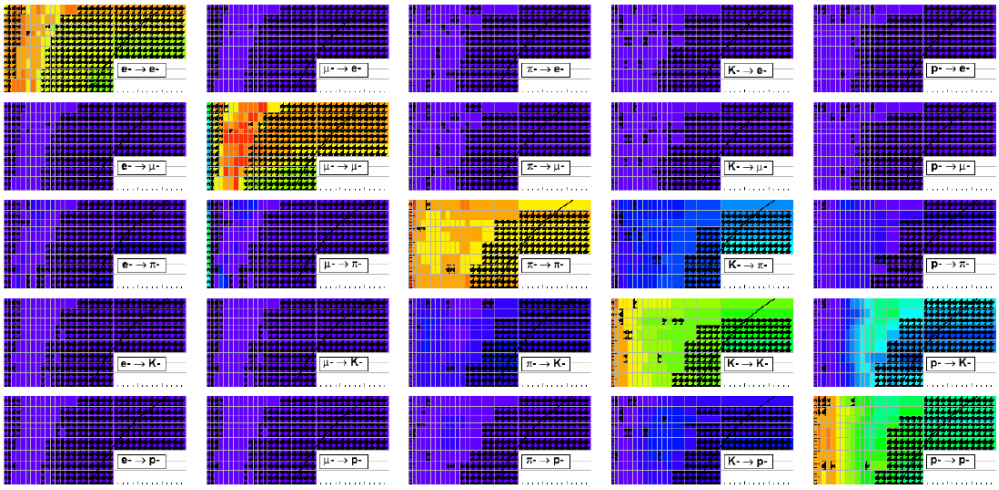


Figure 4.19: Illustrational plot for half of all 7500 PID probabilities from the second set of extrapolated probabilities. The z coordinate color code runs from 0.0 to 1.0. The x axis for each sub-plot shows laboratory frame momenta in GeV/c, the y axis shows consines of laboratory frame scattering angles. Kinematic bins where the probability has been extrapolated with the algorithms described above are plotted with a fill pattern on top of the color coding. No uncertainties are shown.

4.1.6 Inversion of PID Probability Matrices $\mathcal{P}$

After the dual PID probability extrapolation described in the previous Section, two complete sets of probability matrices $\mathcal{P}$ are available for each kinematical bin. Both sets will be inverted independently. The same procedure to invert the matrices and propagate the uncertainties on the single probabilities through the inversion has been applied to both sets.

Inverse Probability Matrix Elements

In order to perform all necessary matrix inversions in the present study, the ROOT-implemented Single Value Decomposition (SVD) inversion algorithm has been used. To obtain inverse probability matrix elements, all 5×5 probability matrices have only been inverted once. All inversions were successful.

Uncertainties of Inverse Probability Matrix Elements

Since the inversion of a matrix can be regarded as solving a system of linear equations, uncertainties on all probabilities can in general be propagated through the inversion, utilizing the usual analytic formulae. The 1-σ uncertainties on the inverse matrix elements should then properly reflect the 1-σ uncertainties on the uninverted matrix elements. However, a study (Ref. [79]) was found which limits the applicability of the usual analytical uncertainty propagation formulae depending on the significance of the matrix determinant.

Extraction of Inverse Element Uncertainties from Matrices with Low Determinant Significance The authors of Ref. [79] focus in their study on uncertainty-endowed matrices with low ratios of $\det(\mathcal{A})/\sigma_{\det(\mathcal{A})}$. Low ratios, or low determinant significances are a feature of matrices with low determinant values (from large off-diagonal elements) and/or large matrix element uncertainties. To illustrate the effect of low determinant significance on inverse matrix element uncertainties, let the matrix $\mathcal{A}_{\text{var}}$ be an instance of the matrix $\mathcal{A}$ where the uninverted elements $[\mathcal{A}]_{mn}$ have been varied randomly according to their 1-σ uncertainties. If a sufficiently large sample of randomly varied matrices $\mathcal{A}_{\text{var}}$ is formed, and the significance of $\det(\mathcal{A})$ is sufficiently low, some matrices $\mathcal{A}_{\text{var}}$ will show negative determinants. This implies that the determinants of other matrices $\mathcal{A}_{\text{var}}$ can fluctuate arbitrarily close to 0. If all varied matrices in the sample are inverted, the matrices with determinants close to 0 will show arbitrarily large inverse matrix elements[11] The authors of Ref. [79] argue that this nonlinear behavior for matrices with low determinant significance is not properly reflected by analytical covariance expressions.

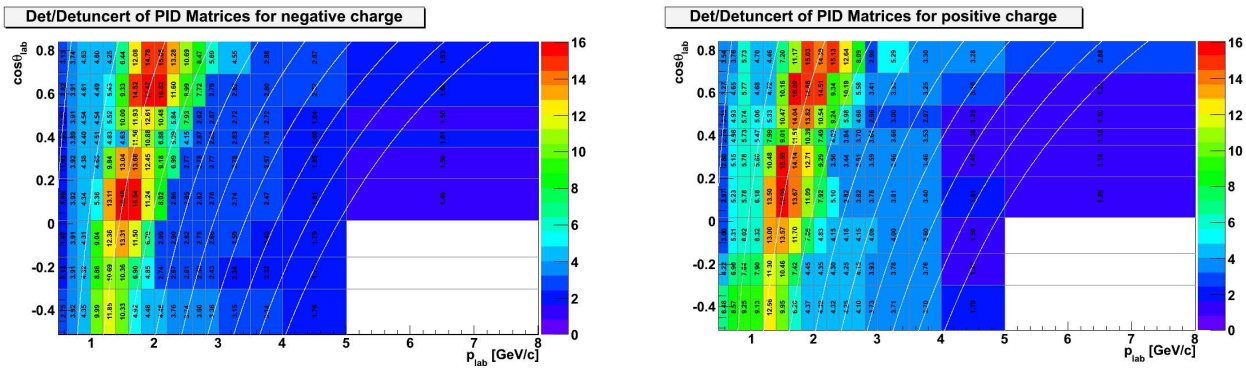
The idea can also be expressed differently by stating that analytical expressions rely on involved observables being distributed normally. However, a distribution of $\det(\mathcal{A}_{\text{var}})$ from a matrices $\mathcal{A}$ with low determinant significance will feature values close to 0, which in turn causes an distribution $1/\det(\mathcal{A}_{\text{var}})$ with tails to very large (in absolute terms) inverse values.

The authors of Ref. [79] study the discrepancy between the analytical and a statistical approach to obtain a single covariance element of the inverse of a given 2×2 matrix with non-zero element uncertainties. They find the discrepancy to rise above 4% once the determinant of the original matrix drops below 10 in units of its uncertainty, with the discrepancy quickly growing larger for even lower determinant significance.

To evaluate the relevance of the statements made in Ref. [79] for the present analysis, Figure 4.20 shows the distributions of ratios $\det(\mathcal{P})/\sigma_{\det(\mathcal{P})}$ for all PID matrices from the second set of extrapolated probabilities. All matrices from the second set show similar values. In the plots, all explicit ratio values are given as well. The uncertainties of the determinants have been evaluated using an expression from Ref. [79].

As can be seen from the Figure, determinants of PID matrices are low and determinant uncertainties are high at very low and high laboratory frame momenta, and ratios of the two numbers are well below 10. This can be explained by the findings in earlier Sections 4.1.3 and 4.1.5: At low momenta for leptons and at high momenta for protons, PID fake-rates reach the same values as the respective diagonal efficiencies or even surpass them. Thus the corresponding probability matrices $\mathcal{P}$ show small determinants in these kinematical regions. Furthermore, the lepton deconvolution and extrapolation algorithms assign conservative, i.e. large uncertainties, and also uncertainties on probabilities extracted from experimental data increase at high

[11] This can be induced from considering simple 2×2 and 3×3 inversions, where each inverse matrix element expression contains a factor $1/\det(\mathcal{A})$.



(a) (b)

Figure 4.20: Ratios of determinants of all probability matrices $\mathcal{P}$ on the $(p_{\mathrm{lab}},\ \cos\theta_{\mathrm{lab}})$ bin grid over their respective determinant uncertainties for negative (a) and positive (b) track charges. White lines indicate lines of equal values of z for charged pions. The rightmost curve, $z = 1.0$, represents the maximum kinematical range at Belle.

momenta due to low statistics. Thus, in addition to low determinant values, also the uncertainties on probability matrix determinants are large in those kinematical bins, overall producing determinants with low significance.

Extraction of Inverse Matrix Element Uncertainties through Monte Carlo Procedure As a result of the reasoning above, it was decided to employ statistical methods for the extraction of reliable inverse matrix element uncertainties. A Monte Carlo procedure has been adopted.

For each probability matrix $\mathcal{P}$, 'random instances' $\mathcal{P}_k^{\mathrm{MC}}$ are created by varying the original matrix components $\mathrm{p}_{i\ \to\ j}$ randomly according to their uncertainties. The components of the k^{th} random instance $\mathcal{P}_k^{\mathrm{MC}}$ of a matrix $\mathcal{P}$ read

$$\mathrm{p}_{i\ \to\ j}^{\mathrm{MC}\ (k)} \ \doteqdot\ \mathrm{p}_{i\ \to\ j} + \sigma_{(\mathrm{p}_{i\ \to\ j})}\Gamma_k, \tag{4.18}$$

where the symbol Γ_k represents a normally distributed pseudorandom variable with a mean value of zero and a standard deviation of 1. Each random instance matrix $\mathcal{P}_k^{\mathrm{MC}}$ is inverted and all elements of the inverse matrix are recorded in separate Single Element Distributions (SEDs).

The SEDs are considered representative of the variation of inverse matrix elements under random variation of the original, uninverted matrix elements within their 1-σ uncertainties. Thus the uncertainties of all inverse matrix elements are extracted as the standard deviations of all recorded values in each SED, with respect to the inverse matrix element values from the one-time analytical inversion of the original matrix. For a given random instance matrix, all inverted elements are correlated by the inversion process. There-

fore, the extracted uncertainties from the SEDs are assumed to also contain contributions from correlations between matrix elements.

The number of MC random instance matrices to be created can be inferred from precision requirements to the uncertainties extracted from the SEDs. The statistical standard deviation of an estimator $\hat{\sigma}$ for the standard deviation of a sample can be approximated by [5]

$$\sigma_{\hat{\sigma}} \;=\; \frac{\sigma_{\text{true}}}{\sqrt{2N}}. \tag{4.19}$$

If the extracted uncertainties from all SEDs should be calculable with a precision of larger than one per mille, a minimum sample size of $N = 5 * 10^5$ is required for each SED. A similar expression to Equation 4.19 for the standard deviation of an estimator for the mean of a sample involves a factor $1/\sqrt{N}$. Also requiring statistical fluctuations of the mean position to be smaller than one per mille, a sample size of $N = 10^6$ has been chosen for each probability matrix.

Since also the MC random instance matrices should represent physical probability matrices, all random instance elements are required to be non-negative. Additionally, the sums of all elements in each matrix column are required to be no further from 1.0 than the column sum uncertainty of the original uninverted probability matrix (analogous to the 'unitarity' requirement used in the probability extrapolation algorithm). However, these requirements cause asymmetric/skewed SEDs, especially for those where the original PID probability is small and/or has large uncertainties. Since the requirements are vital for physicality of all SEDs and also should affect the uncertainties on the final inverse matrix elements, it has been decided to extract the inverse element uncertainties by calculating asymmetric standard deviations.

Figure 4.21 shows all 25 SEDs for the probability matrix of the second set of dual extrapolations in the kinematical bin $p_{\text{lab}} = 4,\ cos\theta_{\text{lab}} = 5$, for negative track charges. Red dashed lines indicate the value of the respective inverse element from the one-time analytical inversion. For illustration, red solid lines indicate matrix element uncertainties evaluated from standard analytical expressions. Black solid lines show the final asymmetric matrix element uncertainties determined from standard deviation calculations from the distribution.

Preservation of Numerical Stability of Extracted Inverse Element Uncertainties Following the reasoning given above, uninverted probability matrices with low determinant significance can potentially cause MC random instance matrices to show negative determimants and thus fluctuations arbitrarily close to 0. As an illustration, Figures 4.22 show determinant (a) and inverse determinant (b) distributions of MC random matrices for a kinematical bin with large determinant significance. As a contrast, Figures 4.22 c)

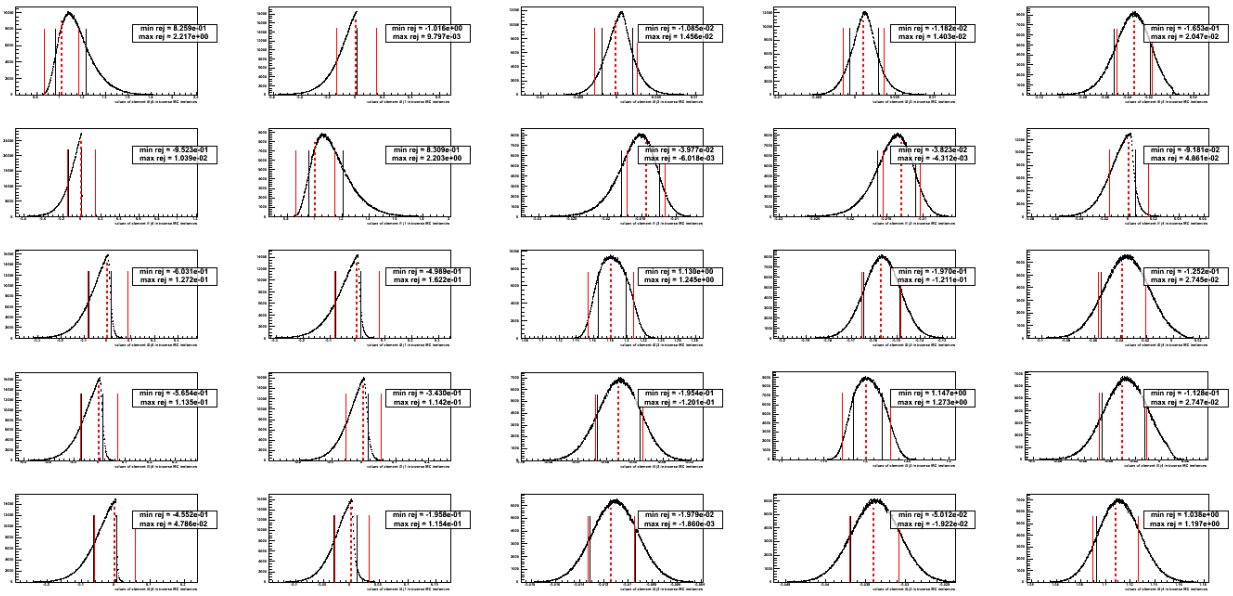
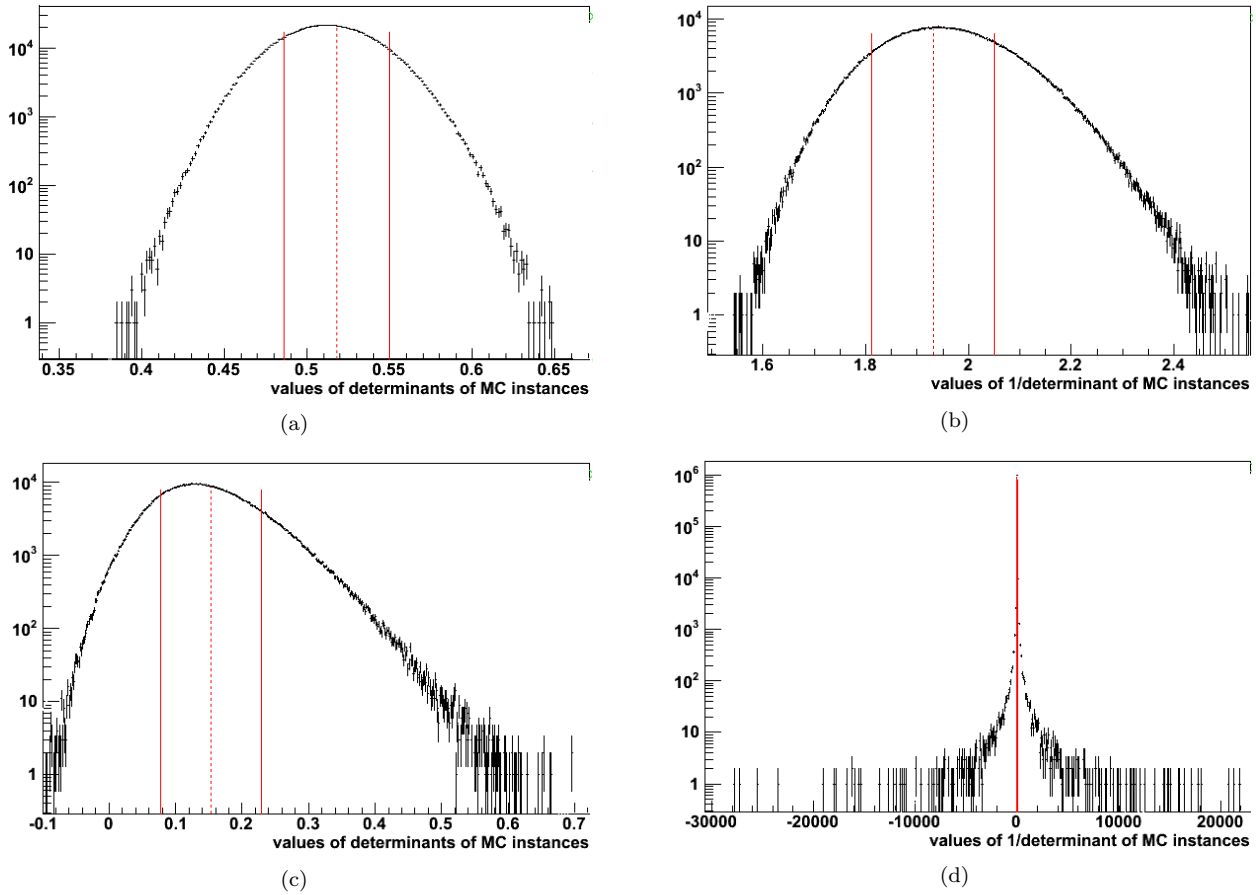


Figure 4.21: Single Element Distributions (SEDs) from the matrix inversion uncertainty propagation for the probability matrix in the kinematical bin $p_{\mathrm{lab}} = 4$, $cos\theta_{\mathrm{lab}} = 5$, for negative track charges. The plots are described in the text.

and d) display analogous plots for a probability matrix with low determinant significance. It can be seen clearly that low determinant significance causes large fluctuations in the inverse determinant distributions, and thus also in the resulting SEDs.

It has been found that those fluctuations of random instance determinants cause very large values in the SEDs, which in turn dominate the uncertainties extracted from the SEDs. While the nonlinear behavior caused by low determinant significance should be accounted for in the extracted uncertainties, very large fluctuations in SEDs do not represent any physical meaning and distort the extracted uncertainty information. Also extracted uncertainties would vary for different inversion samples, since different fluctuations occur.

Thus, for each uninverted probability matrix, it was decided to identify matrices with determinants closest to 0 and remove the closest one percent of the sample from the analysis. The effect of this truncation on the extracted uncertainties is estimated by analyzing a random Gaussian distribution with 10^6 entries. Discarding the largest one per cent of all entries changes the value of the standard deviation relative to the mean by about 3 per cent, which was considered sufficient to retain precision. This procedure renders the extracted uncertainties mostly independent of very large fluctuations, but still preserves the SED standard deviations as being representative uncertainties also for matrices with low determinant significance.



(a) (b) (c) (d)

Figure 4.22: Determinant and inverse determinant distributions of MC random matrices. Plots a) and b) show data for a probability matrix with large ratios of determinant over determinant uncertainty. Plots c) and d) display similar plots, but now for a probability matrix with low determinant significance.

Monte Carlo Seeds and Numerical Precision For the creation of values for the pseudorandom variable Γ_k, a ROOT random generator class based on the Mersenne Twister generator [80] has been used. The periodicity of this generator, $2^{19937} - 1$, is far above the approximate number of different random numbers needed in the present analysis, $\sim \ 2^{33}$. [12] Nevertheless, the generator has been seeded afresh in every kinematic bin. The seed is performed by the creation of a Universally Unique Identifier (UUID) which fills the first 8 integers of the seed array. Due to this procedure, the seed is reportedly guaranteed to be unique in space and time [81].

In order to assure numerical precision of each single matrix inversion, all random instance matrices are required to show a 'condition number' $< \ 10^8$. Reference [82] describes the condition number as an estimate

[12] The number $2^{33} \ \sim \ 7.5 \cdot 10^9$ is a lower bound for all necessary calls to the generator. The number can be calculated from 25 matrix elements in 300 matrices with sample sizes of $N = 10^6$ for each matrix.

for numerical accuracy when solving $Ax = b$ for x. The condition number k for a matrix A is evaluated from

$$k = \|A\|_1 \, \|A^{-1}\|_1, \text{ where } \|A\|_1 \equiv \max_j(\Sigma_j |A_{ij}|). \tag{4.20}$$

For double precision numbers, a matrix condition number of 10^n entails a precision in x of $15 - n$ digits. This ensures 7 digit precision for all calculated elements. As an illustration, Figure 4.23 shows plots of differences between unity matrix elements and the respective elements from product matrices from all random instances $\mathcal{P}_k^{\mathrm{MC}}$ and their respective inverse matrices for negative track charges, integrated over all kinematic bins. All values are similar to 0 down to $< 2 * 10^{-7}$ as expected.

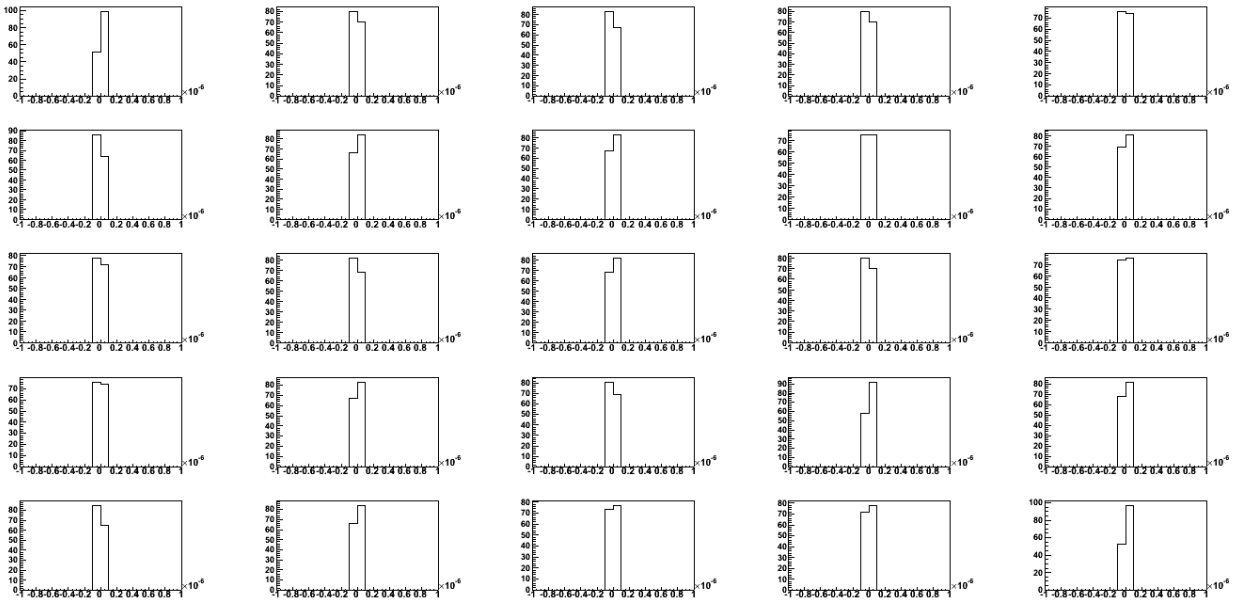


Figure 4.23: Difference between unity matrix elements and matrix elements of the product of analytical inverse and uninverted probability matrices, integrated over all kinematic bins for negative track charge. All differences are $< 2 * 10^-7$ as proposed by the condition number description.

The first set of PID probability matrices from the dual extrapolation described in the previous Section produces similar plots and numbers as shown in this Section for the second set of matrices.

4.1.7 Correction for Particle Misidentification

With all PID matrices from both probability sets after dual extrapolation being inverted, the raw experimental yields shown in Section 3 can be corrected for particle misidentification.

Correction Formalism

The correction can be written down, following Equation 4.1, as

$$N_i^{\text{real ID}}(z) \;=\; \mathcal{P}_{ij}^{-1}(p_{\text{lab}}, cos\theta_{\text{lab}})\; N_j^{\text{likelihood-selected ID}}(p_{\text{lab}}, cos\theta_{\text{lab}}; z_i), \tag{4.21}$$

with an implicit sum over j on the left hand side.

The elements of the inverse probability matrix $\mathcal{P}^{-1}$ cannot be interpreted as probabilities anymore, since in matrix inversions inverse elements larger than 1.0 and also negative elements occur. All inverse matrix elements are correlated with each other by the inversion process. Thus, in contrast to the uninverted probability matrix, single inverse elements cannot be assigned clearly separable interpretations but have to be considered in the context of all other inverted elements.

In Equation 4.21, the kinematic dependencies of all involved quantities are given explicitly. Since the inverted matrices are laboratory frame quantities, whereas the corrected yields $N_i^{\text{real ID}}$ are binned in fractional center-of-mass energy z, a connection has to be established between the two reference frames. This is done by calculating fractional center-of-mass energies z_i for every reconstructed track five times, once for each mass/species hypothesis $i = \{e, \mu, \pi, K, p\}$, to obtain yields $N_j^{\text{likelihood-selected ID}}(p_{\text{lab}}, cos\theta_{\text{lab}}; z_i)$. Then in Equation 4.21, the yields $N_i^{\text{real ID}}$ are formed by adding the z spectrum of each laboratory frame bin yield $N_j^{\text{likelihood-selected ID}}(p_{\text{lab}}, cos\theta_{\text{lab}}; z_i)$, weighted with the corresponding inverse matrix element $\mathcal{P}_{ij}^{-1}$. In total, all possible yields $N_j^{\text{likelihood-selected ID}}(p_{\text{lab}}, cos\theta_{\text{lab}}; z_i)$ are represented by 7500 histograms binned in z.

Uncertainties are propagated through the correction process according to Equations 3.5 and 3.6. No a-priori systematic uncertainties are assumed to exist for the raw measurement sample, $\sigma^{2,\text{ syst.}}_{N_n^{\text{uncorr.}}} = 0$. Following the asymmetric extraction of uncertainties for all inverse matrix elements described in the previous Section, positive and negative systematic uncertainties on the corrected yields are calculated separately.

Discussion of Correction Results

After the dual extrapolation and independent PID matrix inversions, two sets of corrected samples are obtained for each track charge. As a reminder from Section 4.1.5, two sets of probabilities are created to give additional constraints for the probability extrapolation at high track momentum by MCHadronBJ

probabilities, since only few probabilities from experimental data analyses are available there. For the first set of matrices, missing probabilities are extrapolated considering experimental data as well as MCHadronBJ data PID information from neighboring kinematical bins. The second set of matrices contains the same probabilities, except for at very high and very low momenta where MCHadronBJ information is directly used to replace missing probabilities.

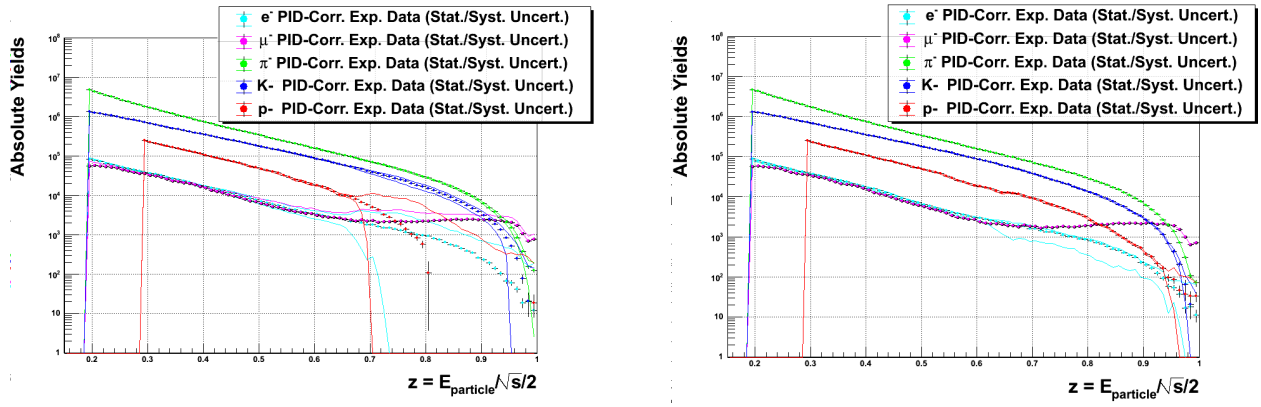


Figure 4.24: Experimental data measurement sample after PID correction with the first set of extrapolated probabilities (a) and second set of probabilities (b), including statistical and systematic uncertainties, for negative track charge.

Figures 4.24 a) and b) show yields from the measurement sample described in Section 3 corrected for PID misidentification with the first and second set of extrapolated probabilities, respectively, for negative track charge. The smaller systematic uncertainties for the correction with the second set of probabilities can be related to the fact that the extrapolation algorithm assigns large probability uncertainties, while MCHadronBJ probability uncertainties are fairly small.

For the first set of extrapolated probabilities, proton yields turn slightly negative above $z \sim 0.8$. However, systematic uncertainties are consistent with positive yields throughout. In an analysis of the contributions producing negative proton yields, inverse elements corresponding to probabilities $p_{K \rightarrow p}$ have been found to show largest negative contributions. Examining the extrapolation results of these probabilities, cf. Figure 4.25, it can be seen that fake-rates $p_{K \rightarrow p}$ from experimental data are systematically larger than MCHadronBJ fake-rates. The extrapolation algorithm reasonably extends this trend to higher momenta where only few probabilities from experimental data are extracted. However, the extrapolation algorithm is only weakly constrained in those kinematical regions, so the fake-rates $p_{K \rightarrow p}$ might be overestimated. Since only pion and kaon multiplicities are extracted in this analysis and no anomalies were observed in these spectra, the proton yields were not further investigated.

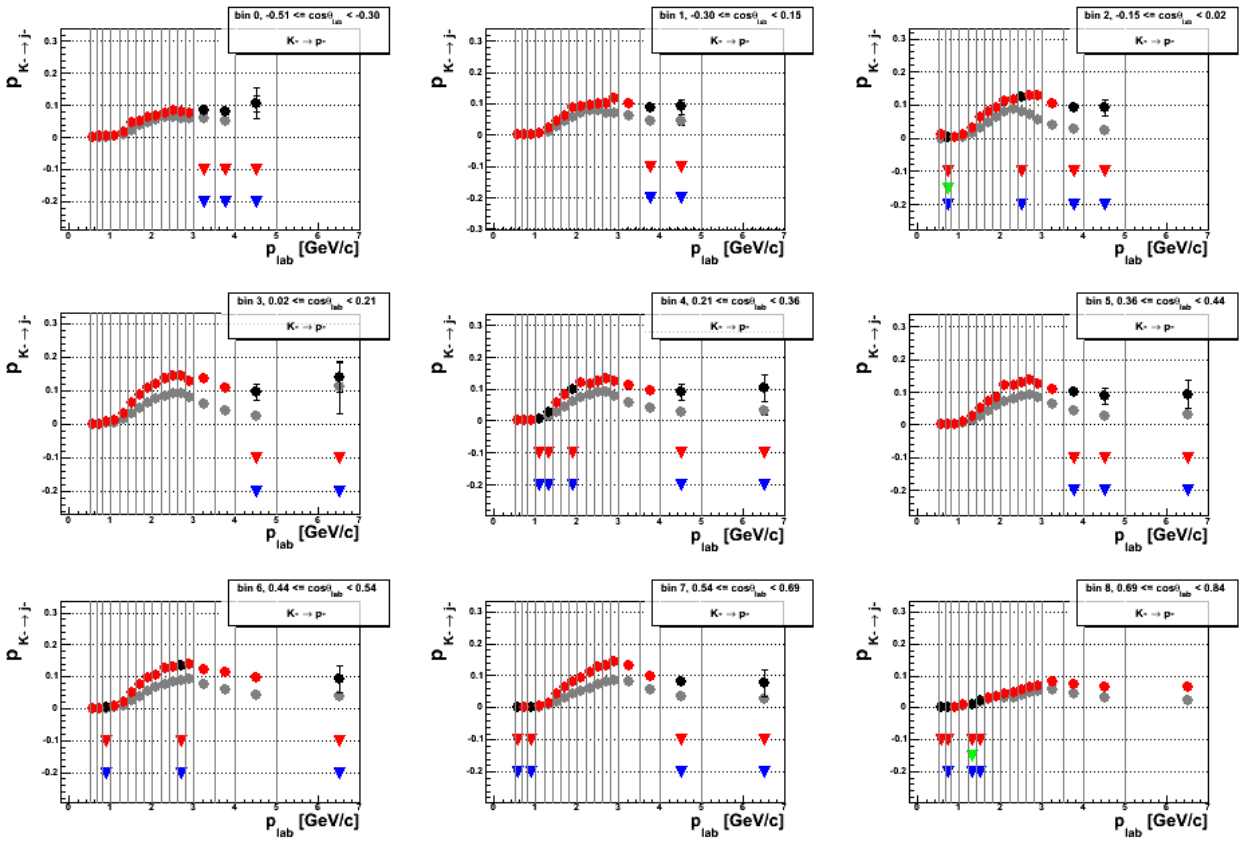
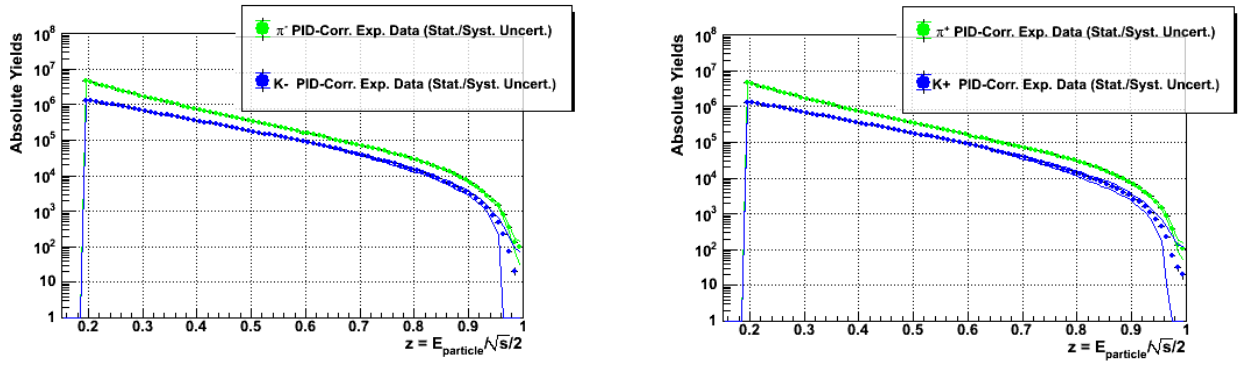


Figure 4.25: Probabilities $p_{K^- \to p}$ extracted from experimental data analysis (red circles), from MCHadronBJ analysis (gray circles) and extrapolated probabilities entering the first dual set of probabilities (black circles). All plot specifications given in Section 4.1.5 hold.

Final PID-corrected Spectra from Average of Dual Corrections

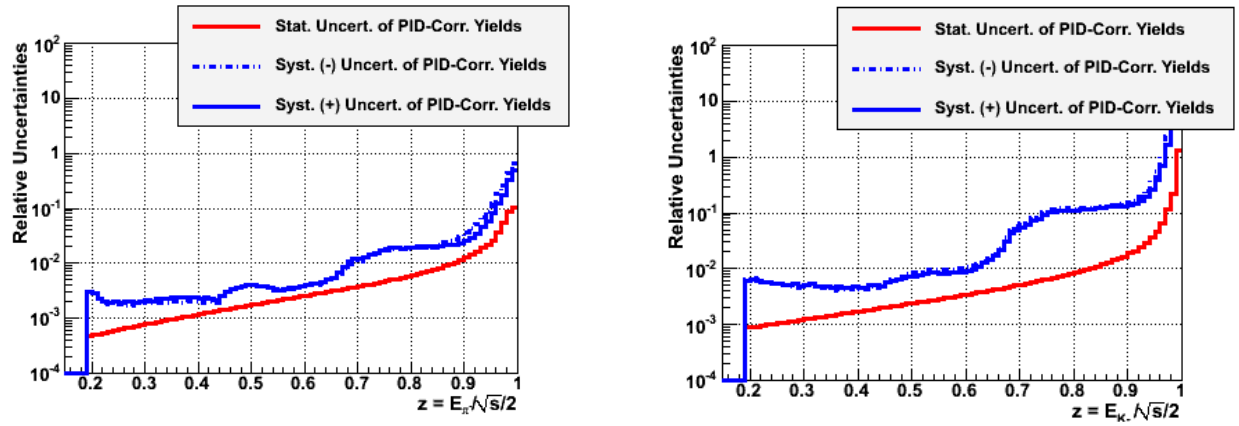
As outlined in Section 4.1.5, final PID-corrected spectra are obtained by averaging the two PID-corrected spectra resulting from the dual probability extrapolation methods. Statistical and systematic uncertainties of both corrected spectra are propagated through the average expression. Half of the difference between the correction results is added as an additional systematic uncertainty in quadrature. Figures 4.26 a) and b) show final PID-corrected pion and kaon spectra for negative and positive track charge, respectively.

Figure 4.27 shows relative statistical and systematic uncertainties on the final PID-corrected pion and kaon yields for negative track charge. As expected, systematic uncertainties are limiting the precision of the measurement. The dominant contribution can be identified to come from inverse matrix element uncertainties. The shape of the curves for relative systematic uncertainties is a result of laboratory frame bin contributions with different uncertainties and varying differences between the two sets of corrections after dual probability extrapolation.



(a) (b)

Figure 4.26: Final PID-corrected spectra after averaging between the two correction results from dual extrapolation with statistical and systematic uncertainties represented by error bars and uncertainty bands, respectively. Figure a) shows spectra for negative track charge, Figure b) contains spectra for positive track charge.



(a) (b)

Figure 4.27: Relative uncertainties per bin for final PID-corrected pion (b) and kaon (b) spectra after averaging between the two correction results from dual extrapolation, for negative track charge.

Discussion of the PID-Correction Significance In Appendix B.7, Figure B.24 are given displaying the ratio between PID-corrected and raw experimental yields for negatively charged pions and kaons. As additional illustration, Figure B.25 in the Appendix shows ratios of all species relative to the kaon yield versus z after PID-correction.

Pion yields are seen to change by about 10% due to the PID correction. This can be motivated by considering the pion-hadron and pion-lepton PID fake-rates extracted in previous Sections. Since significant fake-rates indicate that true/physical pions will contribute to likelihood-selected spectra of other hadron or leptons, pions will 'gain' yield contributions from those spectra in the PID correction. On the other hand, likelihood-selected spectra of pions likewise also contain contributions from other true/physical species caused

by non-zero hadron-pion and lepton-pion fake-rates. Correcting for PID misidentification will remove those contributions from the pion sample.

Figure B.24 a) in the Appendix indicates that the 'gains' for the pion sample are dominant for all values of z. This can be motivated by the relative abundances of pions and all other analyzed species, illustrated by Figure 3.1. Since pions are by far the most abundant particles over all ranges of z, cf. Figure B.25, yield 'losses' caused by other particles being likelihood-selected as pions will not cause large changes in the pion spectra. Rather, only pions misidentified as other particles will be the most significant source of PID-correction-induced yield changes for pions. Following this reasoning, the largest pion-hadron fake-rate can be assumed to be the main contribution in Figure B.24 a). In fact, the laboratory frame momentum dependence of the pion-kaon fake-rate, cf. Figure 4.18 c), is closely similar to the z-dependence of the pion ratio shown in Figure B.24 a).

Analogously to the previous paragraph, the kaon ratio plot in Figure B.24 b) can be analyzed. For low z, PID-corrected kaon spectra contain up to 20% less tracks than the raw kaon spectra. As can be seen again from Figure 4.18 c), the pion-kaon fake-rate of about 10% at low z creates a contribution in the likelihood-selected kaon spectra. Since the relative abundance of pions and kaons at low z reaches a rough factor of 2, cf. Figure B.25, recovering a 10% pion loss from the kaon yields equals a reduction of about 20% of the latter. For higher z, pion-kaon fake-rate are still of the order of $5 - 10\%$. An additional contribution reducing kaon yields for high z arises from high proton-kaon fake-rates (cf. Figure 4.18 e)). However, as can be seen from Figure B.25, pions are less than twice as abundant as kaons and kaons are four times more abundant than protons for $z >\sim 0.5$. Thus the impact of both of those kaon yield-reducing fake-rates is weakened. On the other hand, total kaon-hadron fake-rates rise up to about 30% at high z, which causes significant gains in kaon yields in the PID correction. The gains can be assumed to by far compensate for the yield reducing compenents, thus producing a ratio in Figure B.24 b) of around 1.1 for $z >\sim 0.4$.

4.2 Correction for Sample Impurity- Non-$q\bar{q}$ Events in HadronBJ

As indicated in Section 4.1.5, generic Belle Monte Carlo samples only contain reactions $e^+e^- \rightarrow q\bar{q}$. A better match with experimental yields versus fractional energy z can be achieved by including also other processes, e.g. QED (especially $\tau\tau$) or 2γ events, in the Monte Carlo analyses. All processes considered are grouped together in the MCHadronBJ sample. Since this analysis aims at measuring charged pion and kaon multiplicities only from events $e^+e^- \rightarrow q\bar{q}$ where $q = \{u, d, s, c\}$, pions and kaons being created in other processes have to be removed from the measurement sample.

4.2.1 Extraction of Correction Factors

A common procedure in cross section measurements is to remove unwanted contributions from each measured yield by subtraction. In this analysis, the size of contributions from non-$q\bar{q}$ events is extracted from QED and 2γ MC samples contained in the MCHadronBJ dataset introduced in Section 4.1.5. On-resonance and continuum energy level datasets are analyzed to minimize statistical uncertainties from those samples. The differences in hadron yields from the QED and 2γ samples due to the different center-of-mass energy between on-resonance and continuum is considered negligible. All event and track cuts are applied as described in Section 3.1.

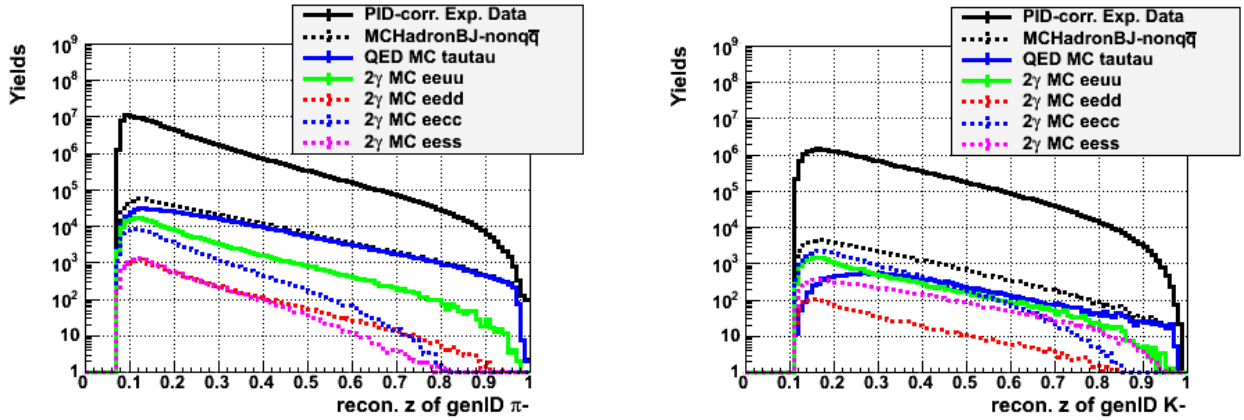
Pion and kaon yields versus reconstructed z but for true generator ID are extracted from a total of 100.5 fb^{-1} of QED and 2γ MCHadronBJ data. To match the luminosity of the experimental measurement sample of 68.0 fb^{-1} given in Ref. [57], all bin contents of non-$q\bar{q}$ MC yields are scaled down by a factor of 1.48. Since the cross sections for all components of the MCHadronBJ datasets describing QED and 2γ events are well known, it is assumed that

$$N_{i;z}^{\text{Exp. Data, } QED,2\gamma} = N_{i;z}^{\text{MC, } QED,2\gamma}. \tag{4.22}$$

Under this assumption, a direct subtraction of hadron yield contributions from non-$q\bar{q}$ events can be performed.

Figure 4.28 shows PID-corrected experimental yields for negatively charged pions (a) and kaons (b) (black solid lines), together with the luminosity-matched sum of all MCHadronBJ QED and 2γ yields (black dashed lines). Additionally, the luminosity-matched yields of all contributing single QED and 2γ processes are given as well.

It can be seen in Figure 4.28 that for pions yields, the most significant non-$q\bar{q}$ contributions arise from QED $\tau\tau$ processes. Also noticeable is a steep fall-off of τ event contributions at maximum z, most likely



(a) (b)

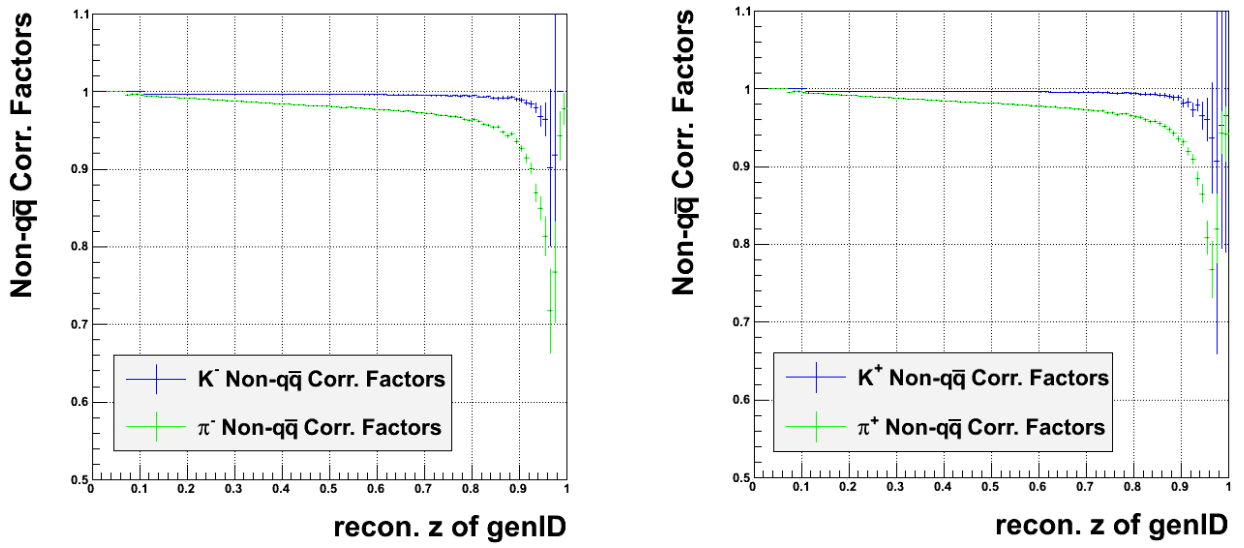
Figure 4.28: PID-corrected experimental yields for negatively charged pions (a) and kaons (b), given by the black solid lines. Additionally shown are luminosity-matched yields from MCHadronBJ QED and 2γ processes for the respective species. The sums of all MCHadronBJ QED and 2γ yields are indicated by the dashed black lines.

related to phase space limits in τ decays. For kaons, τ events contribute most at large z, whereas at lower z contributions from 2γ events are most abundant.

Figure 4.29 shows correction factors $1.0 - n_i^{non-q\bar{q}}$ versus z, where $n_i^{non-q\bar{q}}$ are fractions of non-$q\bar{q}$ yields over PID-corrected experimental yields for species i. These ratios steadily decrease from 100% to a minimum of about 70% for pions and 90% for kaons at $z \sim 0.96$. At higher z, the fractions recover again due to phase space limitations for high z hadron production in non-$q\bar{q}$ events.

Contributions from $b\bar{b}$ resonance decays The measurement sample was recorded with the Belle detector at a collider center-of-mass energy of 10.52 GeV, below the $\Upsilon(4S)$ threshold of 10.58 GeV/c^2 and above the lower $\Upsilon(1S, 2S, 3S)$ resonances at 9.46, 10.02 and 10.36 GeV/c^2, respectively. However, if the electron or positron in the initial state emits a hard photon, the lower Υ resonances can be produced. Since the Υ particles are $b\bar{b}$ states, pions and kaons from such events represent unwanted background to our measurement of multiplicities from events $e^+e^- \to q\bar{q}$ with $q = \{u, d, s, c\}$.

According to Refs. [83] and [84], the production cross sections for $\Upsilon(1S, 2S, 3S)$ in ISR-events at $\sqrt{s} =$ 10.52 GeV are equal to 20.82, 19.20 and 40.79 pb, respectively. Compared to the hadronic cross section of 3.3 nb, the ISR-Υ cross sections are smaller than 2% in relative size. Reference [5] suggests that most $\Upsilon(2S, 3S)$ decay down to $\Upsilon(1S)$. Some of the channels for those decays involve pions. However, the mass differences between the resonances are only a few 100 MeV/c^2, so the pions carry most likely less than $z = 0.2$ of energy and thus do not enter the mesurement sample. Other decay channels of the $\Upsilon(2S, 3S)$ involve decays into τ pairs which can produce high energy light hadrons, but these channels only have branching



(a) (b)

Figure 4.29: Fractions of pions and kaons from luminosity-matched MC non-$q\bar{q}$ events relative to all pions and kaons in the PID-corrected measurement sample versus z, for negative (a) and positive (b) particle charge.

fractions on the order of percent. Reference [5] further suggests that $\Upsilon(1S)$ decay fractions into light hadrons are small as well.

In addition, a sizable fraction of $\Upsilon(1S, 2S, 3S)$ events is expected to fail the applied event selection cuts in the HadronBJ selection criteria and the additional requirement for the visible energy to be larger than 7 GeV. Therefore the fraction of light hadrons from ISR-produced Υ resonances should be even smaller. In summary, the fraction of pions and kaons from ISR-produced Υ decays is assumed to be negligible for all z.

4.2.2 Correction of Measurement Samples

After the PID-correction described in previous Sections, the measurement samples are corrected for non-$q\bar{q}$ contributions in the HadronBJ skim with the correction factors shown in Figure 4.29 in each z bin. All uncertainties are propagated according to Equations 3.2 and 3.3. Since all MC QED and 2γ events are generated according to their well known cross sections, no additional systematic uncertainty was assigned to the correction factors, $\sigma_\epsilon^{\text{syst.}} = 0$.

Figure 4.30 shows pion and kaon yields for negative particle charge after correcting for non-$q\bar{q}$ contributions in the HadronBJ skim with statistical and systematic uncertainties. Relative systematic uncertainties only rise by about 1% at highest z due to this correction.

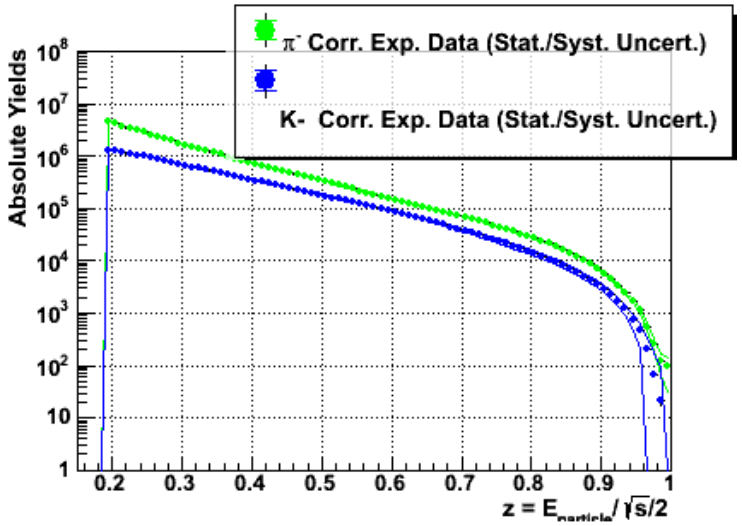


Figure 4.30: Experimental pion and kaon yields corrected for non-$q\bar{q}$ events in HadronBJ for negative particle charge. Statistical as well as systematic uncertainties are shown.

4.3 Correction for Kinematic Reconstruction/ Smearing Effects

Another correction is applied to account for finite precision of detector reconstruction of particle kinematics. The correction transforms yields in bins of reconstructed z into yields in bins of true/physical z.

4.3.1 Correction Formalism

The correction can be written down analogously to the PID correction, cf. Equations 4.1 and 4.21, by

$$N_m^{\text{recon. } z} = \mathcal{S}_{mn}\, N_n^{\text{real } z}, \tag{4.23}$$

and

$$N_m^{\text{real } z} = \mathcal{S}_{mn}^{-1}\, N_n^{\text{recon. } z}, \tag{4.24}$$

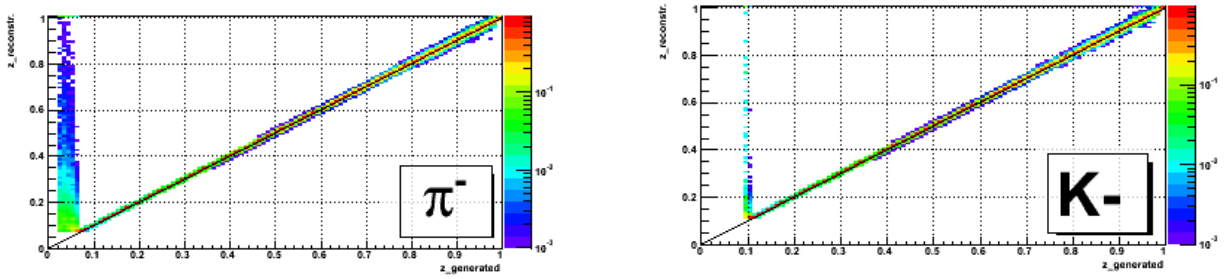
where $\mathcal{S}_{mn}$ is the correction/smearing matrix connecting the two yield distributions $N_m^{\text{recon. } z}$ and $N_n^{\text{real } z}$ and with m and n indicating z bin indices. A matrix-based correction is necessary since smearing effects can correlate any two bins m and n.

4.3.2 Extraction of Correction/Smearing Matrices

At Belle, a GEANT [56]-based detector simulation is used to extract correlations between generated z and reconstructed z for a given particle species. The analyzed data sample consists of 5 streams of on-

resonance and continuum standard Belle Monte Carlo data from events $e^+e^- \rightarrow q\bar{q}$ where $q = \{u, d, s, c\}$ from Experiment 7 through 27, corresponding to about $2 * 10^9$ events. Again, all track and event cuts as described in Section 3 are applied to all tracks and events.

For all reconstructed tracks, MC generator information is available and provides values for the generated z. Pairs of generated and reconstructed values of z are stored for each particle in a two dimensional 100×100 bin histogram. Each entry of the histogram can be normalized by the respective sum of all reconstructed particles for a given generated z bin. This procedure yields a matrix $[\mathcal{S}]_{mn}$ with elements describing the probability that a particle reconstructed in a z bin m was generated in a z bin n. Figures 4.31 show smearing histograms normalized to respective column sums for negatively charged pions (a) and kaons (b).



(a) (b)

Figure 4.31: Smearing histograms for negatively charged pions (a) and kaons (b). Each histogram bin is normalized to the sum of entries for its entire column. For illustration, a black line indicates the diagonal of the histogram.

The low z values of the smearing histograms Figures 4.31 reflect the acceptance cut of $p_{\text{lab}} \geq 0.5$ GeV. Since for the matrix inversion square matrices are needed and pion and kaon yields are only measured above $z = 0.2$, cut-off values have been introduced for low z. It was decided to also include fiducial areas in z to minimize boundary effects in the smearing matrices used for data correction. Pion smearing matrices are set to start at $z = 0.08$ and kaon matrices at $z = 0.12$.

Uncertainties on the smearing matrix elements are calculated from the normalization to the column sum via assuming binomial uncertainties for the histogram bin content and Poisson uncertainties for the column sum, in analogy to PID probability uncertainties. All smearing histogram bins with no entries are given the value 0.0 in the smearing matrix probabilities = 0.0. The uncertainty on those probabilities is given by the inverse of the respective column-sum, providing an upper limit on the maximum value of a non-zero matrix element.

For the inversion and uncertainty propagation of smearing correction matrices, the same formalism is used as for the inversion of PID matrices described in Section 4.1.6. As a reminder, uncertainties on elements

of the uninverted matrix are propagated through the inversion by creating a sample of matrices where all original elements are varied randomly according to their uncertainties. Inverting all matrices in the sample yields Single Element Distributions (SEDs) for each inverse matrix element. Uncertainties on inverse matrix elements can be extracted from SEDs by calculating standard deviations with respect to the analytical inverse element value. As a 'unitarity' requirement, randomly varied matrices with column-sums deviating from 1.0 further than the column sum uncertainty of the original uninverted matrix are rejected. The elements in the extracted smearing matrices corresponding to the highest z bins bear very large uncertainties. For the last bins, the requirement of consistency with 1.0 for the matrix column sum cannot be satisfied anymore for most randomly varied matrices. Thus the smearing matrices had to be limited to only contain information up to $z = 0.98$, which also sets a limit to the accessible z range of this measurement.

4.3.3 Matrix Inversion and Uncertainty Propagation

Inverse Matrix Elements

As mentioned above, the same technical approach to inverting smearing matrices and propagating smearing probability uncertainties has been chosen as for the PID matrix inversions described in Section 4.1.6. Values for inverse smearing matrix elements are calculated from a one-time inversion of the 91×91 pion and 87×87 kaon smearing matrices, for both charges. Those inversions were successful.

Propagation of Uncertainties

In the context of the reasoning given in Section 4.1.6, the necessity of a statistical approach to extract representative uncertainties on inverse smearing matrix elements shall be evaluated. Table 4.9 shows determinants, determinant uncertainties and the ratio of these two values for all four smearing matrices in this correction.

Species	$\det(\mathcal{S})$	$\sigma_{\det(\mathcal{S})}$	$\det(\mathcal{S})/\sigma_{\det(\mathcal{S})}$
π^-	$8.80 * 10^{-13}$	$1.09 * 10^{-13}$	8.05
π^+	$8.44 * 10^{-13}$	$1.04 * 10^{-13}$	8.14
K^-	$2.27 * 10^{-12}$	$3.74 * 10^{-13}$	6.08
K^+	$1.41 * 10^{-12}$	$3.00 * 10^{-13}$	4.71

Table 4.9: Properties of smearing matrices used in the present analysis.

The small values of determinants can be motivated from the size of the smearing matrices, e.g. considering the 91×91 matrix for pions. At high z, the diagonal elements become significantly smaller than 1.0. Assuming diagonal values of 0.9 for the first 70 columns and values of 0.5 for the following 20 columns yields a value

for the dominant determinant contribution of about $6 * 10^{-10}$. This value is similar to the values given in Table 4.9.

Considering the values in Table 4.9, low determinant significance is given for all four smearing matrices. Thus the same statistical approach as employed in the inversion of PID matrices will be used to propagate matrix element uncertainties through the smearing matrix inversion.

Analogously to the PID probability uncertainty propagation, a sample of 10^6 randomly varied matrices is generated and inverted for each smearing matrix. Figure 4.32 shows all 91 Single Element Distributions (SEDs) for row 50 of the inverted π^- smearing matrix. Visible are again skews in the SEDs caused by the positivity and unitarity requirement on the randomly varied matrices. All specifications given for Figure 4.21 hold accordingly. Inverse matrix element uncertainties are extracted from SED standard deviations.

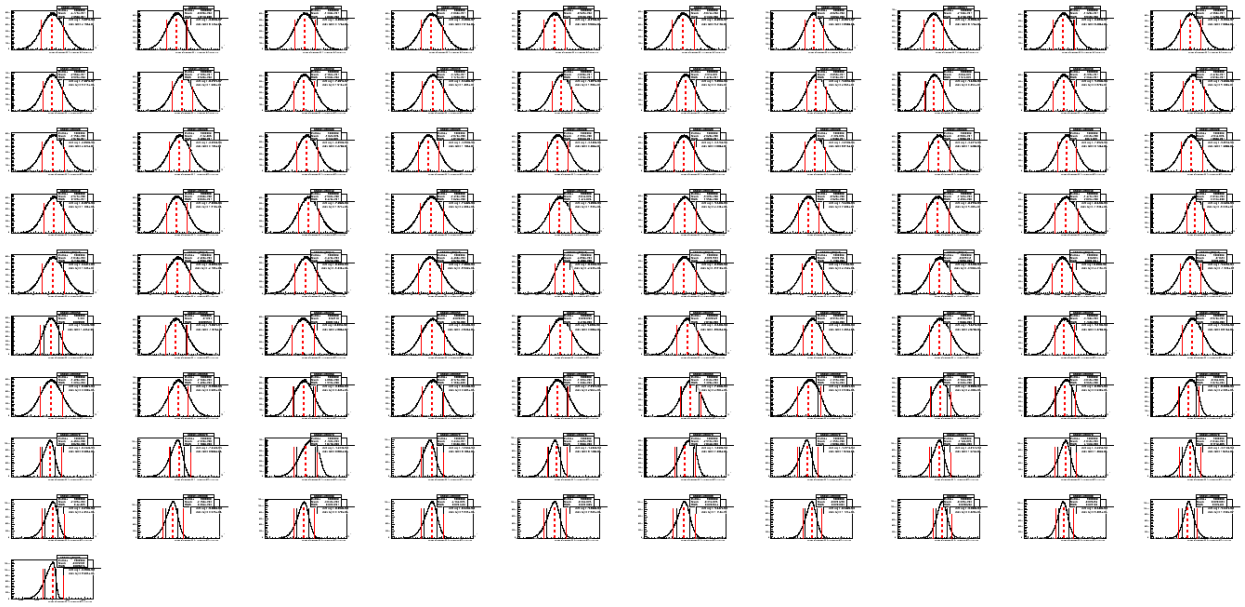
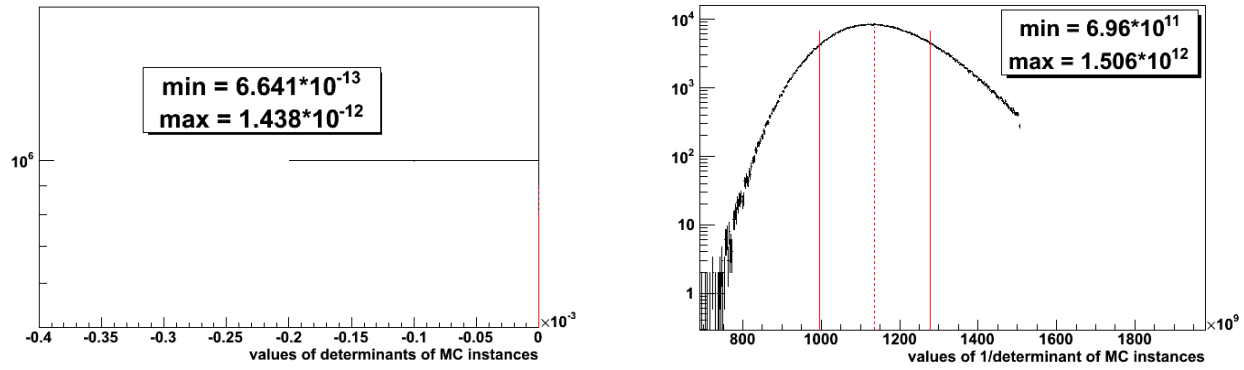


Figure 4.32: All 91 Single Element Distributions (SEDs) of row 50 from the 91×91 inverse π^- smearing matrix. For details on the plot cf. Figure 4.21.

Large fluctuation in the SEDs, which then could dominate extracted SED standard deviations, are again avoided by excluding the one per cent of all randomly varied matrices with determinants closest to 0. Figures 4.33 show the distribution of determinants (a) and reciprocal determinants (b) of all randomly varied matrices for the π^- smearing matrix.

4.3.4 Correction of Measurement Samples

The measurement samples for charged pion and kaon yields has been corrected for kinematic smearing with the inverted matrices and matrix element uncertainties calculated in the previous Section. All uncertainties



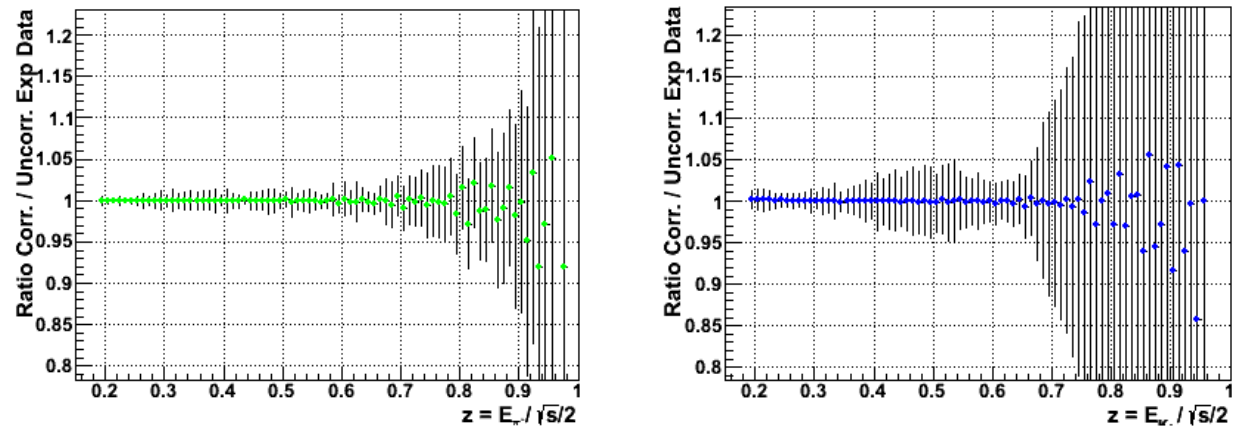
(a) (b)

Figure 4.33: Determinant and inverse determinant distributions of randomly varied matrices for the π^- smearing matrix.

are propagated according to Equations 3.5 and 3.6.

Smearing which causes particles to be reconstructed with $z > 1.0$ has been studied. For most z bins, bin contents are large enough such that the absolute value of all particles smeared to $z > 1.0$ is insignificant. At very high z, systematic uncertainties are found to be larger than the total yield of all particles smeared to $z > 1.0$. Thus no additional uncertainty arising from smearing to $z > 1.0$ is assigned.

Figure 4.34 shows ratios of negatively charged pion (a) and kaon (b) yields versus z, after and before the smearing correction. Statistical and systematic uncertainties for each yield are added in quadrature and propagated through the ratio. As can be seen from the plot, the ratios are consistent with unity for all values z within their uncertainties.



(a) (b)

Figure 4.34: Ratios of corrected over uncorrected negatively charged pion and kaon yields. Before the ratio, statistical and systematic uncertainties are added in quadrature for each plot. Gaps in the plots are present since some ratio values fluctuate outside of the histogram y axis range shown, but which are still consistent with unity within their uncertainties (uncertainty bars are not shown for those points in the plots).

Systematic uncertainties increase significantly due to the smearing correction, especially at high z. This can be attributed to low statistics in the highest z bins of the MC sample from which the smearing histograms are extracted. This causes high uncertainties on the matrix elements, which in turn produces SEDs with large standard deviations.

4.3.5 Consistency and Systematic Studies for Smearing Correction

The impact of the smearing correction on measured yields will be investigated. Furthermore, differences in the momentum resolutions in the MC simulation of the Belle detector and experimental cosmics data are analyzed with respect to impact on measured yields.

Consistency Study of Smearing Correction Impact on Measured Yields

Symmetrical smearing is characterized by generated particles being reconstructed with similar probabilities at higher as well as lower values of z. Even if symmetrical smearing applies, particle bin yields are expected to shift significantly to higher z because of the rapid fall-off of yields versus z. However, Figures 4.34 show that the change in bin yields caused by the smearing correction is consistent with unity within uncertainties. Therefore a consistency study is performed to verify this result.

Smearing Symmetry and Integrated Smearing Probabilities First, the symmetry and integrated probabilities of smearing to higher and lower reconstructed z is evaluated. Figure 4.35 shows integrated smearing probabilities for smearing above and below the diagonal of the smearing histograms in Figures 4.31, versus generated z.

It can be seen that the smearing for negatively charged pions and kaons is almost symmetrical for most z bins, with only a slight preference for smearing to higher reconstructed z. For positive particle charge, smearing to higher reconstructed z is more likely than smearing to lower z by up to 3% for both pions and kaons. This reconstruction charge bias is known at Belle. A general observation is that the total probability of a particle to not be reconstructed in its generated bin is $< 20\%$ for $z < 0.5$ and $< 50\%$ at maximum z.

Smearing Width Next, the typical width of smearing in units of bins from the diagonal generated z is evaluated. Figures 4.36 shows smearing probabilities versus reconstructed z for several fixed generated z, corresponding to vertical slices of the smearing histograms in Figures 4.31. The slices are shown relative to the diagonal bin in positive and negative bin distance versus reconstructed z.

Figures 4.36 show that the dominant reconstruction smearing outside of the corresponding generated z bin occurs into the immediately adjacent bins on both sides. Even for maximum z, the reconstruction

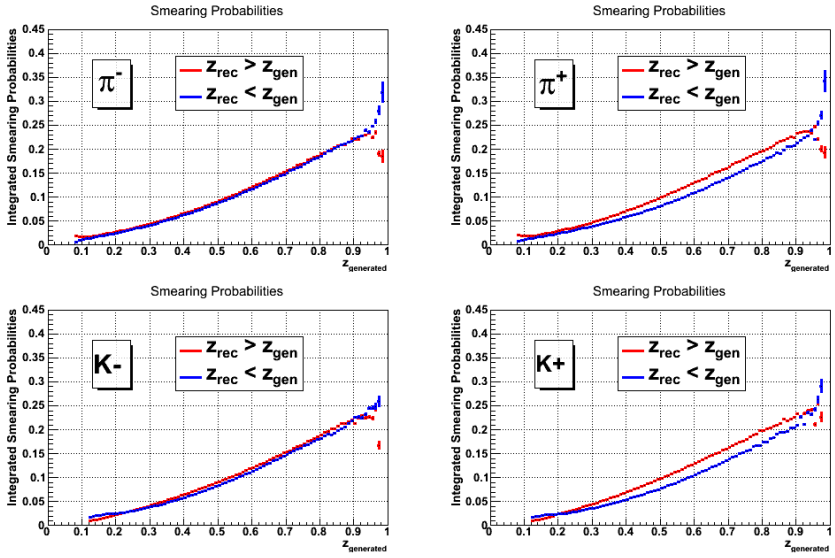


Figure 4.35: Integrated smearing probabilities for smearing above (red) and below (blue) the diagonal generated z against generated z, integrated from the smearing histograms given in Figures 4.31.

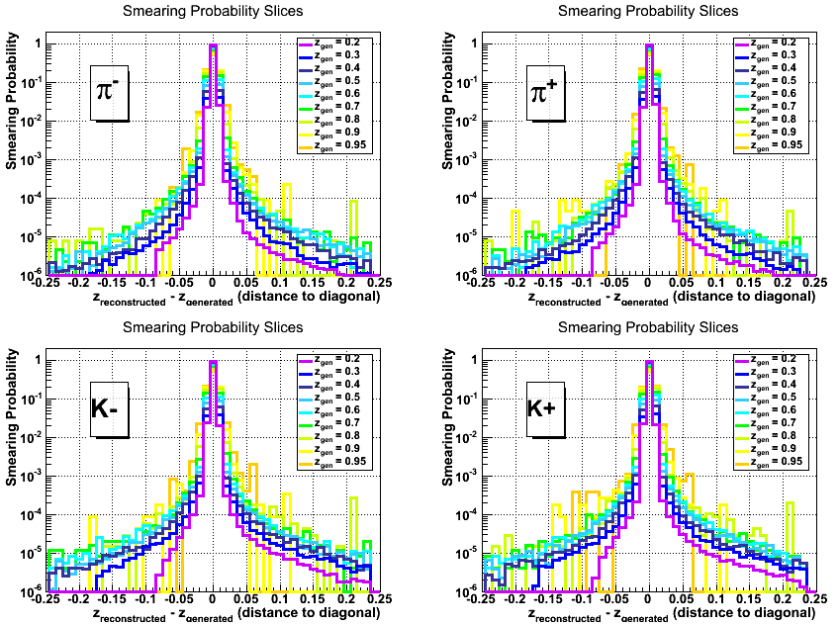


Figure 4.36: Smearing probabilities versus difference of $z_{\text{reconstructed}} - z_{\text{generated}}$ for given $z_{\text{generated}}$. All values are extracted from the smearing histograms in Figures 4.31.

smearing of particles into bins further away from the diagonal generated z than the immediately adjacent bins only occurs at the percent level.

Quantification of the Fall-off of Yields versus z Even if smearing is symmetrical to lower and higher reconstructed z, it still can significantly change bin yields if large bin-to-bin differences due to rapid fall-

off are present. To quantify if this situation applies to the presented measurement, Figure 4.37 shows the fraction of each particle yield bin content relative to its adjacent bin towards lower z.

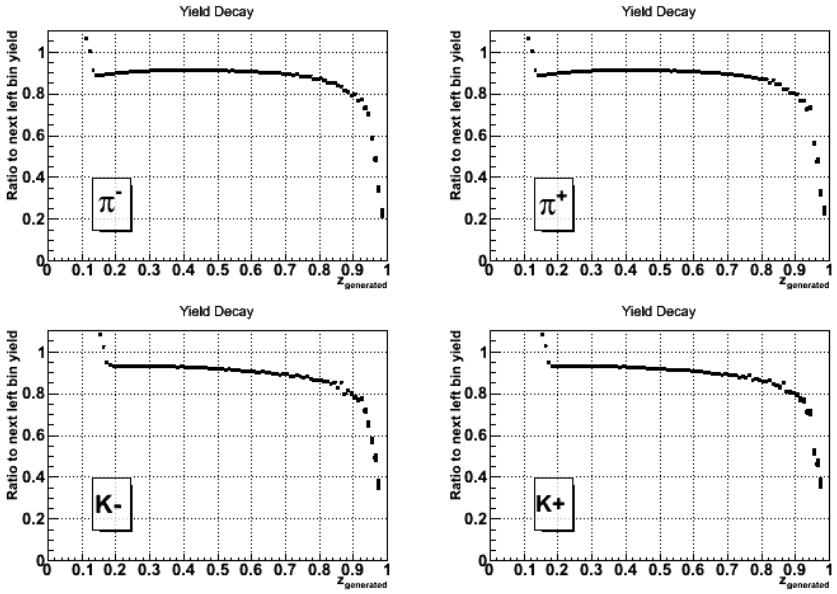


Figure 4.37: Ratios of each bin yield relative to its next neighbor towards lower z. All values are extracted from the generated yields used to fill the smearing matrices in Figures 4.31.

Most bins only have a moderate decay of less than 10% with respect to the next lower bin in z up to $z = 0.7$. This can be attributed to the narrow bin width in z, 0.01, which does not allow the yields to change rapidly from one bin to the next.

Quantification of Impact of Smearing Correction on Measured Yields Combining all information given in the above Sections, the small impact of the smearing correction on measured yields can be cross-checked. As an example, the smearing in the generated z bin $[0.5; 0.51)$ for negatively charged pions shall be discussed.

For simpler and more accessible description, the generated bin yield of z bin $[0.49; 0.50)$ shall be labeled with N_0, the generated bin yield for bin $[0.5; 0.51)$ accordingly by N_1 and the yield for bin $[0.51; 0.52)$ by N_2. Utilizing the fall-off rates from Figures 4.37, N_1 and N_2 can be expressed by $N_0 * 0.91$ and $N_0 * 0.91 * 0.91$, approximately. Following the reasoning with respect to Figure 4.36, it is assumed that all integrated smearing shown in Figure 4.35 only occurs into the nearest adjacent z bins. Therefore bin $[0.5; 0.51)$ will receive smearing contributions from bins $[0.49; 0.5)$ and $[0.51; 0.52)$ for smearing to higher and lower reconstructed z, respectively. These contributions can be approximately quantified with the corresponding numbers from Figure 4.35 as $N_0 * 0.09$ and $N_0 * 0.91 * 0.91 * 0.09$, respectively. On the other hand, bin $[0.5; 0.51)$ loses particles due to smearing to lower and higher reconstructed z, corresponding to $N_0 * 0.91 * 0.088$ and $N_0 * 0.91 * 0.094$,

respectively.

Summing up all changes for bin $[0.5; 0.51)$ gives the approximated bin content for reconstructed z in $[0.5; 0.51)$ as $N_0 * 0.8991$. This corresponds to a change of the bin content for generated z bin $[0.5; 0.51)$, $N_0 * 0.9$, due to smearing equal to about 1 per mille. The scale of the change of bin yields caused by smearing shown in Figure 4.34 is comparable to the one obtained from the simplified treatment in this Section. Similar results are expected for positively charged pions and kaons. Therefore the smearing correction results obtained in this Section are considered consistent.

Study on Impact of Different Momentum Resolutions in MC and Experimental Data

The relative momentum resolution of the Belle tracking was determined from experimental cosmics data and is published in Ref. [46]. The momentum resolution in the MC simulation of the Belle detector can be different to the experimental resolution. The concern is that in such a case, a correction for momentum smearing performed with the MC smearing histograms shown in Figure 4.31 over- or undercorrects the experimentally measured yields. Therefore it is necessary to compare the experimental resolution from Ref. [46] to the resolution in the generic Belle MC.

Extraction of MC Relative Kinematic Resolutions For this study, a sample of generic *uds* and *charm* Belle MC is analyzed with all event and acceptance cuts. Momentum resolutions for PID-selected species of tracks are extracted in 8 bins of $p_{t,\mathrm{lab}}$ by fitting distributions of $p_{t,\mathrm{lab}}^{\mathrm{generated}} - p_{t,\mathrm{lab}}^{\mathrm{reconstructed}}$ with a Gaussian function. The fitted width of the Gaussian is normalized to the value of the respective value of $p_{t,\mathrm{lab}}^{\mathrm{generated}}$ to obtain relative momentum resolutions. Figure 4.38 shows a sample of fitted smearing distributions $p_{t,\mathrm{lab}}^{\mathrm{generated}} - p_{t,\mathrm{lab}}^{\mathrm{reconstructed}}$ for 9 bins of $p_{t,\mathrm{lab}}^{\mathrm{generated}}$ for PID-selected negatively-charged pions. The last bin at $p_{t,\mathrm{lab}}^{\mathrm{generated}} = 5.5$ GeV/c is excluded from the analysis because of insufficient statistics.

Anologous extractions are performed for p_{cms} and z smearing distributions. These additional fits allow to relate potential differences between the MC and experimental momentum resolutions to the smearing correction presented in this analysis perfomred in z. Results of relative $p_{t,\mathrm{lab}}$, p_{cms} and z resolutions are plotted together with the published Belle experimental $p_{t,\mathrm{lab}}$ resolutions in Figure 4.39.

It can be found in Figure 4.39 that the experimental and MC $p_{t,\mathrm{lab}}$ resolutions are largely consistent at low momenta. However, the experimental resolutions display a more rapid increase versus momentum than the MC resolutions.

Smearing Correction with Matrices Generated from Experimental Resolutions To evaluate the impact of this difference on the presented smearing correction of measured yields, it was decided to create

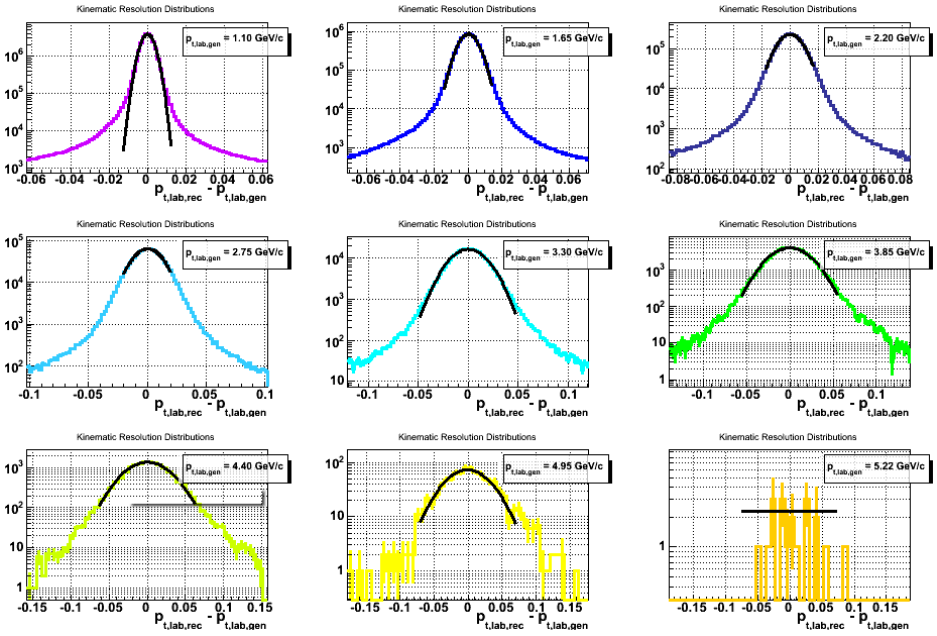


Figure 4.38: Fitted smearing distributions $p_{t,\mathrm{lab}}^{\mathrm{generated}} - p_{t,\mathrm{lab}}^{\mathrm{reconstructed}}$ for 9 bins of $p_{t,\mathrm{lab}}^{\mathrm{generated}}$ for PID-selected negatively-charged pions.

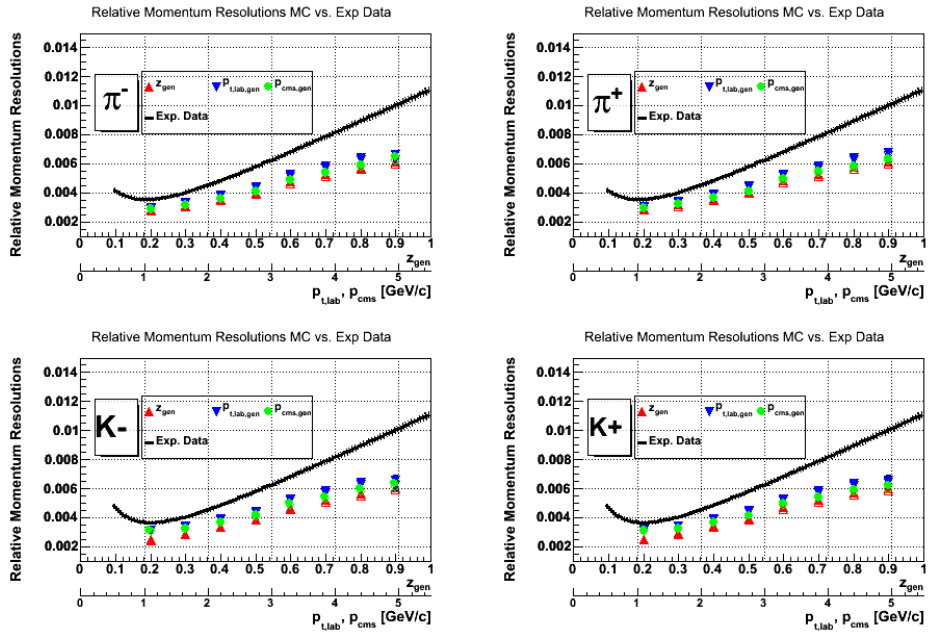


Figure 4.39: Extracted relative MC resolutions of $p_{t,\mathrm{lab}}$, p_{cms} and z plotted versus the respective kinematic variable (symbols). Also plotted are relative Belle experimental $p_{t,\mathrm{lab}}$ resolutions published in Ref. [46] versus $p_{t,\mathrm{lab}}$.

z smearing matrices based on the experimental $p_{t,\mathrm{lab}}$ resolutions. It can be seen in Figure 4.39 that the extracted relative MC resolutions for $p_{t,\mathrm{lab}}$, p_{cms} and z are fairly consistent for each momentum or z value.

In order to obtain a crude parameterization of the experimental resolution in z, the phase space of $p_{t,\mathrm{lab}}$ from 0 to 5.5 GeV/c is mapped on the z phase space of 0 to 1.0. Differences arising from the boost of $p_{t,\mathrm{lab}}$ into the *cms* frame and contributions from non-zero hadron masses are found small in the MC resolutions shown in Figure 4.39 and are negelected.

Smearing matrices with approximated experimental z resolutions are generated by assuming Gaussian smearing distributions with widths equal to the experimental momentum resolutions from Ref. [46]. Corresponding smearing histograms are filled, and smearing matrices are created by normalizing all bin contents to their respective column bin-content sum. Smearing histograms from approximated experimental resolutions are shown in Figure 4.40.

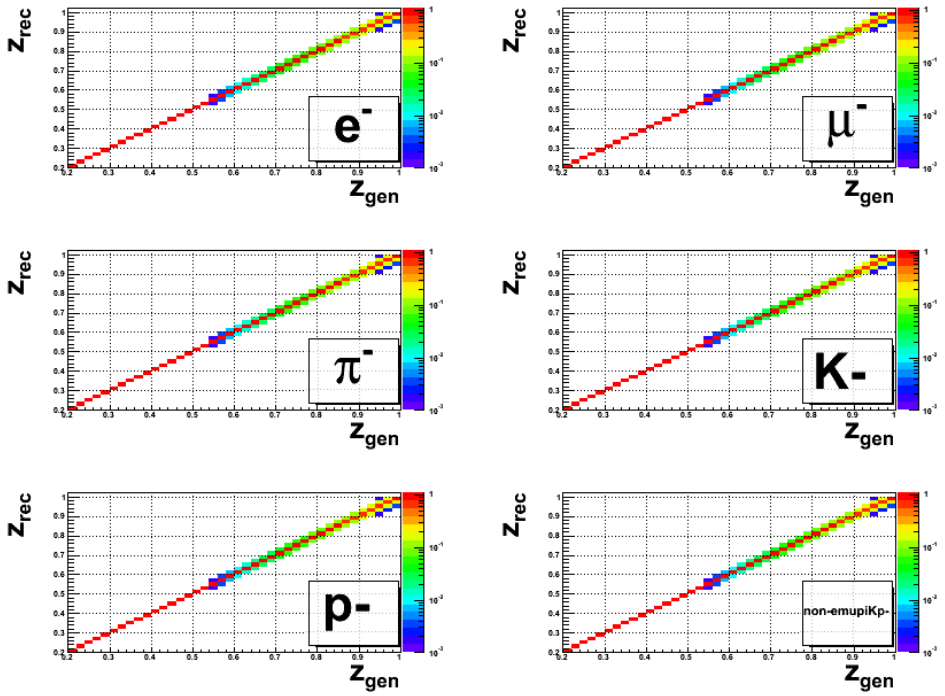


Figure 4.40: Smearing histograms in z generated from experimental $p_{t,\mathrm{lab}}$ resolutions published in Ref. [46] are shown for z bin width of 0.02.

To quantify the impact of the resolution difference between MC and experimental data, measured yields are corrected with inverted smearing matrices extracted from the histograms shown in Figure 4.40. The experimental resolutions surpass the standard z bin width of 0.01 at high z, which causes 'reflexions' in the inverse smearing matrix, characterized by negative off-diagonal values larger in absolute value than the diagonal elements. This causes the corrected yields to show a fluctuating behavior, including fluctuations below 0. Such unphysical properties of the corrected yields can be avoided by increasing the bin width beyond the experimental resolutions. To account for the large experimental resolutions, the bin width in z

is doubled to 0.02 for this consistency study. The histograms in Figure 4.40 are also shown in this binning. Reflection effects for the MC smearing matrices used for the smearing correction are found small and are neglected due to the smaller relative momentum resolutions in MC.

Figures 4.41 display corrected measured yields with MC smearing matrices (full symbols) and yields corrected with generated smearing matrices from experimental resolutions (empty symbols). Figures 4.42 show ratios of yields corrected with MC smearing matrices over yields corrected with smearing matrices from experimental resolutions. No significant deviation from 1.0 is found. However, in the context of a conservative uncertainty treatment, all absolute differences to 1.0 are assigned as additional relative systematic uncertainties.

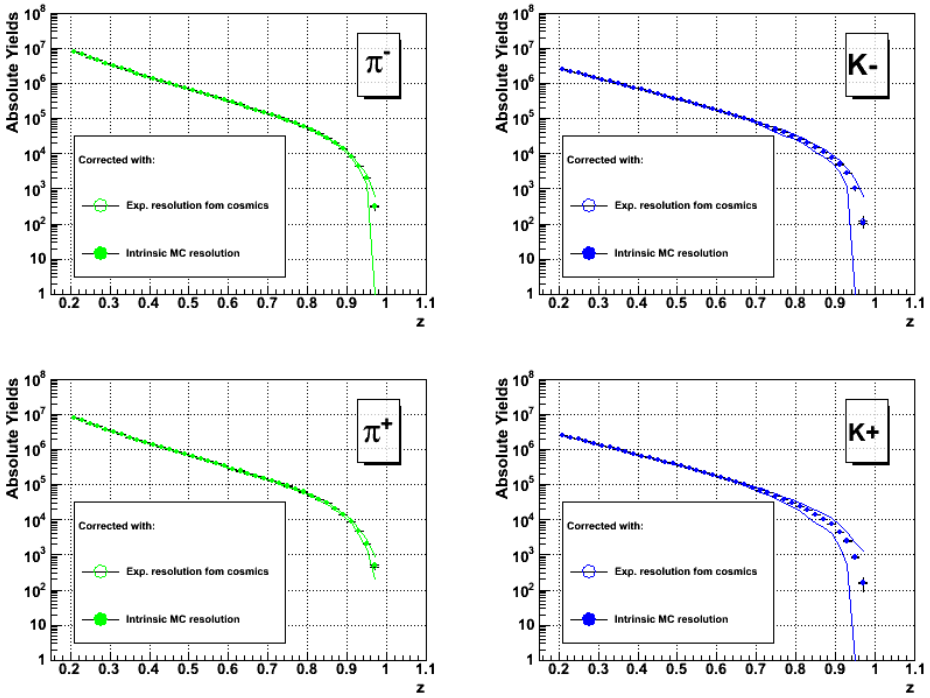


Figure 4.41: Superposition of measured yields corrected with MC smearing matrices (full symbols) and generated smearing matrices from experimental momentum resolutions (empty symbols). For technical reasons described in the text, the bin width is 0.02 in z, which is double the width used for binning the measured yields.

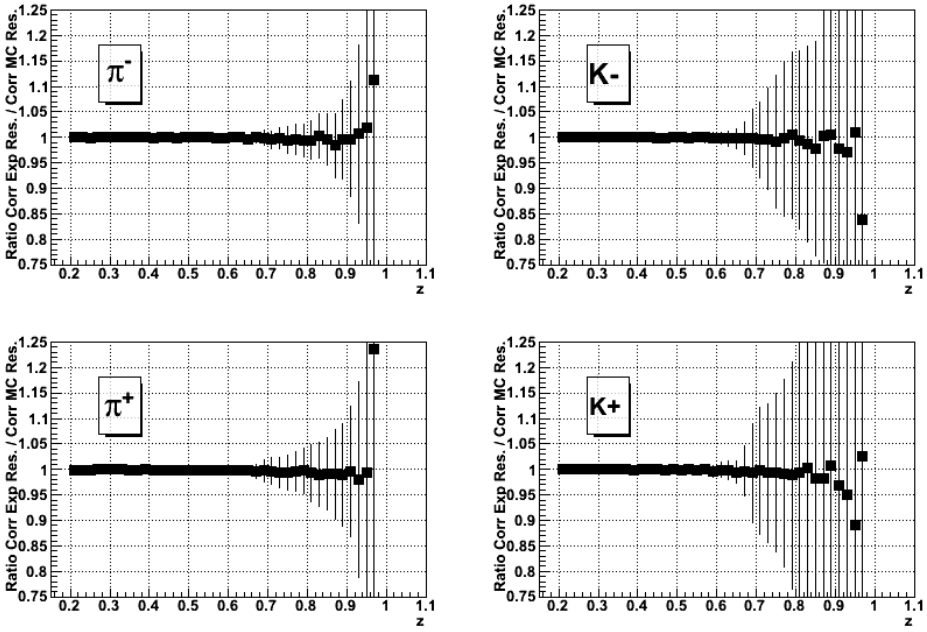


Figure 4.42: Bin-by-bin ratios of measured yields corrected with MC smearing matrices and generated smearing matrices from experimental momentum resolutions from Figure 4.41.

4.4 Joint Correction of Systematic Effects

Some charged pions and kaons produced in the fragmentation of e^+e^- annihilation events decay in flight or are produced in decay-in-flight before they can be properly tracked in the detector. These particles are not present in the experimental reconstructed data samples and have to be recovered. Additionally, fragmentation particles might undergo interaction with detector material and also do not show up in the analysis sample. Other particles in the experimental sample are created from detector interaction by fragmentation particles. These non-fragmentation particles represent undesired contributions to fragmentation pion and kaon yields and have to be removed. Finally, detector and tracking inefficiencies can cause fragmentation particles to fail the helix and SVD hit cuts described in Section 3, which prevents them from entering the analysis sample. All above effects are studied and corrected for in this Section.

4.4.1 Monte Carlo Study Sample and Correction Description

The sample on which all studies of this Section are performed is again consisting of generic Belle Monte Carlo (MC) data, involving the GEANT [56]-based Belle detector simulation. Correction factors are extracted from six streams of continuum energy level MC data with events $e^+e^- \rightarrow q\bar{q}$ where $q = \{u, d, s, c\}$ from Experiment 7 through 55, corresponding to about $1.23 * 10^9$ events. One sixth of this statistics is used to create the illustrational plots shown in this Section. Again, all track and event cuts as described in Section 3

are applied.

As mentioned in the introduction, this correction step transforms particle yields measured in experimental data with 'true' kinematics and PID into yields of particles which solely originate from fragmentation processes. Fragmentation particles which do not enter the analysis sample are recovered, and particles not originating from fragmentation processes are removed. However, it is necessary to check in more detail the validity and consistency of the modeling in the Monte Carlo data and assign additional systematic uncertainties where necessary. In this light, the causes for particle gains and losses named in the beginning of this Chapter are investigated in the following Sections.

4.4.2 Illustration and Study of Particle Losses and Gains

At first, Figure 4.43 shows efficiencies for fragmentation pions and kaons into the reconstructed analysis sample (with 'true' kinematics and PID) for negative particle charge. The plots illustrates that more than about 85% of all desired particles from fragmentation processes are present in the analysis sample for all z. As will be shown in the following Sections, the high-z dependence of these efficiencies is dominated by Silicon Vertex Detector detection efficiencies. The low-z dependence can be motivated from rates of fragmentation particles interacting with the detector.

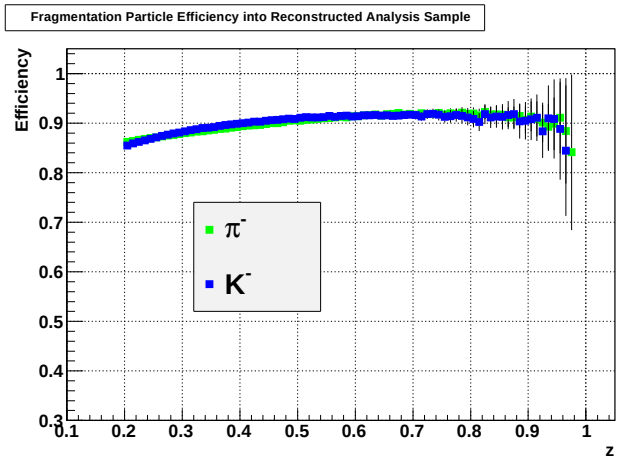


Figure 4.43: Efficiencies into the reconstructed analysis sample for negatively charged primary pions and kaons from $e^+e^- \rightarrow q\bar{q}$ fragmentation events.

Fragmentation Particles Lost to Detector Interactions

If a primary fragmentation particle undergoes interaction with detector hardware, the initial particle might not enter the reconstructed analysis sample. This is influenced by the distance at which the interaction happens, and how the particle interacts with the detector (elastically, hadronically etc). Figure 4.44 shows

numbers of negatively charged fragmentation pions and kaons not entering the analysis sample but interacting with the detector, relative to all fragmentation pions and kaons, respectively. The fraction of lost particles due to detector interaction remains mostly around 10% and only rises to higher values above z 0.9. The high-z behavior will be motivated in the discussion of the Silicon Vertex Detector efficiencies. The difference between the pion and kaon data points is expected to arise from mass and quark-content effects since it decreases for higher energies.

A slight charge dependence is observed for kaons. The rate for K^+ to interact with the detector and not be part of the analysis sample is about half of the rate for K^- at low z, but attains comparable values for $z > 0.7$. This can probably be attributed to different interaction cross sections between protons, which the detector mostly consists of, and kaons.

The Belle Monte Carlo is assumed to describe particle-detector interaction correctly, but the information transfer between GEANT and the Belle data tables is considered incomplete. The lack of event information e.g. leads to apparent energy violations in hadronic reactions with detector material. Thus different kinds of interactions cannot be resolved. Also, the fractions shown in Figure 4.44 contain particles which might not show up in the analysis sample because of detector/tracking inefficiencies. The detector/tracking efficiency is estimated in Section 4.4.2.

Among all kinds of detector interaction, hadronic interactions are assumed to provide the most important contribution. Systematic studies on the dependence of correction factors on the chosen description of hadronic effects in the Monte Carlo data are described in Section 4.4.3.

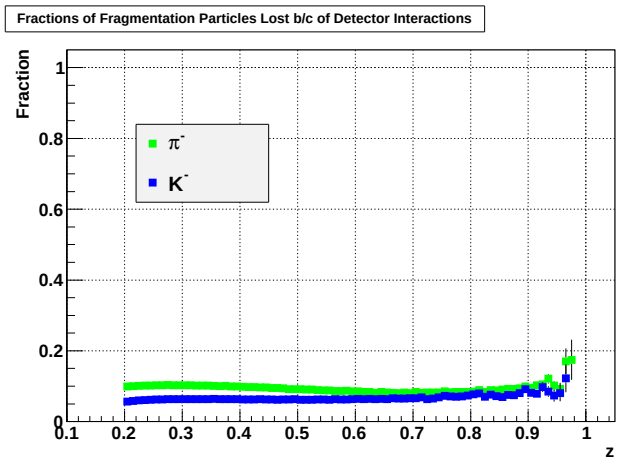


Figure 4.44: Numbers of negatively charged fragmentation pions and kaons interacting with the detector but not entering the analysis sample, relative to all fragmentation pions and kaons, respectively.

Fractions of Fragmentation Particles Lost because of Decay-in-Flight

Apart from interacting with the detector, another cause for fragmentation pions and kaons to not enter the analysis sample is decay-in-flight (DIF). All fragmentation particles undergoing DIF need to be recovered. Additionally, the DIF products of heavier leptons, baryons and mesons contain pions. This analysis considers such pions as valid fragmentation pions. Thus these pions have to be recovered as well if they are not present in the reconstructed analysis sample.

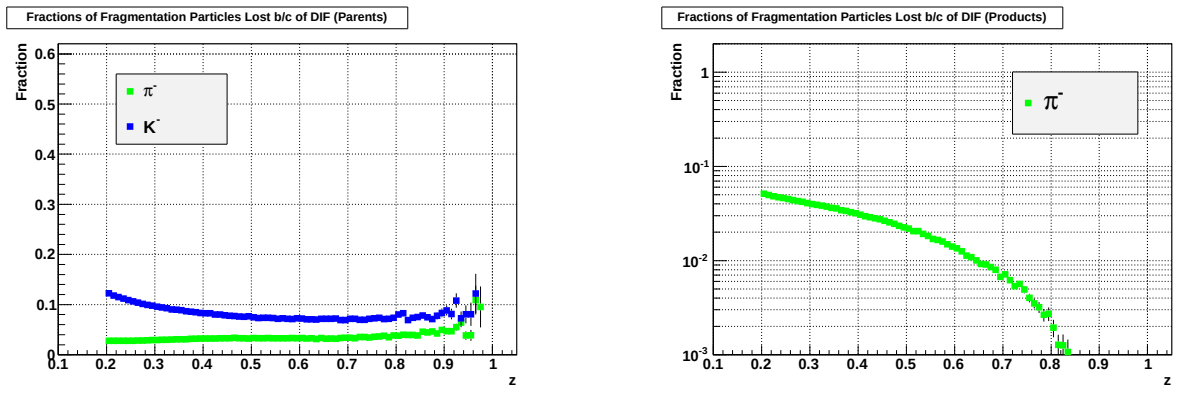
In general, the formalism used in QCD-analyses of hadron multiplicity data does not account for hadrons created in weak decays. However, since it is very difficult to distinguish between pions and kaons produced in strong or weak decays experimentally, this analysis measures hadron contributions from both processes. More detail is given in a later Section.

Figure 4.45 a) displays numbers of negatively charged fragmentation pions and kaons which decay in flight and do not enter the analysis sample, relative to all fragmentation pions and kaons, respectively. Again, these numbers also contain contributions from detector and tracking efficiencies. The fractions remain mostly below 10%. The rise at high z is again assumed to be attributable to SVD detection characteristics. It is expected that the data points for kaons lie above the pion points because of the much shorter kaon decay length, which is verified by the plot. More systematic studies are given in a later Section. A charge dependence for pions is observed. The rate of negatively charged pions decaying in flight and not entering the analysis sample is about half the rate of positvely charged pions at low z, and remains different by about 40% for high z. This behavior will be further studied in the next Section.

Additionally, Figure 4.45 b) shows numbers of negatively charged pions originating from DIF processes and not entering the analysis sample, relative to all fragmentation pions. These fractions start at low z at values of a few percent and drop below one percent at around z 0.6.

Consistency Study on Rates of Decay-in-Flight between Monte Carlo and Analytical Calculation

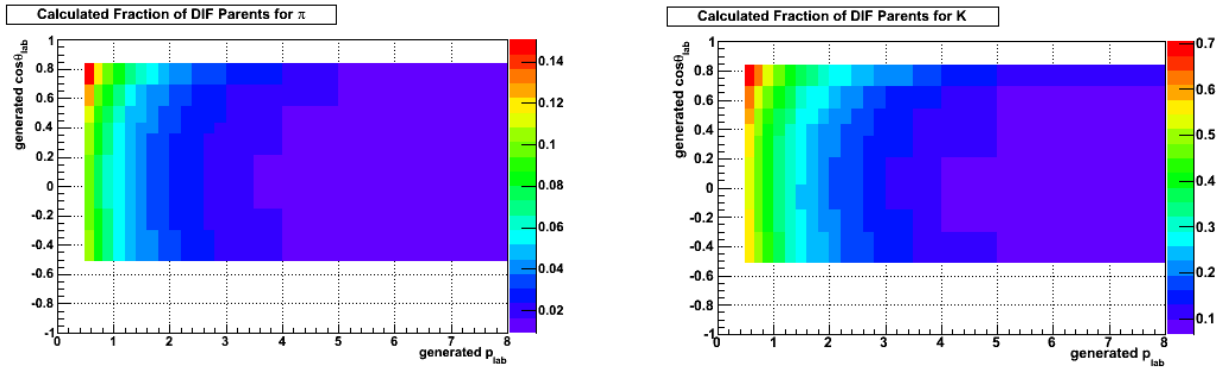
For a given momentum and flight length, DIF rates can be calculated using the common law for radioactive decay (neglecting the track curvature in the 1.5 Tesla magnetic field). In order to cross-check the DIF implementation in the Belle Monte Carlo, analytically calculated DIF rates are compared to rates from the Monte Carlo analysis sample. For illustration, Figure 4.46 shows calculated fractions of particles which will have decayed in flight before the last sensitive K-Long and Muon detector layer at radius of 3335 mm perpendicular to the beam pipe for pions (a) and kaons (b). The nonlinear patterns seen in Figure 4.46 arise from the asymmetric beam energies at KEK-B and the related boosted, asymmetric momentum and



(a) (b)

Figure 4.45: (a) Numbers of negatively charged fragmentation pions and kaons decaying in flight and not entering the reconstructed analysis sample, relative to all fragmentation pions and kaons, respectively. (b) Numbers of negatively charged pions originating from decay-in-flight, relative to all fragmentation pions.

scattering angle distributions of all created particles in the laboratory frame. Rates of particles decaying before being picked up by the Belle tracking for $z > 0.2$ are expected and found to be significantly smaller, cf. Figure 4.45.



(a) (b)

Figure 4.46: Analytically calculated fractions of particles decaying in flight until the last sensitive K-Long and Muon detector layer at radius of 3335 mm perpendicular to the beam pipe, for pions (a) and kaons (b). The ranges for the z coordinate axes are different between the two plots.

To achieve the comparison between calculated and MC DIF rates, distributions of DIF rates versus z need to be extracted for calculated DIF rate. These distributions are created by weighting yields binned in laboratory frame kinematics from the MC study sample with calculated DIF rates. Figures 4.47 shows all reconstructed pions in the analysis sample, for negative (a) and positive (b) particle charge (solid black line). Also shown in each plot are pions which are present in the analysis sample but decay in flight before the last KLM layer according to Monte Carlo information (solid green line) and according to the analytical weights

given in Figures 4.46 (dotted green line). Figures 4.47 c) and d) give charge ratios of Monte Carlo-tagged DIF yield fractions for pions and kaons, respectively, dividing yields for positive by negative particle charge.

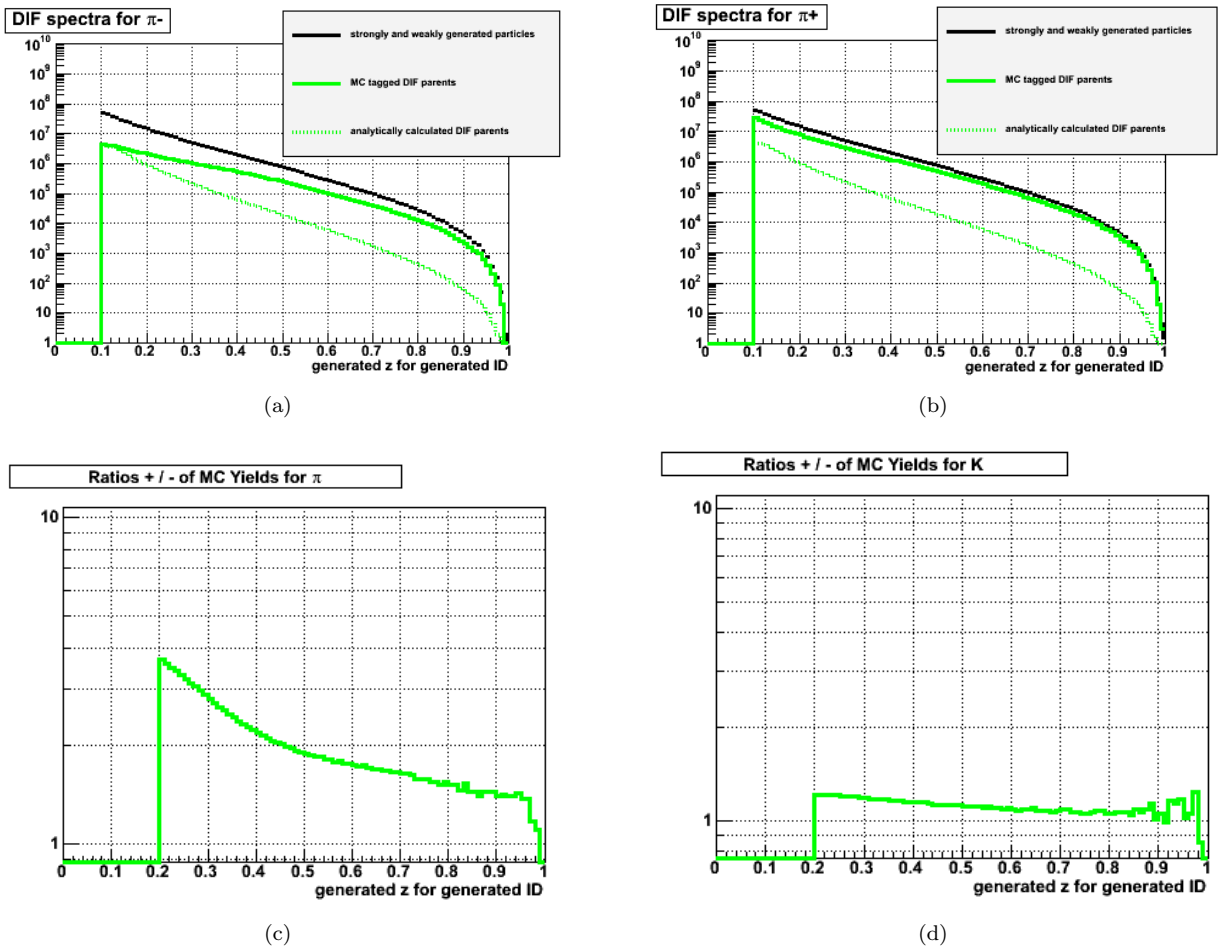


Figure 4.47: Plots a) and b): Yields of pions in analysis sample, decaying in flight according to MC information and decaying in flight according to analytical calculation, for negative and positive charge, respectively. Plots c) and d) show charge ratios of fractions of particles decaying in flight according to MC information for pions (c) and kaons (b).

As can be seen from Figures 4.47, Monte Carlo and analytical rates for particles decaying in flight in the analysis sample do not agree with each other. Additionally, especially for pions, Monte Carlo suggests that positively charged pions are almost four times as likely to decay in flight before the last sensitive layer of the KLM than negatively charged pions. The same trend is visible for kaons, although with a smaller difference between the charges. Similar observations have been quoted in the previous Section when only studying primary fragmentation particles.

In order to confirm correct DIF implementation in the Belle Monte Carlo, plots from Figures 4.47 are re-extracted, but requiring that the trajectories of contributing particles stay inside the beam pipe for at

least 10 cm. Also, to obtain a sample of particles with production vertices close to the interaction point, only particles emerging from strong decay processes are analyzed. Figures 4.48 show MC and analytical DIF fractions from this sample. As can be seen, agreement of MC information with the analytical expectation is achieved, and charge biases are not present anymore.

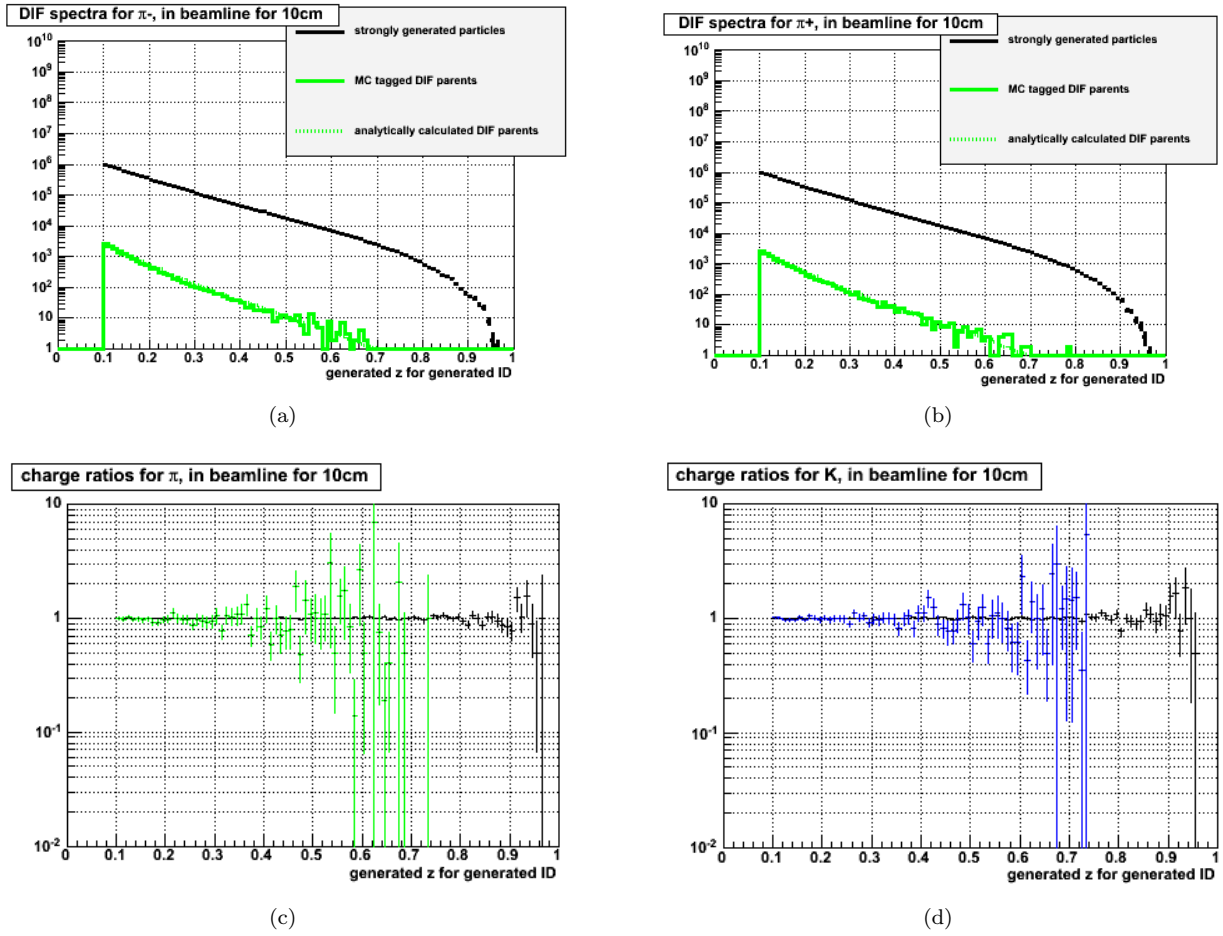


Figure 4.48: Similar to Figures 4.47, but only considering particles produced in strong decay processes and requiring particles to travel inside the beam pipe for at least 10 cm. All plot specifications given for Figures 4.47 hold.

Comparing Figures 4.47 with Figures 4.48, we conclude that the difference in DIF rates arises because of material interaction properties which are not accounted for in the analytical calculation. Additionally, it is assumed that positively and negatively charged pions and kaons interact with material differently. Further study suggests that cross sections between positively and negatively charged pions with protons might explain the charge differences observed in MC data. Figure 4.49 shows the ratio of cross sections of pions with protons, dividing cross sections for positive by cross sections for negative pion charge. The cross section for elastic interaction between protons and positively charged pions is larger by about 10% than the

cross section for negatively charged pions for low to medium laboratory frame momenta.

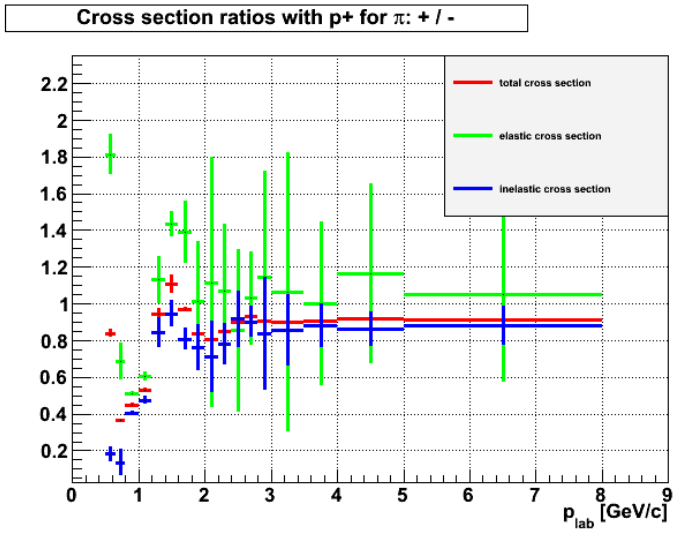


Figure 4.49: Charge ratio of cross sections of pions interacting with protons versus laboratory frame momentum. Cross sections for positive pion charge are divided by cross sections for negative pion charge. All cross sections are taken from Ref. [5].

The charge difference observed in MC-tagged DIF rates can probably be related to the larger elastic proton cross section for positive pions. It seems valid to assume that positively charged pions undergo more elastic interactions with detector material, lose more energy and decay with a higher rate than negatively charged pions.

Unfortunately, the Belle Monte Carlo data tables do not allow for more detailed studies. Only particle information at production time is stored in the 'gen_hepevt' tables, instead of the Pythia/Jetset default of the full history of a given particle through all interactions. Thus no detailed analysis to compare decay-in-flight rates for pions including material interaction effects can be carried out. From the good agreement between Belle Monte Carlo information and analytical calculations for DIF rates inside the beam pipe, it is assumed that the Belle GEANT detector simulation treats DIF correctly, also if interaction with material is considered.

Detector/Tracking Efficiency

Most plots shown above in this Section contain contributions from detector and tracking inefficiencies. Limitations in the Belle Monte Carlo do not allow to isolate these effects, as described in previous Sections. However, it is possible to extract an efficiency estimate, describing the joint quality of the total Belle tracking including contributions from detector and tracking efficiencies.

This general tracking efficiency is extracted from a subsample of the one introduced in the beginning

of this Chapter. The sample solely consists of pions and kaons produced in strong interactions, with the same Monte Carlo production time as the virtual photon in the annihilation event. This ensures that these particles are generated well within the beam pipe. As an additional requirement, all particles in the subsample are required to not decay in flight before reaching the KLM detector at a radius of about 3.30 meters, and thus passing all Belle tracking layers.

The general detection efficiency is determined by dividing the yields of tracks entering the reconstructed analysis sample over all tracks in the strong decay subsample. Figure 4.50 shows general tracking efficiencies for negatively charged pions and kaons. Although there is limited statistics in this sample at high z, the data points indicate that the general tracking efficiency is above 92% for all $z > 0.2$ under application of track quality cuts.

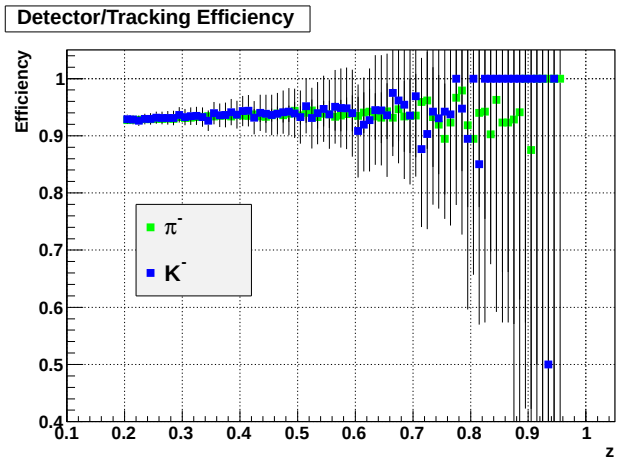


Figure 4.50: Efficiency into the reconstructed analysis sample of negatively charged pions and kaons produced within the beam pipe and not decaying before the KLM detector.

Particles Gained from Detector Interactions

In addition to particles from direct fragmentation, the reconstructed analysis sample also contains further particles which have not been accounted for yet. Some primary fragmentation particles interact with the detector and produce secondary particles. These particles can enter the analysis sample as well. Figure 4.51 shows the number of negatively charged pions and kaons produced in detector interaction processes relative to the number of primary fragmentation pions and kaons. It can be seen that such particles are only present in the analysis in small numbers. Thus it is considered safe to assume that they do not contribute to the present correction significantly.

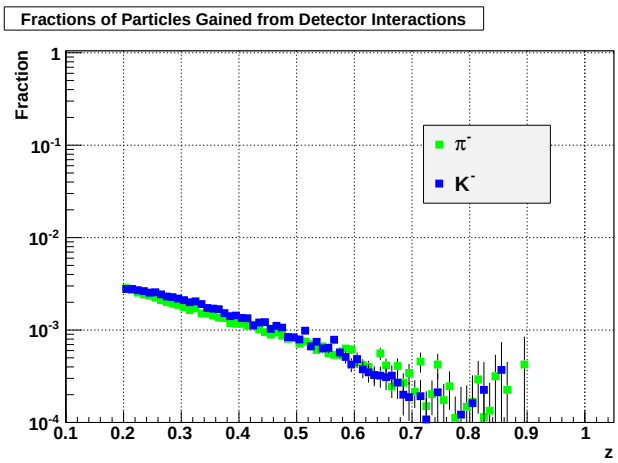


Figure 4.51: Numbers of negatively charged pions and kaons produced in hadronic interactions of fragmentation particles and entering the reconstructed analysis sample, relative to all fragmentation pions and kaons, respectively.

Particles Gained from Multiple Reconstructions

In the track finding process during the reconstruction of events, the coordinates of detector signals are connected and a track fit is performed. Sometimes the tracking algorithm produces more than one track candidate to describe the signals of one physical particle due to detector inefficiencies or occupancy effects. In some cases, these 'fake-tracks' or 'ghosts' enter the reconstructed analysis sample as well. Since fake-tracks are produced in reconstruction, they are studied in reconstructed kinematics with likelihood-selected PID. Figure 4.52 gives the number of fake-tracks with reconstructed negative track charge and likelihood-assigned pion ID relative to the number of all reconstructed pion tracks after analysis cuts. Since these fractions are already well below one per mille for all z, it is assumed that they will not contribute significantly to the present correction either.

4.4.3 Extraction of Correction Efficiency Factors

In the previous Sections in this Chapter, processes have been analyzed and studied which remove primary fragmentation particles from and contribute non-fragmentation particles to the yields in the reconstructed analysis sample. In the following Sections, correction factors are extracted and the correction is performed.

As mentioned in the previous Sections, some of the different causes for particle losses/gains are convoluted in the Belle Monte Carlo data structures and cannot be studied or resolved separately. Only the impact of the applied analysis track cuts can be analyzed on their own. Thus it was decided to calculate the total correction factor from the product between an analysis cut efficiency and a joint efficiency arising from all remaining effects studied above.

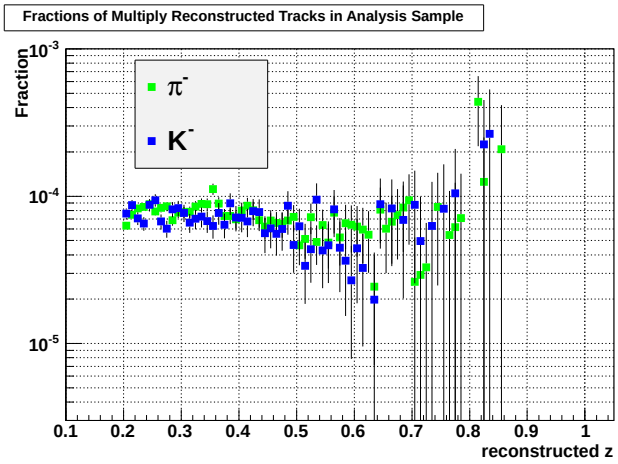
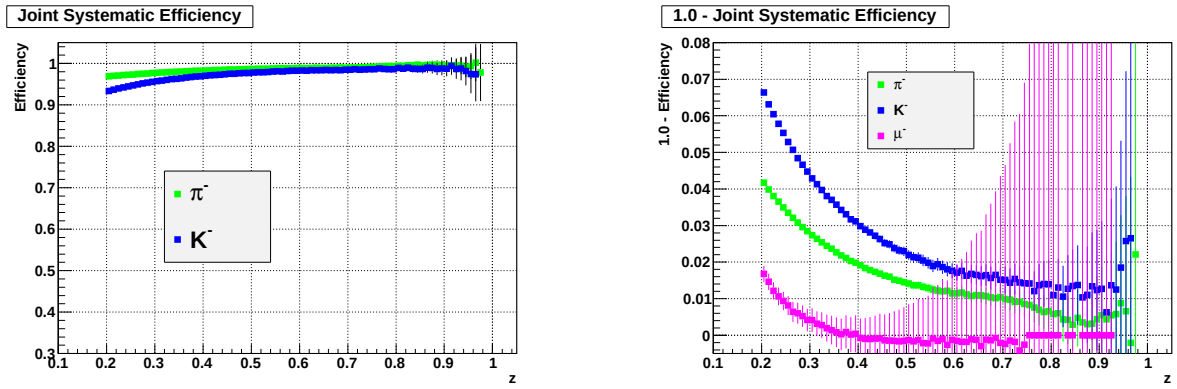


Figure 4.52: Numbers of fake-tracks originating from multiply reconstructed particles in the analysis sample, relative to the number of all reconstructed tracks after analysis cuts, for negative track charge and pion ID. Particle species are assigned by likelihood cuts and reconstructed kinematics are shown.

Joint Efficiency Factors

Extraction of Joint Efficiency Factors The joint efficiencies include decay-in-flight, detector interaction as well as detector/tracking efficiency contributions. The efficiencies are extracted from the analysis sample introduced in the beginning of this Chapter. Figure 4.53 a) shows joint efficiencies for negatively charged fragmentation pions and kaons from the MC analysis sample.



(a) (b)

Figure 4.53: (a) Joint efficiency factors for negatively charged fragmentation pions and kaons from the MC analysis sample. (b) Values of joint efficiencies subtracted from 1.0 for negatively charged pions, kaons and muons from the MC analysis sample.

As can be seen from Figure 4.53 a), the change of measurement yields due to this part of the correction is below 10% for all values of z. The main difference to the lower efficiencies shown in Figure 4.43 is that analysis track cuts are not applied when forming the efficiencies shown in Figure 4.53 a). It is assumed that

the difference between pion and kaon efficiencies can be related to the differences in decay lengths of the particles. More kaons than pions are expected to decay before the tracking can properly detect the particles, and the difference is expected to become smaller with increasing z. This reasoning is verified by the data points.

Estimation of Contributions from DIF and Hadronic Detector Interaction in Joint Efficiencies As a result of the studies of DIF and detector interactions in the Sections above, significant losses of fragmentation particles have been found. Another way to estimate contributions from those two processes to joint efficiencies is given by studying muons.

Pions and muons are very close in mass. However, muons do not interact strongly and have a significantly longer decay length than pions. Based on those differences, comparing pion and muon joint efficiencies provides a means to estimate the size of DIF and hadronic interaction contributions to joint efficiencies. To best enable successful reconstruction, muons produced in DIF from pions and kaons are excluded from the efficiency calculation. Figure 4.53 b) shows the values of joint efficiency factors subtracted from 1.0 for negatively charged fragmentation muons, pions and kaons.

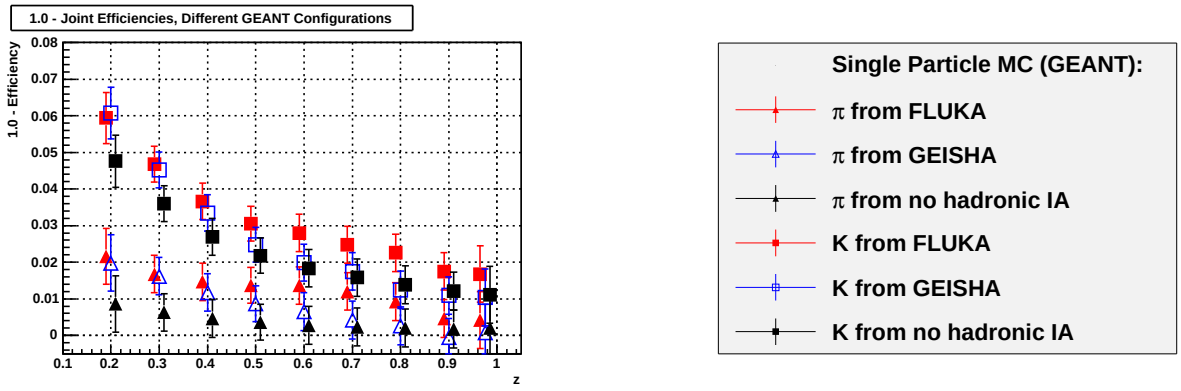
Judging from Figure 4.53 b), joint systematic efficiencies for muons are very close to 100% over the entire range of z. The efficiency values are expected to mostly reflect the general particle detection efficiency. Thus it can be concluded in analogy to the findings in previous Sections that DIF and hadronic interactions largely contribute to the difference of hadron joint efficiencies to 100%. The difference between the general particle detection efficiencies for muons and pions/kaons, Figures 4.50 and 4.53 b), is most likely caused by not applying track quality cuts for the muon joint efficiencies.

Estimation of Systematic Uncertainty on Joint Efficiency Factors In the Belle Monte Carlo GEANT-based detector simulation, hadronic interactions are modeled by the GEANT-FLUKA [85] software package. It is expected that a comparison of efficiencies from different hadronic interaction packages provides an estimate for the systematic uncertainty arising from a specific package choice.

To perform this comparison, a feature of GEANT is used which generates single tracks and propagates them through the detector simulation. Three samples of 10^6 pions and kaons of positive and negative charge are created. The first sample is generated using the Belle default GEANT-FLUKA package, the second one using the GEISHA [86] simulation package. Particles for the third sample are created while disabling hadronic interactions overall in the MC simulation.

Joint efficiencies are calculated for each of the samples in 9 z bins $[0.15; 0.25)$, $[0.25; 0.35)$, ..., $[0.95; 1.0)$. The results can be seen in Figures 4.54. Differences between the GEANT-FLUKA efficiencies shown in

Figures 4.54 to Figure 4.53 b) are expected to arise from occupancy effects. As can be seen from Figures 4.54, differences in joint efficiencies calculated from samples with different or no hadronic interaction simulation packages are small. As additional systematic uncertainty, the difference between GEANT-FLUKA and GEISHA joint efficiencies is added in quadrature to the statistical uncertainty of the joint efficiencies shown in Figure 4.53 a).



(a) (b)

Figure 4.54: Joint efficiencies for negatively charged pions and kaons, extracted from self-generated MC samples using the GEANT-FLUKA package, the GEISHA package, and with no hadronic interaction modeling at all.

Analysis Cuts Efficiency Factors

Extraction of Analysis Cuts Efficiency Factors To extract efficiency factors for the helix and SVD analysis cuts described in Section 3.1.1, the same Belle Monte Carlo analysis sample has been employed as in the Sections above. Ratios are calculated from reconstructed fragmentation pions and kaons before and after applying the helix and SVD cuts. Figure 4.55 shows analysis cut efficiencies for negatively charged fragmentation pions and kaons. The difference to 1.0 of all ratios aside the tracking performance mostly measures the Silicon Vertex Detector (SVD) efficiency.

Study of SVD Hit Association Efficiency The most restrictive track quality cut applied is the requirement that at least three SVD detector elements show hits for a given track. The SVD hit association algorithm in the Belle reconstruction uses the same requirement for SVD hits to be associated with a track. If less than 3 hits are found, the track is fitted with CDC information only.

Figure 4.56 shows the fraction of negatively charged pions without associated SVD hits in the reconstructed analysis sample,for generated kinematics. It can be seen that the black curve is closely related to the complement of the pion curve in Figure 4.55 with respect to 1.0. The black curve in Figure 4.56 shows a

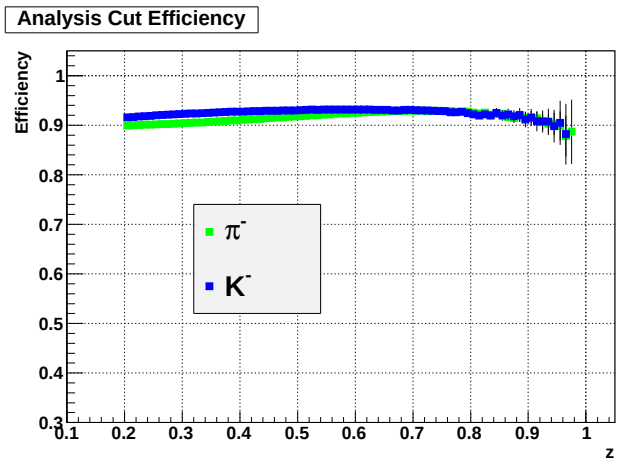


Figure 4.55: Analysis cut efficiencies for negatively charged pions and kaons from the MC analysis sample.

tendency to increase at high z. This is not expected, as it should become more likely with increasing track energy that SVD hits can be associated with a straight track.

After consulting SVD experts, it turns out that there is a known inefficiency in the SVD L0 trigger scheme which causes some events to contain only tracks with no associated SVD hits. This inefficiency affects events containing high z tracks more than events containing low z tracks, thus causing an increase in the fraction of particles without associated SVD hit at high z. The red curve in Figure 4.56 is formed by removing all tracks from SVD-inefficient events, again showing fractions of negatively charged pions without SVD hits in the reconstructed analysis sample, for generated kinematics. The curve is approaching an asymptotic value at high z, as expected for a decreasing likelihood that SVD hits cannot be associated with an essentially straight track.

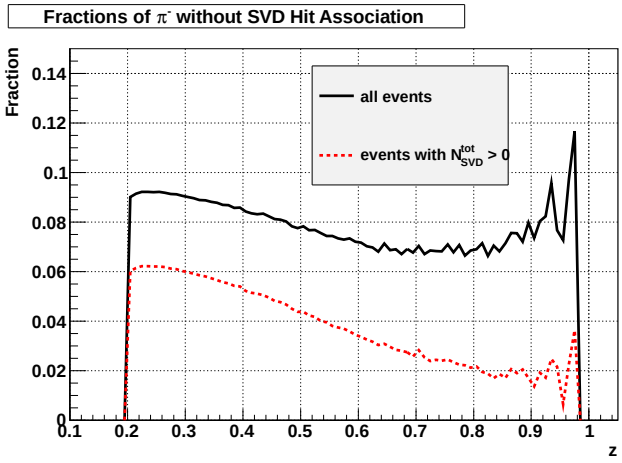


Figure 4.56: Fractions of negatively charged reconstructed pions without associated SVD hits, relative to all reconstructed negatively charged pions from the MC sample, for generated kinematics.

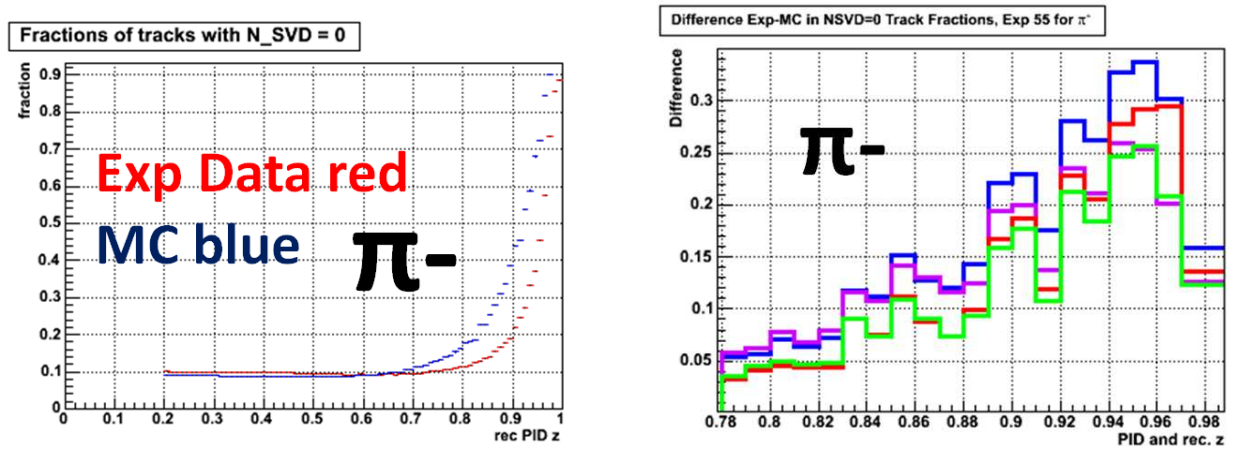
In the course of studies to resolve the issue of increasing particle fractions without associated SVD hit versus z, a discrepancy was found between Monte Carlo and experimental data. Figure 4.57 a) shows fractions for tracks without associated SVD hit from Monte Carlo and experimental data for on-resonance Experiment 55 data, for reconstructed kinematics and likelihood-selected ID. The drastically larger values of the fractions at high z are an effect of momentum smearing. It can be seen from Figure 4.57 a) that both Monte Carlo and experimental data fractions show a very similar z dependence up to about z 0.65. For higher z, the Monte Carlo fractions rises earlier than the experimental data fraction, accumulating a maximum absolute difference of about 0.35 for highest z. Two attempts are made to align the Monte Carlo fractions with the experimental data fractions. The reasoning applied bases on the idea that the Monte Carlo sample lacks well-reconstructed tracks with SVD hits associated with them.

The first attempt consists of combining τ events with the generic Belle Monte Carlo data, to resemble the consistency of experimental data containing also non-q$\bar{q}$ events. Monte Carlo τ events generated according to the Experiment 55 cross section are included analogously to the analysis in Chapter 4.2. Figure 4.57 b) shows the absolute value of the difference between Monte Carlo and experimental data track fractions without associated SVD hit as a blue line. The purple line indicates this difference if τ events are included in the Monte Carlo sample. Including τ events is seen to decrease the difference between Monte Carlo and experimental data at z higher than about 0.85, as expected from the analysis in Chapter 4.2 since τ contributions to pion yields are highest at high z. However, still a sizable difference between Monte Carlo and experimental data remains.

The second attempt to align Monte Carlo with experimental data track fractions without associated SVD hits bases on the analysis results of Chapter 5.3: Comparing multiplicities extracted from generic Belle Monte Carlo with the results of this measurement, it can be seen that the Monte Carlo data undershoot experimental data at high z by more than a factor of two for negatively charged pions. Thus it is attempted to scale the fraction of fragmentation pions in the reconstructed sample of this study by the value of the ratio between Monte Carlo and experimental data multiplicities. The resulting difference between the scaled generic Monte Carlo, including τ contributions, and experimental data track fractions without associated SVD hits is given by the green line in Figure 4.57 b). While the difference to experimental data decreases yet some more, still a fairly large discrepancy remains, largest at about 25% for z 0.95.

Since kinematic smearing has already been corrected for at this point (cf. Section 4.3), the differences between MC and experimental described above have to be assessed for smearing-corrected kinematics. In these kinematics, the experimental data fractions are expected to become almost flat at higher z analogously to the MC fractions for generated kinematics in Figure 4.56. The value of the fractions at higher z is assumed

to be dominated by single detector layer inefficiencies as opposed to lower z, where more difficult tracking of non-straight tracks could cause higher fractions. The Belle SVD single detector layer efficiencies measured from cosmic rays are quoted to lie between 92 and 97%, cf. Ref. [87]. Therefore both experimental and MC fractions of particles with no SVD hits are expected to be around 10% for smearing-corrected kinematics. Correspondingly, the difference between MC and experimental fractions for smearing-corrected kinematics can reasonably be assumed to be much smaller than for reconstructed kinematics. However, it was decided to assign the differences from reconstructed kinematics as systematic uncertainties to the analysis cut efficiency factors. This reflects a conservative approach to account for the stated differences of SVD efficiencies in MC and experimental data.



(a) (b)

Figure 4.57: (a) Comparison between fractions of tracks with no associated SVD hit between Monte Carlo and experimental data for Experiment 55, for reconstructed kinematics, negatively charged tracks and likelihood-selected pion ID. b) Difference between fractions of tracks with no associated SVD hit of generic Belle Monte Carlo and experimental data (blue line), with additional contributions from τ events in the Monte Carlo sample (purple line), and with scaled generic Belle Monte Carlo and τ event contributions (green line). All lines are extracted for reconstructed kinematics, negative track charge and pion likelihood selections.

4.4.4 Correction of Measurement Samples

In order to be able to correct for all effects described in this Section, the joint and analysis cut efficiencies shown in Figures 4.53 a) and 4.55 are multiplied yielding combined efficiencies, with all uncertainties propagated. The combined efficiencies for negative particle charge are shown in Figure 4.58. Finally, the analysis measurement samples are corrected with the inverse of the efficiencies shown in Figure 4.58.

Uncertainties from all involved quantities are again propagated according to Equations 3.2 and 3.3. The dominant contribution to the increase in systematic uncertainties on the measurement samples due to this

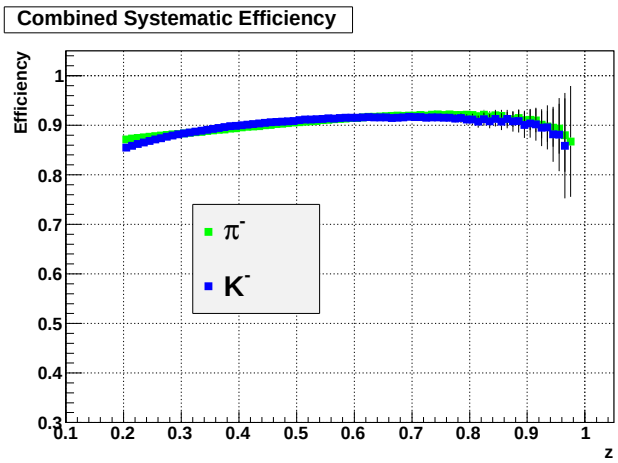


Figure 4.58: Combined correction efficiencies for negatively charged pions and kaons from the MC analysis sample.

correction arises from the differences between Monte Carlo and experimental data track fractions without associated SVD hit. Systematic uncertainties on the charge-integrated measurement sample increase by a maximum of 10% at high z.

4.4.5 Study of Impact of Correlations between PID and Decay-in-Flight Corrections

Study Introduction The presented analysis attempts to measure the complete set of pions and kaons originating from $q\bar{q}$ annihilations under utilization of reconstructed data. Due to DIF, some pions and kaons are not present in the reconstructed data set used as input for the PID correction. So far, the PID and DIF corrections are assumed to independently factorize in the presented analysis. This assumption needs to be tested to assure that all decay-in-flight effects are consistently taken into account in the PID correction.

Qualitatively, pion and kaon in-flight decays which happen inside the SVD detector are of no concern. Their decay products traverse the full detector and are reconstructed and PID-selected as such. The original pions and kaons are fully and consistently recovered in the DIF correction. Hadron DIF events occurring after the last layer of the KLM do not represent an ambiguity between the PID and the DIF correction either. They enter the PID correction as pions and kaons and are not influenced by the DIF correction. For all decays between the SVD and KLM, the PID correction needs to consistently treat all effects of DIF present in the raw data sample.

Comparing Decay-in-Flight Contributions in Raw Experimental and PID Probability Extraction Samples To assess the consistency of DIF treatment in the PID correction, the relation of the

amounts of DIF products in the raw PID-uncorrected sample and the sample used to extract PID probabilities has to be established. In this study, the PID-uncorrected raw data sample is approximated by the MCHadronBJ samples introduced in Section 4.1.5. The observed undershooting of MCHadronBJ yields compared to experimental yields especially at high z is considered irrelevant in this study, as the fractions of pions and kaons decaying in flight is expected to be the same in either sample. As the second data set in this study, the MC D^* skim is analyzed, since only probabilities for true pions and kaons are expected to show a possible influence of DIF effects.

To compare the amount of DIF particles present in both samples, a test observable is chosen. This observable is constructed starting with the yields of fragmentation pions and kaons which decay in flight before the last layer of the KLM detector in each z bin for reconstructed kinematics. Additionally, these pions and kaons or their decay products are required to pass all track quality cuts and enter the analysis sample. Their PID likelihood selections are resolved. Finally, these PID-selected yields are normalized in each z bin to the PID-selected total yields of fragmentation pions and kaons entering the analysis.

Since this study aims to extract DIF contributions for raw PID-uncorrected yields, all cuts are evaluated in reconstructed kinematics and for PID-selected particle species. It shall be noted that when analyzing the MC D^* skim sample, the test observables have to be assembled from the actual sample used to calculate PID probabilities. Therefore, the observables are extracted from all tracks entering the final Δm invariant mass histograms in the D^* PID probability extractions after all cuts.

Plots showing the test observables in each of the two study samples are given in Figures 4.59 and 4.60. The observables from the MC D^* skim are available with sufficient statistical significance only up to $z \sim 0.7$ due to phase space limitations in the production of pions and kaons from D^* decays. The z dependence of the observables is determined by a convolution of PID-detector response and DIF kinematics. Observable differences between particle charges are assumed to be analogous to DIF charge differences investigated in Section 4.4.2.

Ratios between the curves shown in Figures 4.59 and 4.60 describe the difference of DIF presence between the raw PID-uncorrected sample and the sample used to extract PID probabilities. Such ratios are shown in Figure 4.61. It can be seen that the difference in DIF presence between the two samples for 'diagonal' PID-selection (true fragmentation pions PID-selected as pions, similarly for kaons) is less than 5%. Differences for 'off-diagonal' PID-selections are larger.

In the context of the PID correction, a notable feature of the ratios shown in Figure 4.61 is the difference in DIF presence between the two samples for decaying true fragmentation kaons which are PID-selected as protons. This difference indicates that kaon PID probabilities are extracted from samples with less DIF

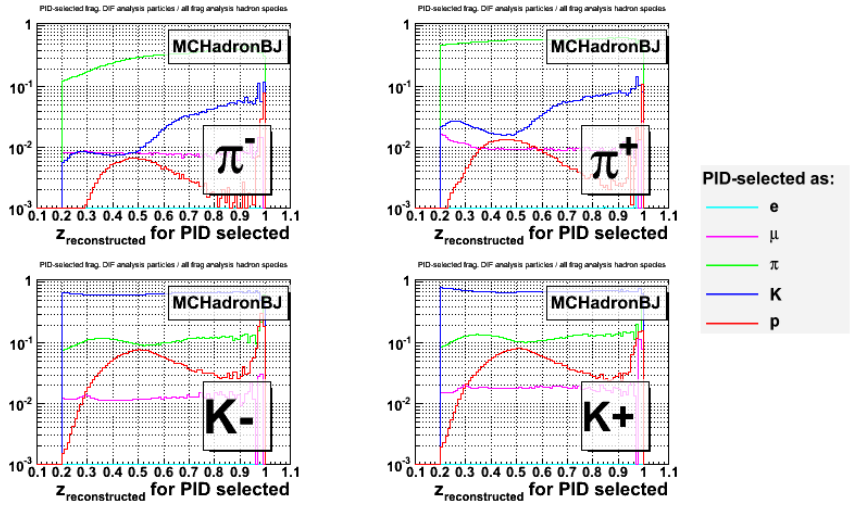


Figure 4.59: Test observables to determine contributions from decay-in-flight processes extracted from the MCHadronBJ data samples. The curve for electrons is not shown due to large statistical fluctuations. The definition of the test observable is described in the text.

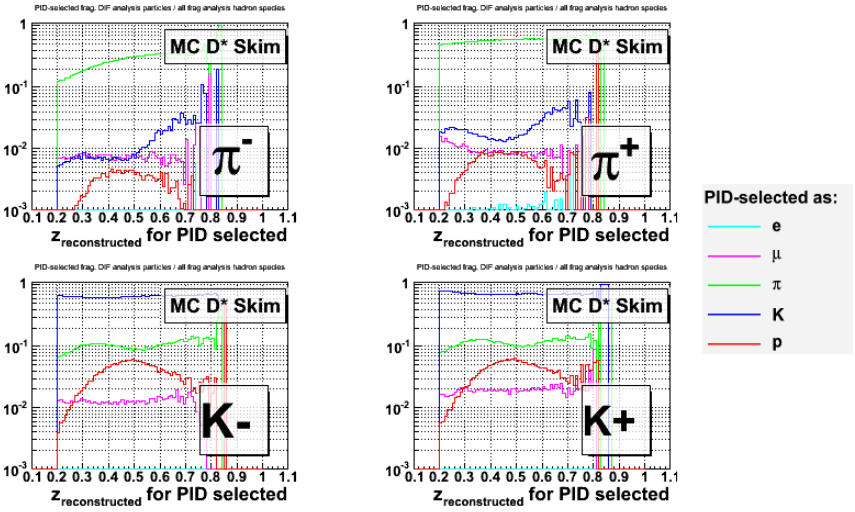


Figure 4.60: Test observables to determine contributions from decay-in-flight processes extracted from the MC D^* skim sample. The curve for electrons is not shown due to large statistical fluctuations. The definition of the test observable is described in the text.

products present than in the experimental raw sample. A possible consequence is that raw proton PID-selected yields have been over-corrected in the PID correction, leading to the negative PID-corrected proton yields observed in Section 4.1.7.

Extraction of Systematic Uncertainties from PID-DIF Correlations The impact of the difference between the presence of DIF products in raw uncorrected and PID probability extraction data shown in Figure 4.61 needs to be assessed further. To establish the relevance of these differences, Figure 4.62 displays

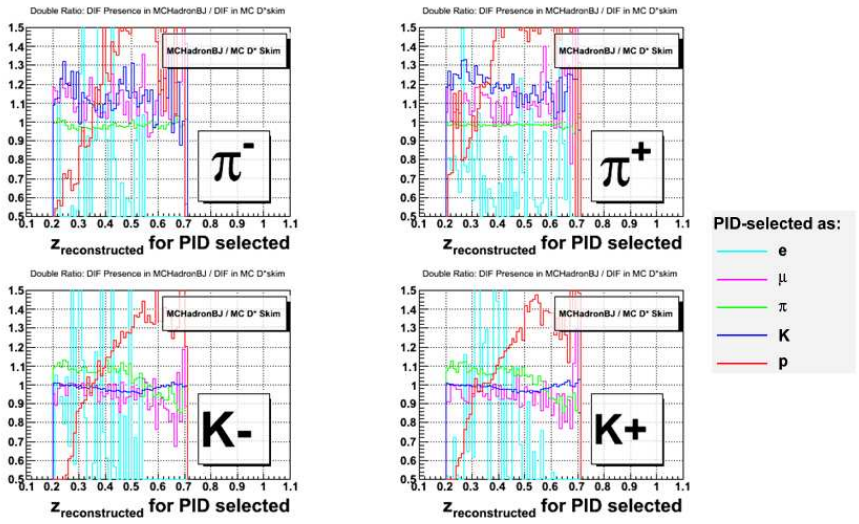


Figure 4.61: Ratios of test observables extracted from the MCHadronBJ samples over observables from the MC D^* skim. The ratios quantify the differences in DIF presence between raw PID-uncorrected yields and the sample used to extract PID probabilities.

fractions of DIF products in total raw PID-selected yields extracted from the MCHadronBJ samples. Again, the z dependence of these fractions is determined by PID detector response and DIF kinematics, and charge differences are motivated in Section 4.4.2.

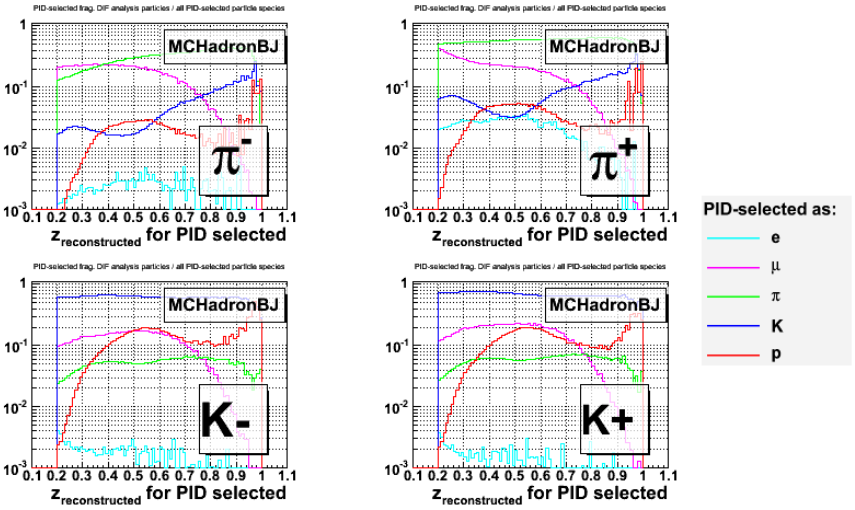


Figure 4.62: Fractions of true fragmentation pions and kaons decaying in flight before the KLM and entering the analysis after all track cuts, for their respective PID selections. The fractions are normalized to all tracks of each PID selection entering the analysis.

The relevance of DIF differences can be visualized by calculating the fraction of DIF contributions to PID-selected yields shown in Figure 4.62 when scaled by the fractions found in Figure 4.61. Such scaled fractions represent changes in raw PID-uncorrected yields corresponding to the difference of DIF presence

between the PID-uncorrected yields and the PID probability extraction sample, therefore measuring the size of PID-DIF correlations. Figure 4.63 displays such changes in raw PID-uncorrected yields. For comparison, the relative systematic uncertainties for PID-corrected yields are plotted as well.

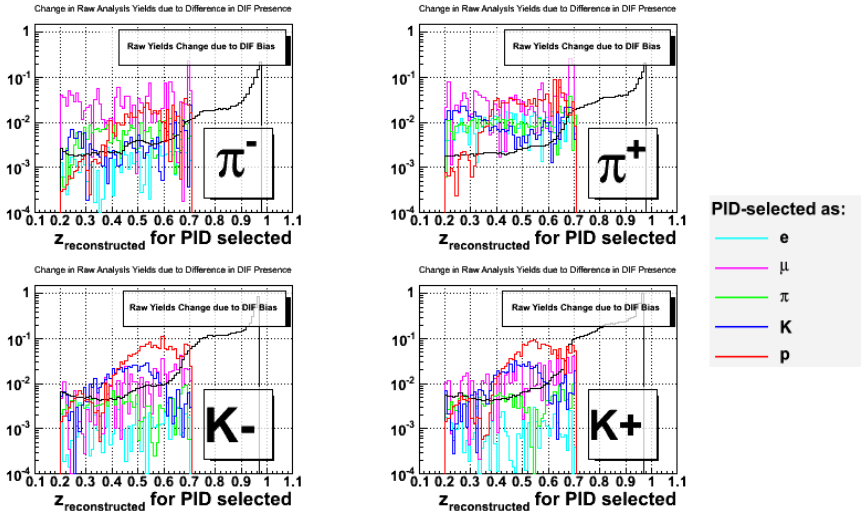


Figure 4.63: Changes in raw PID-uncorrected yields corresponding to size of PID-DIF correlations.

In Section 4.1.7, the PID correction was found to change pion yields by up to 10% and kaon yields by up to 20%. Considering the relative magnitudes of raw PID-uncorrected yields between different PID-selected species shown in Figure 3.1, the impact of changes of the order of 10% in raw PID-selected lepton and proton spectra on pion and kaon yields can be neglected. Therefore, only the changes to diagonal PID selection yields are considered relevant. Figure 4.63 indicates that the effects of PID-DIF correlations correspond to changes in raw PID-uncorrected yields by a few percent for pion and kaon PID-selected yields. The size of these changes are assigned as systematic uncertainties to account for PID-DIF correlations. At high z, present systematic uncertainties of PID-corrected yields are higher already and the graphs in Figure 4.63 do not exhibit a strong increase. Therefore no additional uncertainties are assigned at high z.

4.5 Correction for Visible Energy Analysis Cut

The event cuts introduced in Section 3 contain a cut on the visible reconstructed energy in the Belle detector. The visible energy observable is formed for each event by adding up reconstructed energies in tracking (assuming the pion mass for all tracks) and photon clusters. In the present analysis, the cut $E_{\text{vis}} > 7$ GeV is used in an attempt to reject events from entering the measurement sample. In these events, neutrinos are assumed to carry away a substantial part of the available energy since they are not detectable with the Belle detector. Thus the visible reconstructed energy of such events is expected to be significantly lower than the full center of mass energy.

4.5.1 Correction Factor Extraction

The same standard Belle MC data sample as in the previous correction has been used to extract correction factors, containing about $4.4 * 10^8$ events from two streams of on-resonance standard Belle MC data with events $e^+e^- \to q\bar{q}$, with $q = \{u, d, s, c\}$ from Experiment 55. The assumption for the correction is that in each z bin, the ratio of particles in the ToF acceptance from events with E_{vis} values larger than 7 GeV over the number of particles from all events in the HadronBJ data skim in the ToF acceptance is the same in Monte Carlo as well as experimental data. Again, only particles directly originating from events $e^+e^- \to q\bar{q}$ or decay products of such particles enter the yield ratios. Yield ratios extracted from the Monte Carlo study sample are displayed in Figure 4.64.

The z dependence of the ratios can be motivated by considering that events with high z particles in the ToF acceptance are more likely to achieve more than or equal to 7 GeV of visible energy than events with particles with low z. The difference between pion and kaon efficiencies is assumed to arise from the mass difference between the two species and the fact that for calculation of E_{vis}, always the pion mass is assumed for each track.

4.5.2 Correction of Measurement Samples

The measurement samples are corrected by bin-by-bin multiplication with the inverse of the efficiencies shown in Figure 4.64. The reconstruction and visible energy calculation process is assumed to be sufficiently well modeled in the Belle Monte Carlo. Thus no additional systematic uncertainty is assigned for the correction factors. All uncertainties are propagated according to Equations 3.2 and 3.3. Relative systematic uncertainties of the measurement samples increase by about one per cent for highest z bins due to this correction. Corrected measurement samples for negative particle charge are shown in Figure 4.65.

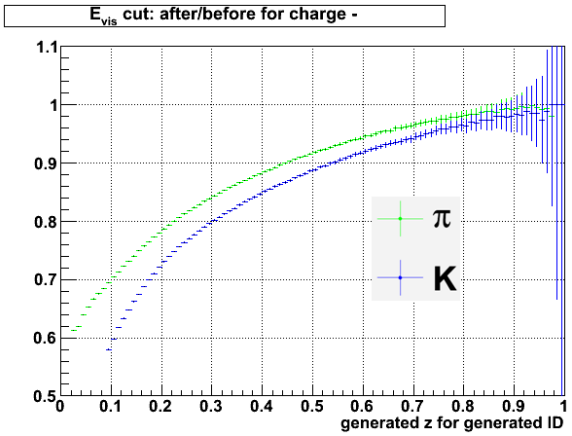


Figure 4.64: Ratios of Yields after applying the E_{vis} analysis cut over yields before applying the cut for negative track charge.

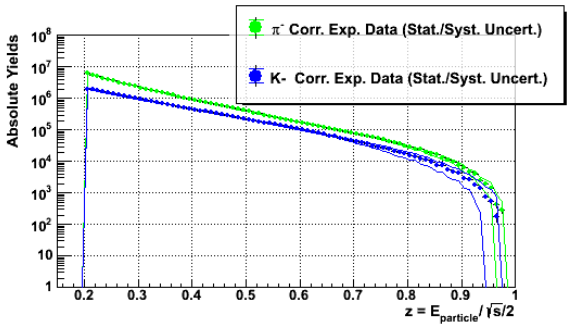


Figure 4.65: Pion and kaon measurement samples for negative particle charge, corrected for the analysis visible energy cut.

4.6 Correction for HadronBJ Data Skim Selection

As introduced in Section 3, each event contributing particles to the measurement sample is required to pass the HadronBJ event selection criteria. In order to include also events $e^+e^- \to q\bar{q}$, with $q = \{u, d, s, c\}$ which did not pass the selection criteria in the measurement, a correction for the HadronBJ skim selection has to be made.

4.6.1 Correction Factor Extraction

HadronBJ Skim Efficiencies

As the sample which correction factors are extracted from, again the same standard Belle MC data sample as in the previous correction has been used. The sample contains about $4.4 * 10^8$ events from two streams of on-resonance standard Belle MC data with events $e^+e^- \to q\bar{q}$ where $q = \{u, d, s, c\}$ from Experiment 55. For this correction, the basic assumption is that for each bin in z, the ratio of particles in the ToF acceptance

from events passing the HadronBJ skimming criteria over particles in the ToF acceptance from all generated events is the same in MC as well as experimental data. Also in this correction, for the ratio only particles directly originating from events $e^+e^- \rightarrow q\bar{q}$ or decay products of such particles are considered.

Figure 4.66 shows yield ratios/ skim efficiencies extracted from the MC analysis sample for negative particle charge. The z dependence of the ratios will be motivated in the next section.

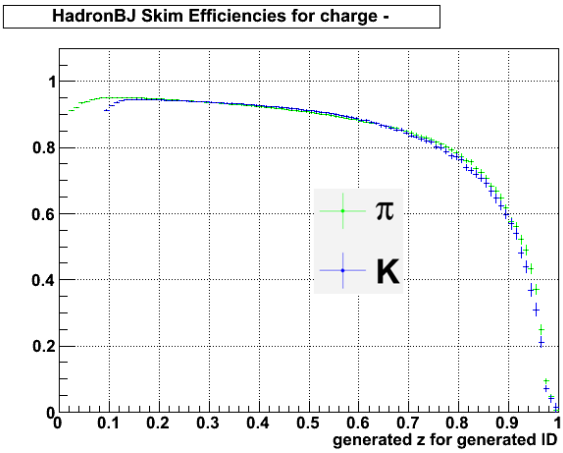
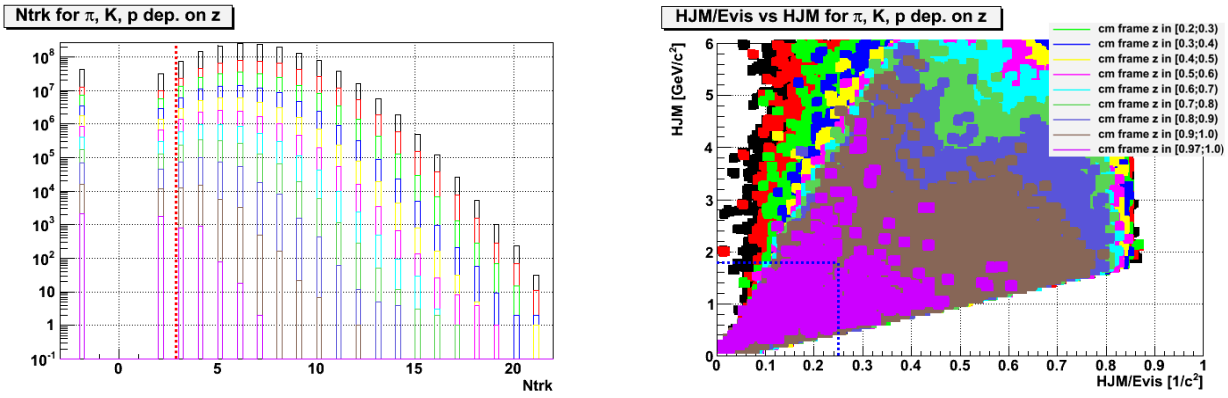


Figure 4.66: HadronBJ data skim efficiencies for negative particle charge.

Study on z-Dependence of HadronBJ Skim Efficiencies

In order to motivate the z dependence of the skimming efficiencies shown in Figure 4.66, especially the fall-off at high z, the selection criteria of the HadronBJ skim are studied in more detail. In particular, the requirement for more than or equal to three reconstructed charged tracks with transverse laboratory frame momentum of more than 100 MeV/c from the interaction point turns out to be highly selective against events containing high z particles. Likewise, the conditional requirement of Heavy-Jet-Mass (HJM) divided by visible energy of larger than 0.25 $1/c^2$ or HJM larger than 1.8 GeV/c^2 is found to reject most events containing very high z particles. Figures 4.67 show distributions of charged reconstructed tracks N_{charged} (a) and correlation plots of HJM over visible energy versus HJM (b) for pions, kaons and protons from different ranges in z.

From Figure 4.67 a), it can be seen that the number of reconstructed charged tracks N_{charged} decreases if the event contains high z particles. This can be expected, since apart from the high z particle, little center of mass energy is available for more charged particles to be created in the fragmentation process. Figure 4.67 b) suggests that events containing very high z particles show low values of visible energy as well as low HJM. This can also be motivated from similar arguments as given in the previous Section, namely that few reconstructed tracks or photons will be produced aside the high z particle. Additionally, the θ



(a) (b)

Figure 4.67: Distributions of reconstructed charged tracks N_{charged} (a) and correlation plot of HJM over visible energy versus HJM (b) for pions, kaons and protons from different ranges in z as indicated in the legend.

distribution of high z particle is expected to be very close to the theoretical $1 + cos^2\theta_{\text{cms}}$ distribution of the final state fragmenting quarks. Thus most high z particles are produced with kinematics outside of the tracking acceptance, which also decreases reconstructed HJM and visible energy.

In addition to the reasoning in the previous paragraph, another property of the HadronBJ skim selection prevents high z events from entering the HadronBJ sample. Negative numbers of reconstructed charged tracks in Figure 4.67 a) represent events which do not pass a hadronic pre-selection in the Belle reconstruction. These events are automatically rejected by the HadronBJ skimming criteria, and are also not shown in Figure 4.67 b). The pre-selection consist of the cuts given in Table 4.10,

Conditional Cut 1

Cuts	**Criterion**
N_{charged}	$>= 3$
$E_{\text{vis}}/\sqrt{s}$	$>= 0.4$
$\text{abs}(P_z^{\text{sum}})/\sqrt{s}/2$	$<= 1.0\ 1/c$
$E_{\text{ECL}}^{\text{sum}}/\sqrt{s}/2$	in $[0.05; 1.8]$

Conditional Cut 2

Cuts	**Criterion**
N_{charged}	$>= 2$
$E_{\text{vis}}/\sqrt{s}$	$>= 0.4$
$\text{abs}(P_z^{\text{sum}})/\sqrt{s}/2$	$<= 1.0\ 1/c$
$E_{\text{ECL}}^{\text{sum}}/\sqrt{s}/2$	in $[0.05; 1.8]$
$P^{\text{sum}}/\sqrt{s}/2$	$<= 1.8\ 1/c$

Table 4.10: Hadronic pre-selection applied in the Belle reconstruction algorithms. All quantities are explained in the text.

where P_z is the sum of all reconstructed center-of-mass momentum components along the beamline, E_{ECL}

the sum of cluster energies in the ECL detector, and P the summed reconstructed center-of-mass momenta of all tracks in the event. Figure 4.68 shows particle ratios from events which do not pass the hadronic pre-selection, events which fail the charged reconstructed track requirement and events which fail the conditional HJM/visible energy cut. From the plot, it can be seen that the z dependence in the HadronBJ efficiencies in Figure 4.66 closely reflects the dependence of the conditional HJM/visible energy requirement efficiency in Figure 4.68.

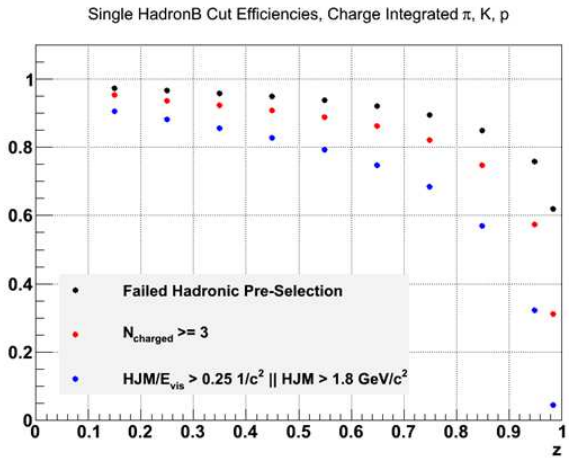


Figure 4.68: Dependence on particle z of single event requirements in HadronB data skim selection for charge-integrated pions, kaons and protons.

4.6.2 Correction of Measurement Samples

The measurement samples are corrected bin-by-bin with the inverse values of efficiencies shown in Figure 4.66. All uncertainties are propagated according to Equations 3.2 and 3.3. Relative systematic uncertainties of the measurement samples increase by about one per cent for highest z bins due to this correction.

Figure 4.69 shows corrected measurement samples for negative particle charge.

4.6.3 Systematic Study: Impact of Differences in Event Shape Variable Distributions between MC and Experimental Data

Introduction for Study The HadronBJ data skim efficiencies shown in Figure 4.66 are extracted from generic Belle MC data. As described in Section 2.4.3, events must pass several cuts on values of event shape variables to enter the HadronBJ skim. The concern is that event shape variable distributions are different between MC and experimental data. This would cause the efficiencies extracted from MC data to either over- or under-correct experimentally measured yields. Thus the existence and size of such differences is

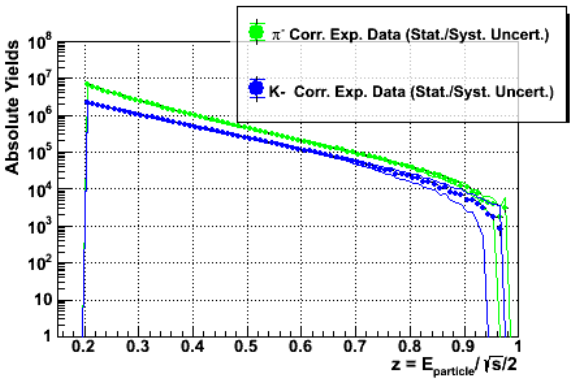


Figure 4.69: Pion and kaon measurement samples for negative particle charge, corrected for the HadronBJ data skim selection cuts.

investigated.

As shown in Figure 4.68, the requirement on the number of charged tracks in each event is one of the most selective cuts in the HadronBJ event selection. Also, the number of charged tracks is part of the cuts in the hadronic preselection described in Section 4.6.1. Therefore the study of the impact of differences in event shape variable distributions between MC and experimental data is performed exemplarily on the N_{charged} distribution. The effects found for this distribution, however, are still assumed to be representative of effects on the full set of cuts in the HadronBJ event selection.

In order to compare MC with experimental distributions, again the MCHadronBJ samples are analyzed and compared to experimental data for equal amounts of luminosity. Additionally, all observables are extracted for reconstructed kinematics and for PID-selected species. Distributions of N_{charged} are shown in Figure 4.70 for the sum of all contributing processes in the MCHadronBJ datasets (red solid line) and for experimental data (black line), only showing events which pass the HadronBJ cuts. Analogous distributions of N_{charged} for MCHadronBJ datasets without applying HadronBJ cuts are shown as red dotted lines. The first pad is given for illustration only and contains contributions from events weighted by the number of tracks in the event with reconstructed $z > 0.2$ which enter the analysis and are PID-selected as hadrons. The remaining 8 pads resolve the reconstructed z of the tracks used to weight the events in 8 bins of width $\Delta z = 0.1$. The HadronBJ cut requirement of $N_{\text{charged}} \geq 3$ is given by the green solid line.

Establish Correspondence between Approximate N_{charged} and Actual HadronBJ Efficiencies
The distributions of N_{charged} are extracted in this study for reconstructed kinematics and PID-selected hadrons. The HadronBJ efficiencies shown in Figure 4.66 are calculated on different kinematics and with generator ID information. However, the suitability of analyzing the MCHadronBJ samples in reconstructed kinematics to deduce a statement about the effect of event shape variables differences between MC and

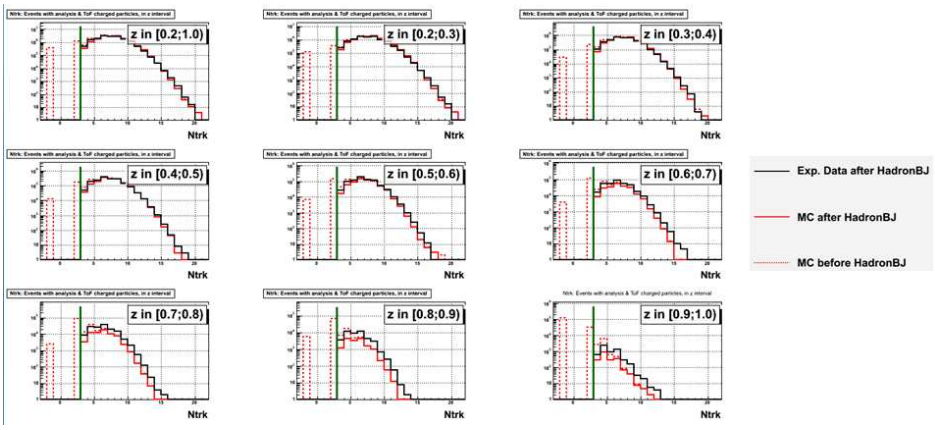


Figure 4.70: Distributions of N_{charged} from MCHadronBJ samples after (solid red line) and before (dotted red line) the HadronBJ event selection, and from experimental data after the HadronBJ event selection (solid black line). The vertical green line indicates the HadronBJ cut $N_{\text{charged}} \geq 3$.

experimental data can be established by the following reasoning. Ratios of entry numbers of the MC histograms shown in Figure 4.70 with and without HadronBJ event selection are calculated in each z bin, yielding approximate HadronBJ efficiencies. These efficiencies are shown in the left plot in Figure 4.71 together with the actual HadronBJ efficiencies given in Figure 4.66. Significant differences are visible. On the other hand, the right plot in Figure 4.71 shows analogous approximate efficiencies, but only for events from the generic *uds* and *charm* datasets in the MCHadronBJ samples. These ratios are largely consistent with the actual HadronBJ efficiencies. Therefore any effects on approximate efficiencies due to event shape variable differences between MCHadronBJ and experimental data are considered to correspond directly to variations in the actual MC HadronBJ efficiencies shown in Figure 4.66.

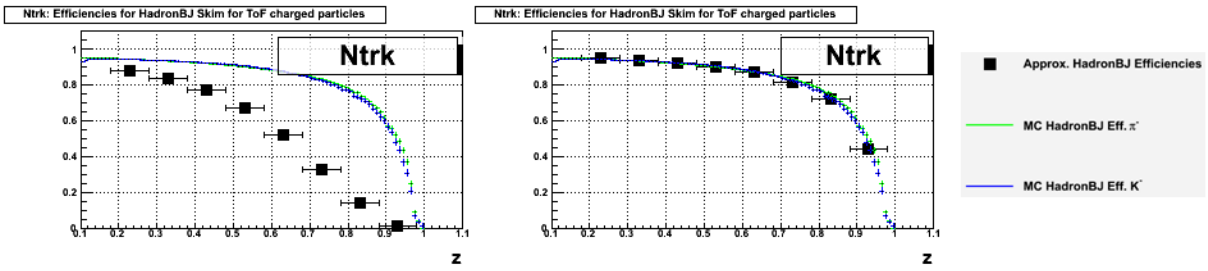


Figure 4.71: Approximate efficiencies from ratios of MC N_{charged} distribution entries, from the full MCHadronBJ sample (left) and from *uds* and *charm* Belle MC samples only (right). For comparison, the efficiencies used in the HadronBJ correction from Figure 4.66 are plotted as well.

Extraction of Systematic Uncertainty from N_{charged} Shape Differences Between MC and Experimental Data In order to quantify the difference in the distributions of N_{charged} between MC and experimental data for events passing the HadronBJ cuts, both distributions are normalized to integrals equal to

1.0. This allows the comparison of the shapes of N_{charged} distributions even though absolute MCHadronBJ yields undershoot experimental data at high z as described in Section 4.1.5. Ratios are formed by dividing the normalized experimental by the normalized MC N_{charged} distributions in each z bin. These ratios are shown in Figure 4.72. Discontinuities of ratios above $N_{\text{charged}} = 12$ are neglected since bin contents drop below 1% of the total number of events. The majority of all ratios is within 0.5 and 2.5.

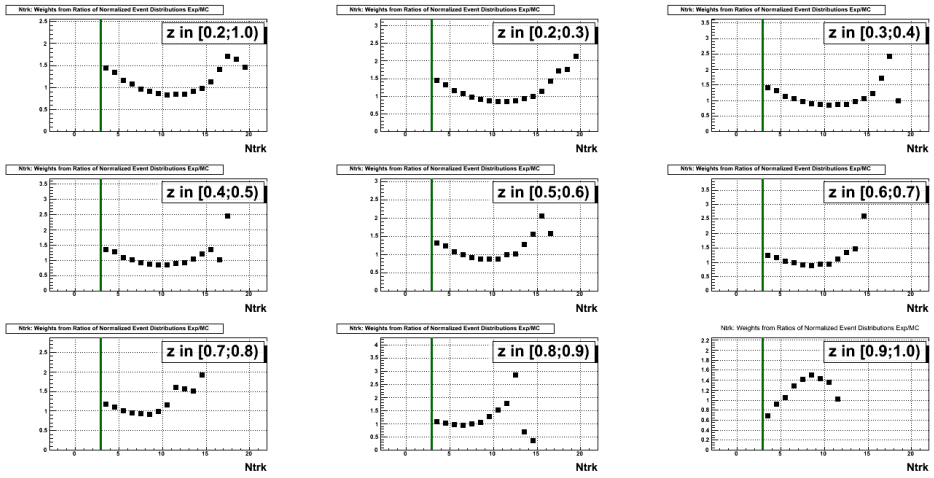


Figure 4.72: Ratios of normalized N_{charged} distributions from experimental data over distributions from MCHadronBJ data. The vertical green line indicates the HadronBJ cut $N_{\text{charged}} \geq 3$.

In the following, the difference in shape of the experimental N_{charged} distribution is transferred into the MC N_{charged} distributions before and after the HadronBJ event selection by reweighting. This allows to extract the change on approximate efficiencies due to differences in distribution shape between MC and experimental data.

All bin contents in the MC N_{charged} distributions after HadronBJ are reweighted with the ratios shown in Figure 4.72. It is assumed that the HadronBJ event selection changes the N_{charged} the same way for experimental as well as for MC data for a given N_{charged}, without limiting generality towards effects of distribution shape differences. Under this assumption, all bin contents of the MC distributions before HadronBJ for $N_{\text{charged}} \geq 3$ are reweighted with the ratios from Figure 4.72 as well. For bins $N_{\text{charged}} < 3$, no weights with input from experimental data are known since no experimental distributions before HadronBJ are available. Estimates of how the difference in distribution shape could influence approximate efficiencies by changing bins $N_{\text{charged}} < 3$ in the MC distributions before HadronBJ are evaluated in two ways. First, these bins are left unweighted (a). In another case, these bins are reweighted with the average of all weights for $N_{\text{charged}} \geq 3$ (b).

Approximate efficiencies are calculated from reweighted distributions after HadronBJ and from reweighted distributions a) and b) before HadronBJ. The resulting efficiencies are plotted in Figure 4.73 along with un-

reweighted approximate efficiencies.

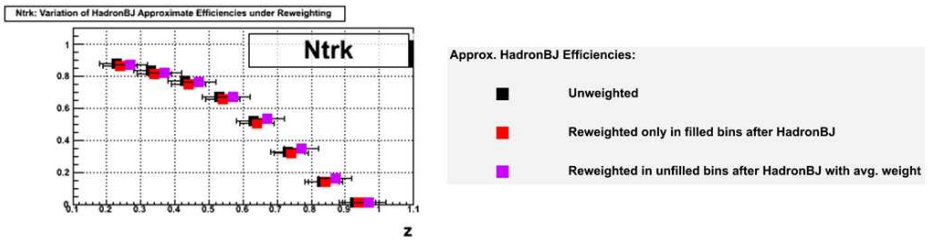


Figure 4.73: Approximate HadronBJ efficiencies from MC N_{charged} distributions unreweighted (black boxes), reweighted for $N_{\text{charged}} \geq 3$ (case (a) in text) and reweighted for all N_{charged} (case (b) in text) versus z.

The different approximate efficiencies are found to only vary at the few-percent level between the reweighted and unweighted cases. It is concluded that differences in the shape of event shape variable distributions between MC and experimental data only cause a small variation in the values of the actual HadronBJ efficiencies. The largest difference of any reweighted to the unweighted approximate efficiencies are assigned as additional systematic uncertainty on the actual HadronBJ efficiencies.

4.7 Correction for ToF θ Acceptance Cut

The last remaining cut on track kinematics given in Section 3 is the selection that all tracks in the measurement sample are required to pass through the ToF acceptance in $\theta_{\rm lab}$. The measurement result should not include any restriction in the scattering angle. Thus a correction is made to extend the measurement sample to full 4π scattering angle acceptance.

4.7.1 Correction Factor Extraction

To extract correction factors in this correction, an on-resonance standard Belle Monte Carlo sample of about $1.2 * 10^9$ events $e^+e^- \rightarrow q\bar{q}$ where $q = \{u, d, s, c\}$ from Experiment 55 from five MC streams has been analyzed. Ratios of particle yields with and without ToF acceptance requirements are extracted in each z bin. Again only particles directly originating from events $e^+e^- \rightarrow q\bar{q}$ or decay products of such particles are considered for the ratios. Acceptance ratios for negative particle charge are given in Figure 4.74.

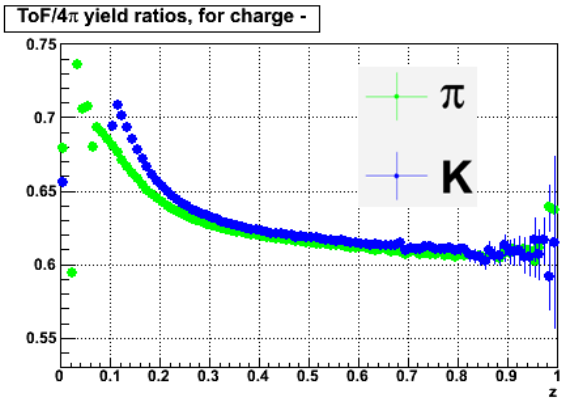


Figure 4.74: Acceptance ratios of particle yields within HadronBJ data skim efficiencies for negative particle charge.

The z dependence of the ratios shown in Figure 4.74 can be motivated as follows: Particles which are formed in the fragmentation process with high z are expected to carry low transverse momentum with respect to the quark-antiquark axis. Thus the scattering angle dependence of such particles is close to the theoretical angular distributions of the partons from e^+e^- annihilation, $1 + \cos^2\theta_{\rm cms}$. Ratios of particle yields versus z are therefore expected to asymptotically approach the ratio calculated from a $1 + \cos^2\theta_{\rm cms}$ distribution for high particle z. This trend is visible in Figure 4.74.

On the other hand, particles with low z are expected to show higher transverse momentum around the quark-antiquark axis than high z particles. Thus low z particles follow an angular distribution which is flatter than the quark-antiquark distribution, producing yield ratios larger than the ones for high z particles. Also

this behavior can be confirmed in Figure 4.74. Differences between ratios for pions and kaons are assumed to originate from mass effects.

4.7.2 Analysis of Monte Carlo-Experimental Data Correspondence

The basic assumption in this correction is that for Monte Carlo as well as experimental data, the ratios of particle yields with and without ToF acceptance requirements are the same. This also assumes that the Monte Carlo description of angular distributions in fragmentation and hadron production is sufficiently precise. To test this assumption, correspondence between Monte Carlo and experimental data has been studied.

In order to analyze the correspondence, distributions of $\cos\theta_{\rm cms}$ have been extracted for 40 bins in z from raw experimental data, Experiment 7 through 55. The experimental data sample contains about $2.7 * 10^9$ events from on-resonance as well as continuum energy levels. To assure good tracking and PID selection, only particles in the barrel part of the Belle detector are accepted in this study. Acceptance ratios between the ToF and 4π acceptance are extracted by fitting all distributions of $\cos\theta_{\rm cms}$ with a fit function containing the quark-antiquark distribution terms $1 + \cos^2\theta_{\rm cms}$ in addition to a linear term $\cos\theta_{\rm cms}$ accounting for γ/Z interference. The scattering angle distributions are fitted for $\cos\theta_{\rm cms}$ within $[-0.6; 0.6)$. Finally, acceptance ratios are formed by dividing the integral of the fitted functions over the ToF acceptance by the integral over 4π. In order to compare the extracted yield ratios to Monte Carlo, the same fit analysis is performed also on the Monte Carlo analysis sample described above. Acceptance ratios from fits have been extracted from Monte Carlo data for each z bin.

Figures 4.75 shows quality plots and sample fits of distributions $\cos\theta_{\rm cms}$. Figures 4.75 a) and b) display reduced χ^2 values for each fit versus z, for MC (a) as well as raw experimental data (b). For Monte Carlo data, the fits achieve high agreement. In the experimental data sample, systematic corrections like PID or smearing corrections are not performed yet- the distributions fitted in this Section are extracted from raw experimental data. Thus the agreement between the histograms and the fits is expected to be not as high as for Monte Carlo data. Figures 4.75 c) and d) show sample fits of distributions $\cos\theta_{\rm cms}$ for Monte Carlo and experimental data. In order to not include events $e^+e^- \to B\bar{B}$ in the ratios, experimental on-resonance data are only used for $z > 0.6$.

Acceptance ratios versus z extracted by fitting MC and raw experimental data scattering angle distributions are shown in Figures 4.76. Judging from the plots, sufficient agreement exists between acceptance ratios from MC as well as experimental data. The deviation in the kaon plot for $z <\sim 0.3$ is assumed to originate from systematic effects which are not corrected for yet in the scattering angle distributions.

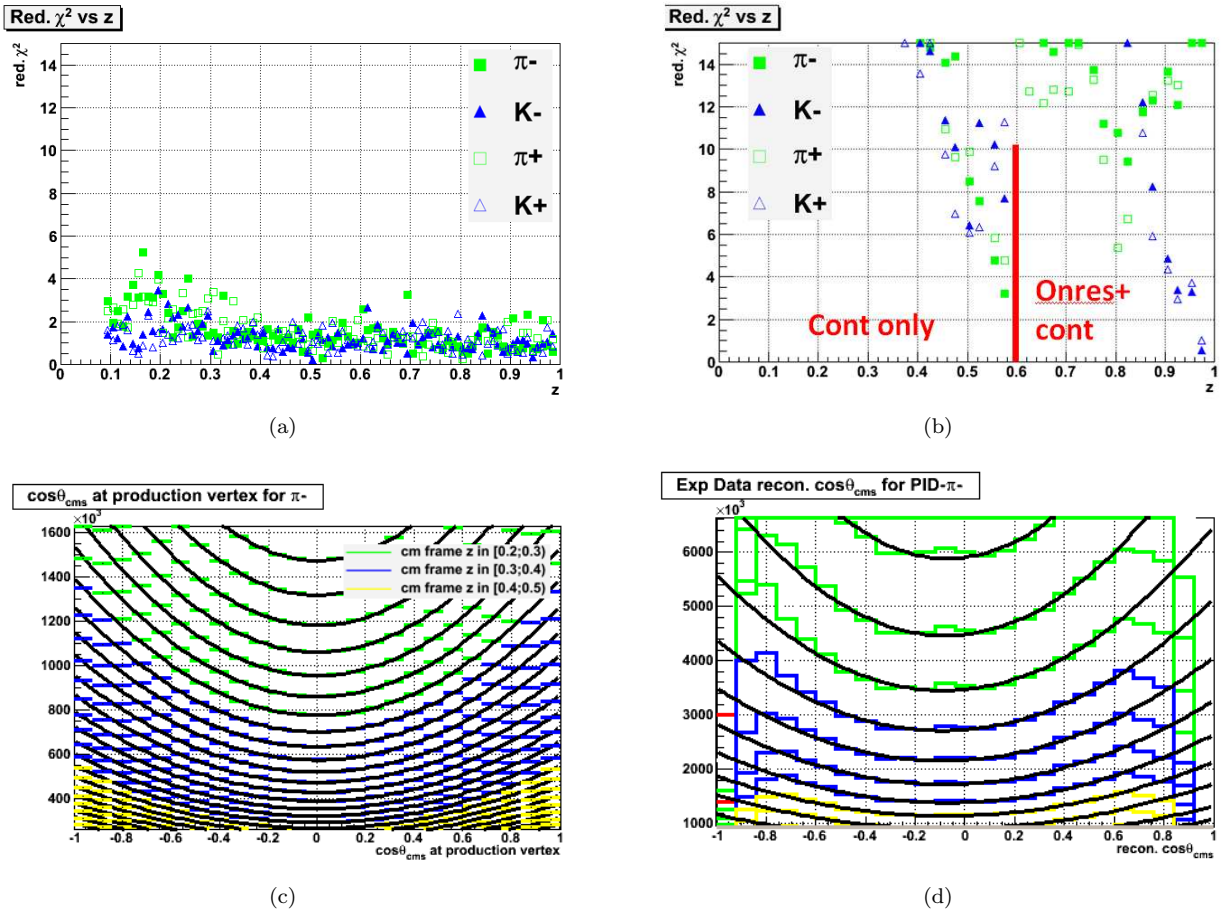


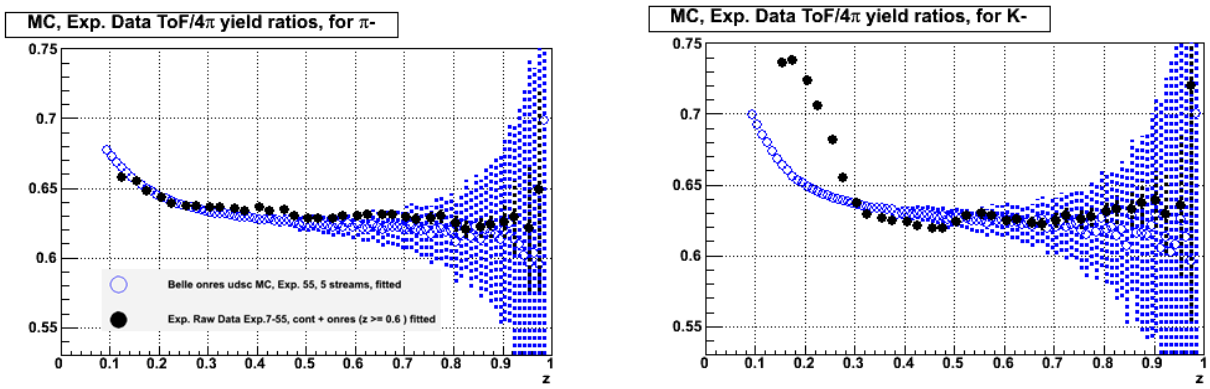
Figure 4.75: Reduced χ^2 plots versus z for fits of $\cos\theta_{\mathrm{cms}}$ from MC data (a) and raw experimental data (b), both for 11 degrees of freedom. Figures c) and d) show sample fits of distributions $\cos\theta_{\mathrm{cms}}$ for MC data (c) and raw experimental data (d).

Since the ratios in Figures 4.76 are sufficiently close, no additional systematic uncertainty is assigned on the acceptance ratios shown in Figure 4.74.

4.7.3 Correction of Measurement Samples

Measurement samples are corrected bin-by-bin with the inverse values of acceptance ratios shown in Figure 4.74. Uncertainties are propagated according to Equations 3.2 and 3.3. Also in this correction, relative systematic uncertainties of the measurement samples increase by less than one per cent for highest z bins.

Corrected measurement samples for negative particle charge are shown in Figure 4.77.



(a) (b)

Figure 4.76: Acceptance ratios versus z, extracted from MC and raw experimental data by fitting and integrating $\cos\theta_{\rm cms}$ distributions, for negatively charged pions (a) and kaons (b).

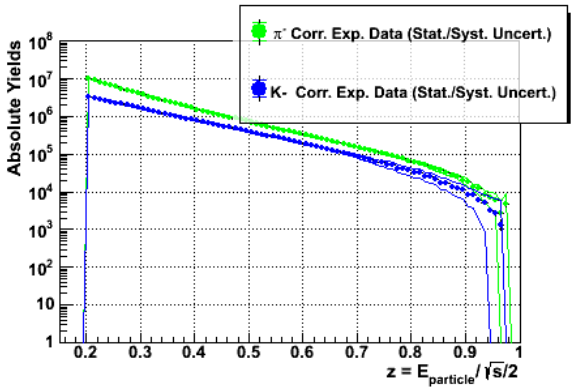


Figure 4.77: Pion and kaon measurement samples for negative particle charge, corrected for the ToF scattering angle acceptance cut.

4.8 Correction for Initial/Final State Radiation Effects

As the last correction in this analysis, initial state radiation (ISR) and final state radiation (FSR) effects are studied.

Initial state radiation events are characterized by one of the initial state leptons emitting a photon with significant energy. Then the center-of-mass energy available for the quark-antiquark pair creation in reactions $e^+e^- \rightarrow q\bar{q}$ is less than the nominal amount. Similarly, a final state quark can radiate a 'hard' photon (relative to the annihilation energy scale) before fragmentation. In this case, the created hadrons originate from a quark with less energy than the nominal quark energy of half of Q as well. Neither (electromagnetic) event is accounted for in the current next-to-leading order QCD theory analyses of experimentally measured hadron multiplicities. Thus the impact of hadrons originating from events with hard quantum electrodynamic effects on measured yields has to be investigated.

4.8.1 Correction Concept and Methods

The occurrence of ISR/FSR changes the fragmentation energy scale of hadronic events. Thus the experimentally measured yields contain contributions from fragmentation processes at several different center-of-mass energies, among which also the nominal Belle center-of-mass energy. These contributions are connected to each other via the DGLAP evolution equations. Therefore a simple bin-by-bin correction for ISR/FSR effects with correction factors from Monte Carlo cannot be performed. On the other hand, an analytical correction for these effects would require more elaborate theoretical procedures than feasible in this analysis. It was decided to limit the hadron contributions from ISR/FSR events and assign the effects of remaining contributions as systematic uncertainty.

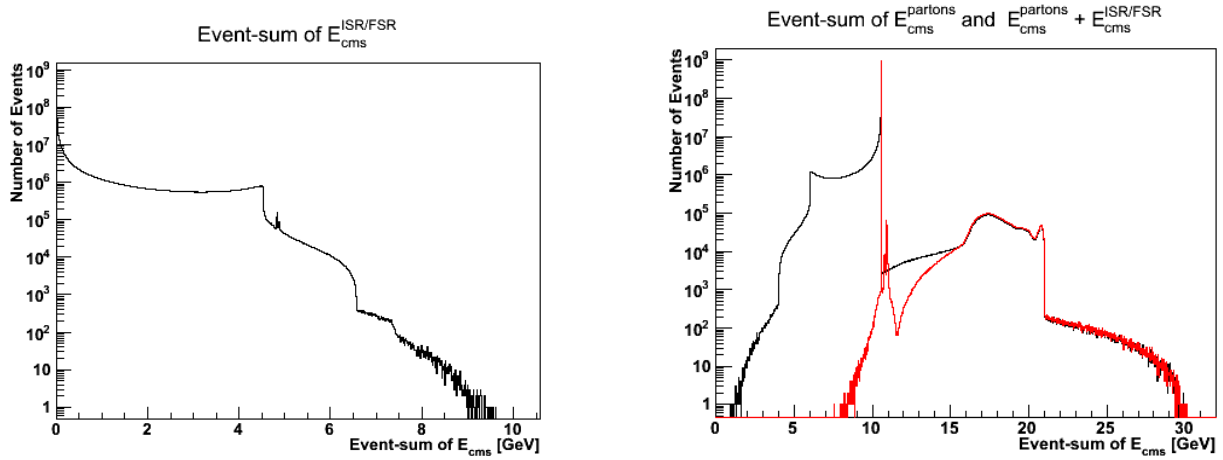
In the Belle Monte Carlo 'gen_hepevt' data table, both ISR and FSR photons are stored as 'decay products' of the virtual photon in each event.[13] Thus contributions from ISR and FSR can be studied and accounted for simultaneously.

A Monte Carlo data sample containing about $1.4 * 10^9$ events $e^+e^- \rightarrow q\bar{q}$ where $q = \{u, d, s, c\}$ from 6 streams of Experiment 55, on-resonance energy level data is analyzed to study ISR/FSR effects. The differences in the correction factors resulting from the energy difference to the measurement sample on the continuum level is assumed to be negligible. For illustration, Figure 4.78 a) shows the distribution of summed ISR/FSR photon energies per event from the above sample. Figure 4.78 b) displays the distribution of summed parton energies per event and the distribution of the sum of parton energies added with all ISR/FSR photon energies per event. The tails below 10.52 GeV in Figure b) (black line) are consistent with ISR events, as they disappear if, event by event, the summed ISR photon energies are added to the summed parton energies (red line). Tails to center-of-mass energies higher than 10.52 GeV in Figure b) originate from multiple generations of fragmenting partons being stored in the Belle MC data tables. Since such events represent only about one per cent of all events analyzed and are assumed to be physically correct fragmentation events, the tails to higher energies in Figure b) are ignored.

Limiting Hadron Contributions from ISR/FSR Events- Exclusion Fractions

The relative precision of next-to-leading order QCD DGLAP evolutions between different center-of-mass energies is estimated to be about 5% of the respective Q [88]. This analysis should provide multiplicities with a range of Q sufficiently smaller than this limit. Thus it was decided to exclude hadron contributions from fragmentation events where ISR/FSR photons carry more than 0.5% of the nominal center-of-mass energy (about 53 MeV).

[13] The connection between photons stored as virtual photon decay products and ISR/FSR photons has been verified on a self-generated sample of $5 * 10^6$ events $e^+e^- \rightarrow q\bar{q}$ with $q = \{u, d, s, c\}$.



(a) (b)

Figure 4.78: Distributions of summed ISR/FSR photon energies (a), and summed parton energies together with summed parton added with ISR/FSR photon energies (b), both per event.

Figure 4.79 shows fractions of negatively charged pions and kaons generated in events where the sum of all ISR/FSR photons is more than 0.5% of the nominal center-of-mass energy, versus z. The plots are extracted from the above-mentioned on-resonance Experiment 55 Monte Carlo data sample. The x-axis shows the 'nominal' values of z, so the hadron center-of-mass energy normalized to $Q/2 = 5.29$ GeV. It is less likely that particles with high nominal z are produced in events whose center-of-mass energy is lower than the nominal Q. Thus the particle fractions are seen to decrease versus z. The difference between the pion and kaon curve can most likely be related to mass effects.

The fractions shown in Figure 4.79 represent the fractions of particles in each z bin which are excluded from the analysis to limit ISR/FSR impact on the measured yields. The assumption in this analysis step is that the fraction of particles from events with hard QED effects is the same in Monte Carlo as in experimental data.

Systematic Uncertainty on Exclusion Fractions due to Monte Carlo Parameterizations

The exclusion fractions shown in Figure 4.79 are extracted from standard Belle Monte Carlo data, generated with the Jetset/Pythia event generator. This generator features a set of variable parameters which can be specified at initialization to improve the description of experimental data. These parameters allow to change how processes like ISR/FSR, fragmentation and others are modeled. Thus it is necessary to investigate the dependence of the extracted exclusion factors on the chosen set of parameters in the Jetset/Pythia event generator.

In order to access and quantify this dependence, twelve different sets of generator parameter sets/'tunes'

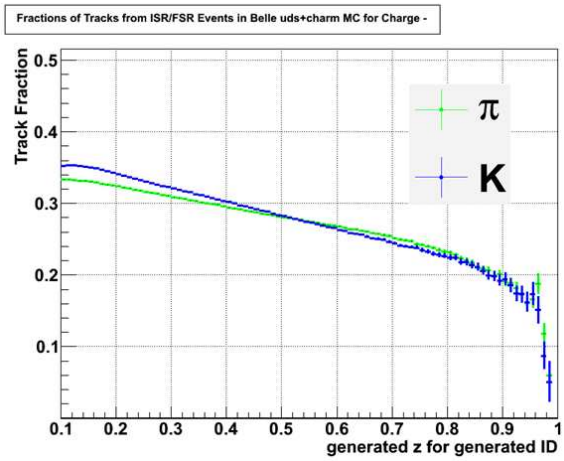


Figure 4.79: Belle Monte Carlo fractions of particles from events with ISR/FSR photon energy of more than 0.5% of Q, for negative particle charge versus nominal z.

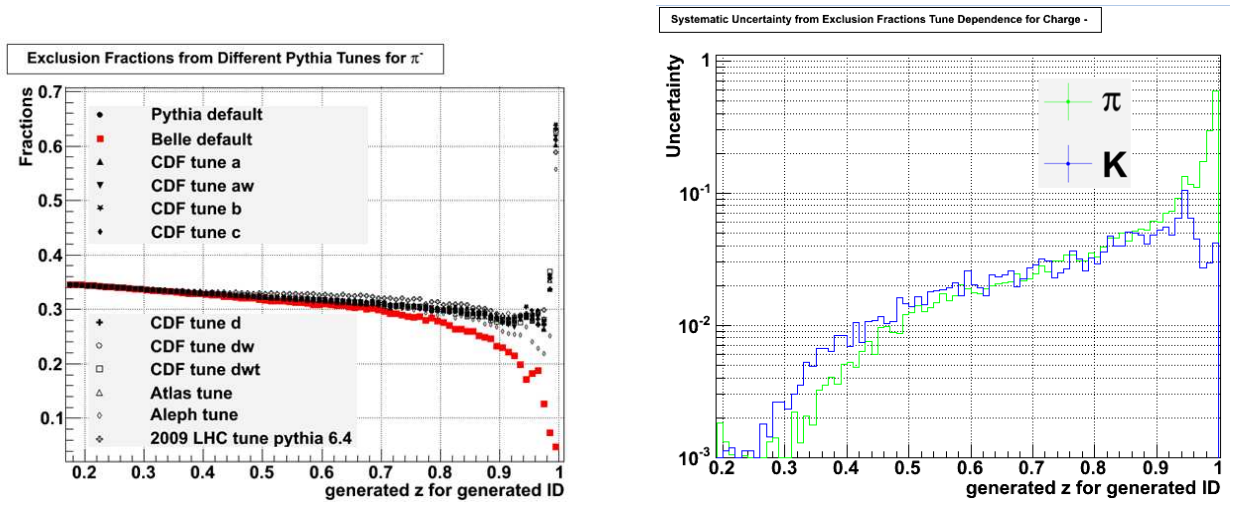
are collected. These are the Jetset/Pythia default tune, the Belle tune and other tunes from the CDF, Aleph and Atlas experiments. For each parameter set, $100 * 10^6$ events $e^+e^- \to q\overline{q}$ where $q = \{u, d, s, c\}$ are created at $Q = 10.6$ GeV and exclusion fractions are calculated. In this study pions, kaons, lambda-baryons etc. are considered stable - differences between Belle exclusion fractions extracted here to the ones shown in Figure 4.79 are assumed to be negligible, as is the effect of the difference in center-of-mass energy by less than 1%. Figure 4.80 a) displays exclusion fractions from all twelve Jetset/Pythia tunes, analogously to the ones shown in Figure 4.79, for negatively charged pions.

To quantify the dependence of exclusion fractions on Monte Carlo parameter choices, the largest difference in each z bin between exclusion fractions from Belle and all other tunes is calculated. These differences are assigned as systematic uncertainty to the exclusion fractions in each bin. The two curves in Figure 4.80 b) show these uncertainties for negatively charged pions and kaons.

Systematic Uncertainty on Yields due to ISR-induced Kinematic Smearing

After excluding hadron contributions from events with hard QED processes, the remaining moderate ISR/FSR effects on measured yields have to be accounted for. It turned out to be difficult to access the remaining effects from FSR events. However it will be shown in Section 4.8.1 that FSR effects after the above exclusion corrections are small and fall below the systematic uncertainties assigned in this Section. Thus this Section focuses on effects on measured yields from ISR only.

The presence of ISR affects the calculation of the normalized hadron energy z in two ways. First, the normalization factor $1/(Q/2)$ overestimates the energy of the fragmenting partons, as mentioned at the beginning of this Chapter. Second, the kinematic variables of each particle are stored in the Belle datatables



(a) (b)

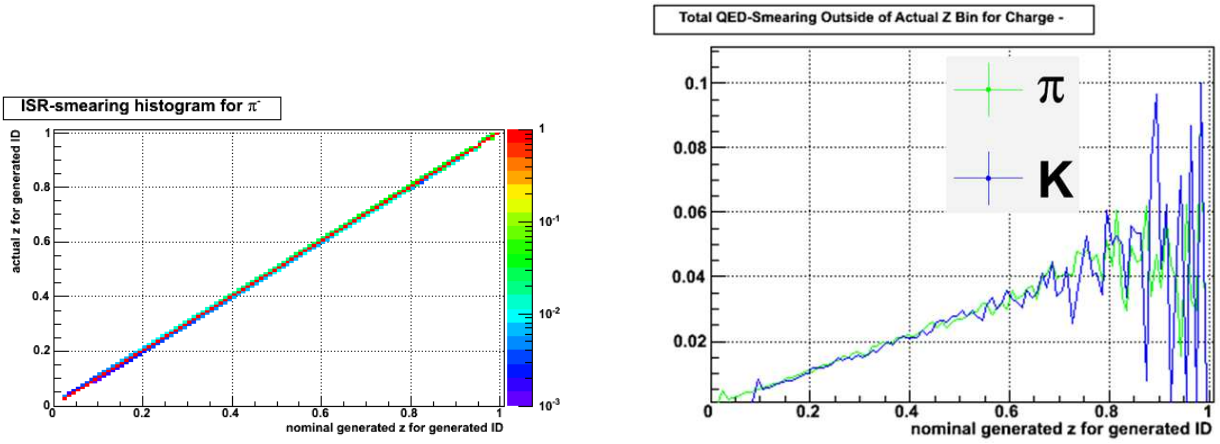
Figure 4.80: Exclusion fractions for particles from events with more than 0.5% of Q loss to ISR/FSR photons, for twelve Monte Carlo tunes (a); maximum difference in exclusion fractions between Belle Monte Carlo and all other tunes (b). Both plots are shown versus nominal z. Uncertainties are not shown for better visibility.

in the laboratory reference frame. Since z is only defined in the e^+e^- center-of-mass frame, all variables have to be boosted. In order to account for ISR, the appropriate e^+e^- center-of-mass frame has to be formed with lepton four-momentum vectors after the last ISR photon has been emitted. For accessing the four-momentum vectors of the annihilating leptons, a sample of $5 * 10^6$ events $e^+e^- \rightarrow q\bar{q}$ with $q = \{u, d, s, c\}$ was generated where more detailed event histories than in the usual Belle Monte Carlo are stored. A general Lorentz transformation is used to connect the ISR-corrected center-of-mass frame to the laboratory frame.

A similar description of the effects of ISR on measured yields can be applied as was used in Chapter 4.3 to study kinematic smearing. Figure 4.81 a) shows an ISR-smearing histogram for negatively charged pions created in events with energy losses to ISR photons of less than 0.5% of Q. The histogram contains pairs of ISR-uncorrected/'nominal' (x-axis) and ISR-corrected/'actual' (y-axis) values of z. The bin yields are normalized by the sum of all particles found in each respective nominal z bin. The histogram is extracted from the self-generated sample described above.

As mentioned in the previous Section, the histogram in Figure 4.81 a) has particle contributions from events with different energy levels Q. This does not allow to use the histogram as a correction matrix along the techniques employed in the PID and kinematic smearing corrections (cf. Chapters 4.1 and 4.3). However, the fraction of particles in a given nominal z bin whose actual z values lie in different z bins can be calculated. To account for kinematic smearing induced by hard QED effects, it was decided to add this fraction in quadrature to the yield systematic uncertainties after the ISR/FSR analysis step. Figure 4.81 b)

shows plots of these fractions versus nominal z for negatively charged pions and kaons.



(a) (b)

Figure 4.81: ISR-induced smearing histogram for negatively charged pions (a), and fraction of particles per z bin whose ISR-corrected/actual z lies in a different bin than their ISR-uncorrected/nominal z, versus nominal z for negatively charged pions and kaons (b). For (b), uncertainties are not shown for better visibility. Both plots are extracted from a sample of events which contain less than 0.5% of Q in ISR photons.

Illustration of Final State Radiation Contributions

As mentioned above, the way to calculate exclusion fractions described in Section 4.8.1 collects contributions from ISR and FSR at the same time. However, it is difficult to access FSR information to assess the impact of remaining FSR contributions on measured yields after exclusion of hadrons from hard ISR/FSR events. Thus the size of remaining FSR contributions is estimated from the relative size between ISR and FSR effects.

In the standard Belle Monte Carlo, FSR and ISR effects cannot be separated due to the way of how information is stored in the data tables. Nevertheless, the standard Belle Jetset/Pythia tune can be employed to study FSR by itself. The standard Belle tune is modified to suppress the emission of hard photons in the initial state. For this modified tune, $100 * 10^6$ events are generated. Again in this study, pions, kaons, lambda-baryons etc. are considered stable, and no impact of this assumption on extracted quantities is expected. Exclusion fractions analogously to the procedure above are calculated for particles created in events with more than 0.5% of Q in FSR photons. These fractions are shown in Figure 4.82 versus nominal z.

From Figure 4.82, it can be seen that the fraction of hadrons originating from events with hard FSR photons remains below 1% in all z bins. Comparing to the combined ISR/FSR exclusion fractions shown in Figure 4.79, FSR fractions are about a factor of 30 lower. Thus it is assumed that after excluding hadron

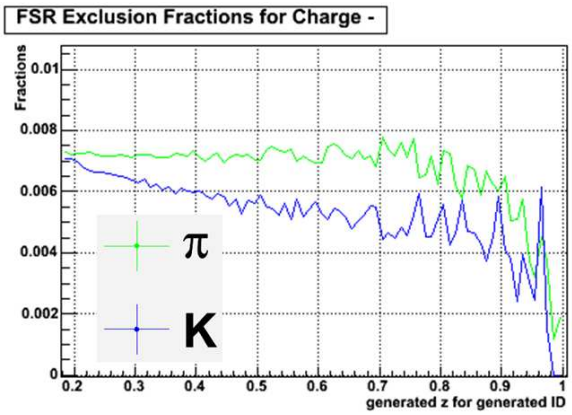


Figure 4.82: Fractions of pions and kaons from events with FSR energy of more than 0.5% of Q, for negative particle charge versus nominal z. Uncertainties are not shown for better visibility.

contributions from hard ISR/FSR events, remaining FSR effects in the measured yields are negligible.

4.8.2 Correction of Measurement Samples

The exclusion of hadron contributions from events with ISR/FSR energy losses larger than 0.5% of Q is performed by multiplying each bin with the above-shown exclusion factors, cf. Figure 4.79. Systematic uncertainties are assigned due to the dependence of the exclusion fractions on the specific choice of Monte Carlo parameter sets. An additional uncertainty is assigned to the final bin yields due to remaining ISR effect. As in other corrections, all uncertainties are propagated according to Equations 3.2 and 3.3. wp Relative systematic uncertainties of the measurement samples increase by about one per cent for highest z bins.

Figure 4.83 shows corrected measurement samples for negative particle charge.

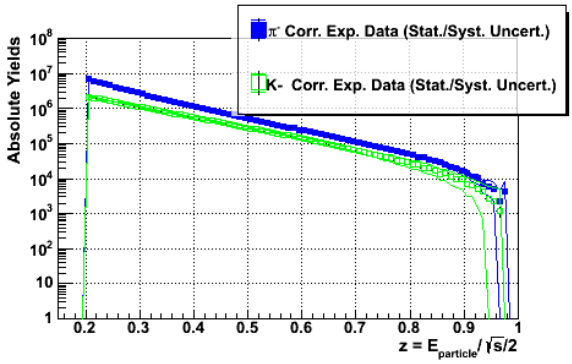


Figure 4.83: Pion and kaon yields for negative particle charge after the correction for hadron contributions from events with hard ISR/FSR processes.

As an effect of excluding hadrons from events with hard ISR/FSR processes, the normalization of mea-

sured yields by the number of analyzed events has to be adjusted as well. The fraction of events with summed ISR/FSR photon energies less than 0.5% of Q is $64.616 \pm 0.003\%$. This fraction will be accounted for in Chapter 4.9.

4.9 Normalization

After all corrections, the experimentally measured pion and kaon yields have to be normalized in order to obtain multiplicities. In global fragmentation function extractions, experimental datasets are fitted with an additional normalization parameter, such that the choice of normalization is, to a certain extent, arbitrary.

To obtain pion and kaon multiplicities, the final measurement yields need to be normalized by the total hadronic cross section $e^+e^- \rightarrow q\bar{q} \rightarrow h + X$ where $q = \{u, d, s, c\}$. Also the exclusion of certain events due to significant initial-state-radiation, as described in the previous Section, has to be accounted for.

Hadronic Cross Section It is possible to measure the total hadronic cross section from Belle data. However this would represent a separate analysis beyond the scope of this measurement. Therefore, the cross section is obtained from a measurement of the 'R' fraction, which is described in the following.

The R fraction is defined as the ratio between the total hadronic cross section divided by the total muon production cross section [5],

$$R(Q) \equiv \frac{\sigma(e^+e^- \rightarrow q\bar{q} \rightarrow h + X, Q^2)}{\sigma(e^+e^- \rightarrow \mu^+\mu^-, Q^2)}. \tag{4.25}$$

The R fraction receives experimental as well as theoretical attention as it allows for direct tests of QCD. At leading order QCD, it is equal to three times the sum of the quark charges squared of all flavors of quarks which can be produced at the studied center-of-mass energy. For the presented analysis, the total hadronic cross section is extracted from the R fraction measured by the CLEO collaboration [89], $R = 3.56 \pm 0.01 \pm 0.07$, at $Q = 10.52$ GeV, which is equal to the energy of the measured data in this analysis. To calculate the muon pair production cross section at $Q = 10.52$ GeV, the usual expression of $\sigma(e^+e^- \rightarrow \mu^+\mu^-)$ [90],

$$\sigma(e^+e^- \rightarrow \mu^+\mu^-, Q^2) = \frac{4\pi\alpha_{\mathrm{EM}}^2}{3s}, \tag{4.26}$$

is used, yielding $\sigma(e^+e^- \rightarrow \mu^+\mu^-, Q = 10.52 \text{ GeV}) = 0.7845$ nb with negligible uncertainties. Substituting this result back into Equation 4.25 together with the CLEO measurement yields a total hadronic cross section of

$$\sigma(e^+e^- \rightarrow q\bar{q} \rightarrow h + X, Q^2 = (10.52 \text{ GeV})^2) = 2.793 \pm 0.056. \tag{4.27}$$

Integrated Luminosity of Analyzed Data As described in Section 3, the statistics analyzed in this measurement include all events from Experiments $7 - 55$ HadronBJ skim data for continuum energy level.

For 'caseA' reconstructed data, the integrated luminosity is given by 68.03 fb^{-1} [57]. A relative uncertainty of 1.4% is quoted for the luminosity, cf. Ref. [57].

Normalization Factor Combining the numbers for the total hadronic cross section and the integrated luminosity of the measurement sample data yields a total number of events of $(190.0 \pm 4.6) * 10^6$ events. In the previous Section, about 35.4% of all events showed sums of ISR photon energies of more than 0.5% different to the nominal center-of-mass energy. Excluding those events also from the normalization leaves $(122.8 \pm 3.0) * 10^6$ events for the normalization factor. The content of each z bin in all measurement sample yields is normalized by the above-calculated number of events.

To transform the measured bin yields on finite binning in z into multiplicities differential in z, the normalized yields are scaled by the inverse bin width of $\Delta z = 0.01$.

Chapter 5

Final Results and Discussion

5.1 Final Measurement Results

5.1.1 Final Charge-Integrated Pion and Kaon Multiplicities

Over the course of the analysis up to this point, yields for pions and kaons resolving the particle charge are extracted. This provides an additional consistency check for the analysis and correction procedures via the calculation of ratios for the charge-dependent yields, as discussed below. To obtain final multiplicities, the charge-separated multiplicities are added. All uncertainties are propagated, yielding a final uncertainty for the sum. Final normalized charge-integrated pion and kaon multiplicities are shown in Figure 5.1.

5.1.2 Relative Uncertainties Contributions by Correction

Figures 5.2 a) and b) show relative statistical and upper systematic uncertainties for the measured charge-integrated pion and kaon multiplicities. Increases in the systematic uncertainties by the contribution of each correction step are shown. Due to the uncertainty determination in the PID and smearing corrections, asymmetric uncertainties are given throughout the analysis. However, the lower systematic uncertainties are of similar magnitude compared to the upper uncertainties.

For both pions and kaons, at low z the uncertainties from several corrections contribute in similar amounts to the total uncertainty. At mid-z, uncertainties from the ISR/FSR and the joint systematic effects corrections are dominant for both hadron species. At maximum z, the uncertainties from the smearing correction dominate the total systematic uncertainties, followed by uncertainties for the HadronBJ event selection correction for pions and the PID correction for kaons. Considering Figures 5.2 a) and b), it can be concluded that the precision of the present measurement is limited by systematic uncertainties.

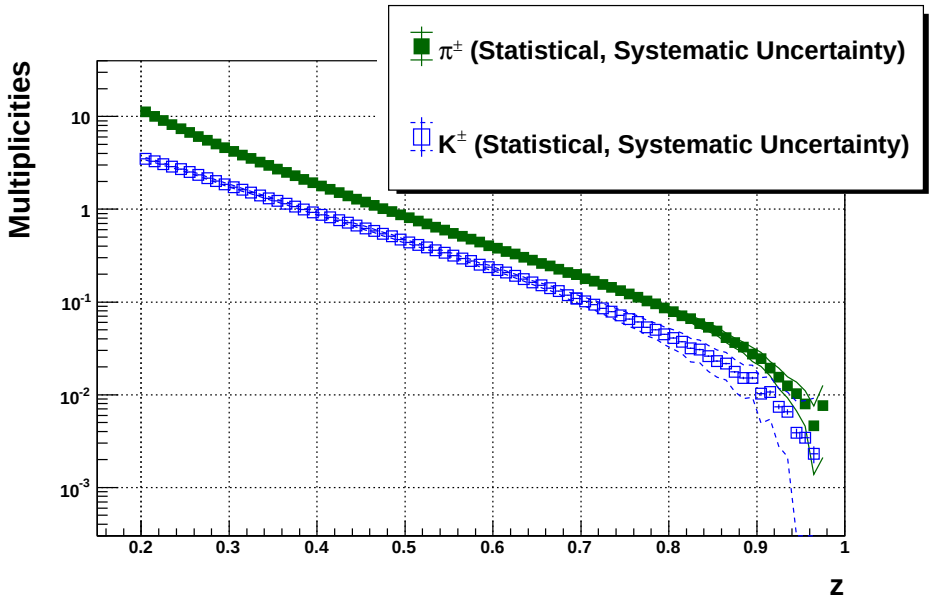
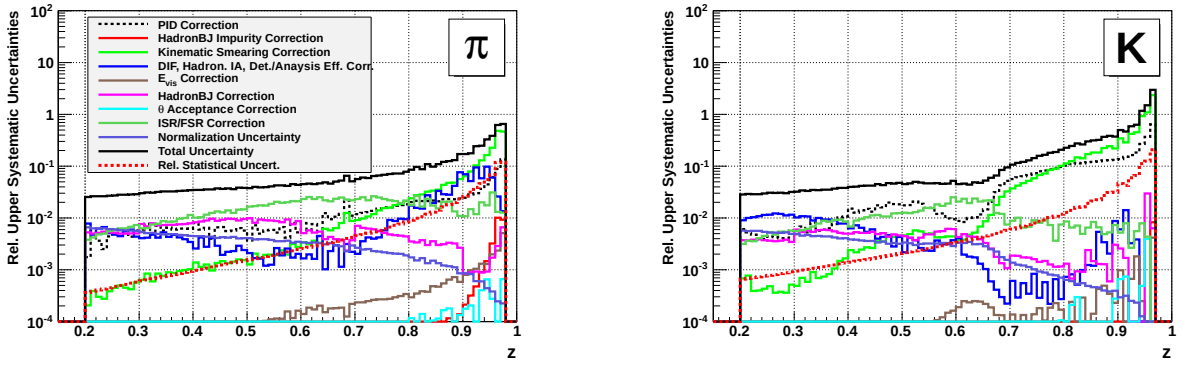


Figure 5.1: Final multiplicities for charge-integrated pions and kaons with full statistical and systematic uncertainties as a function of the fractional hadron energy z.



(a) (b)

Figure 5.2: Relative uncertainties for the final charge-integrated pion (a) and kaon (b) multiplicities as a function of the fractional hadron energy z.

5.1.3 Charge Ratios

As stated above, keeping yields for different particle charges separated during all correction steps allows to confirm validity and consistency of these correction procedures. CPT symmetry requires the charge ratio of particle multiplicities to be equal to 1.0 exactly.

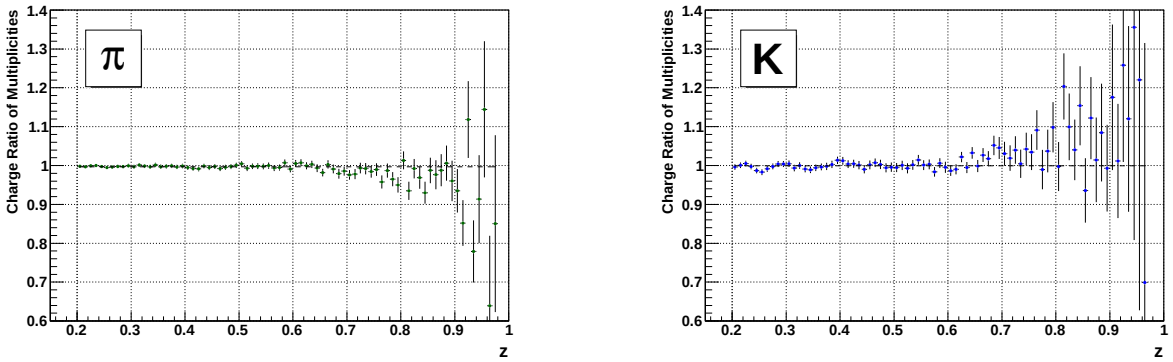
In Figures 5.3 charge ratios of final pion and kaon multiplicities are shown, where multiplicities for negative particle charge are divided by multiplicities for positive particle charge in dependence of z. The

ratios are fitted with constant fit functions on the interval $[0.2; 0.98)$ (π) and $[0.2; 0.97)$ (K), respectively.

The degree of correlation of the systematic multiplicity uncertainties between the two particle charges was not determined. In order to estimate the uncertainty correlation, systematic uncertainties in each z bin are scaled for each multiplicity by global factors λ before quadratic summation with statistical uncertainties and propagation through the charge ratio. The values for λ achieving reduced χ^2 values (χ^2/NDF) of 1.0 to 5% precision are 0.08 (π) and 0.12 (K), respectively. This indicates that systematic uncertainties between multiplicities for negative and positive particle charge are largely correlated.

The results for the constant fit parameters for pions and kaons, using the above-mentioned values for the correlation scale factors λ, are 0.9969 ± 0.0007 (π) and 0.9993 ± 0.0012 (K), respectively. While the kaon ratios are consistent with 1.0 within uncertainties, the pion ratios show a significant (but small) difference to 1.0 of about 3 per mille. Since typical uncertainties on the pion multiplicities are of the order of percent and larger, no additional systematic uncertainties to account for the slight charge asymmetry are assigned.

All fit results are summarized in Table 5.1. The good agreement of the charge ratios with 1.0 considering the typical precision of the analysis per bin is interpreted as a confirmation of the consistency of the presented measurement.



(a) (b)

Figure 5.3: Charge ratios of final multiplicities, negative divided by positive particle charge multiplicities, for pions (a) and kaons (b). The fit result of fits with constant fit functions can be seen as dashed lines near 1.0.

Species	**Syst. Unc. Scale Factor**	**Constant Result**	χ^2**/NDF**
π	0.08	0.9969 ± 0.0007	$80.3/77 = 1.04$
K	0.12	0.9993 ± 0.0012	$79.0/76 = 1.04$

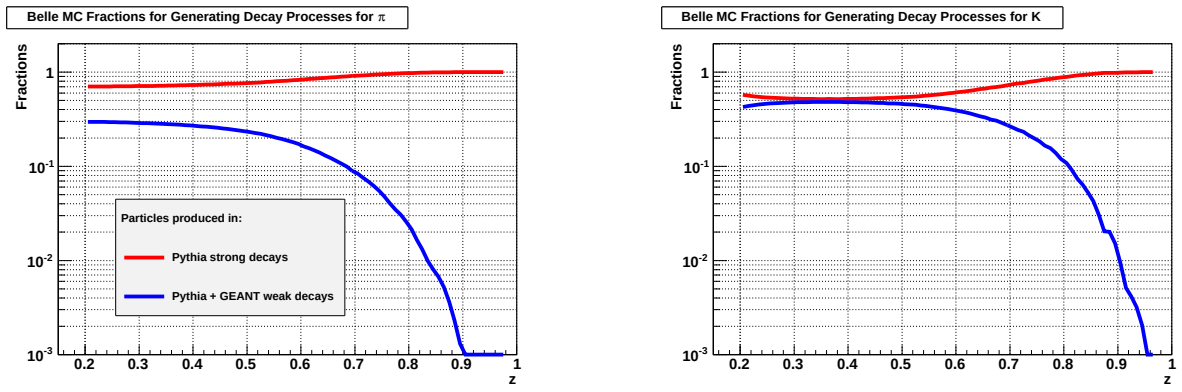
Table 5.1: Fit results for charge ratio constant fits.

5.2 Analysis of Hadrons from Weak and Strong Decays

The general QCD framework used by global analyses usually does not account for fragmentation hadrons created in weak decays. However, these are present in the measured multiplicities as decay products of τ, kaon, Λ and heavier meson and baryon decays. It is considered difficult to extract correction factors from experimental data to remove these pions and kaons from the measured multiplicities. Therefore all weakly produced pions and kaons are kept in the present measurement.

However, we provide fractions of fragmentation pions and kaons originating from strong and weak decays, cf. Ref. [91]. These fractions are extracted from generic Belle MC multiplicities with a treatment of ISR/FSR analogous to the ISR/FSR corrections for the measured multiplicities, for a similar number of events and collision energy as in the experimental dataset.

an analogous ISR/FSR treatment as the measured experimental multiplicities and from the same amount of statistics and energy level. Figures 5.4 a) and b) show fractions of charge-integrated fragmentation pions (a) and kaons (b) produced in strong and weak decays, relative to the sum of the two fractions. It should be pointed out that the sum of the fractions equals the total amount of fragmentation pions and kaons, respectively.



(a) (b)

Figure 5.4: Fractions of charge-integrated pions (a) and kaons (b) produced in strong and weak decays, relative to the sum of the fractions. All data is extracted from a Monte Carlo dataset comparable to the experimental measurement sample. For better visibility, no uncertainties are shown.

Pions and kaons from weak decay processes are seen to drop below 10% above energies z of 0.7 (pions) and 0.8 (kaons). Pions from weak decays represent about 20% of all fragmentation pions at $z < 0.4$, whereas almost half of all fragmentation kaons below $z = 0.6$ originate from weak decays.

5.3 Comparison to Monte Carlo Pythia/Jetset Multiplicities

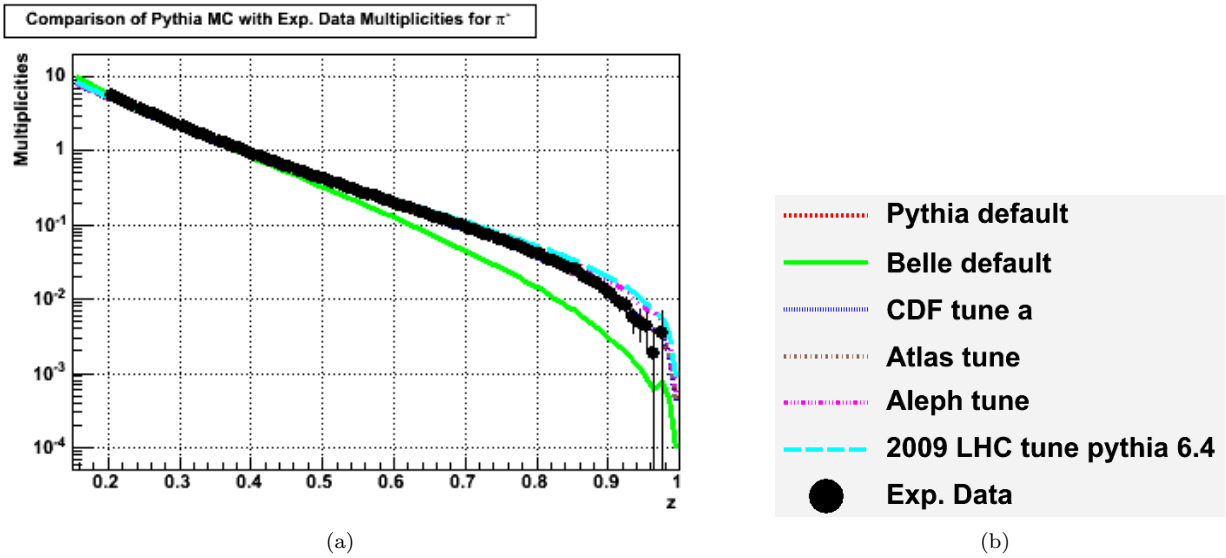
Monte Carlo event generators are used for systematic studies in this measurement and in particle physics experiments in general. However, since fragmentation functions cannot be calculated from first principles, the z dependence of hadron multiplicities in MC event generators is determined by the choice of parameters for fragmentation models employed in the simulation. This Chapter studies the relation between the results of the present measurement and hadron multiplicities obtained from the Pythia/Jetset Monte Carlo generator.

5.3.1 Monte Carlo Comparison Samples

The experimental results are compared to results obtained from a subset of the event generator parameter sets introduced and used in Section 4.8.1. These sets include the generic Belle tune and tunes from the CDF, Aleph and Aleph experiments and finally a general LHC tune. For each of these tunes, a sample of $100 * 10^6$ events $e^+e^- \to q\overline{q}$ where $q = \{u, d, s, c\}$ is created at $\sqrt{s} = 10.6$ GeV, and ISR/FSR contributions are limited analogously to the procedures applied in the experimental measurement, cf. Section 4.8.1. The high precision of the experimental results presented in this thesis might make it possible to improve the description of hadron fragmentation in the different event generator tunes, in particular at high z.

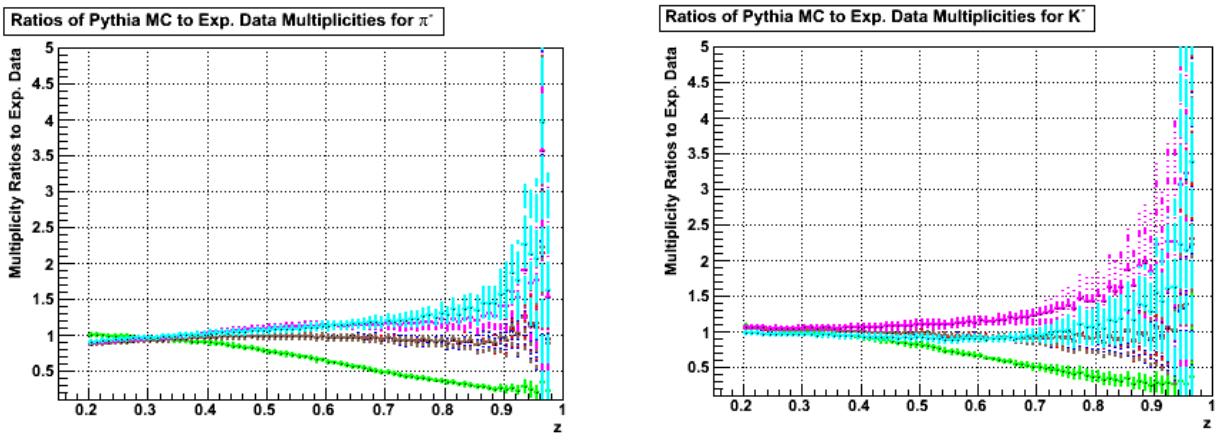
5.3.2 Comparison of Experimental and Generated Pion Multiplicities

In Figures 5.5, pion multiplicities from the different Monte Carlo data samples are compared with the measured pion multiplicities for negative particle charge. Figure 5.6 a) shows ratios of MC multiplicities over experimental data multiplicities. For both Figures, experimental statistical and systematic uncertainties are added in quadrature. As described in Section 4.1.5, the generic Belle Monte Carlo multiplicities exceed experimental data multiplicities at low $z < 0.4$ by about 20%, but then underestimate experimental data by up to a factor of 3 at high z. This behavior results from the fact that the Belle MC was optimized to precisely describe B-meson decay products with $z < 0.5$, while describing hadron production at high z was of lower priority. The general LHC and the Aleph tune describe the present measurement within 10% at low $z < 0.3$, but then exceed Belle experimental data by up to a factor of 5 at maximum z. Interstingly, the Atlas and CDF tune stay close to experimental Belle data within 20% for most of the range of z, and only show larger discrepancies at z 0.9 where uncertainties already grow large.



(a) (b)

Figure 5.5: Experimental multiplicities for negatively-charged pions compared to the multiplicities obtained from different tunes for Jetset/Pythia. For better visibility, uncertainties on MC multiplicities are not shown. Experimental statistical and systematic uncertainties are added in quadrature.



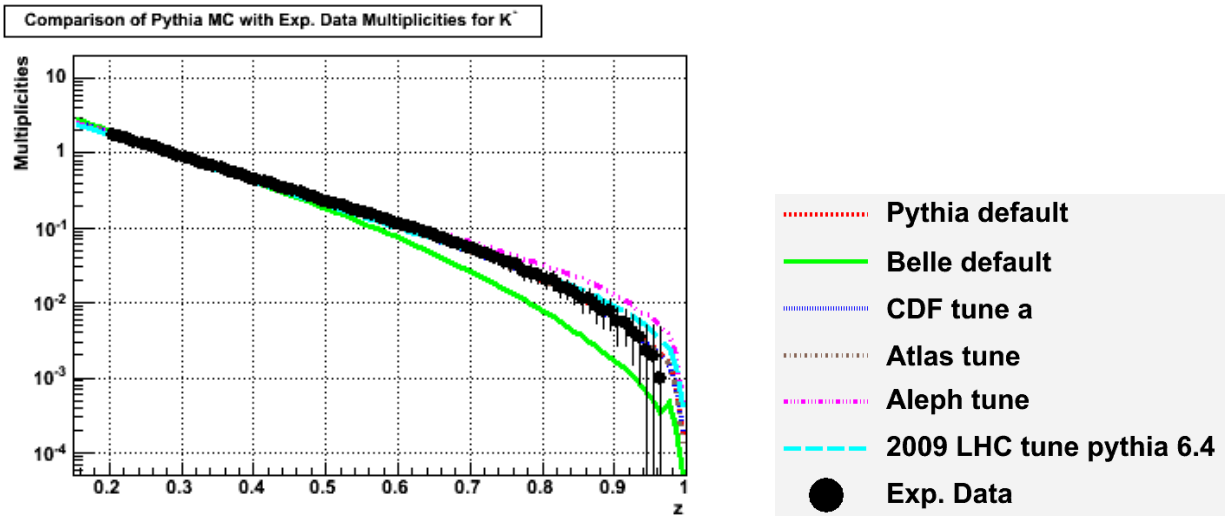
(a) (b)

Figure 5.6: Ratios of multiplicities for negatively-charged pions (a) and kaons (b) from different Jetset/Pythia tunes divided by the experimental results. Experimental statistical and systematic uncertainties are added in quadrature.

5.3.3 Comparison of Kaon Multiplicities

Analogously to the experimental data - Monte Carlo data comparison for pions, Figures 5.7 show kaon multiplicities from Monte Carlo as well as experimental data, for negative particle charge. Figure 5.6 b) shows the corresponding ratios. Experimental statistical and systematic uncertainties are again added in quadrature. Similarly to the pion case, the generic Belle MC data exceed experimental data at low $z < 0.4$

by about 20% and underestimate experimental data by up to a factor of 3 at high z. The Aleph tune overestimates Belle experimental data at low z by about 20% and exceeds the experimental data at high z by up to a factor of 5. However, the general LHC tune describes kaon experimental data fairly well and stays close to the measured multiplicities within 20% for most z. Only at high $z > 0.75$ the tune starts to deviate more. Again the Atlas and CDF tune describe Belle experimental data best within 20% for most of the z range.



(a) (b)

Figure 5.7: Multiplicities for negatively-charged kaons from this measurement and from different Jetset/Pythia tunes. For better visibility, uncertainties on MC multiplicities are not shown. Experimental statistical and systematic uncertainties are added in quadrature.

5.3.4 Conclusion

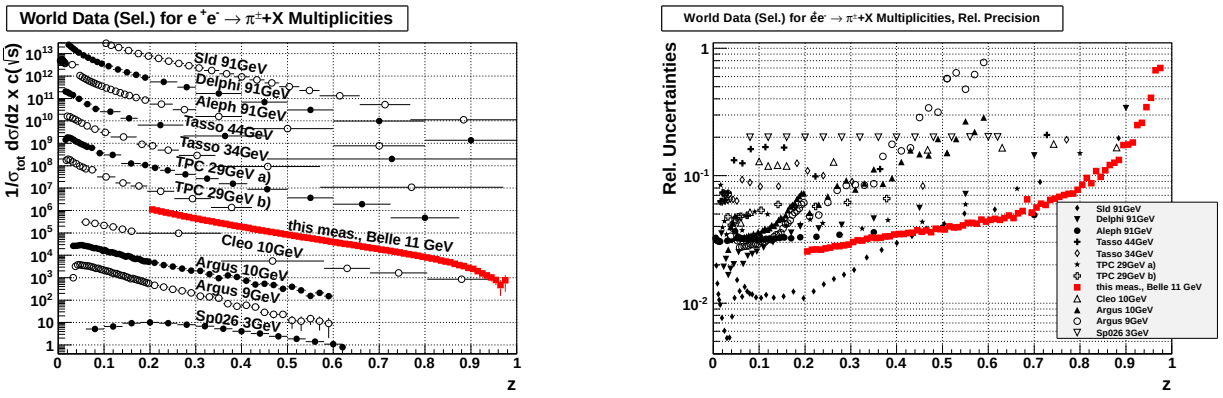
Both extracted pion and kaon multiplicities from the present measurement lie between MC multiplicities generated with different sets of Jetset/Pythia fragmentation parameters. Considering that Jetset/Pythia contains large parts of the current knowledge about fragmentation, this indicates global physical validity of the present measurement and consistency of all analysis and correction procedures. At the same time the significant differences between experimental and Monte Carlo multiplicities suggest that the high precision of the presented experimental results will allow to improve the MC description of hadron fragmentation, in particular at high z.

5.4 Comparison to Previous Experimental Results in e^+e^- Annihilation

In this Chapter, the final results for the pion and kaon multiplicities are compared to existing hadron multiplicities from e^+e^- annihilation measured at other experiments.

5.4.1 Comparison of Pion Multiplicities

Figure 5.8 a) shows charge-integrated pion multiplicities from the SLD, Delphi, Aleph, Tasso, TPC, Cleo, Argus and Sp026 experiments versus normalized hadron energy z[1]. For better visibility, the data sets are scaled with different but constant factors, such that ordering by increasing Q of all measurements is achieved. For all datasets, statistical, systematic and normalization uncertainties are added in quadrature for each data point. Figure 5.8 b) gives relative uncertainties for all included pion data sets.



(a) (b)

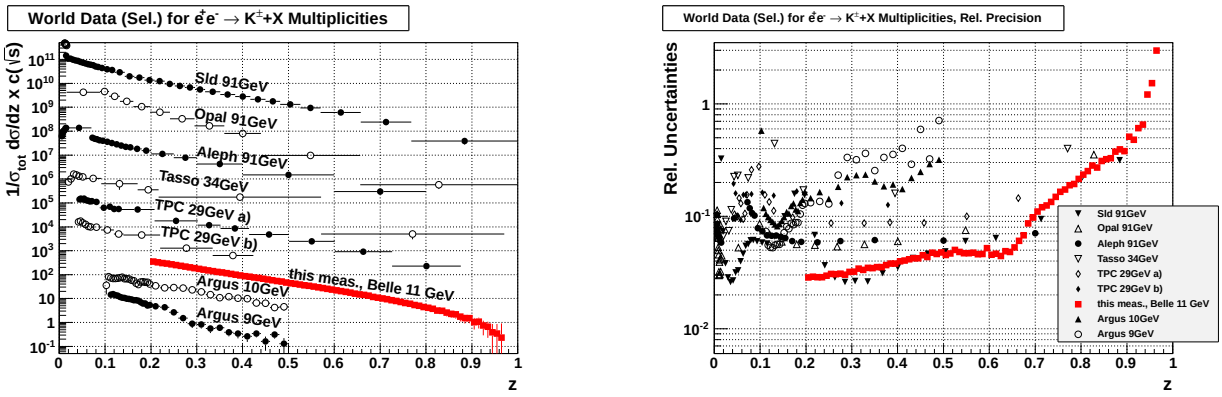
Figure 5.8: Charge-integrated pion multiplicities (a) and relative uncertainties (b) for selected experiments. For better visibility all data sets are scaled with different but constant factors. Statistical and systematic uncertainties are added in quadrature. The new Belle measurement obtained in this thesis is represented by the data points in red (red squares).

Comparing to the existing datasets in Figure 5.8 a), the Belle measurement achieves the goals formulated in the beginning of this thesis. The z dependence of pion multiplicities is measured for the first time at high $z > 0.7$ with unprecedented resolution. In addition, Figures 5.8 b) suggest that the new measurement is a precision measurement with comparable or better precision at all z than achieved in previous measurements. This especially holds for measurements at low Q.

[1]The plots shown in this Section are created following a visualization idea of Otmar Biebel et al.

5.4.2 Comparison of Kaon Multiplicities

Similarly, existing datasets for charge-integrated kaon multiplicities are compared to the new Belle measurement in Figure 5.9 a). Datasets from the SLD, Opal, Aleph, Tasso, TPC and Argus experiments are included. The same uncertainty treatment as for the pion comparison plots are applied, and all datasets are scaled by different but constant factors to achieve ordering in Q. Relative uncertainties for all data samples are shown in Figure 5.9 b).



(a) (b)

Figure 5.9: Charge-integrated kaon multiplicities (a) and relative uncertainties (b) for selected experiments. For better visibility all data sets are scaled with different but constant factors. Statistical and systematic uncertainties are added in quadrature. The new Belle measurement obtained in this thesis is represented by the data points in red (red squares).

Similar to the pion measurement, the new Belle result for kaons represents the first measurement of the z-dependence of kaon multiplicities at high $z > 0.7$. The Belle measurement significantly improves the precision of previously existing data sets at low Q and shows comparable precision to existing measurements at high Q.

Chapter 6

Summary and Outlook

6.1 Summary and Anticipated Impact of Analysis in Field

Multiplicities of identified charged pions and kaons have been measured over a broad range in z with $0.2 \leq z < 0.98$ (π), 0.97 (K) with high relative precision. The analyzed data sample of 68.0 fb^{-1} corresponding to 113×10^6 $e^+e^- \to q\overline{q}$ events, where $q = \{u, d, s, c\}$, has been accumulated at a center-of-mass energy of $Q = 10.52$ GeV. This makes this analysis the first precision measurement on data collected far from LEP/SLC center-of-mass energies. The relative precision of previous hadron multiplicity measurements from e^+e^- annihilation is lower or comparable to the precision achieved in this measurement. In particular, significantly better precision has been reached compared to any of the previous measurements carried out at low energy scale. The resolution in z has been greatly improved compared to previous multiplicity measurements over most of the z range. In addition, no other previous measurement probes the z dependence of multiplicities at high $z > 0.7$. The high statistics and good control of systematic uncertainties will give improved constraints on the z-dependence of hadron FFs in general and constraints at large $z > 0.7$ for the first time.

The results of this thesis have been submitted for publication in Physical Review Letters in January 2013. Several theory groups have expressed their interest to immediately use the new Belle data in their QCD analyses of experimental hadron multiplicity data. Updates of extractions of FFs based on preliminary results from this thesis released at conferences already have been shown during a workshop for fragmentation functions held at RIKEN, Japan (e.g., Refs. [38] and [39]). These analyses suggest significant improvement in precision of the extracted FFs. In addition, resummation effects at high z are studied on preliminary data as well in Ref. [20]. The new Belle data will make it possible to improve the description of hadron production in event generators used for data analysis at the Large Hadron Collider, the Relativistic Heavy-Ion Colider and in SIDIS experiments at CERN and at Jefferson Laboratory.

6.2 Possible Improvements and Extensions of the Presented Analysis

This analysis represents a high precision measurement of hadron multiplicities. However, uncertainties could still be decreased. Since this measurement is systematics-limited in its uncertainties, analyzing larger amounts of data does not increase the measurement precision. Instead, systematic uncertainties would need to be decreased. Especially at high z where uncertainties are largest, decreasing systematic uncertainties on multiplicities would have significant impact on physics analyses focusing on high z kinematics. In the presented analysis, the size of systematic uncertainties are tightly correlated with the statistics available in the MC samples from which correction factors are extracted. Generation of larger MC event samples for tracks with $z > 0.5$ could reduce the uncertainties of this analysis at high z. This approach, however, was not possible given the available computing resources and time limitations.

Future hadron multiplicity measurements at KEK will benefit from the ongoing upgrade of the Belle detector and the KEKB accelerator towards Belle II and SuperKEKB, respectively. The expected improvement in vertex resolution of the upgraded tracking detectors of Belle II will possibly allow separation of weak and strong decay products based on the detection of secondary decay vertices. Higher capacity computing farms will be available for Belle II to produce even larger amounts of MC data to support analyses, which can be used for corrections of systematic effects in multiplicity analyses as well.

Furthermore, this measurement can be extended to also yield charge-integrated proton multiplicities. Through the matrix approach used for the PID correction, PID-corrected proton yields are already available in this analysis. All systematic corrections can, in principle, be extended to also correct proton yields. Similar to the pion and kaon multiplicities, precise results for protons and anti-protons would further improve the description of hadron production from quarks and gluons resulting from high energy particle reactions.

Finally, another extension of the new Belle measurement is a measurement of the dependence of final state hadron multiplicities on the transverse momentum, p_T, of the final state hadrons with respect to the quark-antiquark axis. Knowledge of the p_T-dependence of FFs will make it possible to measure the pQCD non-calculable transverse momentum dependence of quarks inside nucleons in SIDIS by observing the total transverse momentum of the hadron in the final state and then unfolding the p_T contribution from the fragmentation process. In addition, a measurement of di-hadron yields in opposing jet-hemispheres for 2 jet events would contribute to study the flavor dependence of FFs by comparing $\pi - \pi$, $\pi - K$ and $K - K$ yields.

Appendix A

Supplementary Material to Chapter 1

A.1 AKK Parton-to-Pion Fragmentation Functions vs. z

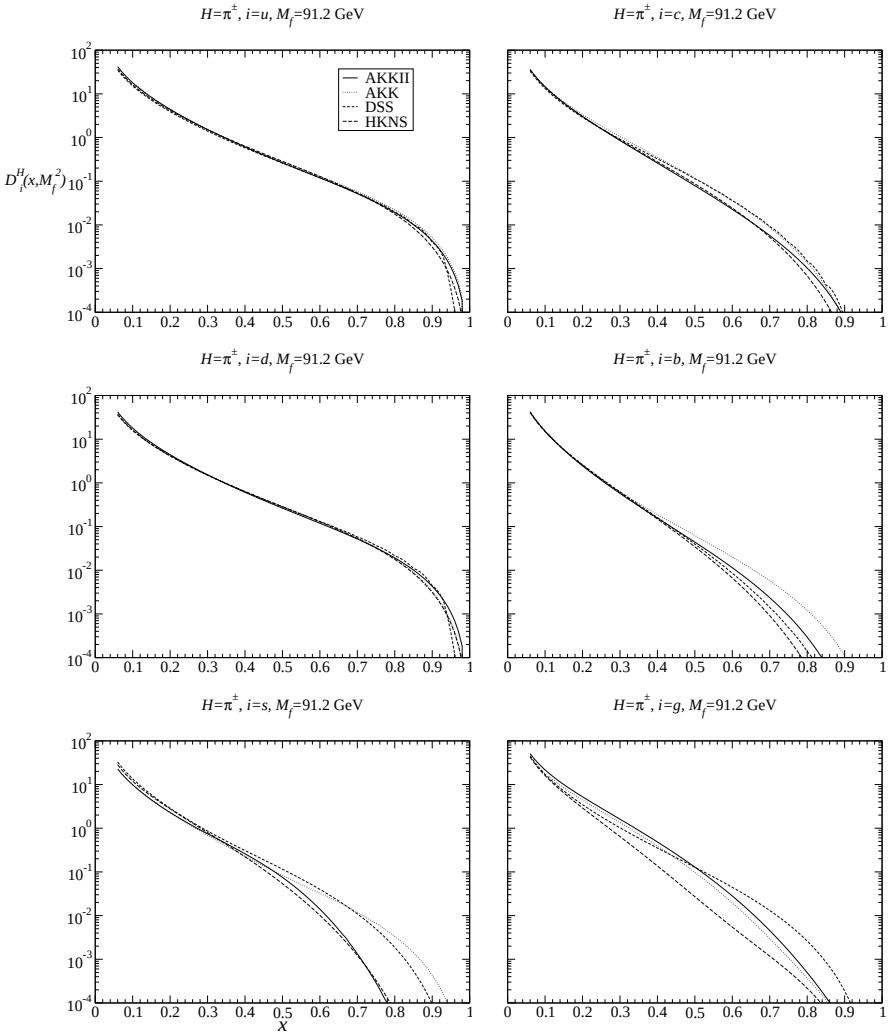


Figure A.1: Extracted parton-to-pion FFs from AKK ('AKKII'), [18]. FFs from an earlier extraction of AKK ('AKK', [92]) and FFs from HKNS and DSS are shown as well for comparison. All FFs are plotted for $Q^2 = M_{Z^0}$. No uncertainties are shown.

A.2 AKK Parton-to-Kaon Fragmentation Functions vs. z

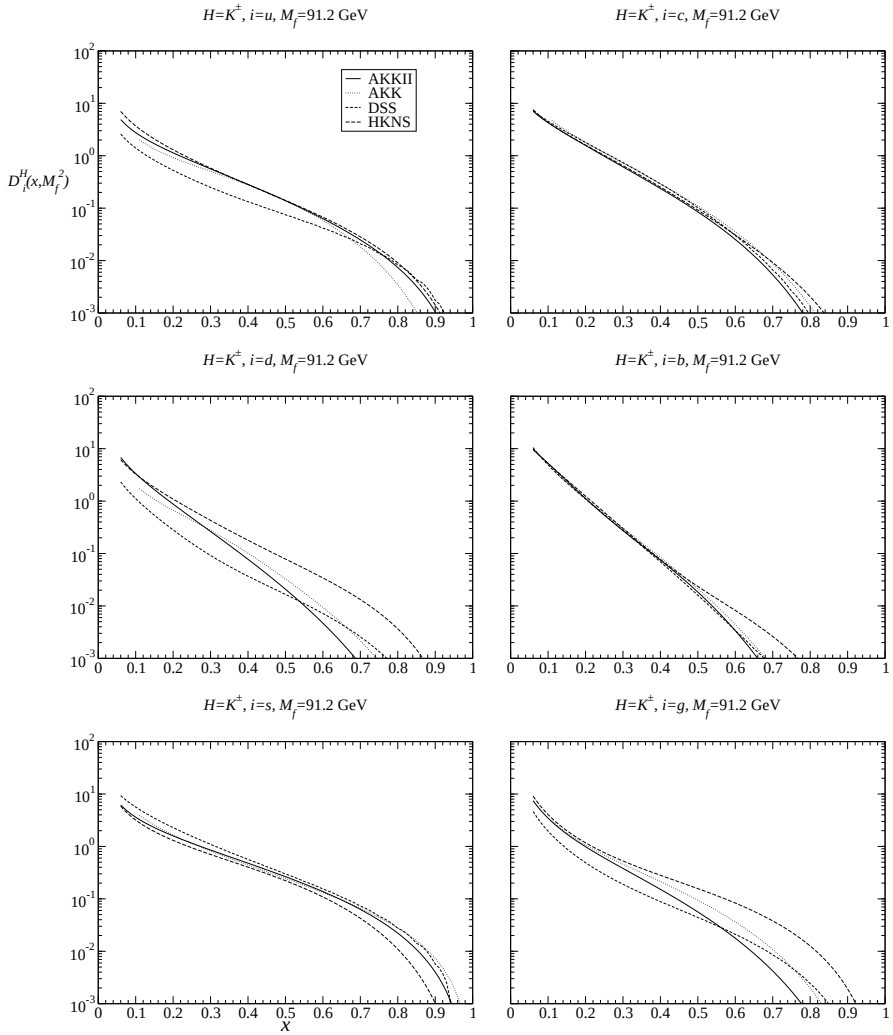


Figure A.2: Extracted parton-to-kaon FFs from AKK ('AKKII'), [18]. FFs from an earlier extraction of AKK ('AKK', [92]) and FFs from HKNS and DSS are shown as well for comparison. All FFs are plotted for $Q^2 = M_{Z^0}$. No uncertainties are shown.

Appendix B

Supplementary Material to Chapter 4

B.1 Concept of Particle Identification (PID) at Belle

Different methods to obtain particle identification (PID) statements about a measured track exist. However, each of these methods only covers a limited section of the total kinematic range of particles which emerge from the BELLE interaction region. Therefore several different subdetectors have been installed in the BELLE experiment, as described in the previous Sections. The different subdetectors produce likelihood statements about the possible species j of a measured track. These values are combined to add up the PID discrimination power of the single subdetectors. The descriptions in the Section for electron/positron, muon and hadron identification follow the information provided in Refs. [48], [49] and [66], respectively.

B.1.1 Electron/Positron Identification

From the track measurements in the drift chambers, the ACC and ECL subdetectors, likelihoods for the track being an electron/positron are calculated. Analogously, likelihood statements describing the possibility that the track measurements have not been caused by an $e^{\pm}$ particle are formed. Several likelihood values, if considered separately, only offer a modest ability to distinguish between electrons/positrons and other species. To maximize the discrimination power, a total normalized electron/positron likelihood ratio L_{eid} is computed by combining the likelihood information of the subdetectors:

$$L_{\mathrm{eid}} \doteqdot \frac{L_e^{\mathrm{sCDC/CDC}} \cdot L_e^{\mathrm{ACC}} \cdot L_e^{\mathrm{ECL}}}{L_e^{\mathrm{sCDC/CDC}} \cdot L_e^{\mathrm{ACC}} \cdot L_e^{\mathrm{ECL}} + L_{\mathrm{no}\ e}^{\mathrm{sCDC/CDC}} \cdot L_{\mathrm{no}\ e}^{\mathrm{ACC}} \cdot L_{\mathrm{no}\ e}^{\mathrm{ECL}}}. \tag{B.1}$$

L_{eid} cannot be interpreted as a probability as correlations between the subdetector likelihoods are not corrected for. The values of L_{eid} for each track can be extracted from the BELLE data tables. A cut on the value of the electron/positron likelihood ratio selects corresponding tracks inside the likelihood limits represented by the cut. In this analysis, likelihoods are extracted by requiring the latest ACC probability distribution functions, no ToF input and default CDC values, $eid.prob(3, -1, 0)$.

B.1.2 Muon Identification

The KLM is the most important subdetector used for muon identification. The extracted likelihoods L_k^{KLM} describe the possibility that a measured track is of a species k, where $k = \{\mu,\ \pi,\ K\}$. These likelihoods can be combined to yield the normalized likelihood ratio for a track to be a muon according to Equation B.2,

$$L_\mu \doteqdot \frac{L_\mu^{\mathrm{KLM}}}{L_\mu^{\mathrm{KLM}} + L_\pi^{\mathrm{KLM}} + L_K^{\mathrm{KLM}}}. \tag{B.2}$$

In order to pre-reject hadron tracks in the muon track selection, several criteria are examined for each track. The normalized muon likelihood ratio in Equation B.2 is not computed if less than 2 KLM resistive plate chamber signals are associated with the track. Other criteria ensure that the difference between the calculated and measured range of a selected track in units of KLM RPC layers is no greater than 5, and that the hit time does not lie outside the event time window of 1 μs.

For the tracks which fail these tests, no muon likelihood ratio is available, and L_μ is set to -5.0 in this analysis. For all other tracks, the respective value for L_μ can be extracted from the BELLE data tables. By applying cuts on the value of L_μ, tracks with likelihood ratios within the range given by the cuts can be selected.

B.1.3 Hadron Identification

Hadron likelihoods are formed from measurements in the drift chambers, the ACC and the ToF subdetectors. A total likelihood L_j^{tot} is created for each track by multiplying the likelihoods from the three mentioned subdetectors,

$$L_j^{\mathrm{tot}} \doteqdot L_j^{\mathrm{sCDC/CDC}} \cdot L_j^{\mathrm{ACC}} \cdot L_j^{\mathrm{ToF}}. \tag{B.3}$$

These total likelihoods can be composed to yield separation likelihood ratios $L_{j^{(1)}:j^{(2)}}$,

$$L_{j^{(1)}:j^{(2)}} \doteqdot \frac{L_{j^{(1)}}}{L_{j^{(1)}} + L_{j^{(2)}}}, \tag{B.4}$$

where $j^{(1)}$ and $j^{(2)}$ are combinatorially taken from $\{\pi,\ K,\ p\}$.[1] These separation ratios are stored in the BELLE data tables for each track. If no statement about the separation between two hadron species has

[1] Reference [66] states that the values of $L_{j^{(1)}:j^{(2)}}$ generally cannot be interpreted as *probabilities* for a track to be of species $j^{(1)}$. The criterion given in Ref. [66] for a probability interpretation is that the considered data sample only consists of tracks originating from only two particle species $j^{(1)}$ and $j^{(2)}$, in equal numbers. In such a case, the a priori probability for tracks to be of species $j^{(1)}$ or $j^{(2)}$ can be given as 0.5. $L_{j^{(1)}:j^{(2)}}$ then represents a Bayesian a posteriori probability for a measured track to be of particle species $j^{(1)}$. However, this reasoning does not apply for general inclusive particle samples.

been achieved, the value of 0.5 is assigned to the respective $L_{j^{(1)};j^{(2)}}$. In physics analyses, cuts can be applied to these separation ratios to select tracks according to the values of their separation likelihood ratios.

For this analysis, all separation ratios have been extracted under the requirement that all three subdetectors, ACC, ToF and CDC, contribute to the total likelihoods as defined in Equation B.3, using the latest ACC probability density functions. Additionally, each z-coordinate of all ToF measurements is required to match with an extrapolated track from the CDC. The syntax used is $atc\_pidsel\_ivsj(3, 1, 0, i, j)$.

B.2 Likelihood Selection Cuts

Selected Species j	Cuts on Likelihood Quantities				
	L_{eid}	L_{μ}	$L_{K:\pi}$	$L_{K:p}$	$L_{\pi:p}$
e	$[0.85; 1.0]$	$[-5.0; 0.9)$	-	-	-
μ	$[0.0; 0.85)$	$[0.9; 1.0]$	-	-	-
π	$[0.0; 0.85)$	$[-5.0; 0.9)$	$[0.0; 0.6)$	$[0.0; 1.0]$	$[0.2; 1.0]$
K	$[0.0; 0.85)$	$[-5.0; 0.9)$	$[0.6; 1.0]$	$[0.2; 1.0]$	$[0.0; 1.0]$
p	$[0.0; 0.85)$	$[-5.0; 0.9)$	$[0.0; 1.0)$	$[0.0; 0.2)$	$[0.0; 0.2)$

Table B.1: The set of PID likelihood cuts used throughout the present analysis to select tracks of likelihood-cut species j. The likelihood quantities L_{eid}, L_{μ}, $L_{K:\pi}$, $L_{K:p}$ and $L_{\pi:p}$ are defined in the Appendix B.1.

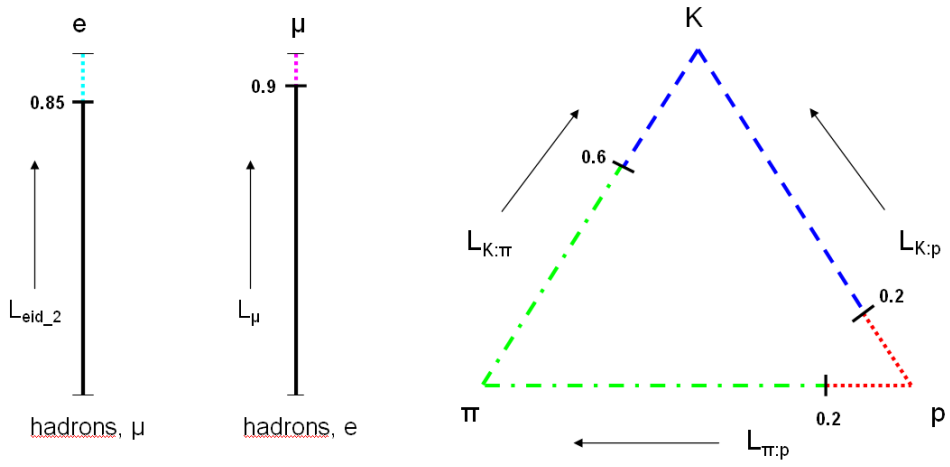


Figure B.1: Illustration of the chosen set of PID likelihood cuts given in Table B.1. The idea of displaying the hadron cuts in an 'identification triangle' has been adopted from Ref. [93].

B.3 Supplementary Material: D^* PID Extraction

B.3.1 Cumulative χ^2 Distribution Plots of Successful Fits

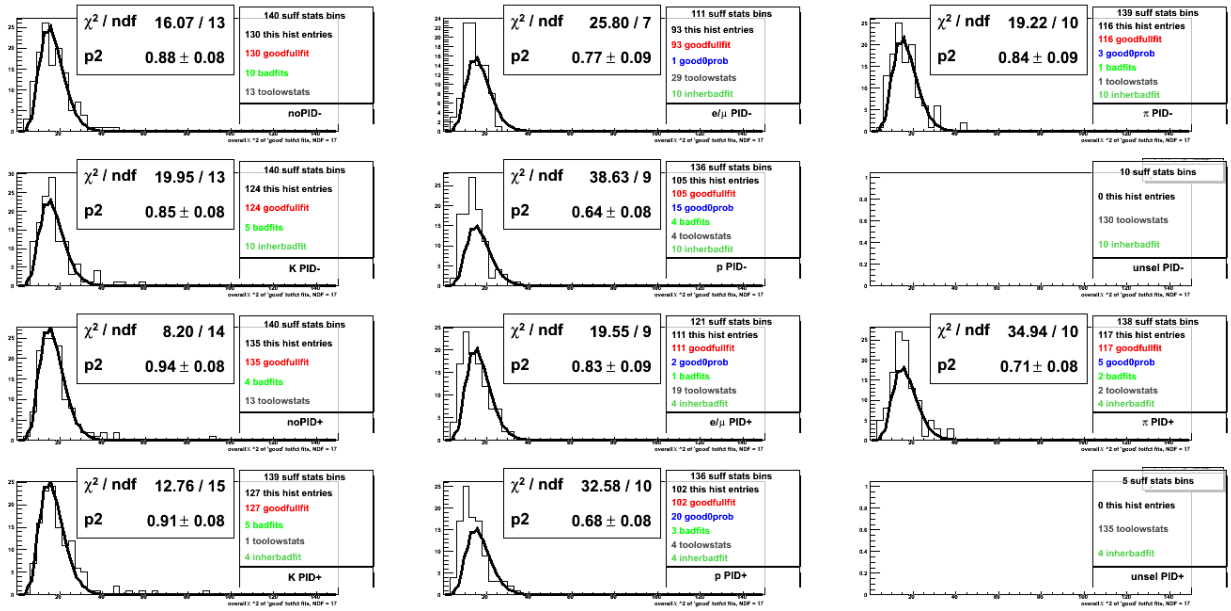


Figure B.2: Cumulative χ^2 distributions for all total fit function fits of histograms filled with kaon track kinematics, without and with additional PID selection cuts. All fits were performed with NDF=17. The solid line is the theoretical χ^2 distribution fitted to the histogram with a scale parameter. The χ^2 of this fit and the value of the scale parameter are shown in each field.

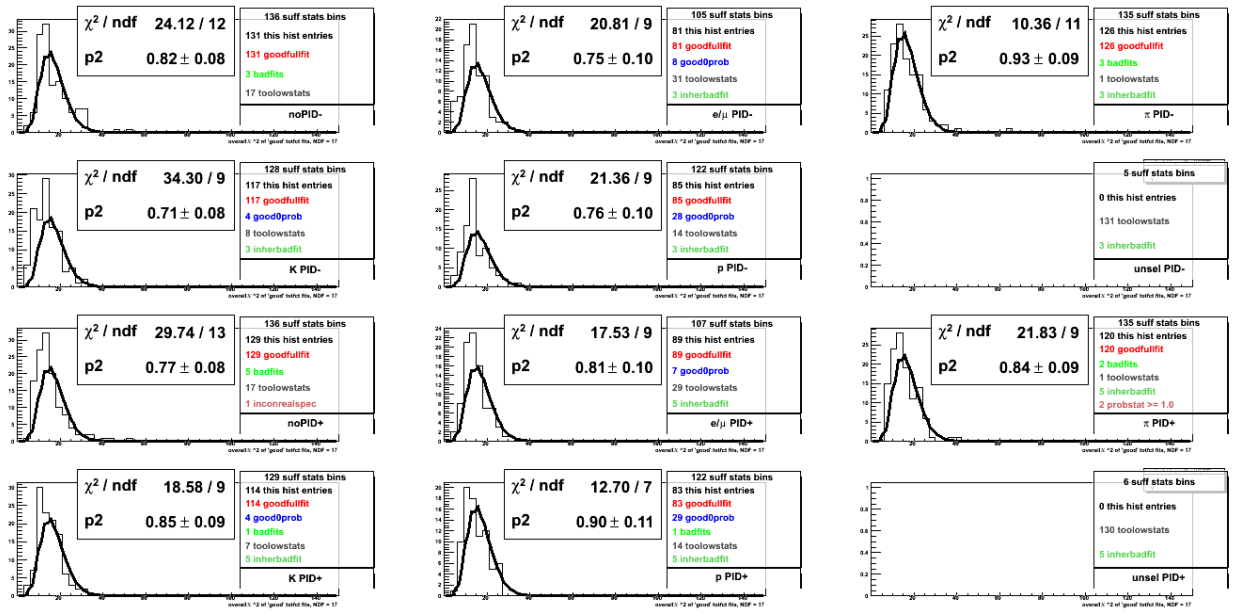


Figure B.3: Cumulative χ^2 distributions for all fits of histograms filled with fast pion track kinematics, without and with additional PID selection cuts. All fits were performed with NDF=17. The solid line is the theoretical χ^2 distribution fitted to the histogram with a scale parameter. The χ^2 of this fit and the value of the scale parameter are shown in each field.

B.3.2 'Kinematic Coverage' Plots for Successful PID Statement Extractions

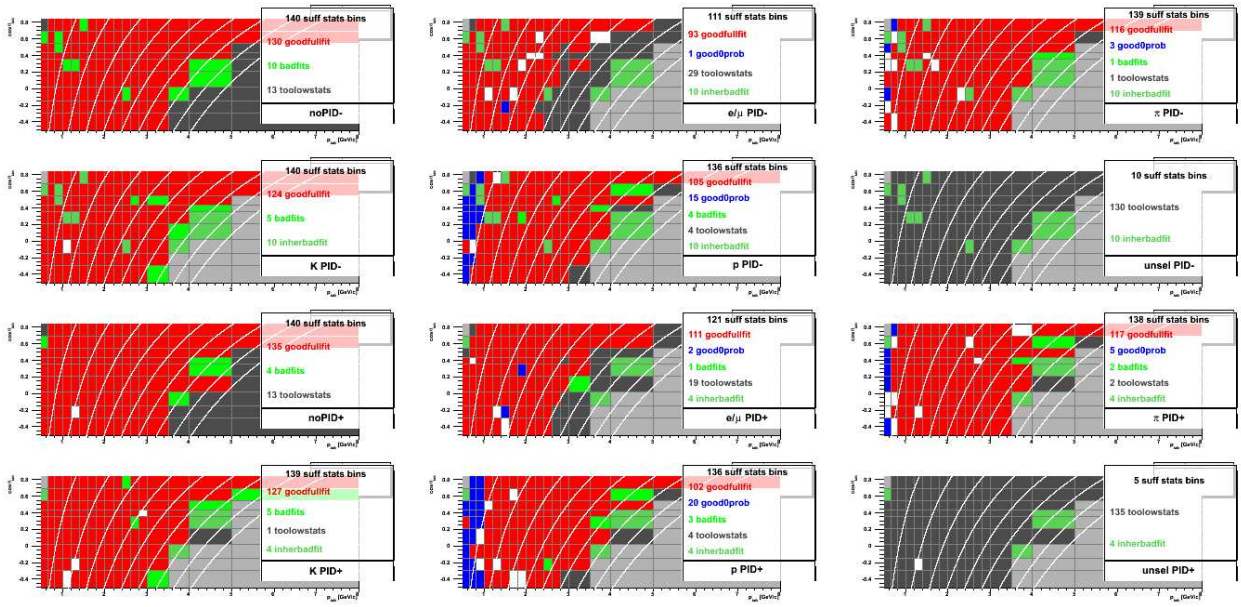


Figure B.4: Two-dimensional plots in the laboratory frame momentum and scattering angle θ of all kinematic bins in the PID correction, for all fitted histograms without and with different PID selections filled from kaon track kinematics. The color coding is explained in the legends of each plot, white kinematic bins indicate that the fit has been rejected by hand because of anomalies. White lines describe kinematic areas with equal normalized cms energy z, the rightmost line indicating $z = 1$.

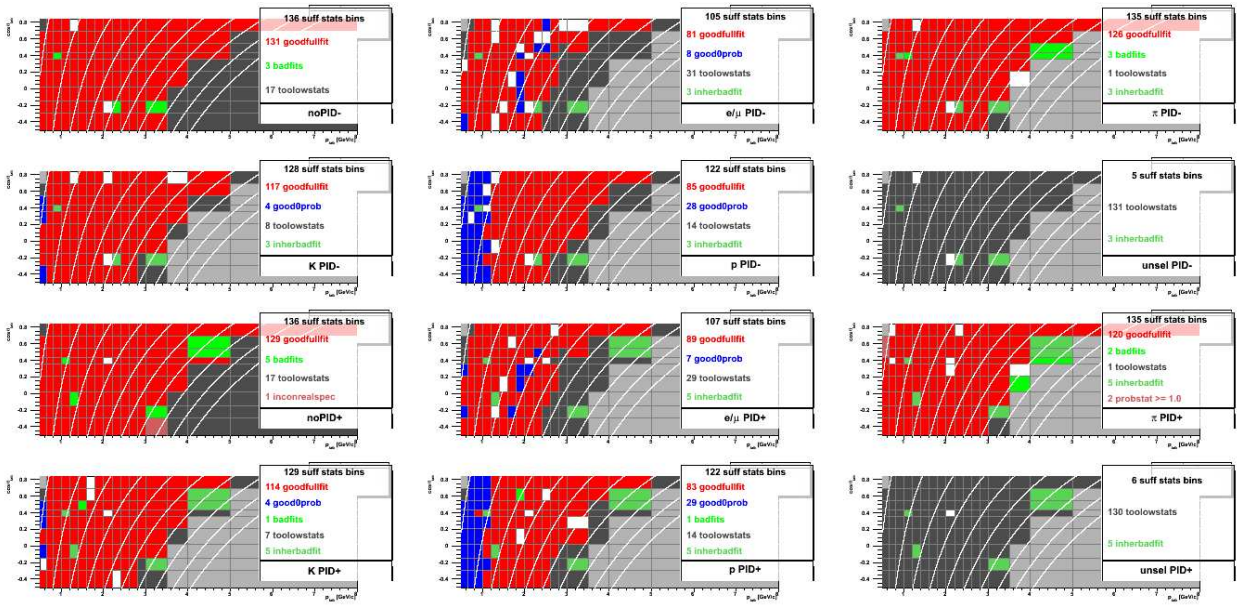


Figure B.5: Two-dimensional plots in the laboratory frame momentum and scattering angle θ of all kinematic bins in the PID correction, for all fitted histograms without and with different PID selections filled from fast pion track kinematics. The color coding is explained in the legends of each plot, white kinematic bins indicate that the fit has been rejected by hand because of anomalies. White lines describe kinematic areas with equal normalized cms energy z, the rightmost line indicating $z = 1$.

B.3.3 Consistency Check: Differences between Probabilities from Fits and Histogram Entries

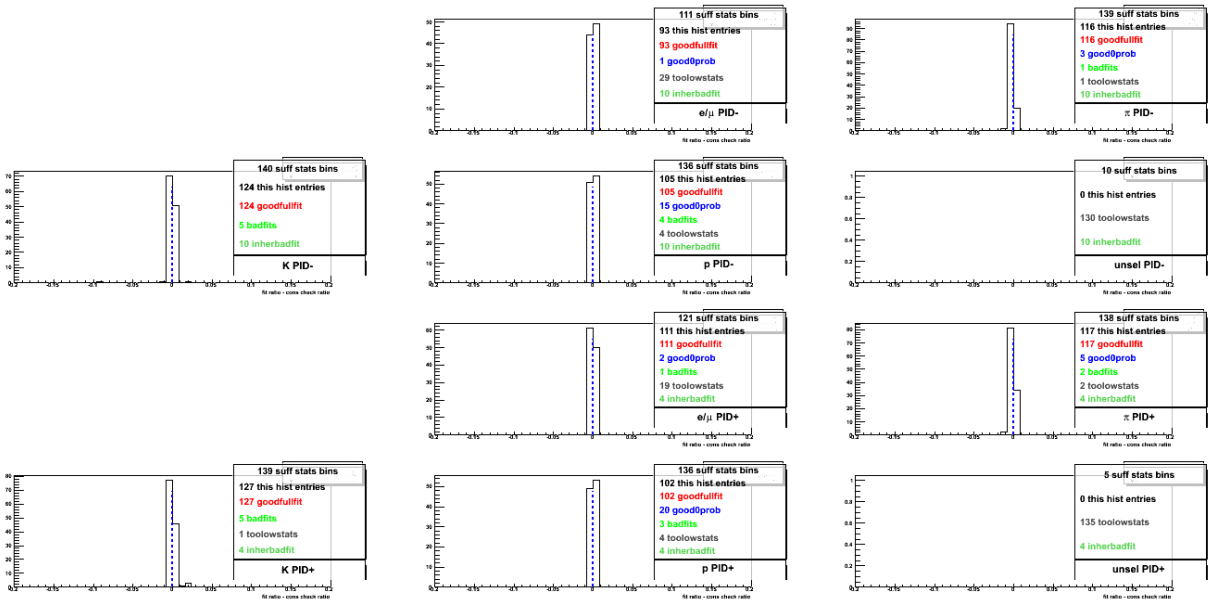


Figure B.6: Absolute differences between probabilities $p_{K \to j}$ from integrated signal fit functions and from histogram signal yields as described in Section 4.1.2. The top five plots compare probabilities $p_{K^- \to j^-}$, where $j = \{e/\mu,\ \pi,\ K,\ p,\ unsel\}$. The lower five plots compare probabilities for positively charged kaons. The center blue line indicates 0.

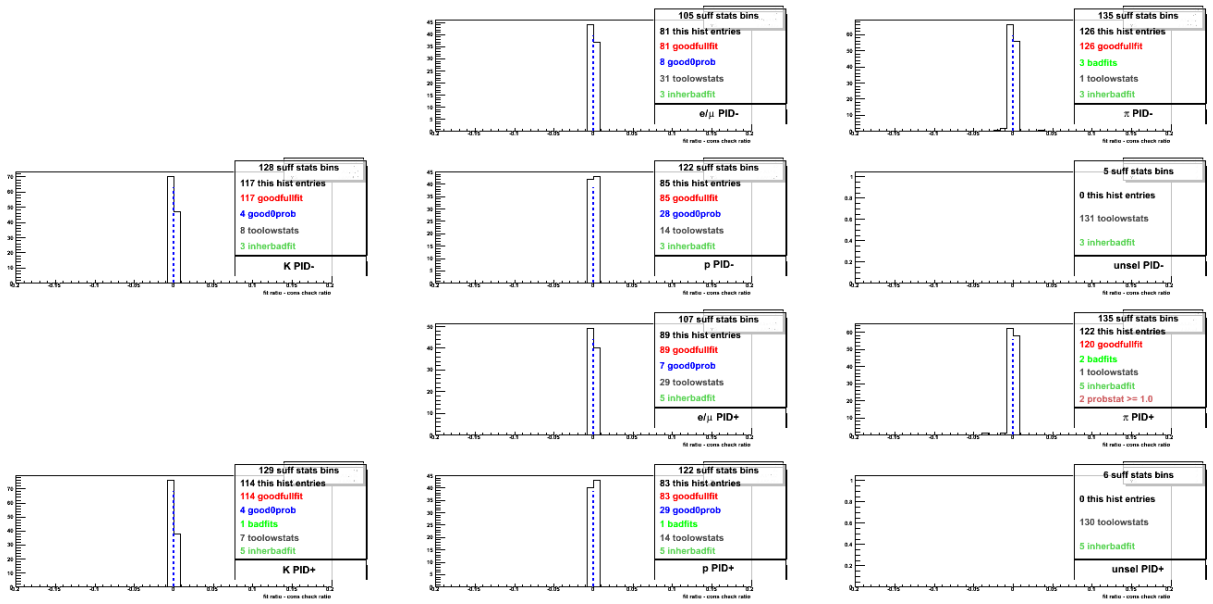


Figure B.7: Absolute differences between probabilities $p_{\pi \to j}$ from integrated signal fit functions and from histogram signal yields as described in Section 4.1.2. The top five plots compare probabilities $p_{\pi^- \to j^-}$, where $j = \{e/\mu,\ \pi,\ K,\ p,\ unsel\}$. The lower five plots compare probabilities for positively charged pions. The center blue line indicates 0.

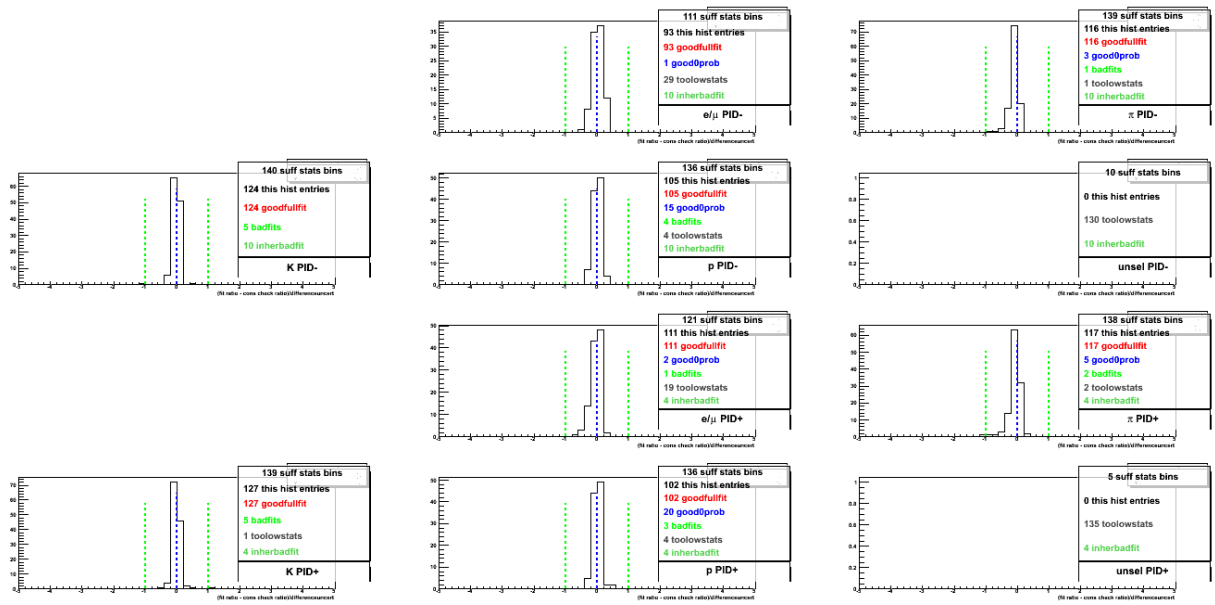


Figure B.8: Difference in units of overall uncertainty between probabilities $p_{K \to j}$ from integrated signal fit functions and from histogram signal yields as described in Section 4.1.2. The top five plots compare probabilities $p_{K^- \to j^-}$, where $j = \{e/\mu,\ \pi,\ K,\ p,\ unsel\}$. The lower five plots compare probabilities for positively charged kaons. The center blue line indicates 0. The green lines indicate 1σ.

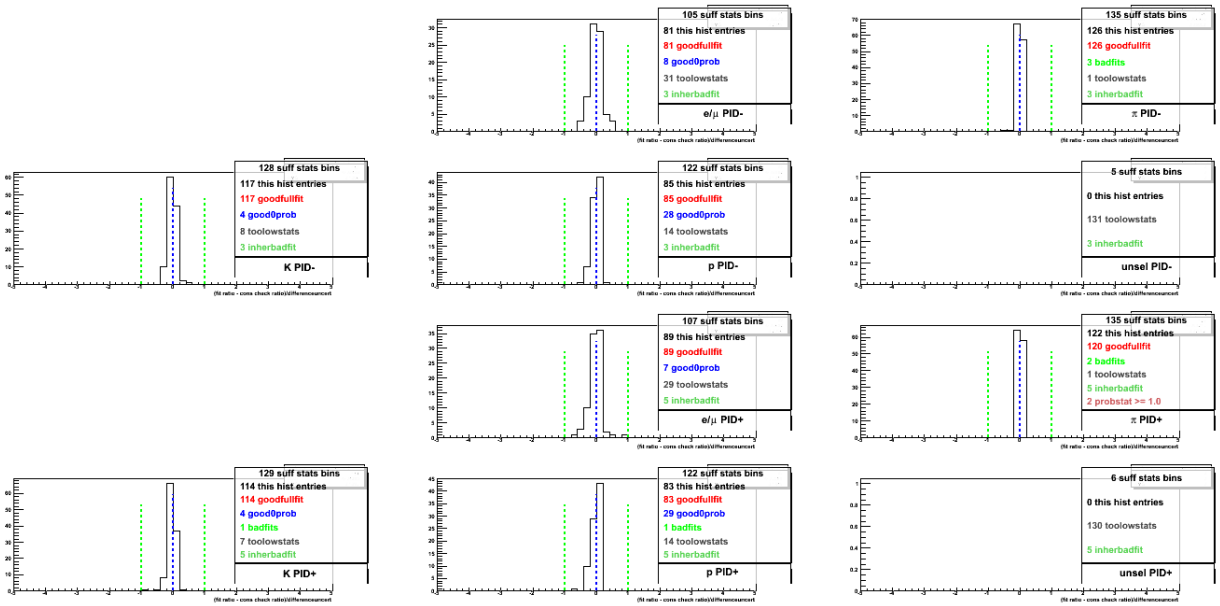


Figure B.9: Difference in units of overall uncertainty between probabilities $p_{\pi \to j}$ from integrated signal fit functions and from histogram signal yields as described in Section 4.1.2. The top five plots compare probabilities $p_{\pi^- \to j^-}$, where $j = \{e/\mu,\ \pi,\ K,\ p,\ unsel\}$. The lower five plots compare probabilities for positively charged pions. The center blue line indicates 0. The green lines indicate 1σ.

B.4 Supplementary Material: Λ PID Extraction

B.4.1 Cumulative χ^2 Distribution Plots of Successful Fits

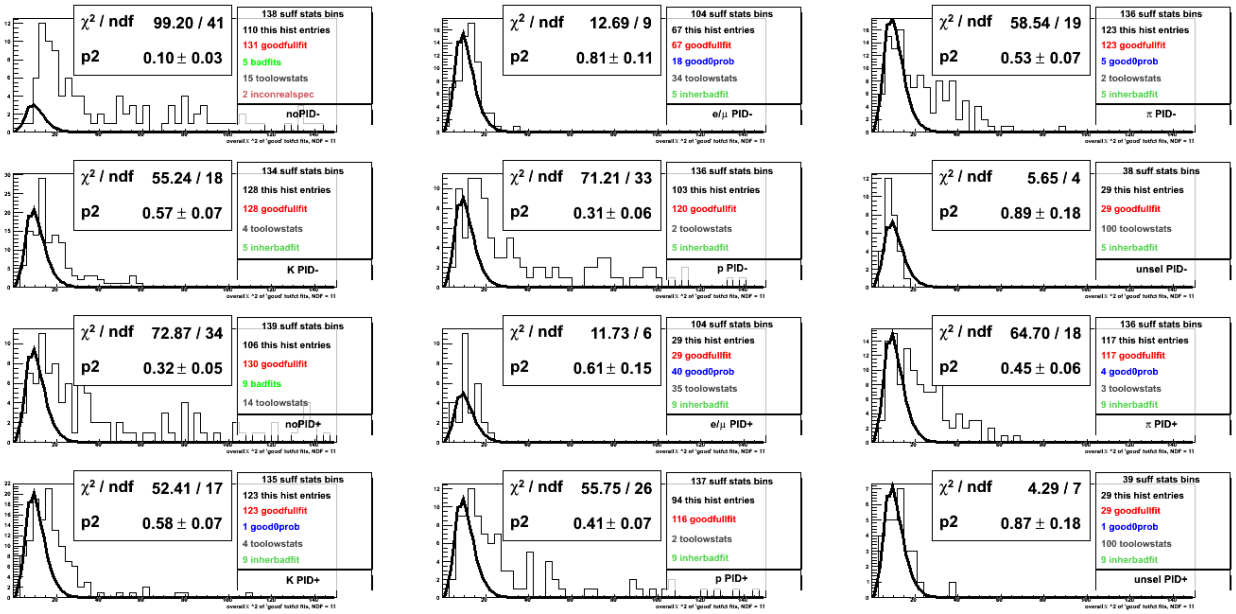


Figure B.10: Cumulative χ^2 distributions for all total fit function fits of histograms filled with proton track kinematics, without and with additional PID selection cuts. All fits were performed with NDF=11. The solid line is the theoretical χ^2 distribution fitted to the histogram with a scale parameter. The χ^2 of this fit and the value of the scale parameter are shown in each field.

B.4.2 'Kinematic Coverage' Plots for Successful PID Statement Extractions

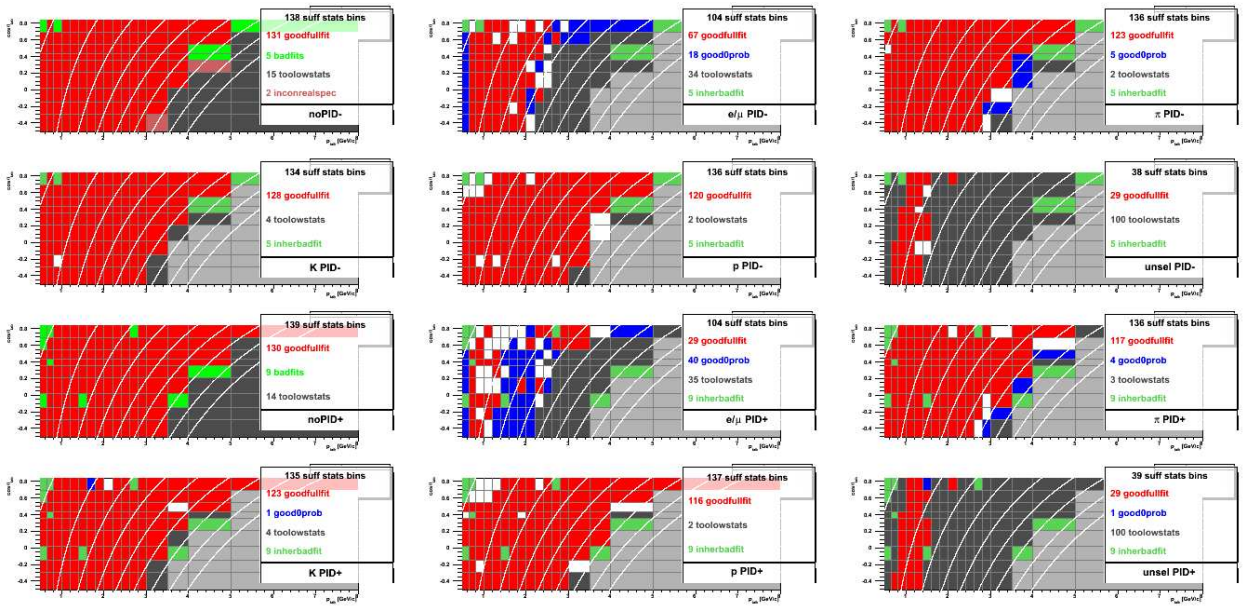


Figure B.11: Two-dimensional plots in the laboratory frame momentum and scattering angle θ of all kinematic bins in the PID correction, for all fitted histograms without and with different PID selections filled from proton track kinematics. The color coding is explained in the legends of each plot, white kinematic bins indicate that the fit has been rejected by hand because of anomalies. White lines describe kinematic areas with equal normalized cms energy z, the rightmost line indicating $z = 1$.

B.4.3 Consistency Check: Differences between Probabilities from Fits and Histogram Entries

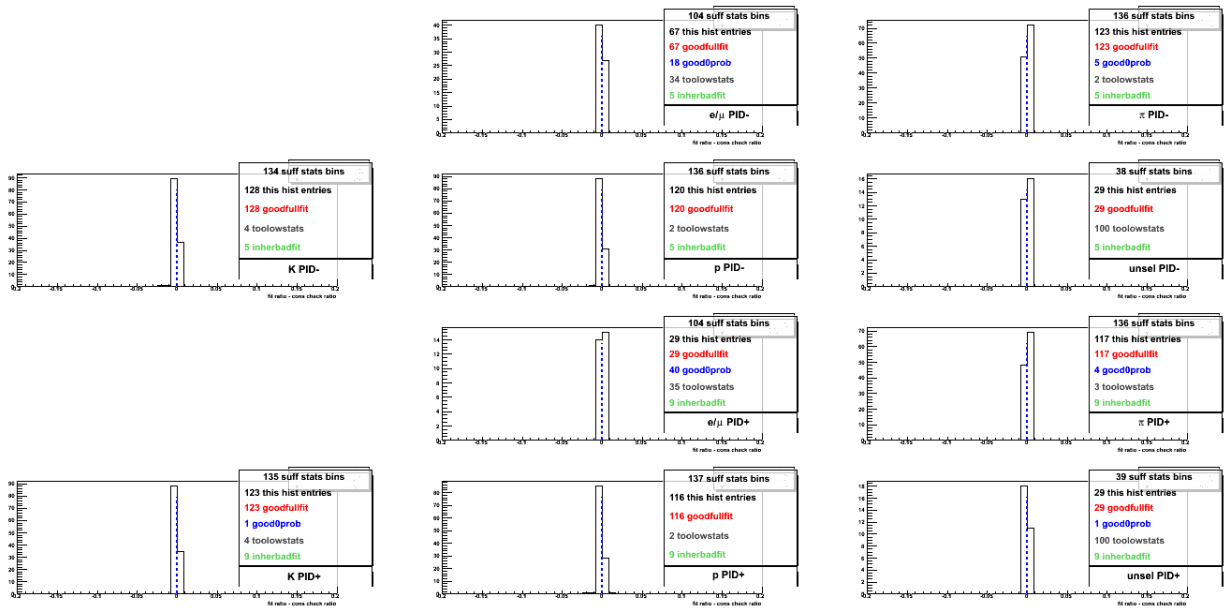


Figure B.12: Absolute differences between probabilities $p_{p \to j}$ from integrated signal fit functions and from histogram signal yields as described in Section 4.1.3. The top five plots compare probabilities $p_{p^- \to j^-}$, where $j = \{e/\mu,\ \pi,\ K,\ p,\ unsel\}$. The lower five plots compare probabilities for positively charged protons. The center blue line indicates 0.

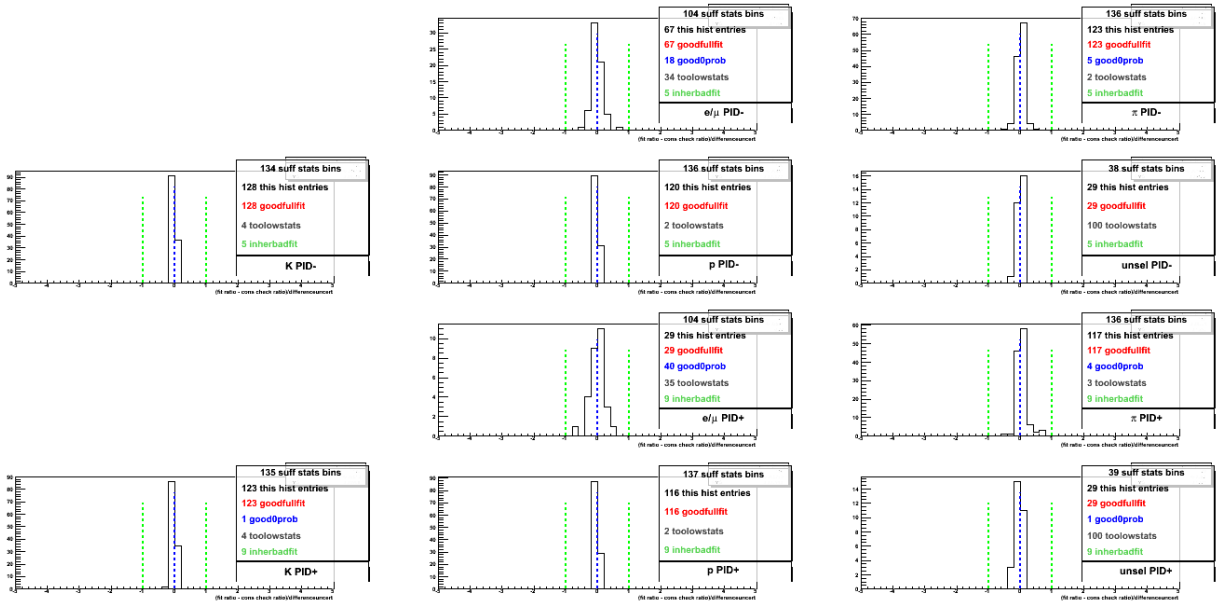


Figure B.13: Difference in units of overall uncertainty between probabilities $p_{p \to j}$ from integrated signal fit functions and from histogram signal yields as described in Section 4.1.3. The top five plots compare probabilities $p_{p^- \to j^-}$, where $j = \{e/\mu,\ \pi,\ K,\ p,\ unsel\}$. The lower five plots compare probabilities for positively charged protons. The center blue line indicates 0. The green lines indicate 1σ.

B.5 Supplementary Material: J/ψ PID Extraction

B.5.1 Cumulative χ^2 Distribution Plots of Successful Fits

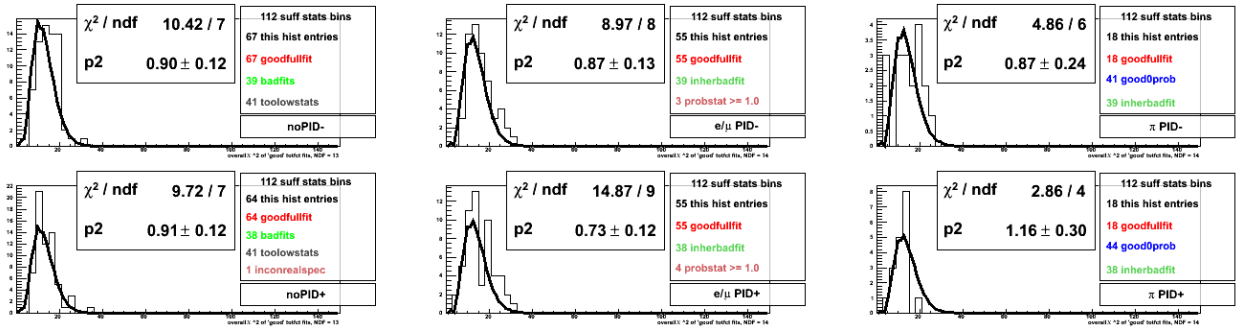


Figure B.14: Cumulative χ^2 distributions for all total fit function fits of histograms filled with e/μ track kinematics, without and with additional PID selection cuts. All 'noPID' fits were performed with NDF=13, all others with NDF=14. Since no fits with signal and background contribution were successful for kaon, proton and 'unsel' PID selections, no χ^2 distributions are given. The solid line is the theoretical χ^2 distribution fitted to the histogram with a scale parameter. The χ^2 of this fit and the value of the scale parameter are shown in each field.

B.5.2 'Kinematic Coverage' Plots for Successful PID Statement Extractions

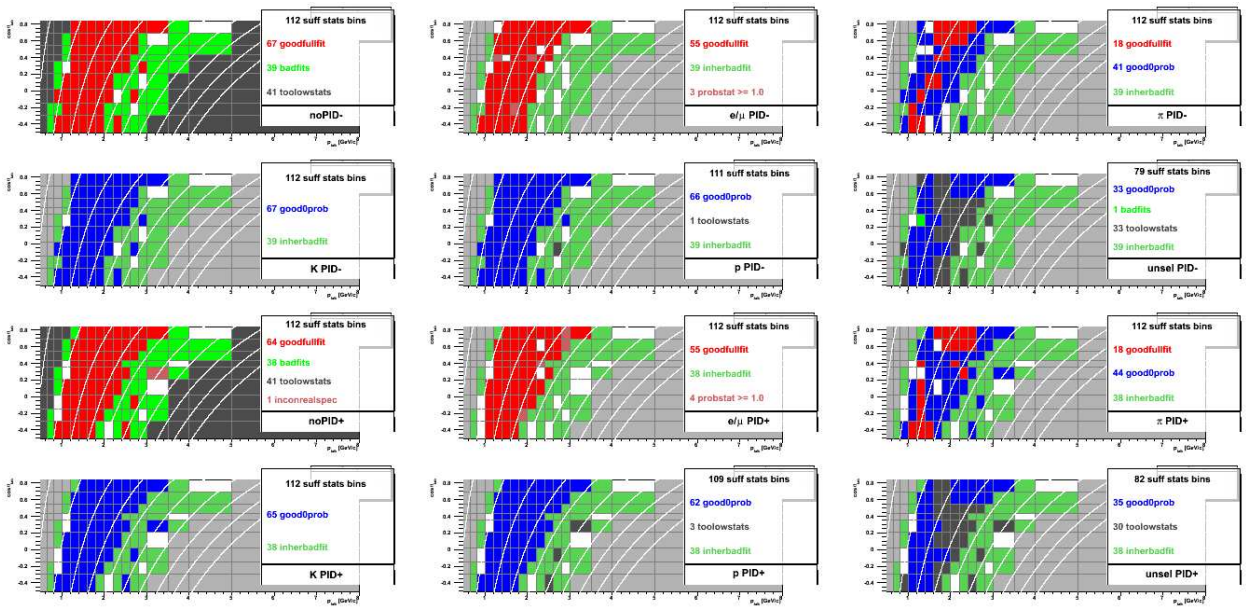


Figure B.15: Two-dimensional plots in the laboratory frame momentum and scattering angle θ of all kinematic bins in the PID correction, for all fitted histograms without and with different PID selections filled from e/μ track kinematics. The color coding is explained in the legends of each plot, white kinematic bins indicate that the fit has been rejected by hand because of anomalies. White lines describe kinematic areas with equal normalized cms energy z, the rightmost line indicating $z = 1$.

B.5.3 Consistency Check: Differences between Probabilities from Fits and Histogram Entries

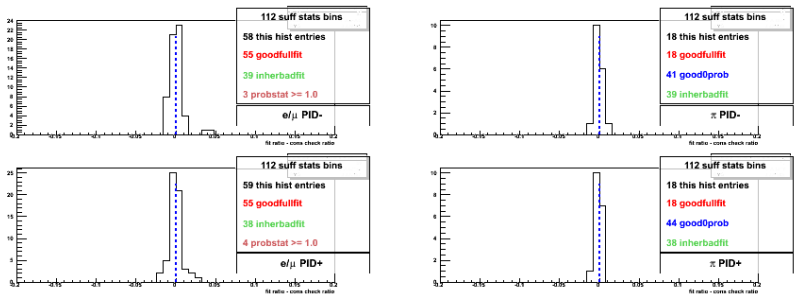


Figure B.16: Absolute differences between probabilities $p_{e/\mu \to j}$ from integrated signal fit functions and from histogram signal yields as described in Section 4.1.4. Since no PID probabilities from total fit function fits were extracted from kaon, proton and 'unsel' PID histograms, no difference distributions are given. The top two plots compare probabilities $p_{e/\mu^- \to j^-}$, where $j = \{e/\mu,\ \pi\}$. The lower two plots compare probabilities for positively charged e/μ. The center blue line indicates 0.

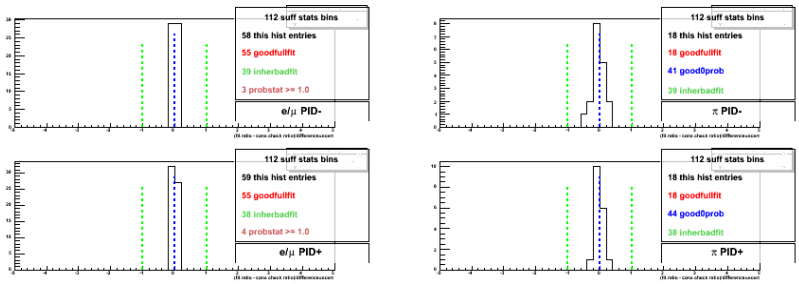


Figure B.17: Difference in units of overall uncertainty between probabilities $p_{e/\mu \to j}$ from integrated signal fit functions and from histogram signal yields as described in Section 4.1.4. The top two plots compare probabilities $p_{e/\mu^- \to j^-}$, where $j = \{e/\mu,\ \pi\}$. The lower two plots compare probabilities for positively charged e/μ. The center blue line indicates 0. The green lines indicate 1σ.

B.6 Supplementary Material: MCHadronBJ PID Extraction

B.6.1 Comparison between Experimental Data, Standard Belle MC and MCHadronBJ Yields vs. z

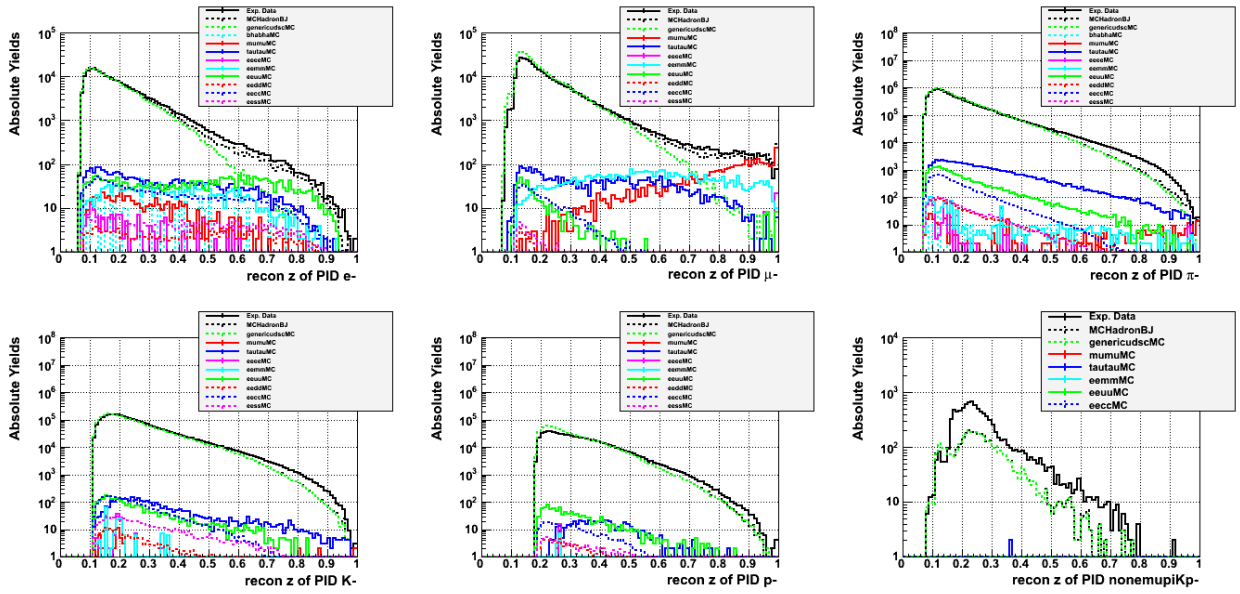


Figure B.18: Comparison between negatively charged track yields from experimental data (solid black line), standard Belle $uds + charm$ MC (dotted green line) circles) and from MCHadronBJ data (total: dotted black line) for 63 runs of Experiment 41 at continuum energy, versus reconstructed z. All shown species are selected by likelihood cuts on track PID likelihood values. The same selection cuts are applied in the analyses of all data samples. No uncertainties are shown.

B.6.2 Sample PID Probabilities Plots

Electron PID Probabilities

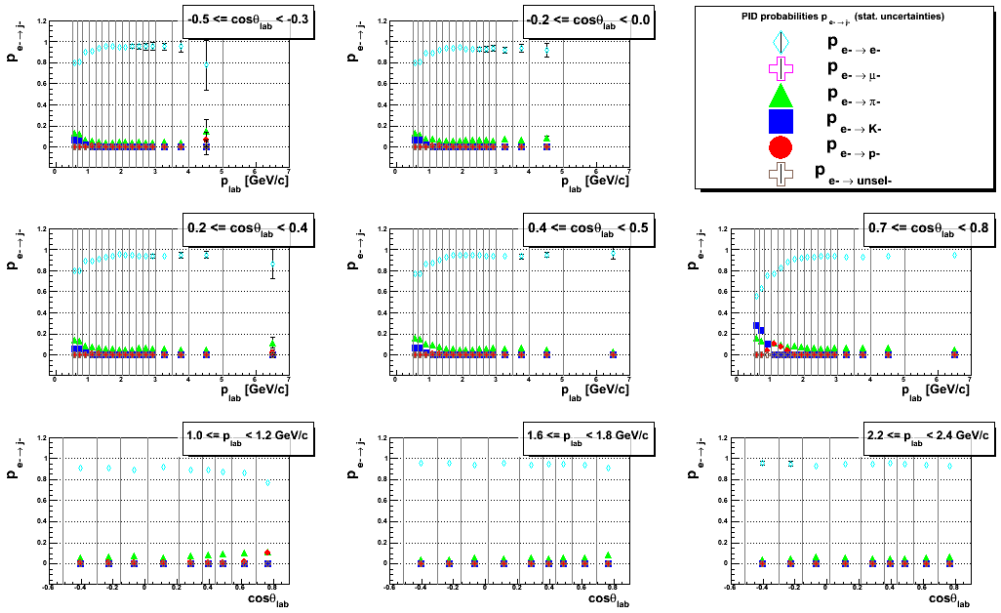


Figure B.19: Samples of extracted PID probabilities $\mathrm{p}_{e^- \to j^-}$ from the MCHadronBJ sample. The different subfigures give plots of probabilities for all likelihood-cut selected species j against p_{lab} in different scattering angle intervals (top 5 pads) and against $\cos\theta_{\mathrm{lab}}$ for different p_{lab} bins (bottom 3 pads) at the center of each bin.

Muon PID Probabilities

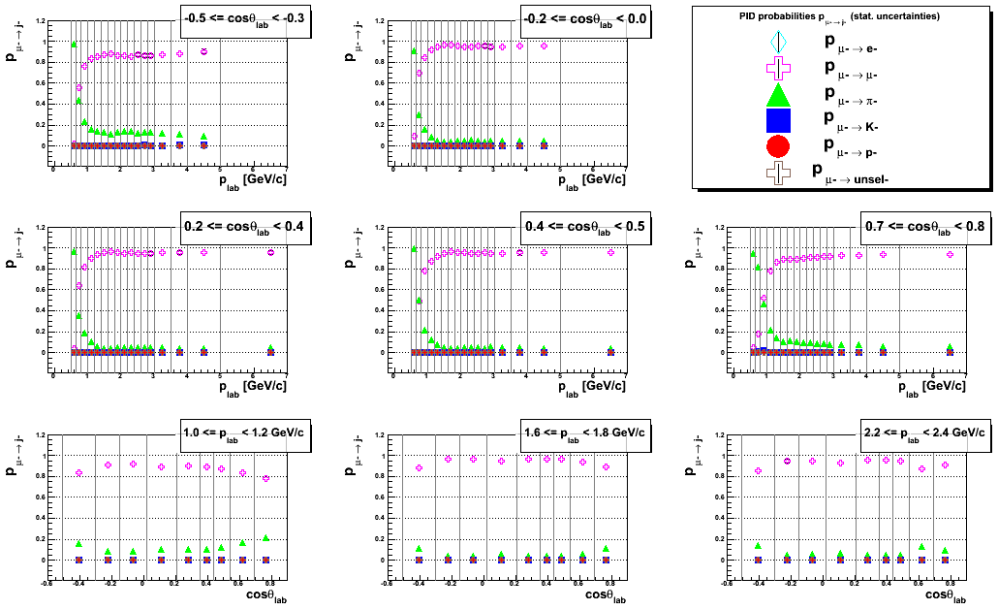


Figure B.20: Samples of extracted PID probabilities $p_{\mu^- \to j^-}$ from the MCHadronBJ sample. See the plot descriptions of Figure B.19.

Pion PID Probabilities

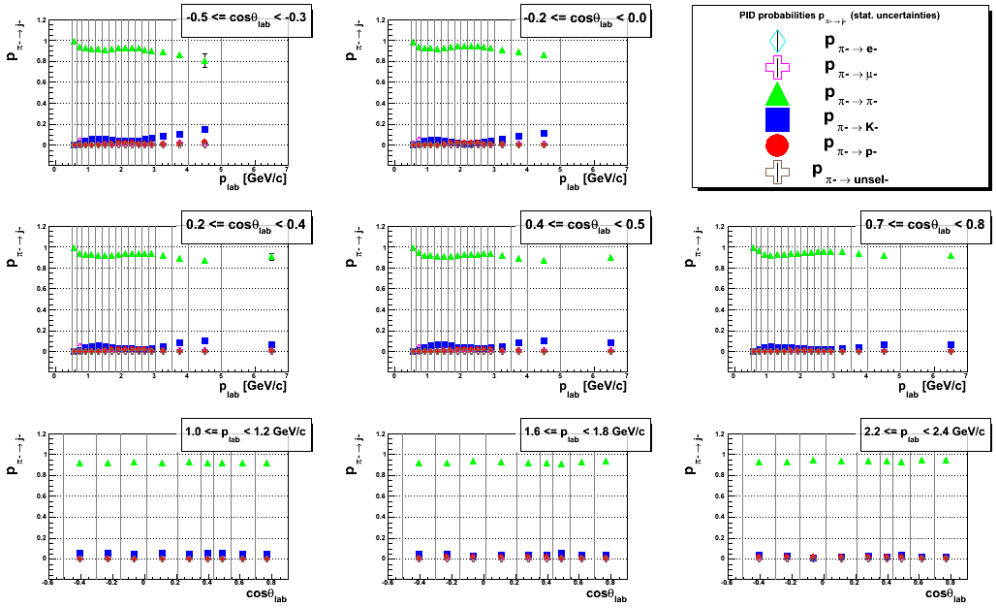


Figure B.21: Samples of extracted PID probabilities $p_{\pi^- \to j^-}$ from the MCHadronBJ sample. See the plot descriptions of Figure B.19.

Kaon PID Probabilities

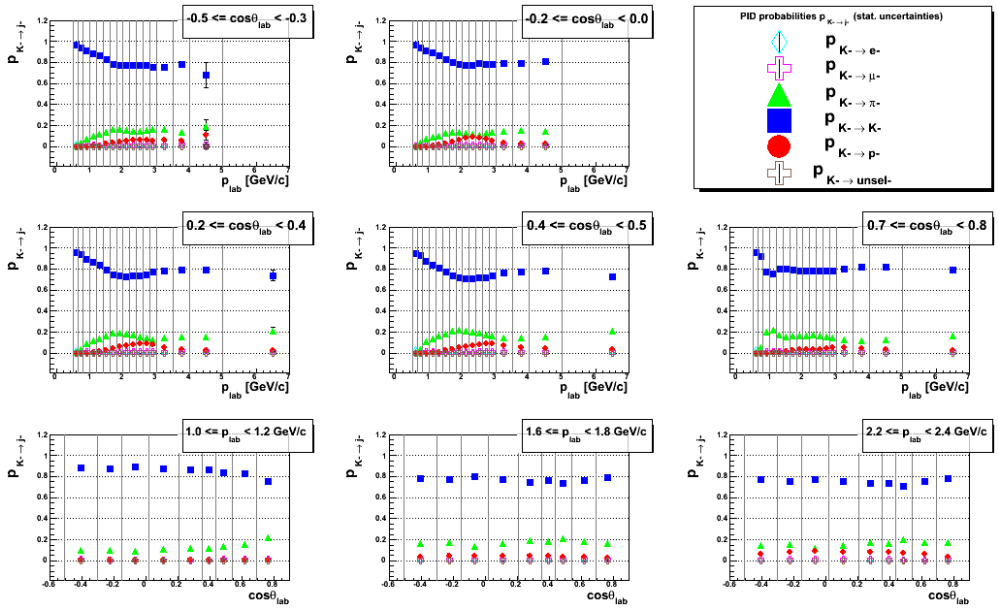


Figure B.22: Samples of extracted PID probabilities $p_{K^- \to j^-}$ from the MCHadronBJ sample. See the plot descriptions of Figure B.19.

Proton PID Probabilities

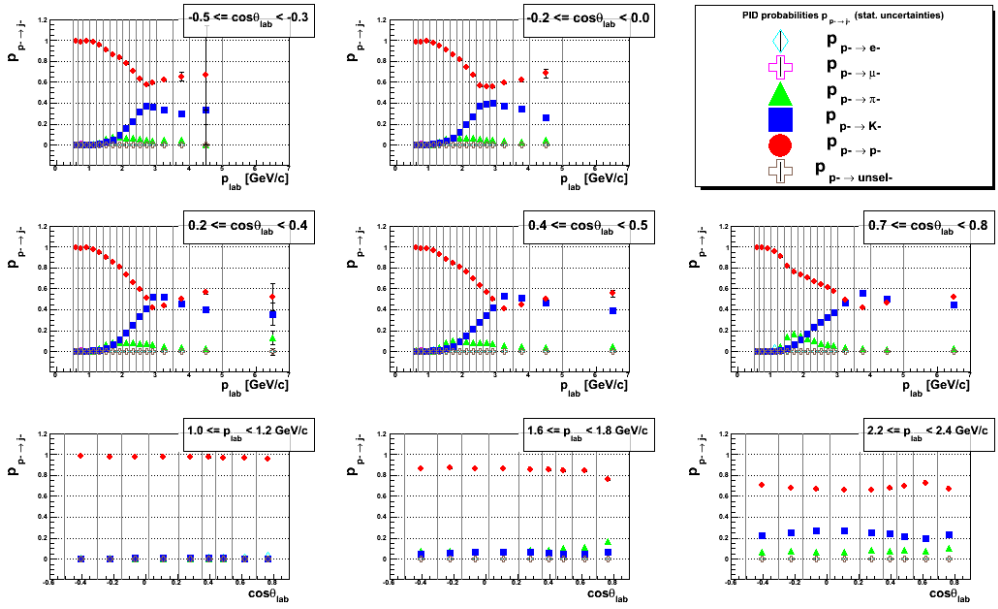
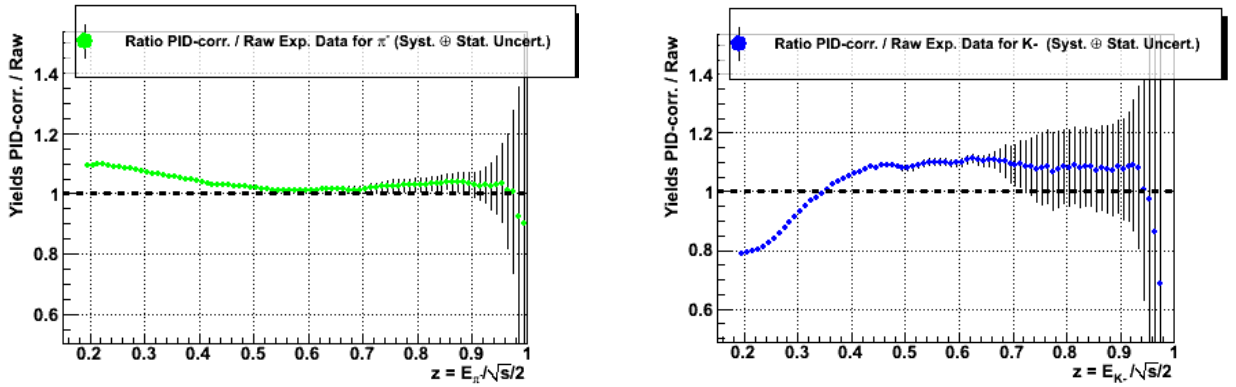


Figure B.23: Samples of extracted PID probabilities $p_{p^- \to j^-}$ from the MCHadronBJ sample. See the plot descriptions of Figure B.19.

B.7 Supplementary Material to PID Correction of Raw Yields

B.7.1 Ratio of PID-Corrected Yields over Raw Yields vs. z



(a) (b)

Figure B.24: Ratio between PID-corrected and raw experimental yields for negatively charged pions (a) and kaons (b). For the ratio, statistical and systematic uncertainties have been added quadratically into a single uncertainty for all PID-corrected spectra.

B.7.2 Ratio of PID-Corrected Yields over Kaon Yields vs. z

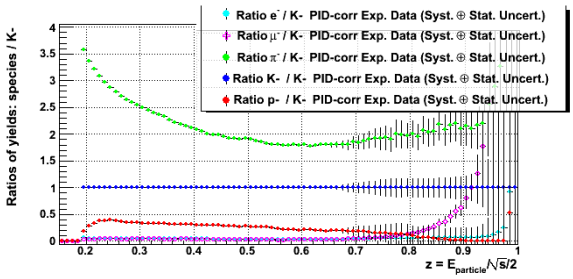


Figure B.25: Ratio between PID-corrected yields and kaon PID-corrected yields for negative track charge. For the ratio, statistical and systematic uncertainties have been added quadratically into a single uncertainty for all PID-corrected spectra.

References

[1] B. Odom *et al.*, Phys. Rev. Lett. **97**, 030801 (2006).

[2] E. D. Bloom *et al.*, Phys. Rev. Lett. **23**, 930 (1969).

[3] M. Breidenbach *et al.*, Phys. Rev. Lett. **23**, 935 (1969).

[4] J. I. Friedman and H. W. Kendall, Ann. Rev. Nucl. Part. Sci. **22**, 203 (1972).

[5] J. Beringer *et al.* (Particle Data Group), Phys. Rev. D **86**, 010001 (2012).

[6] M. L. Perl, E. R. Lee, and D. Lomba, Mod. Phys. Lett. A **19**, 2595 (2004).

[7] X. Artru and G. Mennessier, Nucl. Phys. B **70**, 93 (1974).

[8] B. Andersson *et al.*, Phys. Reports **97**, 31 (1983).

[9] G. Marchesini *et al.*, Comp. Phys. Comm. **67**, 465 (1992); G. Corcella *et al.*, JHEP **0101**, 010 (2001); M. B′ahr *et al.*, Eur. Phys. J. C **58**, 639 (2008).

[10] T. Gleisberg *et al.*, JHEP **0902**, 007 (2009).

[11] O. Biebel, Phys. Rep. **340**, 165 (2001).

[12] J. C. Collins, D. E. Soper, and G. Sterman, Adv. Ser. Direct. High Energy Phys. **5**, 1, 1988.

[13] G. Sterman *et al.*, Rev. Mod. Phys. **67**, 157 (1995).

[14] D. de Florian, R. Sassot, and M. Stratmann, Phys. Rev. D **75**, 094009 (2007).

[15] T. Ito, W. Bentz, I. C. Cloët, A. W. Thomas, and K. Yazaki, Phys. Rev. D **80**, 074008 (2009).

[16] L. N. Lipatov, Yad. Fiz. **20** (1974) 181 [Sov. J. Nucl. Phys. **20** (1975) 94];
V. N. Gribov and L. N. Lipatov, Yad. Fiz. **15** (1972) 781 [Sov. J. Nucl. Phys. **15** (1972) 438];
G. Altarelli and G. Parisi, Nucl. Phys. B **126** (1977) 298;
Yu. L. Dokshitzer, Zh. Eksp. Teor. Fiz. **73** (1977) 1216 [Sov. Phys. JETP **46** (1977) 641].

[17] M. Hirai, S. Kumano, T.-H. Nagai, and K. Sudoh, Phys. Rev. D **75**, 114010 (2007).

[18] S. Albino, B.A. Kniehl, and G. Kramer, Nucl. Phys. B **803**, 42 (2008).

[19] M. Epele, R. Llubaroff, R. Sassot, and M. Stratmann, arXiv:1209.3240 [hep-ph].

[20] D. P. Anderle, F. Ringer, and W. Vogelsang, arXiv:1212.2099 [hep-ph].

[21] I. Helenius, K. J. Eskola, H. Honkanen, and C. A. Salgado, JHEP 1207, 073 (2012); arXiv:1205.5359.

[22] D. de Florian, R. Sassot, M. Stratmann, and W. Vogelsang, Phys. Rev. Lett. **101**, 072001 (2008).

[23] E. Leader, A. V. Sidorov, and D. B. Stamenov, Phys. Rev. D **82**, 114018 (2010).

[24] M. Anselmino and E. Leader, Z. Phys. C **41**, 239 (1988).

[25] D. de Florian, R. Sassot, M. Stratmann, and W. Vogelsang, Phys. Rev. D **80**, 034030 (2009).

[26] D. de Florian, R. Sassot, M. Stratmann, and W. Vogelsang, Phys. Rev. D **80**, 034030 (2009).

[27] W. Melnitchouk and A. W. Thomas, Z. Phys. A **353**, 311 (1995).

[28] ALEPH Collaboration, R. Barate *et al.*, Phys. Rep. **294**, 1 (1998).

[29] DELPHI Collaboration, P. Abreu *et al.*, Eur. Phys. J. C **5**, 585 (1998).

[30] SLD Collaboration, K. Abe *et al.*, Phys. Rev. D **69**, 072003 (2004).

[31] ARGUS Collaboration, H. Albrecht *et al.*, Z. Phys. C **44**, 547 (1989).

[32] CLEO Collaboration, S. Behrends, *et al.*, Phys. Rev. D **31**, 2161 (1985).

[33] M. T. Ronan *et al.*, Phys. Rev. Lett. **44**, 367 (1980).

[34] TASSO Collaboration, W. Braunschweig *et al.*, Z. Phys. C **42**, 189 (1989).

[35] TPC/Two Gamma Collaboration, H. Aihara *et al.*, Phys. Rev. Lett. **52**, 577 (1984).

[36] TPC/Two Gamma Collaboration, H. Aihara *et al.*, Phys. Rev. Lett. **61**, 1263 (1988).

[37] OPAL Collaboration, R. Akers *et al.*, Z. Phys. C **63**, 181 (1994).

[38] H. Kawamura, presentation at Workshop on Fragmentation Functions and QCD 2012 (Fragmentation 2012), RIKEN, Wako, Japan.

[39] M. Stratmann, presentation at Workshop on Fragmentation Functions and QCD 2012 (Fragmentation 2012), RIKEN, Wako, Japan.

[40] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, Phys. Rev. Lett. **13**, 4 (1964).

[41] K. Abe *et al.*, Phys. Rev. Lett. **87**, 091802 (2001).

[42] `http://www-acc.kek.jp/kekb/commissioning/record/History_of_record.html`, last accessed on 01/03/2013.

[43] `http://belle.kek.jp/belle/slides/slide_index.html`, last accessed on 01/03/2013.

[44] H. Aihara *et al.*, Nucl. Instrum. Methods A **568**, 269 (2006).

[45] A. Kibayashi, Nucl. Instrum. Methods A **569**, 5 (2006).

[46] A. Abashian *et al.*; Nucl. Instrum. Methods A **479**, 117 (2002).

[47] E. Nakano, Nucl. Instrum. Methods A **494**, 402 (2002).

[48] K. Hanagaki, H. Kakuno, H. Ikeda, T. Iijima, and T. Tsukamoto, Nucl. Instrum. Methods A **485**, 490 (2002).

[49] A. Abashian *et al.*, Nucl. Instrum. Methods A **491**, 69 (2002).

[50] `http://belle.kek.jp/secured/bgmpdf/2002/0304/haba.ppt`, Belle internal web, last accessed on 01/03/2013.

[51] `http://belle.kek.jp/secured/bgmpdf/2003/0731/haba.ppt`, Belle internal web, last accessed on 01/03/2013.

[52] Personal communication with S. Uno, Belle Collaboration, High Energy Accelerator Research Organization (KEK), Institute of Particle and Nuclear Studies, Japan.

[53] http://belle.kek.jp/secured/kid/other/plot_summer2000/figure.html, Belle internal web, last accessed on 01/03/2013.

[54] For further reference about the EvtGen generator, see http://www.slac.stanford.edu/~lange/EvtGen, last accessed on 01/03/2013.

[55] T. Sjöstrand, Comp. Phys. Commun. **82**, 74 (1994).

[56] For further reference about GEANT see http://wwwasd.web.cern.ch/wwwasd/geant/index.html, last accessed on 01/03/2013.

[57] http://belle.kek.jp/secured/runinfo/, Belle internal web, last accessed on 01/03/2013.

[58] http://belle.kek.jp/group/ecl/private/lum/, Belle internal web, last accessed on 01/03/2013.

[59] For further reference about the PANTHER package, see http://belle.kek.jp/group/software/panther/, Belle internal web, last accessed on 01/03/2013.

[60] T. Higuchi *et al.*, Proceedings tor the International Conference on Computing in High Energy Physics, 1998.

[61] For further information about the ROOT software framework, see http://root.cern.ch/, last accessed on 01/03/2013.

[62] B. Casey, Internal Belle collaboration paper 'Belle Note' (BN) 390 (2001), unpublished.

[63] R. Abe *et al.*, Nucl. Instrum. Methods A **535**, 558 (2004).

[64] http://belle.kek.jp/group/software/slides/EvtCls/soft_EvtCls.html, Belle internal web, last accessed on 01/03/2013.

[65] S. Nishida, Internal Belle collaboration paper 'Belle Note' (BN) 779; Version 0.3, January 17^{th} 2005; unpublished.

[66] KID Group (Belle collaboration), Internal Belle collaboration paper 'Belle Note' (BN) 321 (2000), unpublished.

[67] /sw/belle/belle/b20090127_0910/src/skim/kid_skim/doc/README, Belle internal network, last accessed on 01/03/2013.

[68] G. Majumder, Internal Belle collaboration paper 'Belle Note' (BN) 670 (2003), unpublished.

[69] Y. J. Lee, Internal Belle collaboration paper 'Belle Note' (BN) 684 (2004), unpublished.

[70] W-L. Yeh, Internal Belle collaboration paper 'Belle Note' (BN) 1126 (2009), unpublished.

[71] /sw/belle/belle/b20090127_0910/src/rec/general/v0finder/src/V0CFinder.cc, Belle internal network, last accessed on 01/03/2013.

[72] F. Fang, Internal Belle collaboration paper 'Belle Note' (BN) 323 (2000), unpublished.

[73] /sw/belle/belle/b20090127_0910/src/anal/mdst/src/findKs.cc, Belle internal network, last accessed on 01/03/2013.

[74] http://belle.kek.jp/~hastings/secured/pantherdoc/b20050311_0738/, Belle internal web, last accessed on 01/03/2013.

[75] L. Hinz *et al.*, Internal Belle collaboration paper 'Belle Note' (BN) 777 (2004), unpublished.

[76] L. Hinz, Internal Belle collaboration paper 'Belle Note' (BN) 954 (2006), unpublished.

[77] /sw/belle/belle/b20090127_0910/src/skim/psi_skim/src/PsiSkim.cc, Belle internal network, last accessed on 01/03/2013.

[78] http://www.hepl.phys.nagoya-u.ac.jp/public/tau_ana/wiki.cgi, Nagoya University internal network, last accessed on 01/03/2013.

[79] M. Lefebvre, R. K. Keeler, R. Sobie, and J. White, Nucl. Instrum. Methods A **451**, 520 (2000).

[80] For further reference of the Mersenne Twister algorithm, see http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/ewhat-is-mt.html, last accessed on 01/03/2013.

[81] http://root.cern.ch/root/html/TRandom3.html, last accessed on 01/03/2013.

[82] http://root.cern.ch/download/doc/14LinearAlgebra.pdf, last accessed on 01/03/2013.

[83] Personal communication with C. P. Shen, High Energy Physics Laboratory, Nagoya University, Japan.

[84] E. A. Kuraev and V. S. Fadin; Sov. J. Nucl. Phys. **41**, 466 (1985).

[85] A. Fass, A. Ferrari, J. Ranft, and P.R. Sala, Proc. IV Int. Conf. on Calorimetry in High Energy Physics, La Biodola, Italy (1993); Ed. A. Menzione and A. Scribano, World Scientific, 493.

[86] H. Fesefeldt, RWTH Aachen Report PITHA, 85/02 (1985).

[87] Z. Natkaniec *et al.*, Nucl. Instrum. Methods A **560**, 1 (2006).

[88] Personal communication with M. Stratmann, Physics Department, Brookhaven National Laboratory, USA.

[89] R. Ammar *et al.* (CLEO Collaboration), Phys. Rev. D **57**, 1350 (1998).

[90] e.g. F. Halzen and A. D. Martin, John Wiley & Sons (1984).

[91] The fractions are stored in an ASCII file, available at https://www.box.com/s/os8bts1ht5cxb8hf480a.

[92] S. Albino, B.A. Kniehl, and G. Kramer, Nucl. Phys. B **725**, 181 (2005).

[93] D. Gabbert, Diploma Thesis in Physics at the Technische Universität München (2005).