Dissertation

submitted to the

Combined Faculties of the Natural Sciences and Mathematics

of the Ruperto-Carola-University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Put forward by

Dipl.-Phys. Thomas Nikodem

born in: Paderborn, Germany

Oral examination: 16.12.2015

Angular analysis of $B^0 \to K^{*0}\mu^+\mu^-$ decays using $3\,\text{fb}^{-1}$ of integrated luminosity at LHCb

Referees: Prof. Dr. Stephanie Hansmann-Menzemer

Prof. Dr. Peter Glässel

Abstract: An angular analysis of the decay $B^0 \to K^{*0}(\to K\pi)\mu^+\mu^-$ is performed on a data set from LHCb Run 1 corresponding to a total integrated luminosity of $3\,\text{fb}^{-1}$. A complete set of *CP*-averaged angular observables and their correlations are determined simultaneously in intervals of the dimuon invariant mass squared, q^2. Special care is taken to separate the two contributions in which the kaon and the pion either originate from a spin 1 or a spin 0 resonance.

The measurement reveals a good agreement with Standard Model predictions. However, observables S_5 and S_{6s} show small discrepancies with respect to the predictions in the q^2 region $1-6\,\text{GeV}^2$. The measurement is also performed in a second angular basis, confirming a previously observed local deviation in one of its observables, P_5'. The agreement between the measurement of P_5' and predictions in the q^2 region $0.1-8\,\text{GeV}^2$ corresponds to a p-value of $8.2\cdot10^{-4}$ only. The observations are compatible with theoretical models which predict additional physics contributions at the TeV scale.

Kurzfassung: Eine Winkelanalyse des Zerfalls $B^0 \to K^{*0}(\to K\pi)\mu^+\mu^-$ wird durchgeführt, wobei der Datensatz von LHCb Run 1 verwendet wird, der einer integrierten Luminosität von $3\,\text{fb}^{-1}$ entspricht. Ein kompletter Satz von *CP* gemittelten Observablen der Winkelverteilung sowie deren Korrelation wird simultan in Intervallen der quadratischen invarianten Masse des $\mu^+\mu^-$ Systems, q^2, bestimmt. Besonders wurde darauf geachtet die Beiträge zu unterscheiden, in denen das Kaon und das Pion entweder von einer Spin 1 oder einer Spin 0 Resonanz herrühren.

Die Messungen stimmen gut mit den Standardmodell Vorhersagen überein, aber es gibt kleine Diskrepanzen in der q^2 Region $1-6\,\text{GeV}^2$ für die Observablen S_5 und S_{6s}. Zusätzlich wird die Messung auch in einer zweiten Basis vorgenommen, in der in einer vorangegangenen Messung in der Observablen P_5' eine lokale Diskrepanz mit den Vorhersagen gesehen wurde. Die Diskrepanz ist auch in dieser Messung sichtbar und die Messung von P_5' stimmt mit Vorhersagen in der q^2 Region $0.1-8\,\text{GeV}^2$ nur mit einem p-Wert von $8.2\cdot10^{-4}$ überein. Die Messungen sind mit den Erwartungen theoretischer Modelle kompatibel, die eine zusätzliche physikalische Komponente auf der TeV Skala vorhersagen.

Contents

1 Introduction

Since the early days mankind wondered about what keeps the world together in the innermost. In the age of high energy particle colliders the understanding of the inner structure of matter has reached a new frontier. Elementary particles, such as quarks and leptons, are identified as the smallest building blocks of the universe known today. A sophisticated model of particle physics describes these particles and their interactions, the so called Standard Model.

As the name *model* already suggests, it is only a description of the observed nature and not the final all-explaining theory. Nevertheless, in the current and last century a multitude of particle properties and dynamics were measured each in agreement with Standard Model predictions. On the other hand there are still open questions which cannot be answered with the current understanding. Especially the origin of the matter-antimatter asymmetry observed in the universe remains to this day inexplicable.

A deeper understanding in particle physics can be achieved by finding not yet discovered elements and processes, so called *New Physics*. In the presence of many theories which provide explanations beyond the Standard Model new additional particles are expected to reveal at higher energy scales. The discovery or exclusion of these particles would allow to verify and falsify these theories. Besides directly searching for new heavy particles in the form of resonances it is also advantageous to indirectly search for them. In certain processes virtual particles appear, whose invariant mass can be orders of magnitude larger than the total energy of the system. Therefore energy scales can be tested which are not accessible in direct searches. Deviations of measured observables from their predictions would hint to the existence of New Physics.

One of the most famous of such indirect measurements is the search for the decay $K_S^0 \to \mu^+\mu^-$ done in 1969 [1]. The absence of a signal of that decay was explained by the Glashow-Iliopoulos-Maiani (GIM) mechanism [2]. To the at the time known three light quarks (u, d, s) a fourth quark and a corresponding additional decay amplitude were postulated to destructively interfere with the already known physics. Today, not only this fourth quark is a basic component of the Standard Model, but also two additional quarks which were later also predicted and discovered.

One of the major physics facilities dedicated to the study of particle physics is the European Organization for Nuclear Research (CERN). Today over 10 000 scientists from over 600 universities and over 100 nations are working on research at CERN, especially on measurements related to the Large Hadron Collider (LHC). The LHC is a synchrotron with a circumference of 27 km located near Geneva. To this date it is the most powerful particle collider in the world, concerning both luminosity and centre-of-mass energy of the collisions. There are four major experiments located at the LHC, namely Alice, Atlas, CMS and LHCb, each with a specific physics programme. Already great success is achieved by discovering the predicted Higgs particle [3, 4], which is related to the award of the Nobel price of physics in 2013 to

Peter Higgs and François Englert. Confirmation of the Standard Model predictions is obtained in many further measurements.

However, in recent years several statistically significant deviations from predictions are found. Especially decays, in which a b quark transforms into an s quark, so called $b - s$ transitions, cause excitement in the physics community. The rate of such $b - s$ transitions is suppressed in the Standard Model, and therefore in comparison to the known processes New Physics processes can make a potentially large contribution. One of the most promising recent analyses is the measurement of the angular observables of the decay[1] $B^0 \to K^{*0}\mu^+\mu^-$ performed at LHCb [5]. In one observable a local deviation from the Standard Model prediction of $3.7\,\sigma$ is observed.

In this thesis an update of the $B^0 \to K^{*0}\mu^+\mu^-$ angular analysis will be presented. Not only the event yield will be increased by a factor of about two with respect to the previous analysis, but the complete set of angular observables including their correlation will be measured for the first time. Physics studies at the LHC are performed in large analysis groups. The studies presented in this thesis are related to the work and fruitful discussions with many colleges. Also it is often preferred to cross check obtained results by doing the measurement independently multiple times. The official publication of the update of the $B^0 \to K^{*0}\mu^+\mu^-$ analysis using an integrated luminosity of $3\,\text{fb}^{-1}$ of data, which includes a significant contribution from the author of this thesis, is Ref. [6]. In this thesis the identical event selection (see Sec. 7) and the identical simulation of the full detector (see Sec. 6.1) is used, whereas remaining studies are performed independently. It is checked that within expected statistical fluctuations results agree with the ones from Ref. [6].

In the first sections of this thesis the Standard Model of particle physics (see Sec. 2), the LHCb experiment (see Sec. 3), and its physics programme (see Sec. 4) are introduced. In the remaining part the analysis of the author is discussed in detail. An analysis overview is given in Sec. 5.

[1] The inclusion of charge conjugate processes is implied throughout this thesis.

2 The Standard Model of particle physics

The Standard Model of particle physics describes elementary particles and their interaction. The basic concepts of the Standard Model are briefly introduced in Sec. 2.1. The CKM mechanism which is especially important for the description of the decay $B^0 \to K^{*0}\mu^+\mu^-$ is explained in Sec. 2.2. In recent times the Standard Model is severely tested for validity. Two of these tests are outlined in Sec. 2.3. Nevertheless, there are still open questions which motivate a further deeper investigation of the Standard Model, as discussed in Sec. 2.4. The introduction to the Standard Model given here follows Refs. [7–9].

2.1 Particles and interactions

The Standard Model of particle physics describes the physics at the smallest scales which are known today. The once thought to be indestructible atom is found to be build up out of electrons and the atom nucleus. The nucleus can be further divided into neutrons and protons, consisting of again smaller particles, the so called quarks. Quarks and electrons are smaller than the smallest scale which can be measured today ($< 10^{-18}\,\mathrm{m}$). Both are described in the Standard Model as point like particles.

Nevertheless, the elementary particles itself can be categorized by many different properties. For example each particle has a so called spin. Further properties are different charges corresponding to elementary forces and a property called mass. The spin of a particle can be treated as an intrinsic angular momentum. Its origin is still not completely understood, as in the classical sense, the angular momentum of a point like structure must be identical to zero. The spin of particles has a discrete value, either an integer or an half of an integer[2]. In the first case, corresponding particles are so called bosons, in the latter case so called fermions. Fermions must obey the Pauli exclusion principle, which is not the case for bosons.

Today there are identified several different elementary particles, which differ besides others in the value of their spin. The full list of particles is sketched in Fig. 1. There is one spin 0 particle, the Higgs boson, six quarks and six leptons with spin 1/2 and finally 4 gauge bosons with spin 1. For each fermion there exists additionally a particle with opposite quantum numbers, the so called antiparticle.

The interaction of particles is described by the exchange of the gauge bosons. Each of the gauge bosons belongs to an elementary force. All quarks, and three of the leptons (e, μ, τ) have an electrical charge and thus couple to the photon. The property which describes if a particle couples to the strong interaction is the colour-charge. Only quarks and gluons carry colour-charge. The colour-charge can be red, green or blue. Like mixing light of these three colours results in the colour

[2]For simplicity throughout this thesis it is referred to Planck units defined as: $c = \hbar = 1$.

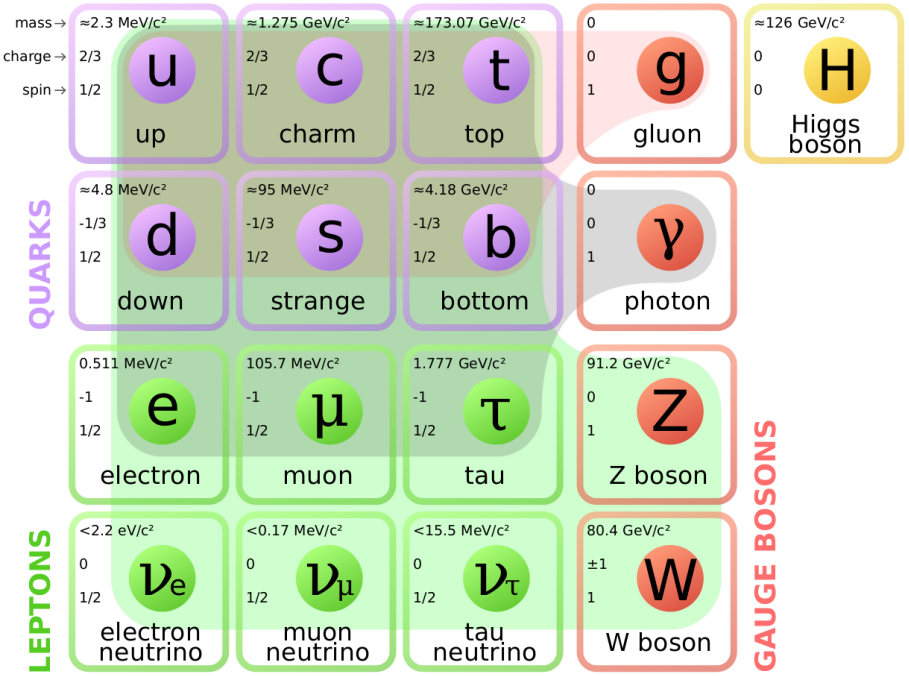

Figure 1: Elementary Particles described by the Standard Model of particle physics. [10]

white, the net colour-charge of an object containing all three colours is zero. The colour-charge of antiquarks is denoted as anticolour. Combining a quark with an anti-quark containing the corresponding anticolour also results in cancelling of the colour-charge. Due to the confinement of quarks no free particle must contain a net colour-charge. Contrary to the photon, which is free of electrical charge, the gluons carry colour-charge, which results in gluon self-interaction. The W and Z bosons couple to all other particles including themselves. All particles containing a spin also have a binary property called chirality. Particles can be either left handed or right handed, denoted with L, R respectively. In the limit of large momentum ($E \gg m$) chirality gets identical with the helicity of a particle, which expresses if the spin of a particle points in the direction of its motion (right handed) or in the opposite direction (left handed). The W boson couples only and the Z boson preferably to particles with left handed chirality and antiparticles with right handed chirality. Therefore in the weak interaction (and only in the weak interaction) physics is not invariant under both charge and parity transformation. Gravity, which is at the scale of quarks and leptons more than 30 orders of magnitude weaker than the other interactions, is not yet described in the context of the Standard Model.

In the framework of Quantum Field Theory (QFT) particles are treated as excitations of a field. There are fermionic fields (also called Dirac fields) which are

related to spin 1/2 particles and bosonic fields related to spin 1 particles. Additionally in the Standard Model there is a scalar field, the so called Higgs field, to which all fermions couple via a so called Yukawa interaction. The coupling strength of the particles to this field is described by a property called mass. Due to spontaneous symmetry breaking the vacuum expectation value of the Higgs field is non-zero, resulting in a non-zero mass of all fermions. A summary of the mass of the different particles is given in Fig. 1.

The Higgs boson is an excitation of the Higgs field, coupling to other particles according to their mass. The particles containing the largest mass and therefore have the largest coupling to the Higgs boson are the W/Z bosons, the t quark and the Higgs boson itself.

The unification of electromagnetic and weak interaction is the so called electroweak theory. The photon and W/Z bosons are linked to the same origin, namely the massless bosons W^{1-3} and B. Due to the spontaneous symmetry breaking of the Higgs field the observed W/Z bosons and the photon are superpositions of these bosons.

All couplings in the Standard Model can be described by different symmetry groups. The electromagnetic interaction is described by $U(1)$, the weak interaction by $SU(2)$ and the strong interaction by $SU(3)$.

As indicated before, quarks usually appear bound in the composition of larger objects. Due to confinement they are not observed alone, but always either in the combination of three (anti)quarks or the combination of a quark and an antiquark. Objects consisting out of three quarks are called baryons, whereas particles made up of a quark and an antiquark are called mesons. In contrast leptons appear in nature also as single particles. Due to the different interactions, if it is allowed, particles decay to lower energetic states. The only stable combination of quarks is the proton (consisting out of two u and a d quark), while the only stable leptons are the electron and the neutrinos[3]. All other baryons, mesons and leptons decay after a certain time into these stable particles. As long as the quark and lepton content of the total system stays identical decays are mainly described by the strong and electromagnetic interaction. However, the transition from quarks and leptons into different quarks and leptons is only allowed by the weak interaction. The mechanism responsible for quark transitions is the CKM mechanism, which is explained in the next subsection. The time-scale of the decays is of the order of 10^{-24} to 10^{-21} s for processes mediated by the strong interaction, 10^{-19} to 10^{-17} s for processes dominated by the electromagnetic interaction, while only of the order of 10^{-13} to 10^{3} s if the only allowed transition is with involvement of the weak interaction.

[3]Due to their oscillation neutrinos mix into each other. Neutrino oscillation is a topic of its own and will not be discussed in this thesis. Instead it is referred to any newer Physics textbook, e.g. Ref. [7]

2.2 The CKM mechanism

The Lagrangian describing the Standard Model consists of three parts. In the flavour basis, which describes the interaction of particles, it can be written as:

$$\mathcal{L}_{SM,flavour} = \mathcal{L}_{kinetic} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa} \tag{1}$$

where $\mathcal{L}_{kinetic}$ describes the different fermion fields, $\mathcal{L}_{Higgs}$ the Higgs field and $\mathcal{L}_{Yukawa}$ the interaction of the fermions with the Higgs field.

In the flavour basis the fermions are Eigenstates of the system and thus also the matrix describing the kinetic Lagrangian has a diagonal form. Also the matrix in the Higgs Lagrangian is diagonal. However, the matrix in the Yukawa part of the Lagrangian is not diagonal, as particles of different generations couple to each other.

The time evolution of a system is described by the mass Eigenstates. Therefore the basis must be changed such that the matrix in the Yukawa part of the Lagrangian, which describes the mass terms, is diagonalized. The corresponding Lagrangian is written as:

$$\mathcal{L}_{SM,mass} = \mathcal{L}_{flavour} + \mathcal{L}_{Higgs} + \mathcal{L}_{mass} \tag{2}$$

where $\mathcal{L}_{mass}$ consists now of a diagonal matrix. The diagonal entries correspond to the mass of the different particles. In the diagonalization process $\mathcal{L}_{Higgs}$ remains diagonal, however, this does not apply for the kinetic part, which is now expressed by $\mathcal{L}_{flavour}$. In the time evolution $\mathcal{L}_{flavour}$ describes the mixing of different flavours. In the leptonic part (assuming massless neutrinos) also here the matrix remains diagonal. However, in the hadronic part there appear non-diagonal entries. The corresponding matrix, which describes the mixing of different quark species into each other, is called CKM matrix.

The CKM matrix is defined as:

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{3}$$

where V_{ij} is a complex scaling factor of the amplitude of the coupling of two quarks of flavour i, j and a $W^{\pm}$ boson. In such a process the quark flavour is not conserved. The size of the matrix elements is such that coupling of quarks in the same generation is strongest and weakest between the 1^{st} and 3^{rd} generation. Quarks coupling to quarks with the same charge, so called Flavour Changing Neutral Currents, are in first order not observed in nature. In the Standard Model this is only allowed in so called loop processes. Heavy virtual particles present in these loops significantly reduce the decay rate of these processes. Two corresponding Feynman diagram examples are shown in Fig. 2.

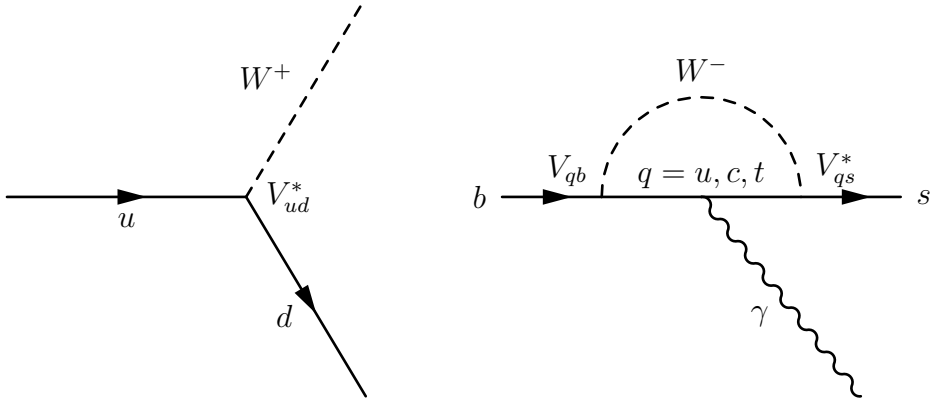

Figure 2: Left: Transition of a u into a d quark, coupling to a W boson. The coupling strength of such a process is proportional to the CKM matrix element V_{ud}^*. Right: An example of a transition of a b into an s quark. Such Flavour Changing Neutral Currents are in the Standard Model only allowed via loop processes. Due to its large mass and the structure of the CKM matrix the exchange of a t quark is favoured in such processes.

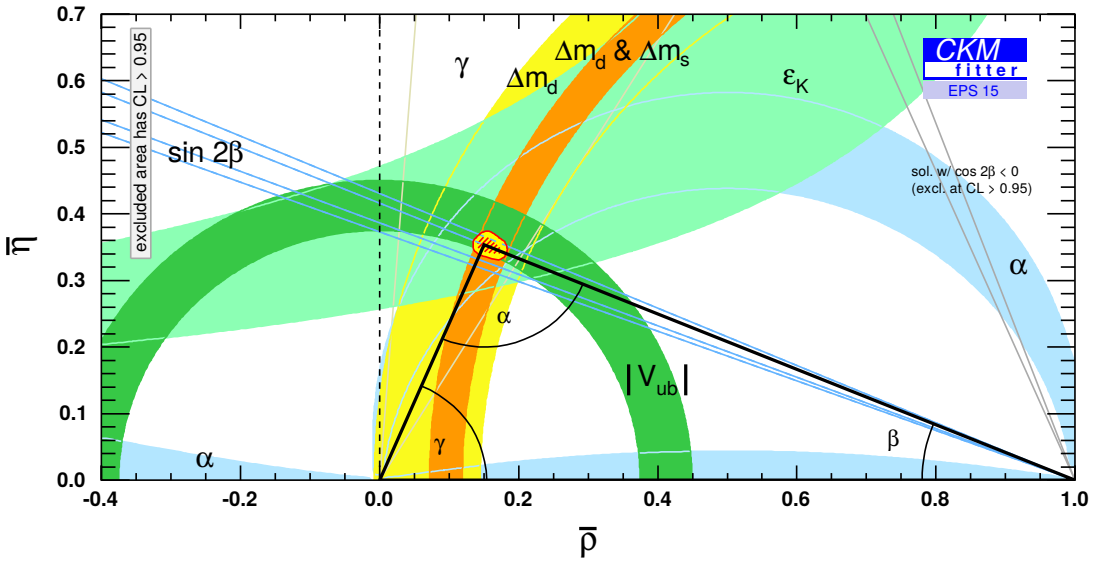

Figure 3: The unitary triangle determined by the CKM Fitter. [11]

2.3 Testing the Standard Model

The Standard Model has been tested over the past years performing a multitude of different experiments. Although both theoretical predictions and experimental measurements got more precise no significant deviations are found. Additionally the completeness of the Standard Model can be tested by verifying the unitarity of the CKM matrix. If there would be an additional contribution (as *i.e.* a light 4th generation of quarks) the different transformation probabilities of the known quarks would not add up to one and thus the CKM matrix would not be unitary.

The unitarity requirement of the CKM matrix can be expressed in several relations. The one where the largest sensitivity is expected is:

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0 \tag{4}$$

The sum of these three complex numbers can be also represented as a triangle. Usually this is done in a complex plane, which is defined such that the upper right side of the triangle is $V_{td}V_{tb}^*/V_{cd}V_{cb}^*$, the upper left side is $V_{ub}V_{ud}^*/V_{cd}V_{cb}^*$ and the lower side is unity. With different measurements this triangle can be over-constrained and the validity of the Standard Model can be tested. The results of different CKM element measurements are shown in Fig. 3. So far all measurements show an amazing compatibility with the Standard Model.

Another success in confirming the Standard Model is the measurement of the oscillation parameters of the B_s^0 meson [12]. Due to the flavour changing nature of the charged weak interaction a B_s^0 particle can transform into its own antiparticle and vice versa (see Fig. 4). This process is described by so called box diagrams which are sensitive to virtual not yet discovered heavy particles, which would interfere and affect the observed physics. Two parameters describing this oscillation, which have a precise theory prediction and which also can be measured precisely at the LHCb experiment, are $\Delta\Gamma_s$ and ϕ_s. In Fig. 5 the theory prediction and also the latest measurements of both parameters are shown. Also here the measurement agrees very well with the predictions.

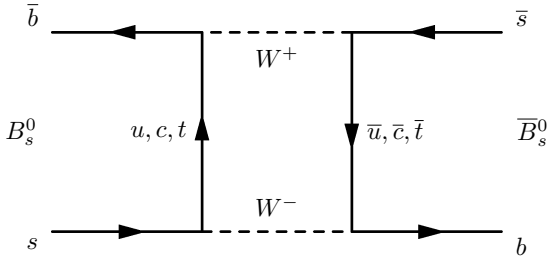

Figure 4: A so called box diagram, by which the B_s^0 meson can oscillate into its own antiparticle. Like the virtual W boson in this process, also new heavy undiscovered particles are expected to contribute.

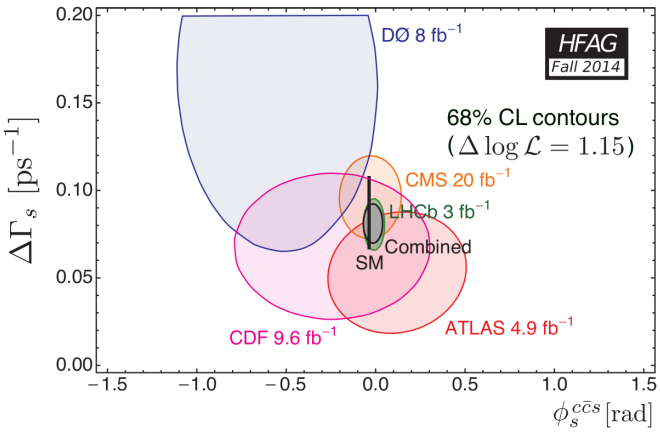

Figure 5: Theory prediction and measurements of B_s^0 mixing parameters. [13]

2.4 Open questions

In spite of the great success of the Standard Model so far there are still several open questions. Especially on the very large scale there is still no final answer how gravity - which can be completely neglected in the subatomic scale - can be included in the theory. Therefore also many observations from astronomy and cosmology are not fully understood in the context of particle physics. For example the origin of the so called Dark Matter has not yet been revealed. Furthermore, the visible and known energy density of the universe is too small to describe the calculated expansion after the Big Bang and therefore there is so called Dark Energy expected. Currently only 5% of the energy content of the universe, which corresponds to light and matter, is understood.

On the other hand there also exist fundamental problems within the Standard Model of particle physics itself. In the following shortly two main arguments against the completeness of the Standard Model are discussed.

2.4.1 The hierarchy problem

The question is how probable something must be that one believes that it happened by chance and how improbable it would have to be that one is certain that there is a reason behind it. Although this is in some sense more a philosophical question, since the late 70s this is discussed also in the physical community under the topic of naturalness. For example as stated by 't Hooft:

> At any energy scale μ, a physical parameter or set of physical parameters $\alpha_i(\mu)$ is allowed to be very small only if the replacement $\alpha_i(\mu) = 0$ would increase the symmetry of the system. - 't Hooft [14].

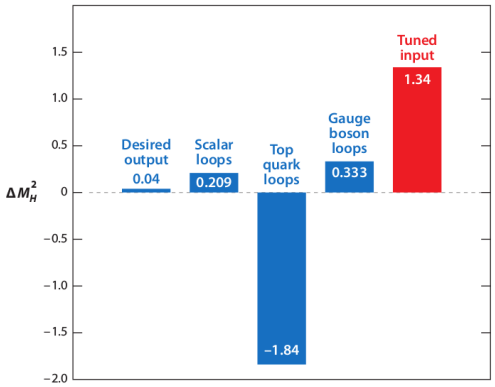

Figure 6: Quantum corrections (in the units of TeV^2) contributing to the Higgs boson mass, M_H, with a cut-off scale of 5 TeV . The fine tuning needed to match with the observed Higgs boson mass is coloured in red. [9]

If there are many different effects of order one present it is *unnatural* that the sum of these effects is very small just by chance.

The Hierarchy Problem in the Standard Model is the relative smallness of the Higgs boson mass. Certain quantum corrections affecting the Higgs boson mass are known and also can be calculated, like for example loop corrections due to gauge bosons and the t quark. There are also unknown corrections, as for example contributions from quantum gravity. As the Higgs boson mass is measured the total size of all quantum corrections can be estimated. The calculations of the Standard Model are valid up to a certain, however, unknown scale. When calculating the quantum corrections to the Higgs boson mass they have to be integrated up to this scale. In Fig. 6 the known quantum corrections (blue) and also the still unknown corrections (red) to the Higgs boson mass are shown for a scale of 5 TeV[4]. Compared to the observed value of the Higgs boson mass the size of quantum corrections is larger by more than two orders of magnitude. In the Standard Model there is no explanation why these quantum corrections should cancel each other out. Furthermore, there does not exist a reason why the cut-off scale should be 5 TeV. If there would be no New Physics a natural choice of the cut-off scale would be either the Planck scale[5] ($\sim 10^{19}$ GeV), or the GUT scale[6] ($\sim 10^{15}$ GeV). If in the calculations one would integrate not until 5 TeV but a scale several of orders larger also the size of the quantum corrections would get accordingly larger making it highly unlikely that all independent quantum corrections cancel each other out nearly perfectly. Thus it can

[4]For simplicity throughout this thesis it is referred to Planck units defined as: $c = \hbar = 1$.

[5]At the Planck scale the effect of gravity can no longer be neglected in the description of subatomic particles. Therefore beyond this scale physics can no longer be described by the Standard Model.

[6]In the Grand Unified Theory (GUT) is is assumed that the electromagnetic, weak and strong force unify at a certain scale. Beyond this scale physics must be described by only one unified force.

be assumed that something new (as *i.e.* a new heavy particle) either limits the upper limit of the integral or the integrands are reduced significantly above a certain value.

2.4.2 The flavour problem

The second hint for New Physics is that in nature there is usually a reason why patterns appear. For example in 1885 Balmer observed patterns in spectral lines of hydrogen [15]. These patterns could later be related to the quantized energy levels of hydrogen, and can nowadays be explained by the spin structure of the electron and proton, a theory which was not known at that time.

In the Standard Model there are a lot of free parameters. In the electroweak theory there are the coupling constants, quark and lepton masses, the structure of the CKM matrix and the shape of the Higgs potential which cannot be calculated but must be plugged in by hand. Similarly to patterns in the spectral lines of hydrogen in these parameters patterns are visible, which cannot yet be explained.

The mass of the fermions can be explained by the spontaneous symmetry breaking of the Higgs field, however, the coupling of the fermions to this field and thus the value of the their mass is completely arbitrary. Nevertheless, both, quarks and leptons are arranged in three generations with rising masses. In addition also a pattern is visible in the CKM matrix, which is also related to the sorting of generations. The CKM matrix elements corresponding to a transition in the same quark generation are largest, while they are smallest for transitions including the first and the third generation. The current Standard Model theory does not predict any of these parameters nor give an explanation for the observed patterns.

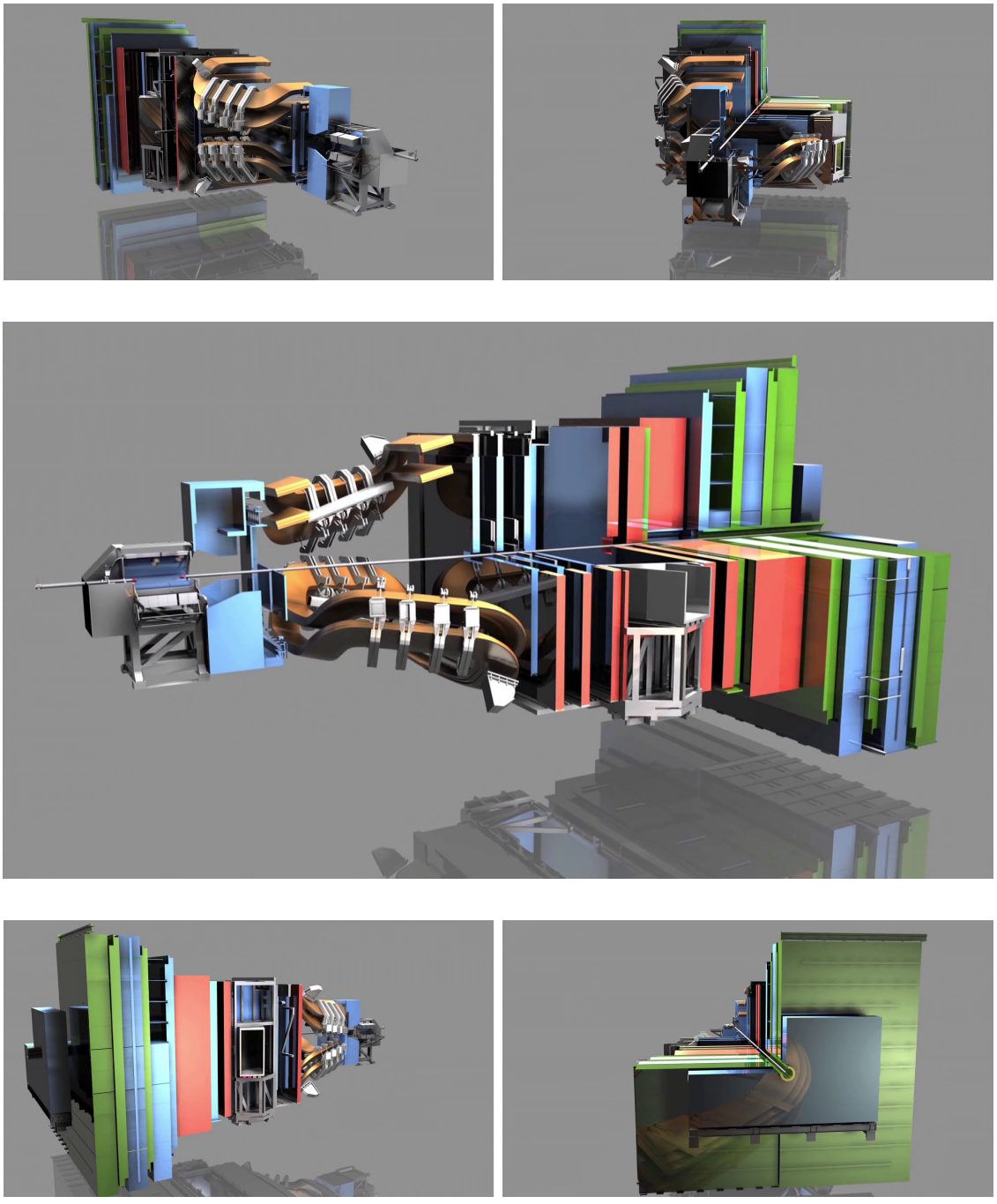

Figure 7: Cut-away drawings of the LHCb forward spectrometer. [16]

3 The LHCb experiment

The LHCb experiment is located at the Large Hadron Collider (LHC), which is a synchrotron with a circumference of 27 km positioned near Geneva, about 100 m underground. It is the most powerful particle collider in form of centre-of-mass energy and luminosity in the world, colliding protons with a centre-of-mass energy of up to 13 TeV.

Two proton beams are accelerated in opposite directions and they collide at four points, where at each point an experiment is located. Two experiments, Atlas [17] and CMS [18], are multi-purpose detectors besides others searching for new heavy particles. The discovery of the Higgs boson can be attributed to these two experiments [3, 4]. The third experiment, Alice [19], is dedicated to the study of quark gluon plasma. Besides protons also heavy lead ions are collided at LHC producing a very hot and dense quark gluon mixture. At last there is also the LHCb experiment investigating especially b physics.

The protons colliding at the LHC are made up of the valence quarks (uud), sea quarks (quarks and antiquarks) and gluons. The leading contribution of heavy quark production is the fusion of a quark and an antiquark and gluon scattering (see. Fig. 8). In contrast to electron-positron collisions, only a fraction of the total energy of the protons is available. Due to the shape of the parton distribution function of the proton it is highly unlikely that both scattering particles carry exactly the same momentum. Therefore the produced b hadrons are in general heavily boosted in forward direction.

The maximum luminosity which can be provided by the LHC is larger than that which can be efficiently processed by the LHCb detector. The beams are slightly shifted transversely at the collision point to reduce the collision rate. During data taking this shift is reduced again such that in spite of the decreasing number of

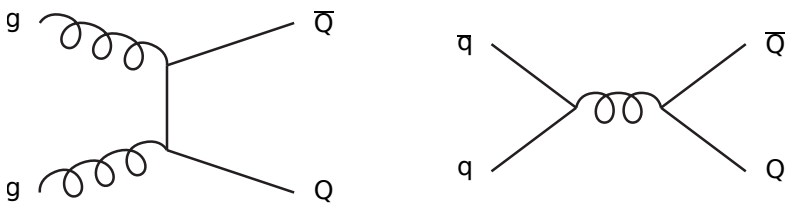

Figure 8: Leading order contributions to heavy quark production at the LHC: Scattering of two gluons (left) and quark antiquark fusion (right). [20]

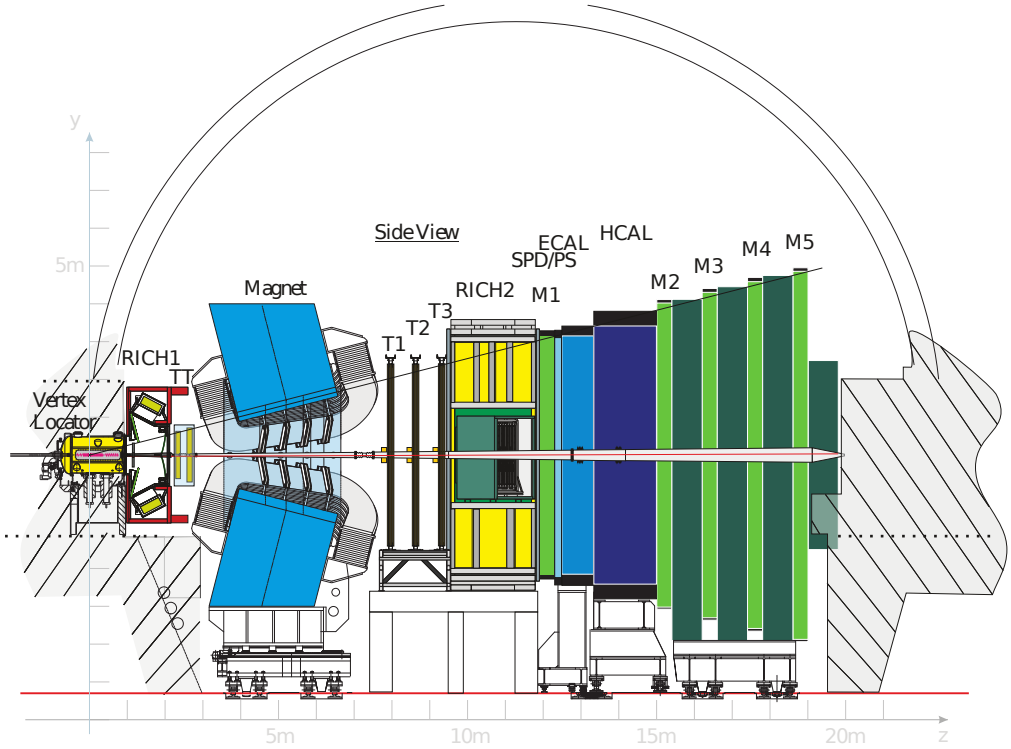

Figure 9: Sketch of the LHCb detector (side view). The collision takes place in the Vertex Locator on the left side of the picture. [21]

protons in the beam, a constant instantaneous luminosity is ensured. Operating always with a similar detector occupancy facilitates to use a time independent trigger strategy and reduce largely systematic effects related to data taking conditions.

During the Run 1 at LHC (in the years 2011 and 2012) proton collisions predominately occurred at a rate of 20 MHz. The centre-of-mass energy of the collisions was 7 TeV in 2011 and 8 TeV in 2012. The data were taken at LHCb with instantaneous luminosities of about $3.5 \cdot 10^{32}\,\mathrm{cm^2 s^{-1}}$ (2011) and $4.0 \cdot 10^{32}\,\mathrm{cm^2 s^{-1}}$ (2012), which are both larger than the design luminosity by up to a factor two. The integrated luminosity is about $1\,\mathrm{fb^{-1}}$ (2011) and $2\,\mathrm{fb^{-1}}$ (2012) corresponding to in total about 10^{12} $b\bar{b}$ quark pairs which are produced.

Despite working at a *messy*[7] hadron collider, at the LHCb experiment it is possible to produce large clean samples of b hadron decays. To achieve this a large

[7]In comparison to the electron-positron colliders, in hadron colliders the exact centre-of-mass energy of the two interacting partons is unknown and there are a lot of particles created in the collision unrelated to the signal decay.

amount of background, several orders of magnitude larger than the signal, must be identified and rejected. A precise detector with a good resolution is needed, but also a trigger system which fast and efficiently processes the information. The LHCb detector is presented in Sec. 3.1 and the trigger is discussed in Sec. 3.2. In Sec. 3.3 the framework of the data processing used at the LHCb experiment is shortly outlined.

3.1 The LHCb detector

A general overview over the LHCb detector is given, only focusing on details important for the presented analysis. A full detailed description can be found in Ref. [21] while performance measurements are summarized in Ref. [22].

The LHCb detector is a single-arm forward spectrometer optimized for the detection and identification of particles in the forward region with a pseudorapidity of 1.6 to 4.9. It is in this forward region in which a large amount of the produced b hadrons traverse the detector. Although the collisions are occurring symmetrically, due to reasons of cost efficiency it is convenient to only build the detector in one direction. In the LHCb detector acceptance there are about 27% of all produced particles containing a b quark.

Cut-away drawings of the LHCb detector are shown in Fig. 7 and a sketch of the side-view is presented in Fig. 9. The proton-proton collision point is inside the Vertex Locator, located on the very left side of the detector. The produced b hadrons usually decay before interacting with any material and are not measured directly. The stable decay products (muons, pions, kaons) then traverse the different sub-detectors and their kinetic properties are precisely determined.

As mentioned earlier due to the collision of two hadrons there is a large amount of background which must be distinguished from the interesting b physics. This is achieved by mainly two aspects of the produced b hadrons. Due to the large mass of the b hadrons the final state decay products have in general a large transverse momentum. As the b quark can only decay into another quark generation the lifetime of the b hadrons is long (flight distance in laboratory of order of a few mm). Therefore the decay vertex separates from the proton collision point enough to be nicely measured.

The LHCb detector is designed to exploit these features by measuring the corresponding quantities as precise as possible. Furthermore, a sophisticated particle identification system takes an important role in suppressing background. The purpose of the for this analysis important sub-detectors is explained in the following[8].

[8]The calorimeter system (ECAL/HCAL) is not used in this analysis.

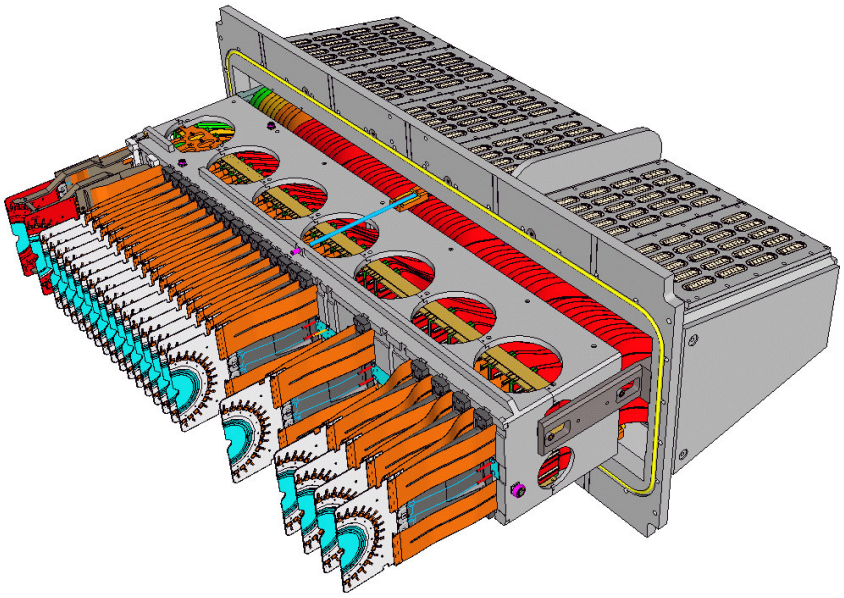

Figure 10: Sketch of the right half of the Vertex Locator. During data taking the sensitive area (blue) is moved up to 8 mm close to the beam axis to provide the best possible vertex resolution. [23]

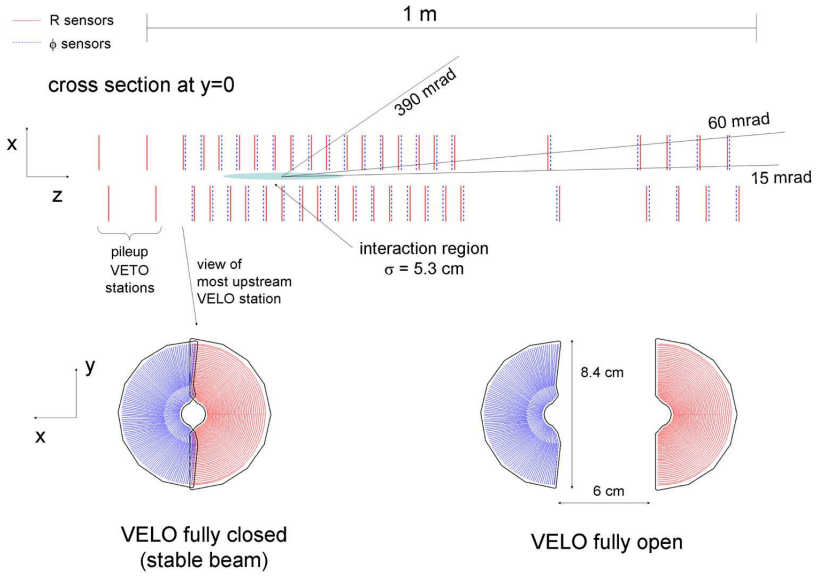

Figure 11: R (red) and ϕ (blue) modules of the VELO, side view (top), front view (bottom). The VELO is closed during data taking (bottom left) and open for beam initialization (bottom right). [21]

3.1.1 Vertex measurement

The detector positioned nearest to the proton collision point is the VErtex LOcator (VELO) [24]. It is a silicon strip detector, consisting out of 46 modules distributed over about 1 m (see Fig. 10 and Fig. 11). The flight distance of b hadrons is in the LHCb laboratory frame of order of a few mm, after which they decay into the more stable kaons, pions and muons. The trajectory of these charged particles, also called track, is measured first by the VELO.

There are two types of modules, R modules and ϕ modules, respectively measuring the distance from the track to the beam axis and the pseudoazimuthal angle of the track. In the ϕ sensors the silicon strips are placed about orthogonal to the beam axis, whereas in the R sensors strips are curved such they are centred around the beam axis. Combining measurements of both module types a 3D spatial track can be reconstructed. The magnetic field inside the VELO is negligible. Therefore, as most particles originate from very close to the beam axis, the projection of a track in the R coordinates is in good approximation for most of the tracks a straight line. The design with R and ϕ sensors in principle offers a good possibility to quickly identify and reconstruct tracks.

To measure precisely the vertices it is necessary to reduce as much as possible the distance which must be extrapolated from the first measurement to the vertex point. The limiting factor for the position of the detector modules is the spread of the beam during its initialization. Therefore the VELO consists out of two movable parts which are separated during beam initialization and are moved up to 8 mm close to the beam axis if there is stable beam. The VELO is placed inside a vacuum chamber directly around the beam, which is at this point not protected by a beam-pipe.

The average cluster finding efficiency is about 99.5%. The single hit resolution depends on the position in the detector and is between 5 µm and 25 µm. The track finding efficiency depends on the position of the track in the detector and is typically above 98%. For high momentum particles the impact parameter (distance extrapolated track to vertex) resolution is up to 13 µm orthogonal to the beam axis and even for very low momentum tracks below 1 GeV it stays smaller than 80 µm.

3.1.2 Momentum measurement

The momentum of particles is determined by measuring their bending inside the magnetic field. A dipole magnet with a field strength of $\int B\mathrm{d}l = 4\,\mathrm{Tm}$ is installed between the VELO and the tracking stations T1-T3. The particles are bended horizontally. During data taking the polarization of the magnetic field is regularly reversed to reduce systematic uncertainties related to the reconstruction efficiency.

After the tracks are precisely determined in the VELO there needs to be a precise measurement after the magnet to measure the deflection of particles. This is done by the Outer Tracker [25], a straw tube detector, and mainly for the high momentum tracks near the beam pipe by the Inner Tracker [26], a silicon strip detector. Both detectors are placed in three stations about $8 - 10$m away from the VELO.

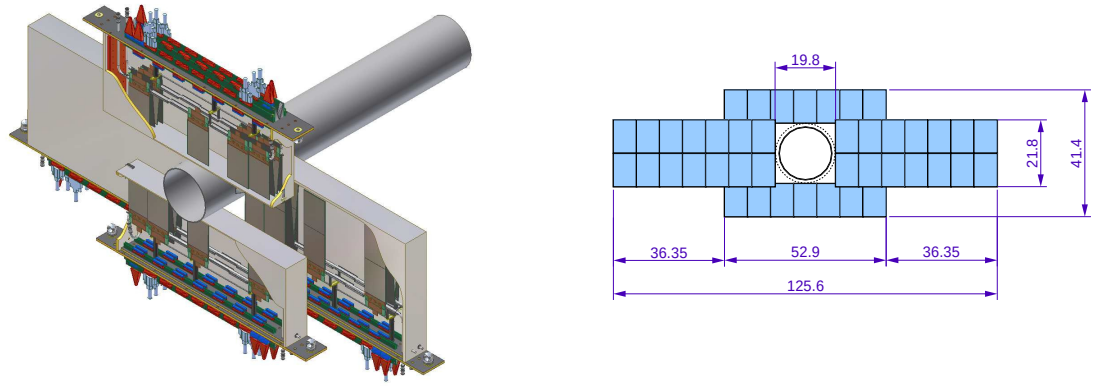

Figure 12: 3D cut-away drawing of the second Inner Tracker station (left). A sketch of the X-modules (right). [21]

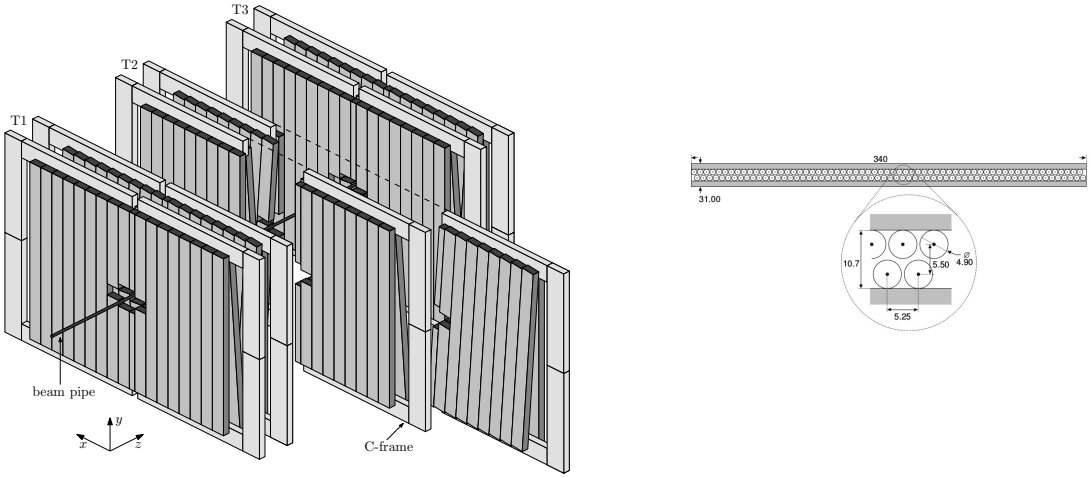

Figure 13: 3D drawing of all three Outer Tracker modules (left). Each module is made out of a double layer structure of straw tubes (right). [21]

In all stations there are four layers (arranged as X-U-V-X), where in X-layers straws/strips are oriented vertically, and in U/V-layers straws/strips are tilted by $+/-5$ degrees with respect to the X-layers. This configuration offers the best resolution in the horizontal direction, in which the particles are bended. Additionally there is the Tracker Turicensis (TT), which consists out of the same technology as the Inner Tracker, and is placed before the magnet to further improve the tracking performance.

The Inner Tracker has a dimension of about 126×40 cm and is placed closest around the beam pipe (see Fig. 12). Due to being a silicon strip detector it offers a good spatial resolution and radiation hardness. Measuring the charge spread inside the strips the resolution can be determined slightly better than the separation of the strips and is about 50 µm. The detection efficiency is about 99.8%.

The Outer Tracker layers (see Fig. 13 (left)) are placed behind the Inner Tracker and have dimensions of $6 \times 5\,\mathrm{m}$ (width $\times$ height). A module consists out of tubes (made out of Kapton®-XC and Kapton®-aluminium) with a diameter of 5 mm and a length of 2.5 m. Inside each tube there is a gold-plated tungsten wire anode. The working principle is similar to a Geiger counter, however, additionally also the drift time of the signal is measured. Due to the timing measurements the position of a particle inside a tube can be determined with a resolution of about 200 µm. The detection efficiency in the central part of the tubes is 99.2%. The straw tubes are arranged in a double-layer structure (see Fig. 13 (right)) such that a nearly perfect detection efficiency is reached in each Outer Tracker module.

The resolution of the momentum measurement is determined mainly by multiple scattering of the particles in the VELO exit window, RICH1 and the TT and the knowledge of the magnetic field. The good spatial resolution of the detectors is, however, important for the track finding algorithms. The momentum resolution depends on the momentum of particles and is $\mathrm{d}p/p = 0.5\%(@20\,\mathrm{GeV})$ to $0.8\%(@100\,\mathrm{GeV})$. The track finding efficiency is above 96% for particles in between 5 to 200 GeV.

One of the main reasons for the good momentum resolution of the LHCb detector is the long distance of about 4 m inside the magnet where particles do not interact with any material. As in this region particles are not detected this also creates a challenge for the track finding algorithms. Detector responses of different independent particles can be wrongly associated to one so called *ghost* track. The fraction of *ghost* tracks in the track reconstruction process is in average about 13%. In the further event processing this number is strongly reduced and it is in general only a (handleable) challenge in the trigger.

3.1.3 Particle identification

The particle identification system at LHCb mainly consists out of muon chambers to detect muons and the Ring Imaging Cherenkov detectors to identify the species of charged hadrons. Especially the muon system is important for the implementation of an efficient trigger (see Sec. 3.2).

The muon system

The muon system [27] consists out of five stations M1-M5. The detector modules in each station are mainly Multi Wire Proportional Chambers. In the first station in the inner region due to the high radiation environment instead GEMs are used. The first station (M1), which is the only one of the five muon stations which is placed before the calorimeter, is only used in the very first so called L0 trigger stage (trigger stages will be explained in Sec. 3.2). In this trigger track reconstruction has not yet been performed and additional information to identify muon candidates is necessary. The other four stations are located behind the calorimeters and are separated each by 80 cm thick iron layers to absorb high ionizing particles. Only low ionizing muons pass through and can be identified as such.

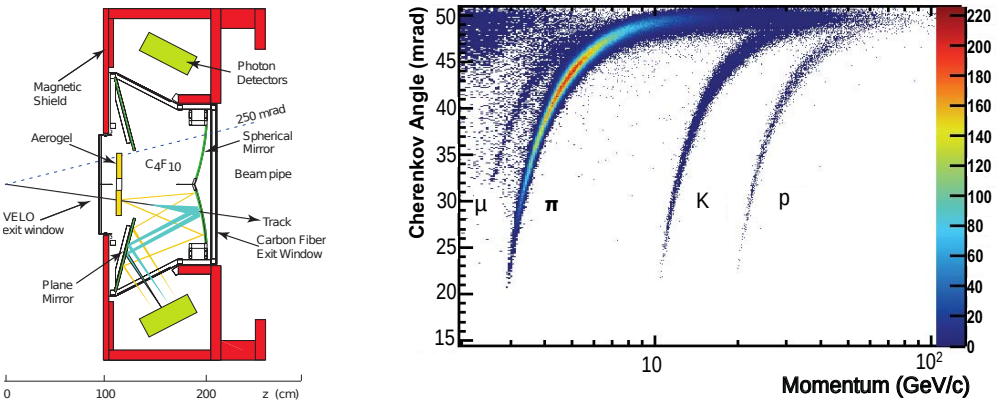

Figure 14: Sketch of the RICH1 particle identification system (side-view, left) and Cherenkov angle measured for various particles for C_4F_{10} used for calibration (right) [21, 28]

The muon system is used in two ways in this analysis. In the L0 trigger a standalone track reconstruction is done in the muon chambers. Large momentum muons can be identified and corresponding events passed to the next trigger stage. The momentum resolution of the muon candidates is about $\mathrm{d}p/p = 20-30\%$ and thus much worse than what is obtained with the tracking stations. Therefore, after tracks have been reconstructed in the tracking stations, the muon stations are only used to verify the muon hypothesis of particles. The track of a particle is extrapolated into the muon stations and it is checked for compatible detector responses.

The muon detection efficiency in the muon station is well above 99%. In the verification process the main cause for misidentification of pions and kaons is their decay in flight into a muon. The misidentification as muons is in general less than 1% for protons and smaller 2% for kaons and pions, while keeping the muon efficiency above 95%. Only for very low momentum pions the misidentification goes up to 20%. The muon identification performance is improved by including information from the Ring Imaging Cherenkov detector.

Ring Imaging Cherenkov Detector

Two Ring Imaging Cherenkov detectors [28] (RICH1 and RICH2) are installed in the LHCb detector. The RICH systems are used to differentiate charged final state particles, especially pions and kaons. Due to the large amount of background from pions it is essential to have a good identification of kaons and pions.

If a particle moves faster than the speed of light in a medium Cherenkov light is emitted under a certain angle. This angle depends on the velocity of the particle and the refractive index of the medium. As the light is spread equally around the particle a ring of light can be detected.

RICH1 (see Fig. 14 (left)) is placed directly after the VELO and shares also the VELO exit window to reduce material inside the path of particles as much as possible. It is optimized for particles with a momentum between 2 and 40 GeV. The angular acceptance is $\pm(25 - 300)$ mrad in horizontal direction (which is the bending plane) and ± 250 mrad in vertical direction covering the full nominal detection region. The gas used as optical medium is C_4F_{10} which has a refractive index of[9] 1.0014. The light produced by a particle is reflected by a spherical mirror at the end of RICH1 to the outside of the detector. An array of photon detectors is used to identify the rings of light. The Cherenkov angle of the different particles (extracted from the radii of the rings) for C_4F_{10} measured in RICH1 is shown in Fig. 14 (right). There is a clear separation between pions, kaons and protons with a momentum below 40 GeV. A distinction of muons and pions is also visible, however, the main separation power is established in this case from the muon chambers. For very low momentum particles below 10 GeV additional a layer of silica aerogel is placed in the beginning of RICH1, which has a refractive index of 1.03 and thus is operating also for particles with a momentum of up to 2 GeV.

The RICH2 detector is placed after the tracking stations. The working principle of RICH2 is similar to the one of RICH1, however, a different optical medium is used. The gas used is CF_4 with a refractive index of 1.0005. It is optimized for high momentum particles with a momentum between 15 and 100 GeV. As these particles do not separate so much from the beam axis the angular coverage is only $\pm(15 - 120)$ mrad in horizontal direction and ± 100 mrad in vertical direction.

Due to the multitude of particles in an event the identification and measurement of the overlapping Cherenkov rings is highly non-trivial. A sophisticated software is used demanding a large amount of computing resources. Instead of measuring the radii of the rings, a particle hypothesis is assumed for each track and it is checked if there are photons corresponding to the expected Cherenkov ring. The position of a particle inside the RICH detectors is known by extrapolations from the tracking stations. For all tracks in the event a particle hypothesis is applied and the combined likelihood, L, is calculated. The different hypothesis are switched for each particle successively and it is searched for the maximum likelihood. The quantities which are assigned to each track is the difference of the logarithm of the likelihood of two particle hypothesis.

The resolution of the determination of the Cherenkov angles is 1.618 ± 0.002 mrad and 0.68 ± 0.02 mrad respectively for RICH1 and RICH2. For a loose particle identification requirement and averaging over the momentum range from $2 - 100$ GeV the efficiency to identify a kaon is $\sim 95\%$ while keeping the pion misidentification below $\sim 10\%$. For a harder requirement the pion misidentification can be reduced to $\sim 3\%$ keeping the kaon efficiency by $\sim 85\%$. The particle identification performance depends on the momentum and phase space of the particle, however, especially also on the number of tracks in the event.

[9] Measured at 101.325 kPa at 0 °C and for a wavelength of 400 nm.

3.2 The LHCb trigger

In the following an overview over the trigger system implemented at LHCb [29, 30] is given. The explicit selection requirements relevant for this analysis are given in Sec. 7.1.

During the Run 1 data taking the proton-proton collision rate was about 20 MHz. This rate is much larger than 5 kHz, which is the rate that can be saved on storage. In several stages the event rate is reduced to an acceptable number. The first hardware based stage, the L0 Trigger, reduces the rate so much that it is processable by the software based High Level Trigger stage.

L0 trigger

The L0 trigger is based on FPGAs which are directly connected to a detector subsystem. It is triggered by a particle candidate with large transverse momentum in either the calorimeters (ECAL, HCAL) or in the muon chambers.

The trigger strategy used for this analysis is based on triggering muon candidates with high transverse momentum. In a fast algorithm it is searched in the muon chambers for tracks with the largest momentum. The maximal output rate of the L0 trigger is limited to 1 MHz.

High level trigger

The second trigger stage is software based and it is running on a large computer cluster with about $3 \cdot 10^4$ CPU cores. As it is based on software it is frequently adjusted to offer the best performance for the current running conditions and priorities.

A full event reconstruction is done, however, using software which is optimized to pass tight computing time requirements. Therefore, it is slightly different to the later reprocessing of the data, where a more accurate alignment and calibration of the sub-detectors is used, as well as more sophisticated versions of the reconstruction software.

The High Level Trigger is divided into two trigger stages. In the first stage information from the tracking stations is used to reconstruct tracks and get a more precise momentum estimation of the particle candidates. It is again searched for particles with a large transverse momentum, in particular if they can also be associated to detector responses in the muon chamber.

In the second stage of the High Level Trigger a more complex and more precise track reconstruction is done. Also information from the particle identification system is used. Several trigger strategies are implemented to match the requirement of different analysis. Due to the clean muon identification in the muon chambers one set of strategies relies on selection high momentum muons. Especially it is also checked if two oppositely charged muons come from the same vertex, which is separated from the proton collision point. As also decays are studied, in which there is no muon present, a different set of trigger strategies relies on the combination of tracks

in an N-body decay, where tracks are not required to be muon candidates. In a multivariate analysis decays with a large transverse momentum and good vertex quality are selected.

If the output rate of a trigger strategy is too large, it is artificially reduced by only selecting each n^{th} event. The total event rate is reduced to about 5 kHz, which is small enough to be saved to storage for later reprocessing and analysis.

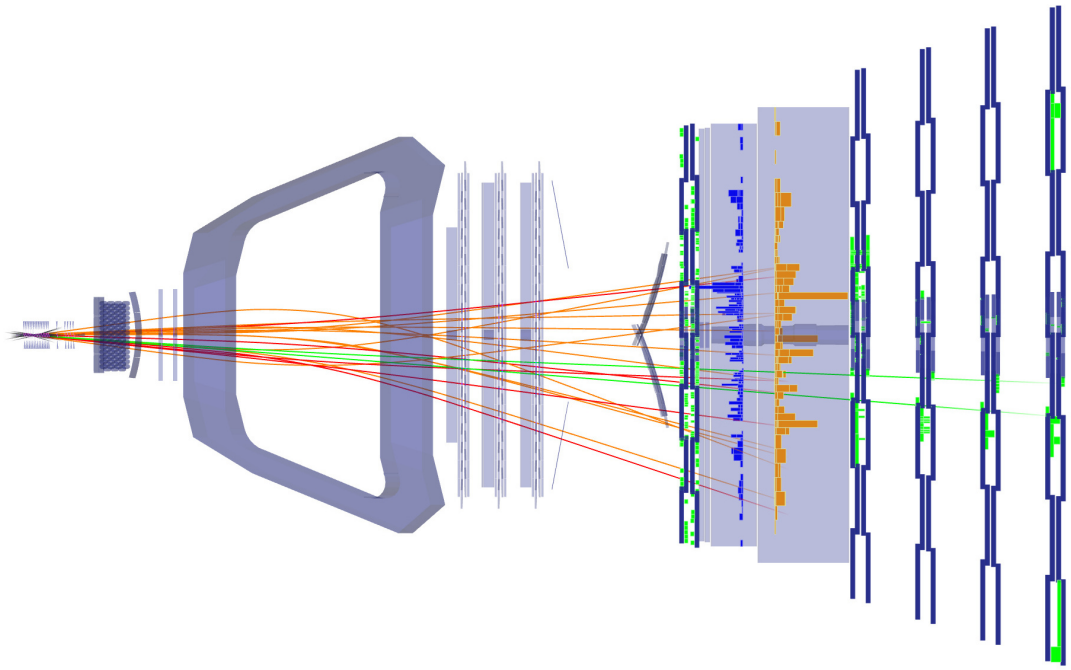

Figure 15: Event display of an $B^0 \to K^{*0}\mu^+\mu^-$ decay candidate (top view) recorded on 03.05.2012 23:45:00. Detector responses of the muon chamber (green), HCAL (orange) and ECAL (blue) are shown proportional to the signal strength. Measurements of the tracking stations are not visible from the top view. Tracks of pion (orange), kaon (red) and muon (green) candidates are formed out of the single detector responses. Clearly visible are two high energetic muon candidates which are used for triggering such events. (made with LHCb event display [31])

3.3 The LHCb software framework

A typical event of a $B^0 \to K^{*0}\mu^+\mu^-$ candidate decay is shown in Fig. 15. A multitude of detector responses must be combined to first form tracks, apply particle hypothesis to the tracks and then reconstruct the complete decay chain. The software framework used at the LHCb experiment consists out of several software packages and is based on the `ROOT` framework [32].

In the High Level Trigger `Moore`[10] [33] is used to process the responses of the LHCb detector and to do a full event reconstruction. A large part of the program code is identical to the one used for later reprocessing, however, the configuration is optimized for speed.

In the later reprocessing the software framework is split into two parts. In the first part `Brunel` [34] combines the different detector responses to tracks. Several (also redundant) pattern recognition algorithms are implemented. The track candidates are fitted with a Kalman Filter. Afterwards information from the particle identification system is read out and a particle hypothesis is assigned to each track. In the second step of the processing `DaVinci` [35] reconstructs the complete decay chain out of the single particle candidates. The decay chain is fitted with a `DecayTreeFitter` [36] determining the quality of the matching of the single particles. Furthermore, selection requirements are applied (see Sec. 7) to reduce the background yield. The versions used for the reprocessing of the data used in this analysis are `DaVinci` v33r6p1 and `Brunel` v43r2p10.

The final parameter estimation methods are usually based on self-produced customized implementations to meet the specific requirements of each analysis.

[10] The `Moore` versions used for data taking between 2010 and 2013 are v10r1 to v14r12.

4 Physics at the LHCb experiment

As explained in the last section the main physics focus at the LHCb experiment is the study of hadrons containing either a b or c quark. In these systems it is especially interesting to indirectly search for New Physics, as outlined in Sec. 4.1. Decays of b hadrons, $b - s$ transitions and their effective description are explained in Sec 4.2. The theoretical framework of the decay $B^0 \to K^{*0}\mu^+\mu^-$, which contains a $b - s$ transition, is detailed in Sec. 4.3. At the end in Sec. 4.4 several New Physics scenarios are discussed, to whose effects the current measurement might be sensitive. The theoretical description follows Refs. [7–9, 37].

4.1 Indirect search for new physics phenomena in b decays

As detailed in the last section the LHCb experiment is excellent suited to do precision measurements of b hadron properties. New Physics different from the Standard Model can be detected by measuring deviations from its predictions. This ansatz is called indirect search, as new undiscovered particles will not be measured directly, but only indirectly by virtual contributions of these particles in the processes which modify the measured observables. To be sensitive to small deviations it is important to have both a precise measurement and a precise theory prediction. The measurement and the observables are therefore chosen accordingly.

Especially interesting to constrain possible New Physics scenarios is the b system. The b quark is the heaviest quark appearing in hadronized particles. The only quark which is heavier is the t quark, which decays before it hadronizes. Due to its mass the b quark is dominating the physical properties of the hadron, making it easier to calculate these systems. Furthermore, as the b hadrons are heavier than most other hadrons, there is a large amount of different decays which can be studied.

As the b quark is lighter than the other quark of the third generation, the t quark, it cannot decay into a quark of the same generation. The decays to another quark generation are, however, suppressed by the small absolute values of the corresponding CKM matrix elements. Looking at $b - s$ and $b - d$ transitions, these decays can in the Standard Model furthermore only occur by loop processes. In processes which only have a small decay rate in the Standard Model New Physics contributions could potentially contribute significantly.

From the experimental point of view the large mass of the b quark results in a large transverse momentum of its decay products. Therefore b decays can be well distinguished from the low momentum background of particles originating from the underlying event. As a result of only decaying via suppressed processes the lifetime of b hadrons is rather long providing a large separation of its production and decay vertex for boosted b hadrons, which is of the order of a few mm in the laboratory frame of the LHCb experiment. Another advantage of b hadrons is the large $b\overline{b}$ quark production cross section. All these features together result in a large and clean sample of b hadrons available at the LHCb experiment.

With this data a large variety of different analyses is done to search for New Physics [38]. As already stated most interesting observables to study are those where there is both a precise theory prediction and also a precise measurement. This is given especially for the measurement of ratios, as many theoretical and also experimental uncertainties cancel out. There are several possible ratios that can be measured. The different behaviour of particles and antiparticles can be accessed by measuring *CP*-violation. Also the measurement of the ratio of branching ratios of different decays provides potential to discover New Physics. Looking only at a single decay it is interesting to look at the ratio of the branching ratio into different phase space regions, like *i.e.* the forward-backward asymmetry. More information as in a simple phase space ratio measurement lies in the measurement of the normalized angular distribution of a process. The analysis done in this thesis is the angular measurement of the decay $B^0 \to K^{*0}\mu^+\mu^-$.

4.2 Weak b decays and the $b - s$ transition

The decays of b hadrons can be classified by the final state particles appearing in these decays. Usually decays can be mainly separated into three groups: leptonic, semi-leptonic and hadronic decays. In these groups the final state particles are, respectively, only leptons, leptons and hadrons and only hadrons. Although in all groups the physics of the b quark is identically, the potential to discover New Physics in each group differs.

This is mainly related to the theoretical description of the weak and strong interaction. The intrinsic difference between the strong and the weak interaction is the strength of the coupling. At low q^2 the coupling constant of the strong interaction, α_s,

is of order one, whereas the weak coupling constant, α_w, has a size of about 1/30 [7]. In addition to these raw coupling constants at low energies the large mass of the W/Z boson in comparison to the massless gluons further reduces the effective coupling strength of the weak interaction by several orders of magnitude. Therefore in calculations higher order diagrams, which describe multiple couplings, can be expected to be small in the weak interaction and thus a perturbative approach is applicable. As the coupling strength of the strong interaction is of order one this approach is not valid here. Therefore precision calculations which include the strong interaction are highly non-trivial.

As a result theoretical predictions for purely hadronic decays are in comparison to the other two groups much more imprecise. Thus even with a precise measurement the potential to discover New Physics is limited. In contrast calculations of purely leptonical decays are more simple and theoretical predictions more precise. However, with $B^0_{(s)} \to \mu^+\mu^-$, which provides a yield just large enough to be discovered [39], the feasible b physics measurements are limited.

In comparison there is a multitude of semi-leptonic decays which offer a large variety of observables. Their theoretical description is nevertheless more demanding and often also less precise than the ones of purely leptonic decays. It is convenient to study the leptonic part of semi-leptonic decays. Especially interesting are decays in which the main physics process is driven by the decay of a heavy b quark and the other quarks of the hadron remain unchanged, being so called spectator quarks. This is also the case for the decay $B^0 \to K^{*0}\mu^+\mu^-$, in which the b quark undergoes a $b-s$ transition and the other quark of the B^0 meson, the d quark, stays in first order untouched (see Fig. 16).

There are many decays which are similar to the decay $B^0 \to K^{*0}\mu^+\mu^-$. The main difference is the flavour of the spectator quark(s) and if it is a $b-s$ or $b-d$ transition. In general for all of these decays, for which it is possible at the LHCb experiment, a corresponding analysis is done or in preparation. The advantage of the decay $B^0 \to K^{*0}\mu^+\mu^-$ is first that a $b-s$ transition is more frequent than a $b-d$ transition due to the absolute value of the CKM matrix elements. Second, because the spectator quark is a d quark, it can be chosen to study decays via the $K^{*0}(892)$ resonance. As the $K^{*0}(892)$ meson is a vector meson its polarization can be assessed in its decay, allowing the study of a large set of different observables.

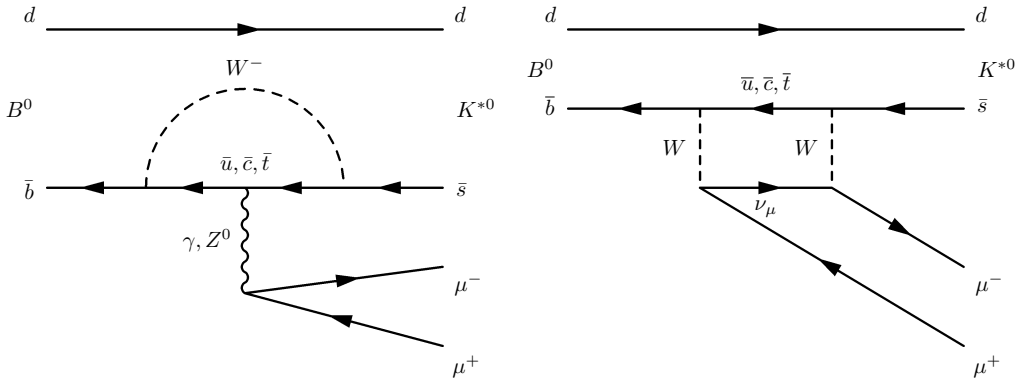

Figure 16: Feynman diagrams for the decay $B^0 \to K^{*0}\mu^+\mu^-$. The decay is possible both over so called penguin diagrams (left) and also over so called box diagrams (right). [40]

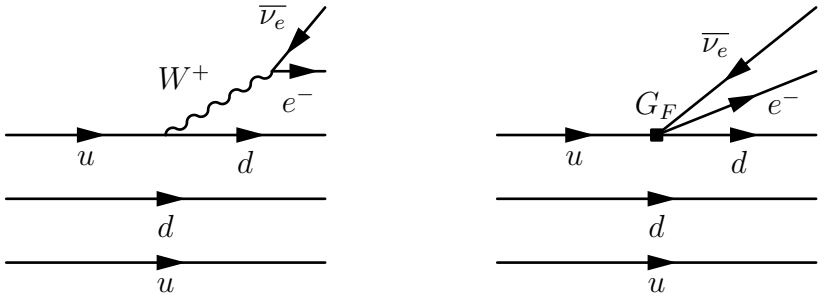

Figure 17: Neutron beta decay described by the full electroweak theory (left); Effective description of this process, described by a 4-point interaction with coupling constant G_F (right).

4.2.1 Effective description of weak b decays

One of the great successes of Albert Einstein was the replacement of Newton's laws of motion with the theory of relativity. Nevertheless, nowadays it is still common to use calculations based on classical mechanics for problems in which the energy and velocity are small enough that no relativistic effects get visible.

Similarly weak b decays, in which New Physics effects can be expected to get visible, are happening at an energy scale much smaller than the mass of the W/Z-bosons. Therefore these heavy degrees of freedom can be *integrated out* from the formulas. The physics is split up into a low energy long distance part and a high energy short distance part. Both parts can be handled with different theoretical tools, each suited for the demands appearing in the specific case. The separation of the low energy and high energy regime is done at an arbitrary physics scale μ, which is usually set to about 2 GeV for b decays. In such an effective theory the amplitude of a weak decay is written as

$$A = \langle H_{eff} \rangle \propto \sum C_i(\mu) \langle O_i(\mu) \rangle \tag{5}$$

where H_{eff} is the effective Hamiltonian, C_i are the so called Wilson coefficients and O_i are local operators. The Wilson coefficients and the operators depend each on the physics scale μ, however, in the calculation of the amplitude μ must drop out.

This ansatz is similar to Fermi's theory of beta decay. Beta decay is the transition of a neutron into a proton, an electron and a corresponding anti-neutrino (see Fig. 17 (left)). The amplitude of this decay is proportional to $1/\left(M_W^2 + q^2\right)$, where M_W is the mass of the W boson and q^2 is in this case the invariant mass squared of the $e\overline{\nu_e}$ system. Fermi, who didn't know about the W boson, described the process by a 4-point interaction of the u quark, d quark, electron and anti-neutrino with a coupling strength proportional to G_F (see Fig. 17 (right)). As $q^2 \ll M_W^2$ the dynamics of the W boson can be neglected in the calculations and Fermi's effective theory is fully sufficient to describe this decay. Also in the case of the weak b decay all involved energies are much smaller than the W mass and a similar strategy can be applied.

The local operators, O_i, describe the low energy (long distance) physics. They contain only information about the light particles. A collection of the important operators for this analysis is shown in Sec. 4.2.2. As the calculations are non-perturbative the precise handling of these operators is still one of the main challenges in the determination of theoretical predictions (see Sec. 4.3.5).

The coupling strength of each of the operators is described by Wilson coefficients. They contain the high energy (short distance) physics like for example the description of the virtual W boson. Due to the asymptotic freedom of quarks the strong coupling constant, α_s, gets small enough to in principle allow perturbative calculations here. However, there are additional challenges appearing in the calculations which must be taken care off (see Sec. 4.3.5).

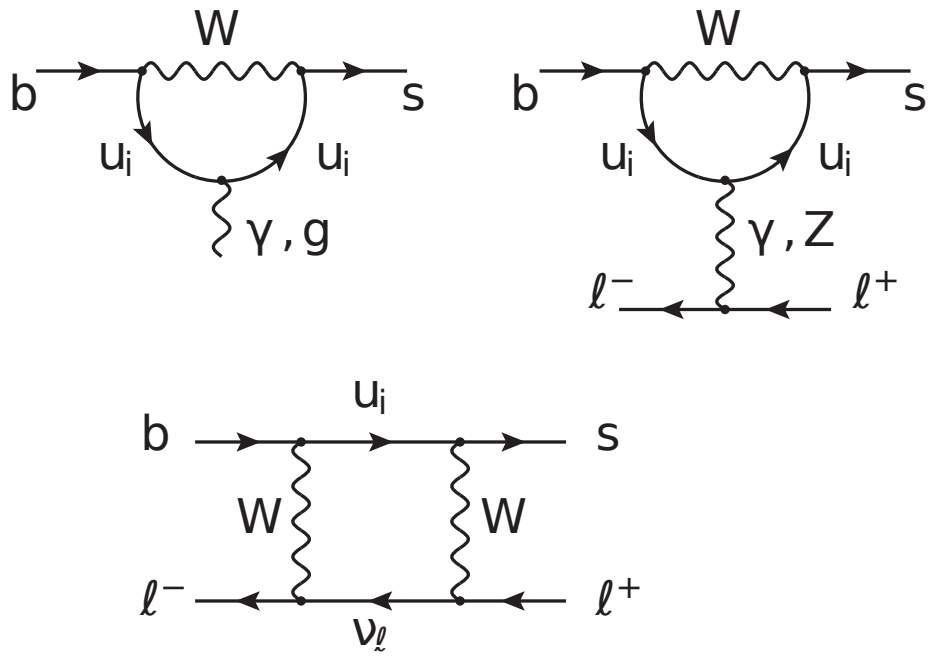

Figure 18: Feynman diagrams represented by different Operators, O_i, as explained in the text. Up-type quarks with flavour i are denoted by u_i. [41]

4.2.2 Wilson coefficients

In this analysis, physics related to the so called *electroweak* Wilson coefficients is tested. Most of the other Wilson coefficients are already strongly constrained by measurements of tree-level processes. It is in the electroweak Wilson coefficients where New Physics is expected to show up (see Sec. 4.4).

The electroweak Wilson coefficients give the coupling strength of the operators with the index $7 - 10$, which are defined as follows [42]:

$$O_7 = \frac{e}{g^2} m_b \left(\overline{s}\sigma_{\mu\nu}\tfrac{1}{2}\left(1+\gamma_5\right)b\right) F^{\mu\nu} \tag{6}$$

$$O_8 = \frac{1}{g} m_b \left(\overline{s}\sigma_{\mu\nu}T^a\tfrac{1}{2}\left(1+\gamma_5\right)b\right) G^{\mu\nu a} \tag{7}$$

$$O_9 = \frac{e^2}{g^2} \left(\overline{s}\gamma_\mu\tfrac{1}{2}\left(1-\gamma_5\right)b\right) \left(\overline{\mu}\gamma^\mu\mu\right) \tag{8}$$

$$O_{10} = \frac{e^2}{g^2} \left(\overline{s}\gamma_\mu\tfrac{1}{2}\left(1-\gamma_5\right)b\right) \left(\overline{\mu}\gamma^\mu\gamma_5\mu\right) \tag{9}$$

where g is the strong coupling constant, m_b is a theoretical input parameter related to the b quark mass, $F^{\mu\nu}$ is the electromagnetic tensor and $G^{\mu\nu a}$ is the gluon field strength tensor.

Feynman diagrams related to these Wilson coefficients are shown in Fig. 18. The Wilson coefficients C_7 and C_8 belong to the top left diagram, respectively to the coupling of a photon and a gluon. If a semi-leptonic decay is studied only the photonic operator O_7 contributes to the process. In the case of the $B^0 \to K^{*0}\mu^+\mu^-$ decay the photon then couples to two muons. For $q^2 = 0$ the photon (gluon) is on shell, being the reason for a resonant contribution at low $q^2 = 0$, the so called photon pole (see Sec. 4.3.3).

The top right diagram, a so called penguin diagram, is related to both C_9 and C_{10}. The Wilson coefficient C_9 is related to vector like coupling, whereas C_{10} is related to axial vector like coupling. Therefore the decay described by a virtual photon is only possible in the case of C_9. The bottom diagram, a so called box diagram, is associated to both C_9 and C_{10}.

4.3 The decay $B^0 \to K^{*0}\mu^+\mu^-$

One of the most promising decays for studying $b-s$ transitions is the decay $B^0 \to K^*\ell^+\ell^-$. In this decay of a pseudoscalar to two vector particles[11] there is a multitude of observables which can be measured. By studying the decay of the K^{*0} into $K\pi$ also the polarization can be accessed.

Looking at the experimental side especially decays containing muons in the final state, thus the decay $B^0 \to K^{*0}\mu^+\mu^-$, give a clean detector signature and can be efficiently triggered at the LHCb experiment (see Sec. 3.2). The decay $B^0 \to K^{*0}e^+e^-$ is from the theoretical point of view in the Standard Model very similar to the decay $B^0 \to K^{*0}\mu^+\mu^-$. However, due to the significant larger Bremsstrahlung of electrons compared to muons this decay is difficult to detect, which is the reason why in this thesis only the decay $B^0 \to K^{*0}\mu^+\mu^-$ is studied.

In the LHCb angular analysis of $B^0 \to K^{*0}\mu^+\mu^-$ using $1\,\text{fb}^{-1}$ of data a discrepancy to the Standard Model has been observed [43], which motivated this PhD thesis.

4.3.1 Definition of the angular basis

The decay $B^0 \to K^{*0}\mu^+\mu^-$ can be fully described by the dimuon invariant mass squared, q^2, and three decay angles. Several definitions of these three decay angles exists. The definition used at LHCb is presented in the following. The idea is that the differential decay rate of the B^0 and $\overline{B}^0$ is described by the same formula. Therefore the definition of the angles for the B^0 decay is the *CP*-transformed definition of the $\overline{B}^0$ decay.

The relations are sketched in Fig. 19 to 21. The angle θ_l is defined as the angle between the flight direction of the μ^- (μ^+) in the $\mu^+\mu^-$ rest frame and the flight direction of the $\mu^+\mu^-$ in the $\overline{B}^0$ (B^0) rest frame. The angle θ_K is defined similarly for both decays. It is the angle between the flight direction of the kaon in the K^* rest frame and the flight direction of the K^* in the $\overline{B}^0$ (B^0) rest frame. The angle ϕ is the angle between the $\mu^+\mu^-$ plane and the plane of the kaon and pion. For the $\overline{B}^0$ decay it is defined as the angle between the normal vector of the kaon and pion and the negative normal vector of the $\mu^+\mu^-$. In the decay of the B^0 it is defined as the angle between the normal vector of the $\mu^+\mu^-$ and the normal vector of the kaon and pion.

[11] The leptons originate from a spin 1 boson.

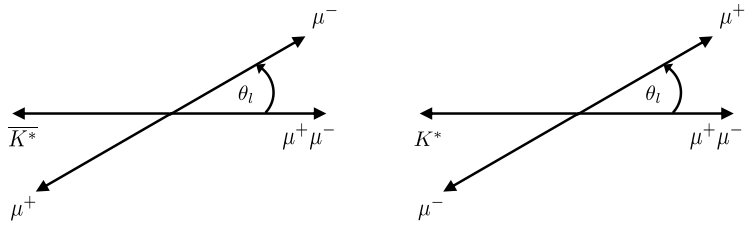

Figure 19: Definition of the angle θ_l for the decay $\overline{B}^0 \to \overline{K}^*\mu^+\mu^-$ (left) and the decay $B^0 \to K^{*0}\mu^+\mu^-$ (right).

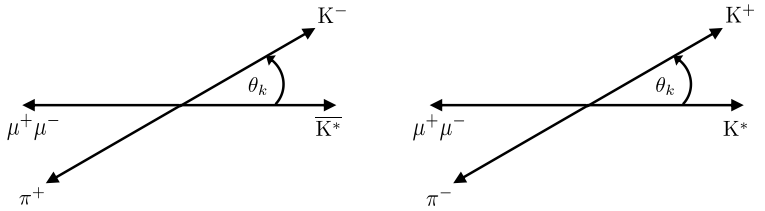

Figure 20: Definition of the angle θ_K for the decay $\overline{B}^0 \to \overline{K}^*\mu^+\mu^-$ (left) and the decay $B^0 \to K^{*0}\mu^+\mu^-$ (right).

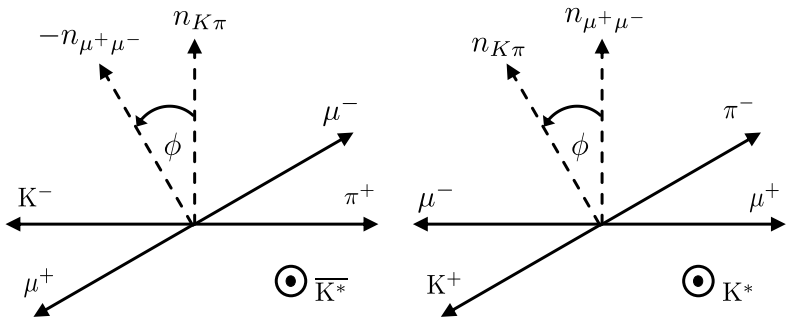

Figure 21: Definition of the angle ϕ for the decay $\overline{B}^0 \to \overline{K}^*\mu^+\mu^-$ (left) and the decay $B^0 \to K^{*0}\mu^+\mu^-$ (right). The flight direction of the K^{*0} is in direction of the reader.

4.3.2 Differential decay rate

The angular observables which are measured in this analysis are defined in the following. The theoretical calculation of these quantities is outlined in Sec. 4.3.4. The following relations are valid if the $K\pi$ decay occurs via the $K^*(892)$ resonance and the lepton mass is neglected in the calculations. Due to the spin structure of the decay $B^0 \to K^{*0}\mu^+\mu^-$ its differential decay rate can be expressed as:

$$\frac{\mathrm{d}^4\Gamma}{\mathrm{d}\cos\theta_l\,\mathrm{d}\cos\theta_K\,\mathrm{d}\phi\,\mathrm{d}q^2} = \frac{9}{32\pi}\sum_i J_i(q^2) f_i(\cos\theta_l, \cos\theta_K, \phi) \tag{10}$$

where f_i are angular terms only depending on the three decay angles and J_i are angular observables depending on the dimuon invariant mass squared, q^2. The terms f_i are given purely by the spin structure of the decay. The observables J_i express how much the different terms f_i contribute to the differential decay rate. In the case of certain New Physics models the size of the J_i is expected to be different from the Standard Model predictions.

Writing the angular terms f_i explicitly the differential decay rate is:

$$\begin{aligned}\frac{\mathrm{d}^4\Gamma}{\mathrm{d}\cos\theta_l\,\mathrm{d}\cos\theta_K\,\mathrm{d}\phi\,\mathrm{d}q^2} = \frac{9}{32\pi}\Big\{ & J_{1s}\sin^2\theta_K \\ & + J_{1c}\cos^2\theta_K \\ & + J_{2s}\sin^2\theta_K\cos 2\theta_l \\ & + J_{2c}\cos^2\theta_K\cos 2\theta_l \\ & + J_3\sin^2\theta_K\sin^2\theta_l\cos 2\phi \\ & + J_4\sin 2\theta_K\sin 2\theta_l\cos\phi \\ & + J_5\sin 2\theta_K\sin\theta_l\cos\phi \\ & + J_{6s}\sin^2\theta_K\cos\theta_l \\ & + J_7\sin 2\theta_K\sin\theta_l\sin\phi \\ & + J_8\sin 2\theta_K\sin 2\theta_l\sin\phi \\ & + J_9\sin^2\theta_K\sin^2\theta_l\sin 2\phi\Big\}\end{aligned} \tag{11}$$

Analogously the differential decay rate for the decay $\overline{B}^0 \to \overline{K}^*\mu^+\mu^-$ can be expressed by the angular observables $\overline{J_i}$. It is convenient to also define the *CP*-averaged angular observables:

$$S_i = \frac{J_i + \overline{J_i}}{\Gamma + \overline{\Gamma}} \tag{12}$$

and the *CP*-asymmetric ones:

$$A_i = \frac{J_i - \overline{J_i}}{\Gamma + \overline{\Gamma}}. \tag{13}$$

In the Standard Model the *CP* asymmetries are expected to be negligible small. Therefore, due to the limited event yield which is available, not the J_i but the S_i are measured in this analysis, effectively doubling the event yield. The influence of the *CP* asymmetries is neglected in this measurement and will be part of future analyses.

In the past also different observables were defined which are related to the observables measured in this analysis. The observable S_{1s} can be expressed as the fraction of the longitudinal polarization of the K^*, F_L. The observable S_{6s} is related to the forward-backward asymmetry, A_{FB}. A linear dependence exists between these observables:

$$\begin{aligned} F_L &= 1 - \frac{4}{3} S_{1s}, \\ A_{FB} &= \frac{3}{4} S_{6s}. \end{aligned} \tag{14}$$

Furthermore, there also exists a basis to which there does not exists a linear transformation for all observables. In this basis certain theoretical uncertainties (related to the form factors, see Sec. 4.3.5) of some observables cancel out to a certain extent. These uncertainties are then shifted to other observables. Therefore this basis has advantages if looking only at one single observable. If the full set of observables including their correlation is taken into account this basis transformation is no longer advantageously. Nevertheless, for a comparison to previous results also a measurement of the observable P_5' from this basis is performed in the following. The basis consists out of F_L and seven so called $P_i^{(\prime)}$ observables:

$$\begin{aligned} P_1 &= \frac{S_3}{1 - F_L}, \\ P_2 &= \frac{S_6^s}{1 - F_L}, \\ P_3 &= \frac{S_9}{1 - F_L}, \\ P_4' &= \frac{S_4}{\sqrt{F_L(1 - F_L)}}, \\ P_5' &= \frac{S_5}{\sqrt{F_L(1 - F_L)}}, \\ P_6' &= \frac{S_7}{\sqrt{F_L(1 - F_L)}}, \\ P_8' &= \frac{S_8}{\sqrt{F_L(1 - F_L)}}. \end{aligned} \tag{15}$$

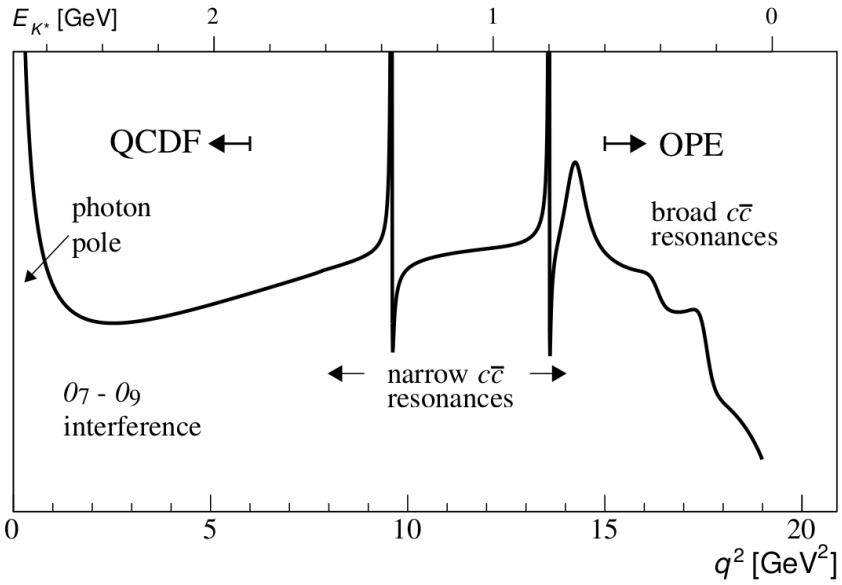

Figure 22: Sketch of the differential decay rate of $B^0 \to K^{*0}\mu^+\mu^-$ vs. q^2. Interesting for discovering New Physics is the region from q^2 $1-6\,\mathrm{GeV}^2$, where the interference of O_7 and O_9 can be measured. Both a precise theory prediction exists here, and also New Physics is expected to be significant. [44]

Table 1: Bins in q^2 used in this analysis.

Bin	$q^2[\,\mathrm{GeV}^2]$
1	$0.10 - 0.98$
2	$1.1 - 2.5$
3	$2.5 - 4.0$
4	$4.0 - 6.0$
5	$6.0 - 8.0$
6	$11.0 - 12.5$
7	$15.0 - 17.0$
8	$17.0 - 19.0$

4.3.3 Dependence on dimuon invariant mass

As sketched in Fig. 22 the differential branching ratio of the decay $B^0 \to K^{*0}\mu^+\mu^-$ changes with q^2. In the different q^2 regions different physics contributions are dominating the decay rate and thus the physics that can be measured.

At $q^2 \approx 0$ the contribution of the operator O_7 is dominating, as it is this operator which describes the decay over a resonant photon. The so called photon pole is visible. Right of this contribution in the region $1\,\mathrm{GeV}^2 < q^2 < 8\,\mathrm{GeV}^2$ physics is mainly described by the interference of the operators O_7 and O_9. Above the charmonium resonances ($q^2 > 15\,\mathrm{GeV}^2$) the contribution of O_7 becomes negligible and physics related to O_9 and O_{10} can be measured.

Therefore the analysis is done q^2 dependent. The observables are measured in bins of q^2, as defined in Tab. 1. In each bin the q^2-averaged observables are defined as:

$$\langle S_i \rangle (q_{min}, q_{max}) = \int_{q_{min}}^{q_{max}} \mathrm{d}q^2 \left(J_i + \overline{J_i}\right) / \int_{q_{min}}^{q_{max}} \mathrm{d}q^2 \frac{\mathrm{d}\left(\Gamma + \overline{\Gamma}\right)}{\mathrm{d}q^2} \tag{16}$$

The brackets are omitted in the following.

Concerning theory predictions different assumptions are valid in the low and high q^2 region. At low q^2 the energy of the K^{*0} is much larger than Λ_{QCD} and calculations can be based on QCD factorization (QCDF) [45]. On the other hand if q^2 is large enough to be used as expansion parameter ($q^2 > 15\,\mathrm{GeV}^2$) the Operator Product Expansion (OPE) [46] is the preferred theoretical tool.

4.3.4 Theoretical description

The observables J_i (see Sec. 4.3.2) contain the full information accessible from the angular distribution of the decay $B^0 \to K^{*0}\mu^+\mu^-$. From the theoretical point of view the Standard Model and also the effect of New Physics is described in form of the Wilson coefficients. It is convenient to do the relation of the Wilson coefficients and the observables J_i in a two-step process. In a first step the decay amplitudes are calculated from the Wilson coefficients and form factors. Then, in a second step, the observables are calculated from the decay amplitudes. In the following these two steps will be outlined in more detail. The advantage of the measurement of the J_i in comparison to directly measuring the Wilson coefficients is that these observables can be determined independently of the form factors, which would otherwise introduce a model dependence in the measurement.

As explained in the last subsection different approaches how to calculate the differential branching fraction of the decay $B^0 \to K^{*0}\mu^+\mu^-$ exists. In the following the results using the QCD factorization are discussed, as presented in Ref. [42]. These calculations are valid in the q^2 range $1 - 6\,\mathrm{GeV}^2$. The description of the differential decay rate used in the measurement is the same in the whole q^2 region.

The leading contribution to the matrix element for the decay $B^0 \to K^{*0}\mu^+\mu^-$ in the Standard Model can be expressed as follows:

$$\begin{aligned}\mathcal{M} = \frac{G_F\alpha}{\sqrt{2}\pi}V_{tb}V_{ts}^*\{ \quad & C_{10}^{eff} \quad \left\langle K\pi \left| \overline{s}\gamma^\mu \tfrac{1}{2}(1-\gamma_5)\, b \right| \overline{B} \right\rangle (\overline{\mu}\gamma_\mu\gamma_5\mu) \\ & +C_9^{eff} \quad \left\langle K\pi \left| \overline{s}\gamma^\mu \tfrac{1}{2}(1-\gamma_5)\, b \right| \overline{B} \right\rangle (\overline{\mu}\gamma_\mu\mu) \\ & -C_7^{eff}\frac{2m_b}{q^2}\left\langle K\pi \left| \overline{s}i\sigma^{\mu\nu}q_\nu \tfrac{1}{2}(1+\gamma_5)\, b \right| \overline{B} \right\rangle (\overline{\mu}\gamma_\mu\mu)\}\end{aligned} \tag{17}$$

where the effective Wilson coefficients C_i^{eff} are a linear combination of the nominal Wilson coefficients including also higher order processes and are defined in [42] and q_ν is the four momentum transfer between the B^0 and the K^{*0}. Besides the $b-s$ transition in this matrix element also hadronic effects of the spectator quarks are included. From this matrix element the decay amplitudes can be calculated.

There are different definitions of the decay amplitudes. In this analysis the transversity basis is used. The decay can be described by seven amplitudes. There are three space like polarizations possible, which are shown in Fig. 23. The two amplitudes belonging to longitudinal polarization are $A_{0L,R}$, where L and R denotes left or right handed chirality. If both particles are polarized transversal, their polarization can be either orthogonal or perpendicular to each other, which corresponds to $A_{\perp L,R}$ and $A_{\parallel L,R}$, respectively. Due to one of both vector particles being virtual there is also the possibility of a time like polarization which is described by A_t.

For the decay $B^0 \to K^{*0}\mu^+\mu^-$ the decay amplitudes can be expressed by the effective Wilson coefficients:

$$A_{\perp L,R} = N\sqrt{2}\lambda^{1/2}\left[\left[C_9^{eff} \mp C_{10}^{eff}\right]\frac{V(q^2)}{m_B+m_{K^*}} + \frac{2m_B}{q^2}C_7^{eff}T_1(q^2)\right] \tag{18}$$

$$A_{\parallel L,R} = -N\sqrt{2}(m_B^2-m_{K^*}^2)\left[\left[C_9^{eff} \mp C_{10}^{eff}\right]\frac{A_1(q^2)}{m_B-m_{K^*}} + \frac{2m_b}{q^2}C_7^{eff}T_2(q^2)\right] \tag{19}$$

$$\begin{aligned}A_{0L,R} = -\frac{N}{2m_{K^*}\sqrt{q^2}}\Bigg(& \\ \left[C_9^{eff} \mp C_{10}^{eff}\right]\left[\left(m_B^2-m_{K^*}^2-q^2\right)(m_B+m_{K^*})\,A_1(q^2) - \lambda\frac{A_2(q^2)}{m_B+m_{K^*}}\right] & \\ + 2m_bC_7^{eff}\left[(m_B^2+3m_{K^*}-q^2)T_2(q^2) - \frac{\lambda}{m_B^2-m_{K^*}^2}T_3(q^2)\right]\Bigg) &\end{aligned} \tag{20}$$

$$A_t = \frac{2N}{\sqrt{q^2}}\lambda^{1/2}C_{10}^{eff}A_0(q^2) \tag{21}$$

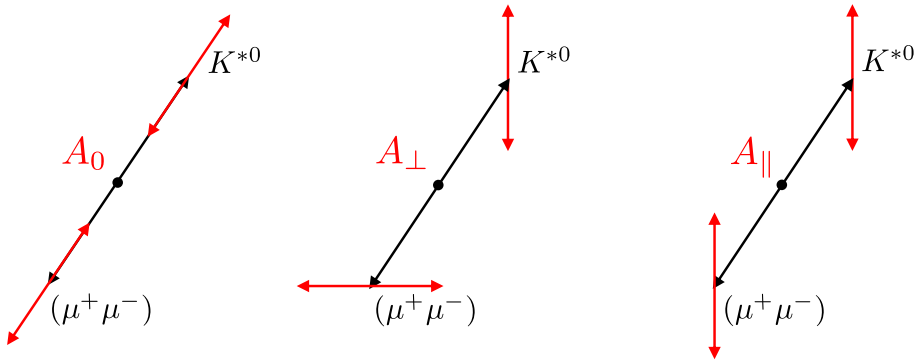

Figure 23: Three polarizations of the $K^*\mu^+\mu^-$ system, longitudinal (left) and transversal (middle, right) and the corresponding amplitudes.

with

$$N = V_{tb}V_{ts}^* \left[\frac{G_F^2 \alpha_s^2}{3.\, 2^1 0 \pi^5 m_B^3} q^2 \lambda^{1/2} \beta_\mu \right]^{1/2} \tag{22}$$

$$\lambda = m_B^4 + m_{K^*}^4 + q^4 - 2\left(m_B^2 m_{K^*}^2 + m_{K^*}^2 q^2 + m_B^2 q^2\right) \tag{23}$$

$$\beta_\mu = \sqrt{1 - 4m_\mu^2/q^2} \tag{24}$$

Additional to the Wilson coefficients there are seven form factors appearing in the formulas ($A_{0,1,2}(q^2)$, $T_{1,2,3}(q^2)$, $V(q^2)$). The form factors describe the transition of the hadronic initial state to the hadronic final state. The precise knowledge of the form factors is necessary to relate the measured values of the amplitudes to the Wilson coefficients. Their calculation is to this day one of the larger challenges in the theoretical description of this decay (see Sec. 4.3.5).

The amplitudes can be related to the measured angular coefficients as follows[12]:

$$J_1^s = \frac{2+\beta_\mu^2}{4}\left[\left|A_\perp^L\right|^2 + \left|A_\parallel^L\right|^2 + (L \to R)\right] + \frac{4m_\mu^2}{q^2}\mathrm{Re}\left(A_\perp^L A_\perp^{R*} + A_\parallel^L A_\parallel^{R*}\right) \tag{25}$$

$$J_1^c = \left|A_0^L\right|^2 + \left|A_0^R\right|^2 + \frac{4m_\mu^2}{q^2}\left[|A_t|^2 + \mathrm{Re}\left(A_0^L A_0^{R*}\right)\right] \tag{26}$$

$$J_2^s = \frac{\beta_\mu^2}{4}\left[\left|A_\perp^L\right|^2 + \left|A_\parallel^L\right|^2 + (L \to R)\right] \tag{27}$$

[12]The expression $(L \to R)$ means the same equation again, however, now with the right handed amplitudes.

$$J_2^c = -\beta_\mu^2 \left(\left| A_0^L \right|^2 + \left| A_0^R \right|^2 \right) \tag{28}$$

$$J_3 = \frac{1}{2}\beta_\mu^2 \left[\left| A_\perp^L \right|^2 - \left| A_\parallel^L \right|^2 + (L \to R) \right] \tag{29}$$

$$J_4 = \frac{1}{\sqrt{2}}\beta_\mu^2 \left[\mathrm{Re}\left(A_0^L A_\parallel^{L*} \right) + (L \to R) \right] \tag{30}$$

$$J_5 = \sqrt{2}\beta_\mu \left[\mathrm{Re}\left(A_0^L A_\perp^{L*} \right) - (L \to R) \right] \tag{31}$$

$$J_6^s = 2\beta_\mu \left[\mathrm{Re}\left(A_\parallel^L A_\perp^{L*} \right) - (L \to R) \right] \tag{32}$$

$$J_7 = \sqrt{2}\beta_\mu \left[\mathrm{Im}\left(A_0^L A_\parallel^{L*} \right) - (L \to R) \right] \tag{33}$$

$$J_8 = \frac{1}{\sqrt{2}}\beta_\mu^2 \left[\mathrm{Im}\left(A_0^L A_\perp^{L*} \right) + (L \to R) \right] \tag{34}$$

$$J_9 = \beta_\mu^2 \left[\mathrm{Im}\left(A_\parallel^{L*} A_\perp^L \right) + (L \to R) \right] \tag{35}$$

A measurement of these observables and their correlations contains the full physical information of the angular distribution. The observables are not fully independent from each other. To simplify the measurement the lepton masses are neglected, which is a good assumption for $q^2 > 1\,\mathrm{GeV}$. Terms including m_μ drop out and $\beta_\mu \approx 1$. It is then easy to extract the following relations:

$$\begin{aligned} J_1^c &= 1 - \frac{4}{3} J_1^s \\ J_2^s &= \frac{1}{3} J_1^s \\ J_2^c &= \frac{4}{3} J_1^s - 1 \end{aligned} \tag{36}$$

There exists also a relation of J_8 and the other J_i. It is, however, not convenient to use this relation in the measurement. Instead J_8 is measured independently, over-constraining the measured system.

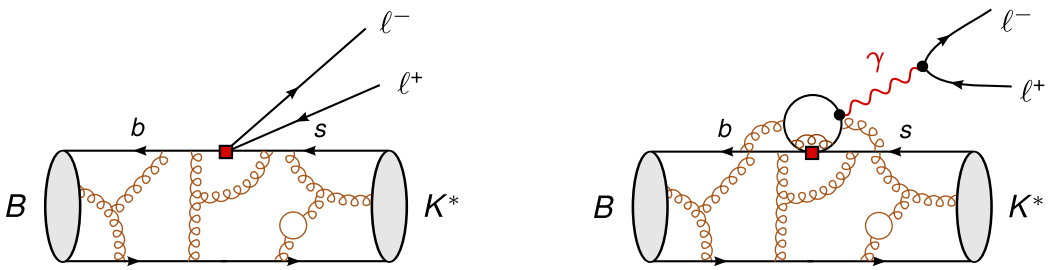

Figure 24: Main theoretical challenges for calculating precise theory predictions: The effect of gluons in the B^0 to K^{*0} transition is treated by form factors (left). Additional so called non-factorisable contributions, as *i.e.* the charm-loop (right), must be treated. The red rectangle symbolizes the four point interaction described by the Wilson coefficients. [47]

4.3.5 Theoretical challenges

There are two main challenges concerning the calculation of the differential decay rate of the decay $B^0 \to K^{*0}\mu^+\mu^-$ [47], also sketched in Fig. 24.

On the one hand there are form factors which due to their non-perturbative nature are difficult to calculate. The form factors describe the transition of a B-meson to a K^{*0}-meson. Uncertainties were especially large in the higher q^2 region above the charmonium resonances, until only recently calculations based on lattice QCD significantly improve the precision of these predictions [48]. Nevertheless, uncertainties related to form factors are still the dominating uncertainty of the predictions.

On the other hand there also exist hadronic effects which are not related to the form factors. In the calculations the separation of the high energetic and low energetic part only works to a certain extent, which gives rise to so called non-factorizable contributions. The uncertainty of these effects is in general predicted to be smaller than the ones of the form factors. However, until now no full theoretical study of the hadronic effects exists and estimations are based on an educated guess. One of these non-factorizable effects is the so called charm-loop [49], which is sketched in the right part of Fig. 24. The b and s quark couple to a $c\bar{c}$ loop, to which the muons couple via a virtual photon. As further gluons can couple *i.e.* between the b and the c quark, calculations get rather complicated. Due to the virtual photon the charm-loop is related to a vector like coupling, which is also the coupling described by the C_9 Wilson coefficient. It is therefore utterly important to not mix up hadronic effects with effects of New Physics.

There are different independent Standard Model predictions available [50–54]. While the central values of all predictions are in good agreement, there is still a lot of discussion about the estimated uncertainties. Especially in Ref. [54] uncertainties of the sub-leading order are treated more conservatively and therefore the total uncertainty is larger by about 400% compared to other predictions.

4.3.6 S-wave pollution

Besides the decay $B^0 \to K^{*0}\mu^+\mu^-$ via the K^{*0} (892) resonance, which is the so called P-wave contribution, also decays via other resonances occur. The so called S-wave contribution are decays via resonances with spin 0, as for example the $K_0^*(1430)$. The angular distribution of the P-wave and S-wave contribution are different, which is why it is important to separate both contributions in the measurement. The P-wave and S-wave also differ in the $K\pi$ mass distribution, which can be used to disentangle both parts (see Sec. 8.5). Including both the S-wave contribution and also the $K\pi$ mass description the $B^0 \to K^{*0}\mu^+\mu^-$ differential decay rate can be written as [55]:

$$
\begin{aligned}
\frac{\mathrm{d}^5\Gamma}{\mathrm{d}q^2\,\mathrm{d}m_{K\pi}\,\mathrm{d}\cos\theta_l\,\mathrm{d}\cos\theta_K\,\mathrm{d}\phi} =& (1-F_S)\frac{9}{32\pi}\left|BW_{P-wave}(m^2_{K\pi})\right|^2 \times \\
& \Big\{ J_{1s}\sin^2\theta_K \\
& + J_{1c}\cos^2\theta_K \\
& + J_{2s}\sin^2\theta_K\cos 2\theta_l \\
& + J_{2c}\cos^2\theta_K\cos 2\theta_l \\
& + J_3\sin^2\theta_K\sin^2\theta_l\cos 2\phi \\
& + J_4\sin 2\theta_K\sin 2\theta_l\cos\phi \\
& + J_5\sin 2\theta_K\sin\theta_l\cos\phi \\
& + J_{6s}\sin^2\theta_K\cos\theta_l \\
& + J_7\sin 2\theta_K\sin\theta_l\sin\phi \\
& + J_8\sin 2\theta_K\sin 2\theta_l\sin\phi \\
& + J_9\sin^2\theta_K\sin^2\theta_l\sin 2\phi\Big\} \\
& + \frac{3}{16\pi}F_S\left|BW_{S-wave}(m^2_{K\pi})\right|^2\sin^2\theta_K \\
& + \frac{3}{16\pi}Re\left[BW_{S-wave}(m^2_{K\pi})BW^\dagger_{P-wave}(m^2_{K\pi})\right] \times \\
& \Big\{ J_{S1}\cos\theta_K\sin^2\theta_l \\
& + J_{S2}\sin\theta_K\sin 2\theta_l\cos\phi \\
& + J_{S3}\sin\theta_K\sin\theta_l\cos\phi\Big\} \\
& + \frac{3}{16\pi}Im\left[BW_{S-wave}(m^2_{K\pi})BW^\dagger_{P-wave}(m^2_{K\pi})\right] \times \\
& \Big\{ J_{S4}\sin\theta_K\sin\theta_l\sin\phi \\
& + J_{S5}\sin\theta_K\sin 2\theta_l\sin\phi\Big\}
\end{aligned}
\tag{37}
$$

where $BW_R(m^2_{K\pi})$ are complex functions corresponding to Breit Wigner functions, as defined by Eq. 48 in Sec. 8.1.1.

Until now the S-wave contribution is neglected in measurements and only accessed in systematic studies. With the data sample available for this analysis and the expected small statistical uncertainty such a procedure is no longer valid. Instead the S-wave contribution is included in the measurement and treated as a nuisance parameters. Including the S-wave there are six additional parameters compared to Eq. 11. Besides the fraction of the S-wave compared to the total decay rate, F_S, also five parameters describing the interference terms, J_{S1-S5}, must be determined.

Thus in total there are now $2 \cdot 17$ observables (J_i and $\overline{J}_i$) which must be measured to determine the differential decay rate. As explained earlier in this analysis the *CP*-averaged observables are measured (see Eg. 12) and it is assumed that the lepton mass can be neglected (see Eq. 36). With this simplifications the number of observables which must be measured reduces to eight P-wave observables and six nuisance parameters defining the S-wave and S/P-wave interference.

4.4 New Physics in *b*-*s* transitions

The search for New Physics is not only limited to one decay channel, but there is a large pool of measurements available constraining the different Wilson coefficients. The Wilson coefficients which are expected to be most sensitive to New Physics are $C_7^{(\prime)}$, $C_9^{(\prime)}$, $C_{10}^{(\prime)}$ and $C_{S,P}^{(\prime)}$. Hereby $C_i^{\prime}$ correspond to the chirality flipped operators of the ones defined in Sec. 4.2.2 and $C_{S,P}$ are related to a scalar and pseudoscalar contribution. The following decays are sensitive to these Wilson coefficients [47] (where X_S means inclusive decays with an s quark):

Decay	$C_7^{(\prime)}$	$C_9^{(\prime)}$	$C_{10}^{(\prime)}$	$C_{S,P}^{(\prime)}$
$B \to (X_S, K^*)\,\gamma$	X	O	O	O
$B \to \left(X_S, K^{(*)}\right)\,\ell^+\ell^-$	X	X	X	O
$B_s \to \mu^+\mu^-$	O	O	X	X

The important role of the decay $B^0 \to K^{*0}\mu^+\mu^-$ to search for New Physics becomes visible. It is in this decay where most of the coefficients can be constrained in one single measurement. Nevertheless, it is only one (large) piece in the puzzle.

Important in constraining New Physics are besides others the measurements of the branching ratio of $B_s^0 \to \mu^+\mu^-$ [39], the branching ratio of the inclusive radiative decay $B \to X_s\gamma$ [56], the *CP*-asymmetry of $B^0 \to K^{*0}\gamma$ [57] and the Isospin asymmetry of $B \to K^{*0}\gamma$ [58]. In the multitude of measurements there exists in general a good agreement with the Standard Model prediction, however, in few

cases a tension compared to the predictions is found. In the angular analysis of the decay $B^0 \to K^{*0}\mu^+\mu^-$ using $1\,\text{fb}^{-1}$ of data in one observable a local $3.7\,\sigma$ deviation compared to the Standard Model prediction is observed [5]. The branching ratios of the decay $B^0 \to K^{*0}\mu^+\mu^-$ [59] and also $B_s^0 \to \phi\mu^+\mu^-$ [60] are both smaller than expected. Especially interesting concerning lepton flavour universality (which is predicted by the Standard Model) is the measurement of the ratio $R_K = BR(B^+ \to K^+\mu^+\mu^-)/BR(B^+ \to K^+e^+e^-)$, which is measured $2.6\,\sigma$ smaller than expected [61].

The results of 81 independent measurements are used to estimate the value of the Wilson coefficients [50]. Varying one or two Wilson coefficients at a time the χ^2 of the measured values minus its predictions divided by the uncertainties is minimized. Varying the Wilson coefficients C_9 and C_{10} the best fit point is more than $2\,\sigma$ away from the Standard Model prediction (see Fig. 25). The authors of that paper have several possible explanations for this observation, supposing it is not only a statistical fluctuation. On the one hand the theoretical uncertainties could have been underestimated. Scaling the uncertainties of the non-factorizable corrections up by a factor of four would indeed lower the tension with the Standard Model prediction. However, it still could not explain a different coupling of muons and electrons as indicated by the measurement of R_K. On the other hand the differences could be also explained by a New Physics contribution. A large negative contribution to C_9, or a positive contribution to C_{10} or a combined contribution with $\Delta C_9 = -\Delta C_{10}$ could reduce significantly the observed discrepancies.

There are several different ideas for New Physics models which can explain additional contributions to the Wilson coefficients. One possible New Physics Model is the Minimal Supersymmetric Standard Model (MSSM). In the MSSM it is not possible to generate a large negative contribution to C_9, however, reducing the observed tension by affecting C_{10} would be possible. Another physics model would be a new heavy neutral gauge boson, *i.e.* a so called Z', due to which there would be a first order Flavour Changing Neutral Current. This boson would have to have a mass of $\mathcal{O}(\,\text{TeV})$ to meet constraints from other measurements. A lighter boson would be in contradiction with $t\bar{t}$ [62], $\tau\overline{\tau}$ [63,64] and $e\overline{e}$, $\mu\overline{\mu}$ [65,66] production cross section measurements done at Atlas and CMS. Also measurements of B_s^0 oscillation parameters done lately by LHCb [12] strongly constrain the properties of the Z' boson. At the current time it is still possible to explain a shift in C_9 (and C_{10}) and at the same time being compatible with other constraints.

Nevertheless, all these New Physics models are still only speculation. Although with recent measurements there are already strong constraints for New Physics models present, only with new more precise measurements the puzzle can be solved piece by piece. One of the central pieces is the update of the angular measurement of $B^0 \to K^{*0}\mu^+\mu^-$ using the full data set of LHCb Run 1 presented in this thesis.

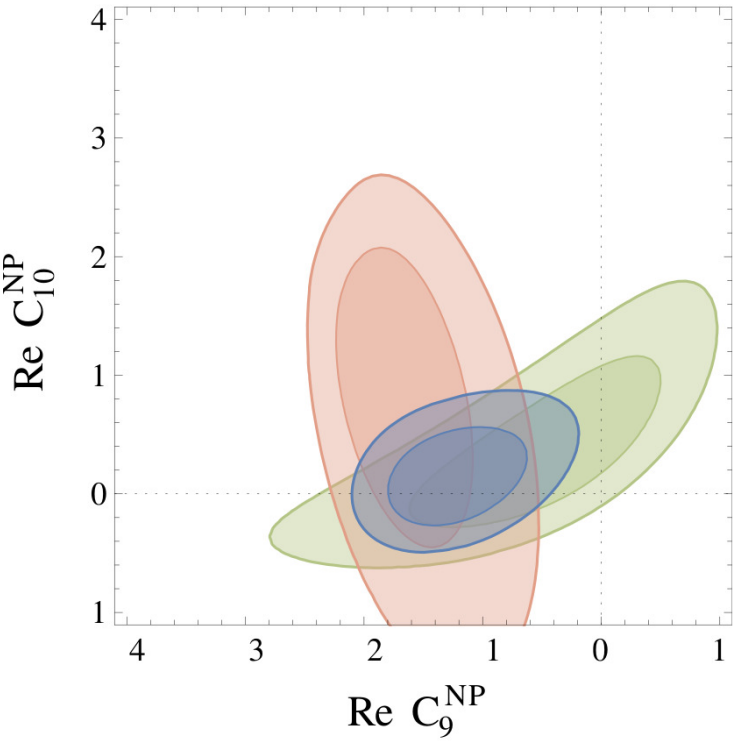

Figure 25: Combination of 81 independent measurements to estimate the most probable value of the Wilson coefficients [50]. It is tested for contributions additional to the Standard Model ($C_i^{obs} = C_i^{SM} + C_i^{NP}$). Always two Wilson coefficients are varied at a time. Shown are the 1 and $2\,\sigma$ confidence levels of a fit of C_9 and C_{10}, including only the angular measurements (red), only branching ratio measurements (green) and the combination of all measurements (blue). The best fit point is more than $2\,\sigma$ away from the Standard Model prediction.

5 Analysis overview

5.1 Summary

The goal of this analysis is the determination of the *CP*-averaged normalized observables, S_i, describing the angular distribution of the decay $B^0 \to K^{*0}\mu^+\mu^-$. The measurement is done in bins of the dimuon invariant mass squared, q^2, as observables are expected to vary with respect to this variable.

In the first part of the analysis the signal decay is selected from the data (see Sec. 7). Remaining background contributions are identified and studied. Then the best measurement method for the angular observables is searched for. Two different methods, the maximum likelihood fit and the Method of Moments, are tested on simulated events and their performances are compared (see Sec. 8). In the measurement detector effects must be accounted for. Combinatorial background must be described. Especially decays, in which the $K\pi$ decay products origin from a spin 0 resonances, the so called S-wave must be treated. Besides the eight angular observables S_i also six nuisance parameters describing this S-wave and S/P-wave interference are included in the measurement. Due to this, the expected statistical precision of the measurement would reduce significantly. As a countermeasure not only the angular distribution but also the $K\pi$ invariant mass distribution is included in the measurement. The maximum likelihood fit shows a better precision in comparison to the Method of Moments. Therefore the maximum likelihood fit is chosen to be the nominal method and the Method of Moments is used as a cross-check.

The validation of these methods (see Sec. 9) is done with two different simulations (see Sec. 6), each testing a different part of the measurement methods. Additionally the full measurement setup is tested on data on the decay $B^0 \to J/\psi(\to \mu^+\mu^-)K^{*0}$, which is theoretically well understood. This decay has the same final state and the same decay topology as the signal decay $B^0 \to K^{*0}\mu^+\mu^-$. Using this decay channel the complete measurement setup is tested in a realistic scenario and results are compared to those of a previous analysis.

Since New Physics is expected to be visible in more than one observable, it is important to analyse the complete set of observables. This requires to also determine the correlation between the measured observables. The correlation coefficients are extracted from the likelihood profile of the maximum likelihood fit (see Sec. 10). Systematic uncertainties in this analysis are determined on simulation mostly independent of the final measurement (see Sec. 11). As the decay $B^0 \to K^{*0}\mu^+\mu^-$ is a rare decay with a small event yield, in comparison to the statistical uncertainty all systematic uncertainties are small.

Finally results of the angular measurement are presented (see Sec. 12). The results are compared to previous measurements and to the Standard Model prediction. A deviation previously observed in one of the observables is verified, rising hope that an indication of New Physics is finally observed.

5.2 Angular observables

In this analysis eight *CP*-averaged (see Eq. 12) P-wave observables are determined in each bin of q^2. Furthermore, six nuisance parameters due to the S-wave and S/P-wave interference are also left floating in the fit. The measurement of these observables allows a model independent analysis of the angular distribution of the decay $B^0 \to K^{*0}\mu^+\mu^-$. In the measurement it is assumed that the *CP* asymmetries are negligibly small, thus that $J_i = \overline{J}_i$, and that lepton masses can be neglected in the angular description (see Eq. 36). The *CP*-averaged normalized decay rate can then be expressed as (compare to Sec. 4.3.6):

$$
\begin{aligned}
\frac{1}{\Gamma+\overline{\Gamma}}\frac{\mathrm{d}^5\left(\Gamma+\overline{\Gamma}\right)}{\mathrm{d}q^2\,\mathrm{d}m_{K\pi}\,\mathrm{dcos}\theta_l\,\mathrm{dcos}\theta_K\,\mathrm{d}\phi} =& (1-F_S)\frac{9}{32\pi}\left|BW_{P-wave}(m^2_{K\pi})\right|^2 \times \\
& \Big\{ S_{1s}\sin^2\theta_K \\
& + (1-4/3S_{1s})\cos^2\theta_K \\
& + 1/3S_{1s}\sin^2\theta_K\cos 2\theta_l \\
& + (4/3S_{1s}-1)\cos^2\theta_K\cos 2\theta_l \\
& + S_3\sin^2\theta_K\sin^2\theta_l\cos 2\phi \\
& + S_4\sin 2\theta_K\sin 2\theta_l\cos\phi \\
& + S_5\sin 2\theta_K\sin\theta_l\cos\phi \\
& + S_{6s}\sin^2\theta_K\cos\theta_l \\
& + S_7\sin 2\theta_K\sin\theta_l\sin\phi \\
& + S_8\sin 2\theta_K\sin 2\theta_l\sin\phi \\
& + S_9\sin^2\theta_K\sin^2\theta_l\sin 2\phi\Big\} \\
& + \frac{3}{16\pi}F_S\left|BW_{S-wave}(m^2_{K\pi})\right|^2\sin^2\theta_K \\
& + \frac{3}{16\pi}Re\left[BW_{S-wave}(m^2_{K\pi})BW^{\dagger}_{P-wave}(m^2_{K\pi})\right] \times \\
& \Big\{ S_{S1}\cos\theta_K\sin^2\theta_l \\
& + S_{S2}\sin\theta_K\sin 2\theta_l\cos\phi \\
& + S_{S3}\sin\theta_K\sin\theta_l\cos\phi\Big\} \\
& + \frac{3}{16\pi}Im\left[BW_{S-wave}(m^2_{K\pi})BW^{\dagger}_{P-wave}(m^2_{K\pi})\right] \times \\
& \Big\{ S_{S4}\sin\theta_K\sin\theta_l\sin\phi \\
& + S_{S5}\sin\theta_K\sin 2\theta_l\sin\phi\Big\}
\end{aligned}
\tag{38}
$$

6 Simulation

6.1 Full detector simulation

Proton-proton collisions in the LHCb detector are simulated, in order to study effects which are related to the detector response. The simulation is based on software which is now for over thirty years in development; with each increment in version more perfecting the description of high energy physics processes. On the other hand this software and also the detailed description of the detector has gotten so complex, that the production of this simulation is only possible with the worldwide LHC computing Grid [67]. The software packages used for the simulation are `Gauss` version v45r4, v45r7 [68] and `Boole` version v26r3 [69]. About an equal amount of the simulation is produced with version v45r4 and v45r7 of `Gauss`.

`Gauss` is used for the description of physics processes and consists out of `Pythia`, respectively, version 6.427.2 [70] and 8.175 [71], `Geant4` version v9.5.p02 [72, 73] and `EvtGen` version v13r6 [74]. Using a Monte Carlo simulation technique `Pythia` simulates proton-proton collisions and randomly generates particles according to the Standard Model description. The decay of these particles is simulated accordingly. Physical processes of the heavy B mesons are described by `EvtGen`. The simulation of all possible physics decays is out of reach for this analysis. To reduce the computing time events not containing the signal decay are discarded right in the beginning. The interaction of the final state particles with the detector is implemented with `Geant4` v9.5.p02, by which both multiple scattering with material and also the description of the response in active detector material is simulated.

The further processing is done by `Boole`: Noise is added to the detector response. The readout electronic and hardware trigger (L0) is simulated. The output of `Boole` are data samples which can be used in the further LHCb framework (see Sec. 3.3) similarly to the real data recorded with the LHCb detector. Out of the simulated detector responses tracks are reconstructed. All further High Level trigger and final offline selection requirements are applied identical to real data taking. Therefore, the simulation provides a realistic description of the reconstruction and also selection efficiency.

There are several different simulation samples generated for this analysis:

1. The decay $B^0 \to J/\psi\, K^{*0}$: 107k reconstructed and selected events
2. The decay $B^0 \to K^{*0}\mu^+\mu^-$: 49k reconstructed and selected events
3. The decay $B^0 \to K^{*0}\mu^+\mu^-$ (PHSP) : 1.5M reconstructed and selected events
4. The decay $\Lambda_b^0 \to pK^-\mu^+\mu^-$ (PHSP) : 1M generated events
5. The decay $B_s^0 \to \phi\mu^+\mu^-$: 600k generated events
6. The decay $B^+ \to K^+\mu^+\mu^-$: 1M generated events

In all cases only the signal P-wave decay is implemented in the physical description. In the so called PHase SPace (PHSP) simulation samples the spin structure of the particles is neglected and decays are generated only according to their kinematic properties. Otherwise physics is described according to the Standard Model.

The full detector simulation is used to measure the reconstruction and selection efficiency (see Sec. 8.3), to validate the corresponding correction method (see Sec. 9) and to assign appropriate systematic uncertainties to the measurement (see Sec. 11).

6.2 High statistics simulation

A high statistics simulation containing both background and also the S and P-wave signal decay is produced. With this simulation the measurement methods can be tested for the correct background description, which is not possible with the simulation described in the last subsection. The generation process of the simulation is speed up by implementing the detector response and event selection in a simplified way. Therefore 10^7 reconstructed and selected events can be generated within 24 hours on a computing cluster with about 100 CPU cores.

The signal P-wave decay is described with the help of the program `EOS` (version from 2015-03-18) [75, 76]. The program `EOS` calculates the 4D differential decay rate (q^2, three decay angles) of the decay $B^0 \to K^{*0}\mu^+\mu^-$ using latest Standard Model predictions of the Wilson coefficients and form factors. Drawing random numbers with a Monte Carlo technique events are generated according to the calculated differential decay rate. With `EOS` it is not possible to calculate the differential decay rate of the S-wave contribution. Therefore, randomly S-wave decays are generated according to Eq. 37 in Sec. 4.3.6. The fraction of the S-wave, F_S, is set to 5%, orientating on the measurement of the control channel (see Sec. 9.3). The S/P-wave interference terms cannot be generated using `EOS` and are neglected in the generation, as their impact in the measurement is small. Similarly to the signal decay, randomly background events are generated according to the description used in the maximum likelihood fit (see Sec. 8.1.1). The signal fraction and the slope of the $K\pi\mu\mu$ invariant mass of the combinatorial background are taken from data (see Tab. 2 in Sec. 8.6). A flat distribution is simulated for the background in the decay angles and in the $K\pi$ invariant mass. The detector response is only implemented in a simplified way with a pass-and-fail method. Based on the simulation explained in the last subsection the reconstruction and selection efficiency is measured (see Sec. 8.3). An event is randomly taken (or rejected) with a probability proportional to its expected reconstruction and selection efficiency. Therefore, reconstruction and selection effects are simulated, however, the simulation cannot be used to test the efficiency correction.

The simulation is used to test the full measurement method (see Sec. 8.1). Multiple simulation samples are generated to study systematic effects (see Sec. 11). The total size of each sample is about 10^7 events, large enough that statistical fluctuations become negligible.

7 Selection and background

Besides the signal decay $B^0 \to K^{*0}\mu^+\mu^-$ also other (physical) processes are recorded, which must be distinguished from the signal decay. The decay topology of the decay $B^0 \to K^{*0}\mu^+\mu^-$ is sketched in Fig. 26. The B^0 meson is produced in the proton-proton collision, and decays after flying few mm away from the production vertex. The final state particles are two muons and a kaon and a pion. The kaon and pion originate from a K^{*0} resonance and the invariant mass of the $K\pi$ system is therefore near the K^{*0} mass. This is not the case for the two muons, which do not origin from any intermediate resonance.

The main source of background is the wrong association of four independent tracks to a fake B^0 meson, so-called combinatorial background. This kind of background shows an exponential behaviour in the $K\pi\mu\mu$ invariant mass and thus can be nicely separated from the peaking signal decay in this variable.

Another source of background are B decays, which are similar to the signal decay and in which one of the final state particles is misidentified. For example there is the decay $B_s^0 \to \phi\mu^+\mu^-$, which has a similar branching ratio and similar decay topology as the signal decay. The ϕ meson usually decays into two kaons. If one kaon is misidentified as a pion, the hole decay can be misidentified as the decay $B^0 \to K^{*0}\mu^+\mu^-$. Since a pion is lighter than a kaon the reconstructed invariant mass of the system is usually slightly smaller. Nevertheless, this kind of background has a broad peaking structure near the B^0 resonant mass. Therefore it is also called a peaking background.

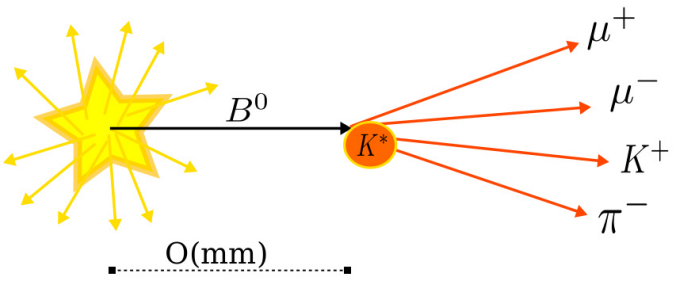

Figure 26: Sketch of the $B^0 \to K^{*0}\mu^+\mu^-$ decays. The B^0 meson is produced at the proton-proton collision point (yellow) and decays after flying few mm away from it. The lifetime of intermediate resonances is negligible, and the final state particles (kaon, pion, two muons) can be treated as originating from the same decay vertex.

Additionally to the signal decay $B^0 \to K^{*0}\mu^+\mu^-$, there are also decays in which the two muons originate from a charmonium resonance. The two prominent resonances in the observed q^2 window are J/ψ and $\Psi(2S)$. These decays have exactly the same decay topology as the signal decay and cannot be separated from it. They occur several orders of magnitude more frequent than the signal decay, however, they only appear at well-defined values of the dimuon invariant mass. Therefore the decay $B^0 \to K^{*0}\mu^+\mu^-$ cannot be studied in the whole q^2 range, but only in regions where the contribution from charmonium resonances is negligible. The q^2 region of the decay $B^0 \to J/\psi\, K^{*0}$ is used as a control channel, since the angular distribution of this decay is well understood.

The data set used in this analysis is the full available data set from LHCb Run 1, corresponding to an integrated luminosity of $3\,\text{fb}^{-1}$. The selection requirements applied to this data set to extract the $B^0 \to K^{*0}\mu^+\mu^-$ signal decay are explained in the following. At first the trigger strategies are detailed (see Sec. 7.1). The complete recorded data set is reduced in a preselection (see Sec. 7.2). Specific physical background is removed by corresponding veto requirements (see Sec. 7.3). A final selection is applied using a multivariate analysis technique (see Sec. 7.4). At last veto requirements to distinguish between the signal decay and the control channel are discussed (see Sec. 7.5).

7.1 Trigger

Several selection requirements are applied to the events by the trigger. As decisions in the trigger have to be computed fast they are mainly based on kinematic quantities and a clear separation of final state tracks from the proton-proton collision point.

In the analysis only events are included for which the signal decay $B^0 \to K^{*0}\mu^+\mu^-$ caused the trigger[13]. As the trigger and reconstruction efficiency is studied on simulated samples, this requirement reduces the complexity of the simulation minimizing related systematic uncertainties.

The trigger in the first stage (L0) requires the identification of at least one muon. If at least one muon candidate with a transverse momentum of 1.48 GeV (in 2011) and 1.76 GeV (in 2012) is found the event is passed to the High Level Trigger.

The first stage of the High Level Trigger searches for a particle candidate with good track quality requirements and a transverse momentum[14] of at least $p_T > 1.6\,\text{GeV}$. If the track matches with the detector responses in the muon chambers and it is not pointing to the proton-proton collision point this requirement is relaxed to $p_T > 1\,\text{GeV}$.

[13] It is allowed, that additionally also an independent particle triggered. However, events in which the decay $B^0 \to K^{*0}\mu^+\mu^-$ is only identified after the reprocessing are explicitly removed from this analysis.

[14] The momentum measurement using the tracking stations is used which is much more precise than the one of the muon stations used in the L0 trigger.

In the second stage of the High Level Trigger two strategies are exploited, a topological trigger and a trigger searching for (di)muon candidates.

The first strategy is based on a topological analysis of the tracks. In a multivariate analysis it is checked if tracks can be combined to a N-body decay, where $N = 2, 3, 4$. The invariant mass of the mother particle is corrected for missing energy, estimated by requiring (transverse) momentum conservation. Thus, even if in a 4 body decay (*e.g.* $B^0 \to K^{*0}\mu^+\mu^-$) not all daughter tracks are correctly identified, events may be triggered. The input variables are the distance of closest approach of the tracks, the separation from the proton-proton collision point, the corrected mass and further kinematic variables. To reduce fluctuations the variables are discretized before using them. In a modified version of this strategy at least one of the tracks is required to be identified as muon. Due to this the other selection requirements can be relaxed.

The second strategy is based on the search for single or dimuon candidates. Single muon candidate, which are separated from the proton-proton collision point which fulfil strong track quality requirements, are selected. For dimuon candidates, being differently charged and originating from a common vertex that is separated from the proton-proton collision point, the track quality requirements are less strict compared to the single muon trigger.

A fraction of 78% of the recorded and selected events are triggered by the single or dimuon trigger. The topological trigger selected 92% of the recorded events and is thus responsible for the dominant contribution. No significant increase in signal yield can be obtained using additional trigger strategies.

7.2 Preselection

In the preselection decays with the decay topology of the decay $B^0 \to K^{*0}\mu^+\mu^-$ are selected. The preselection requirements are divided into different groups. In the first group global event properties are checked. The probability of an event to consist of combinatorial background increases with the activity inside an event. Therefore events with a high activity (number of tracks) are rejected from the analysis.

Furthermore, selection requirements are applied to the final state particles, the combination of these final state particles (dimuon, K^{*0} resonance) and also to the B^0 itself. All final state particles are required to have a reasonable track quality and are separated from the proton-proton interaction point. The output of a multivariate classifier is used to identify and remove ghost candidates, which are candidates that do not belong to a real particle but only are artefacts of the reconstruction. Furthermore, tracks must be separated from other tracks in the event to prevent the wrong association from detector responses to the track.

The muon candidates must be matched to detector responses in the muon chamber. Also information from the RICH particle identification system is used. If the likelihood of a muon candidate is large to be a pion, the corresponding candidate is rejected. If the likelihood of a kaon (pion) candidate to be a kaon (pion) is

significant smaller than to be a pion (kaon) these particles are also removed. The dimuon and K^{*0} vertex are required to have a reasonable quality. The dimuon invariant mass must be smaller than 7.1 GeV and the $K\pi$ invariant mass must be smaller than 6.2 GeV.

The B^0 vertex quality must be reasonable. The B^0 meson is required to be close to at least one proton-proton interaction point. The momentum vector must point in a similar direction as the vector pointing from the proton-proton collision point to the B^0 decay vertex. The flight distance of the B^0 divided by the uncertainty of this measurement must be significantly larger than zero. At last the $K\pi\mu\mu$ invariant mass must be in-between 4.8 and 7 GeV. The final angular analysis is done in the invariant mass range from 5.17 to 5.70 GeV.

7.3 Physical background

There are several decays, which, if wrongly reconstructed, can be misidentified as the signal decay. In contrast to combinatorial background these decays have certain structures in phase space. Thus they violate the assumption that the angular distribution of the background is independent of the $K\pi\mu\mu$ invariant mass. In an extensive study the major contributions are identified and additional veto requirements are implemented in the selection in order to suppress the corresponding background decays.

A large contribution in the observed invariant mass range is due to the decay $\Lambda_b^0 \to pK^-\mu^+\mu^-$. If the proton is misidentified as a pion the decay topology is similar to the signal decay. The invariant mass is recalculated assigning to the pion the proton hypothesis. If the invariant mass is close (in between 5575 and 5665 MeV) to the Λ_b^0 mass and the probability due to the particle identification system of the pion to be a proton is non-negligible these events are rejected.

Another larger source of peaking background is the signal decay itself, when the kaon is misinterpreted as being the pion and vice versa. This background is further referred to as *signal swap*. Although the invariant mass difference between a kaon and a pion is small, the effect on the angular distribution is large due to the definition of the decay angles (see Sec. 4.3.1). If the likelihood of this swap is very large, because of the measurements from the particle identification system, events are removed. The particle hypotheses are switched and the invariant mass is recalculated. Events are rejected if the invariant $K\pi$ mass is between 795 and 995 MeV and the particle identification system indicates a large probability for a swap.

A further source of peaking background is the decay $B_s^0 \to \phi\mu^+\mu^-$ ($\phi \to K^+K^-$), for which a kaon is misidentified as a pion. Assigning to the pion the kaon particle hypothesis the $KK\mu\mu$ invariant mass is calculated. It is tested if the $KK\mu\mu$ invariant mass is comparable with the B_s^0 mass (between 5321 and 5411 MeV) and if the corresponding KK invariant mass is also close to the ϕ mass (between 1010 and 1030 MeV). If further the likelihood of a particle misidentification is large,

the candidate is rejected. If the recalculated KK mass is only in a larger mass window of 1030 and 1075 MeV, but the particle identification system indicates a large probability for a pion-kaon misidentification the candidate is also removed. The decay $B^+ \to \rho^0 (\to \pi^+\pi^-)\, \mu^+\mu^-$ can be mis-reconstructed as a signal decay if the pion is identified as a kaon. The transition from the third to the first quark generation ($b - d$ transition) is less frequent than the transition from the third to the second quark generation ($b - s$ transition) by a factor 20. Therefore combined with the small misidentification probability this decay is considered to be negligible.

A source of peaking background, in which all final state particles hypotheses are correctly identified, is the decay $B^+ \to K^+\mu^+\mu^-$, when an additional random pion is combined to this decay. This type of background cannot be reduced by harder selection requirements of the particle identification system. Due to the additional energy the $K\pi\mu\mu$ invariant mass is larger than the B^0 mass and these decays have a peaking structure right next to the signal peak. Events are removed if the $K\pi\mu\mu$ invariant is larger 5380 MeV and the $K\mu\mu$ invariant mass (without the pion) is between 5220 and 5340 MeV. This selection requirement has a direct impact on the $\cos\theta_K$ distribution and its description is explicitly implemented in the measurement methods (see Sec. 8.1.2).

The discussed veto selection requirements significantly reduce the contribution from peaking backgrounds. After also applying the final selection, explained in the next subsection, the remaining dominant contributions relative to the signal decay are from:

- $\Lambda_b^0 \to pK^-\mu^+\mu^-$ $(1.0 \pm 0.4\%)$
- signal swaps $(0.64 \pm 0.06\%)$
- $B_s^0 \to \phi\mu^+\mu^-$ $(0.33 \pm 0.12\%)$
- $B^+ \to K^+\mu^+\mu^-$ $(0.031 \pm 0.006\%)$

The systematic impact of these decays on the measurement is analysed in Section 11.3.

7.4 Final selection

The final selection is based on the response of a Boosted Decision Tree (BDT) [77]. A single decision tree is the combination of several conditions (e.g. if $x > a\ AND\ y > b\ OR\ z < c$) to obtain a binary decision providing the best guess if an event is background or signal. *Boosting* means the combination of many ($\mathcal{O}(1000)$) different of those decisions trees to one weighted response. The response of a BDT is a value between -1 (background) and 1 (signal). Even if each decision tree has only a limited power to separate signal and background the combination of many trees provides a significantly larger separation power than a cut-based-only analysis. In comparison to a cut-based selection in the BDT (also nonlinear) correlations of the

input variables are efficiently exploited. The BDT implementation provided by the Toolkit for Multivariate Data Analysis [78] is used.

The BDT must be trained on realistic signal and background data samples to work efficiently. The control channel $B^0 \to J/\psi K^{*0}$, which looks very similar to the signal decay is used to train the BDT in order to identify signal events. The unfolding of the signal events from combinatorial background in the control channel is done using the sPlot technique (see Sec. 8.2.2). As background training sample the $K\pi\mu\mu$ invariant mass region from $5350 - 7000\,\text{MeV}$ of the $B_s^0 \to K^{*0}\mu^+\mu^-$ sample is used. Background below the B^0 mass peak also contains partially reconstructed decays, which are excluded from this analysis. The region $5350 - 5700\,\text{MeV}$ is identical to the one used later in the angular measurement, corresponding to about 17% of the events of the measurement.

It is not advisable to train a BDT on events which will be later analysed by this BDT. In this case the BDT response is usually biased and the performance of the BDT is overestimated. In this analysis this double usage is circumvented by using the k-folding technique [79]. The samples are split into 10 equal parts. The BDT for one part is trained on the nine different parts. Therefore each BDT is statistically independent from the training data.

The input variables used to train the BDT are the B^0 lifetime, B^0 vertex quality, B^0 (transverse) momentum and the pointing-angle between the measured momentum of the B^0 and the vector from the proton-proton collision point to the B^0 decay vertex. Furthermore, variables of the particle identification system are used. For the kaon and pion it is the logarithm of the ratio of likelihoods to be a kaon or a pion. The ratio of the likelihood to be a muon and to be a pion is used for the muon candidates. At last also the spatial separation from the muons and hadrons to other particles in the events is used. All these variables show a good separation power between signal and background.

In the final selection a single selection requirement on the BDT response is applied. This requirement is tuned on the Figure of Merit, given as,

$$FoM = \frac{S}{\sqrt{S+B}}, \tag{39}$$

where S and B are the number of signal and background candidates, respectively, in the $K\pi\mu\mu$ invariant mass signal window $5230 - 5330\,\text{MeV}$ (about $\pm 3\,\sigma$ around B^0 mass). The number of signal candidates is estimated from the control channel scaled with the expected branching ratio from the PDG [80] and the ratio of reconstruction efficiencies taken from simulation. The background yield is determined by extrapolating the background of the decay $B^0 \to K^{*0}\mu^+\mu^-$ from the outer $K\pi\mu\mu$ invariant mass regions ($5000 - 5180\,\text{MeV}$ and $5500 - 7000\,\text{MeV}$) to the centre of the signal window.

The $K\pi\mu\mu$ invariant mass distribution of the control channel after the preselection and after the final selection are shown in Fig. 27. A clean mass peak is visible after the full selection.

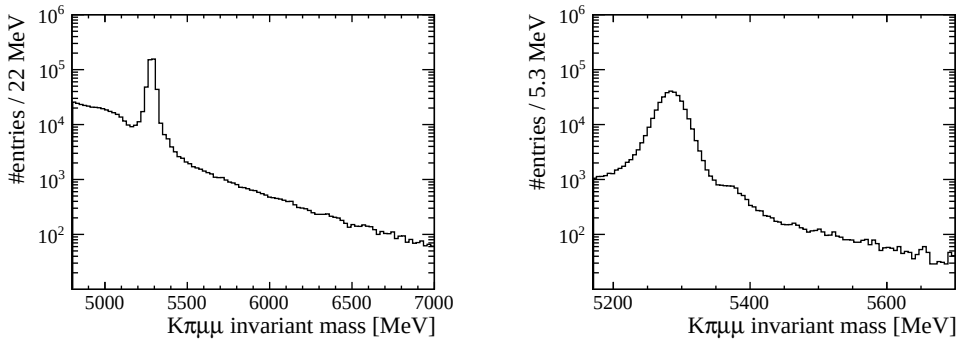

Figure 27: The $K\pi\mu\mu$ invariant mass distribution of the control channel $B^0 \to J/\psi\, K^{*0}$ after the preselection (left) and after the final selection (right). By using the BDT the background level can be reduced significantly. Partial reconstructed decays are visible for smaller masses, which is the reason for omitting this mass region in the angular analysis. Right of the mass peak a small bump related to the decay $B_s^0 \to J/\psi\, K^{*0}$ is visible (see Sec. 9.3).

7.5 Charmonium resonances

The separation of the decay $B^0 \to K^{*0}\mu^+\mu^-$ and the decay $B^0 \to J/\psi\, K^{*0}$ is done by looking at the dimuon invariant mass. In Fig. 28 a 2D plot of the dimuon invariant mass and the $K\pi\mu\mu$ invariant mass is shown. The J/ψ and $\Psi(2s)$ resonances are clearly visible. Therefore the q^2 binning (see Tab. 1) is chosen such that the two resonances are excluded from the investigated regions. If the measured invariant mass of the dimuon system is too small (e.g. one of the muons emitted Bremsstrahlung before being measured) also the $K\pi\mu\mu$ invariant mass is underestimated and vice versa. This can be seen as diagonal bands in the plot. To allow measuring up to a higher q^2 region a rather hard selection requirement on the lower $K\pi\mu\mu$ invariant mass is applied. As can be seen with the chosen invariant mass windows used in this analysis, the decays over charmonium resonances are completely removed. It is also checked using the full detector simulation (see Sec. 6.1) that the decay $B^0 \to J/\psi\, K^{*0}$ does not leak into the measured q^2 range.

For studies of the control channel $B^0 \to J/\psi\, K^{*0}$ the region in q^2 from 8 to 11 GeV^2 ($m_{\mu\mu} = 2828 - 3317$ MeV) is explicitly chosen.

7.6 Event yield

The total event yield after the final selection and the signal fraction, f_{sig}, of the $B^0 \to K^{*0}\mu^+\mu^-$ decay are listed in Tab. 2. The signal fraction is determined with a 1D maximum likelihood fit to the $K\pi\mu\mu$ distribution (as will be explained in Sec. 8.1). The total number of signal candidates which is analysed in the angular measurement is 2342. In comparison in the control channel $B^0 \to J/\psi\, K^{*0}$ there are about 300 000 signal candidates after the final selection.

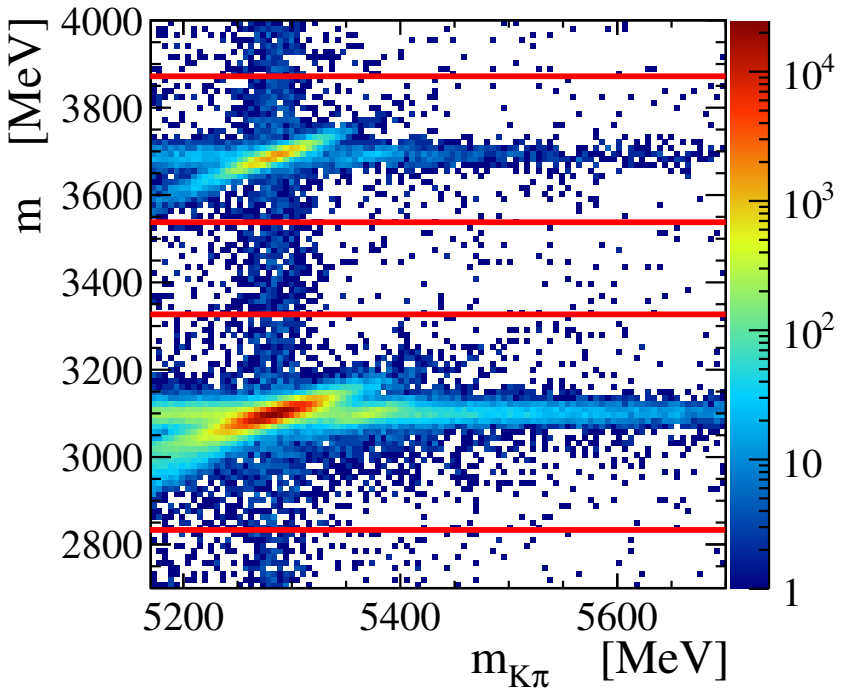

Figure 28: Dimuon invariant mass, $\sqrt{q^2}$, vs. $K\pi\mu\mu$ invariant mass. The horizontal bands show the J/ψ resonance at about 3100 MeV and the $\Psi(2s)$ at about 3700 MeV. These resonances are removed from the measurement but used as control channels for validation. The vertical band shows the signal decay $B^0 \to K^{*0}\mu^+\mu^-$.

Table 2: 1D maximum likelihood fit to the $K\pi\mu\mu$ distribution of $B^0 \to K^{*0}\mu^+\mu^-$ to extract the signal fraction, $f_{\rm sig}$. Also shown is the total number of events in each q^2 bin present in data.

q^2 bin $[\,\mathrm{GeV}^{-1}]$	#events	$f_{\rm sig}$
$0.1 - 0.98$	395	0.851 ± 0.026
$1.1 - 2.5$	300	0.584 ± 0.039
$2.5 - 4.0$	366	0.409 ± 0.037
$4.0 - 6.0$	569	0.486 ± 0.029
$6.0 - 8.0$	678	0.453 ± 0.027
$11.0 - 12.5$	535	0.626 ± 0.029
$15.0 - 17.0$	632	0.718 ± 0.025
$17.0 - 19.0$	439	0.702 ± 0.031

8 Parameter estimation

There are several different estimators available, which can be used to extract the most probable set of physics parameters from data. One of these techniques is the Maximum Likelihood Estimator (MLE) which is discussed in Sec. 8.1. Another estimator is the Method of Moments (MoM) which is introduced in Sec. 8.2. In Sec. 8.3 the correction of the non-uniform reconstruction and selection efficiency is detailed. The determination of the confidence intervals of the estimated values is explained in Sec. 8.4. The performance of the different methods is compared in Sec. 8.5 and Sec. 8.6.

8.1 Maximum likelihood fit

The likelihood $L(\theta; x)$ is defined as the probability density to measure a set of events x given a physical model and a set of physical parameters θ. The best estimate for the true values of the parameters θ are according to the MLE [80–82] the ones which maximize L. Indeed this assumption seems quite reasonable and naively one would expect this to be the best choice. However, throughout history there has been a long discussion about if the MLE really is the best estimator [83].

The MLE is a consistent estimator, thus with enough statistics its estimates converge asymptotically to the true value. Furthermore, it is asymptotically normal. With sufficient data the estimates are normally distributed around the true value. The variance of the estimates asymptotically converges to the Cramér Rao Lower bound. There does not exist any unbiased estimator which has a better performance (smaller spread of the estimates) than given by this bound. Especially the last point is one of the reasons why the MLE is so successful. However, all these properties are valid only asymptotically for a large number of measurements. In practice the MLE is known to be heavily biased in some cases especially for small data sets.

The likelihood is calculated using the Probability Density Functions (PDF) explained in the next subsection. The maximum of the likelihood (or minimum of the negative log likelihood) is determined with `Minuit` [84], which has several routines implemented to minimize multidimensional functions. In this analysis the method `MIGRAD` is chosen, which uses information of the (numerically calculated) first derivatives to reliably find the global minimum of a function. The implementation of the PDF is done with a `ROOT` based framework.

Usually the uncertainty of one single estimate is estimated by the form of the likelihood. Either the second derivative `HESSE` matrix can be used or the likelihood can be scanned using `MINOS` (application of `Minuit`). These techniques do not provide satisfying results if the data sample is small or when the best fit point is near physical boundaries. The Bootstrap method does provide reliable uncertainty estimates also in these cases. It works by re-sampling of the data set multiple times and is explained in detail in Sec. 8.4. The uncertainties of the results presented in this analysis are based on the Bootstrap method.

8.1.1 Implementation

In this analysis a five dimensional fit is used to extract the physical parameters in each q^2 bin. The five observables are the $K\pi\mu\mu$ invariant mass, the $K\pi$ invariant mass and the three decay angles fully describing the decay $B^0 \to K^{*0}\mu^+\mu^-$ at a given value of q^2. The $K\pi\mu\mu$ invariant mass is used to separate the signal decay from combinatorial background. A better constraint on the S-wave fraction is possible by fitting also the $K\pi$ invariant mass. The physical parameters of interest, S_i, are extracted with the decay angles.

As all measurements are statistically independent from each other the total likelihood is calculated as the product of the single PDFs. Concerning computing precision it is more convenient to sum the logarithms of the PDF. The quantity which is maximized in the fit is:

$$\ln L(x;\theta) = \sum_{\mathrm{ev}=1}^{N} w_{\mathrm{ev}} \ln PDF_{\mathrm{ev}}(x;\theta), \tag{40}$$

where w_{ev} is the weight to correct for the angular acceptance for the event ev (see Sec. 8.3) and the PDF of one event is calculated as follows:

$$PDF_{\mathrm{ev}}(x;\theta) = f_{\mathrm{sig}} PDF_{\mathrm{sig}}(x;\theta) + (1 - f_{\mathrm{sig}}) PDF_{\mathrm{bkg}}(x;\theta), \tag{41}$$

where f_{sig} is the fraction of signal candidates in the sample, and PDF_{sig} and PDF_{bkg} correspond to the probability for the event to be, respectively, signal and background.

Signal Component

The signal component of the PDF is further split up:

$$PDF_{\mathrm{sig}}(x;\theta) = PDF_{mass}(x;\theta) \cdot PDF_{ang}(x;\theta), \tag{42}$$

where PDF_{mass} is the probability density function describing the $K\pi\mu\mu$ invariant mass and PDF_{ang} the one describing the distribution in the three decay angles and $K\pi$ invariant mass. The 4D differential decay rate described in Sec. 4.3.6 by Eq. 37 is transformed into the probability density function PDF_{ang}[15]. The eight angular parameters, S_i, are the heart of the measurement and are left floating in the fit. In addition there are six nuisance parameters describing the angular distribution of the S-wave and the S/P-wave interference, F_S, S_{S1-S5}. These observables are of no interest for the analysis, however, they are also left floating in the fit as precise predictions for these parameters are not available and a q^2 dependence of these parameters is expected. The descriptions of the $K\pi\mu\mu$ and $K\pi$ invariant mass distributions used in the signal PDF are detailed in the following.

[15] As the CP-averaged angular distribution is measured (see Eq.12) the J_i in Eq. 37 are replaced by S_i and the total decay rate must be normalized to unity. Furthermore, the relations in Eq. 36 are used to express S_{1c}, S_{2s}, S_{2c} by S_{1s}.

$K\pi\mu\mu$ invariant mass

There are two main effects which must be accounted for describing the invariant $K\pi\mu\mu$ mass distribution. First, due to the resolution of the detector the measured mass shows a Gaussian behaviour. The second effect is that the final state particles can do Bremsstrahlung before being detected. The soft radiated photon is neglected in the reconstruction and therefore the reconstructed mass in these cases is systematically too small. There is no significant effect on the angular distribution due to the missing photon and these decays can also be treated as signal.

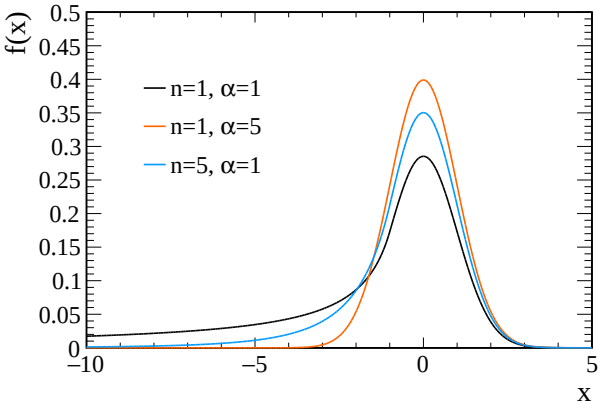

Figure 29: The Crystal Ball function as defined in the text for three different sets of parameters. The mean and σ is in all cases zero and one respectively.

An empirical function describing both effects is the so called Crystal Ball[16] (CB) function [85]. The CB function for three different sets of parameters is plotted in Fig. 29. A CB function is in the core a Gaussian function describing the finite detector resolution with an additional Power Law behaviour on one side of the Gaussian describing the so called *radiative tail*:

$$f(x;x',\sigma,\alpha,n) = N \begin{cases} \exp\left(-0.5\frac{(x-x')^2}{\sigma^2}\right), & x > -\alpha \\ a\left(b - \frac{x-x'}{\sigma}\right)^{-n}, & x \leq -\alpha \end{cases} \tag{43}$$

with N being the normalization, and

$$a = \left(\frac{n}{|\alpha|}\right)^n \exp\left(-0.5\alpha^2\right) \tag{44}$$

$$b = \frac{n}{|\alpha|} - |\alpha|\,. \tag{45}$$

As there is not only one single effect being responsible for the resolution, usually one CB function is not enough to properly describe the detector response. Instead

[16]The naming is due to being developed in the Crystal Ball collaboration.

the invariant mass is described by the sum of two CB functions. The parameter n and the mean of both functions is the same, however, they have different parameters α and σ. The PDF describing the $K\pi\mu\mu$ invariant mass is:

$$PDF_{mass}(x;\theta) = f_{\text{res}} \cdot CB1\,(m; m_{B^0}, \sigma_{m,1}, \alpha_1, n) + (1 - f_{\text{res}}) \cdot CB2\,(m; m_{B^0}, \sigma_{m,2}, \alpha_2, n) \tag{46}$$

where f_{res} is a factor to express the relative contribution of both components. The systematic uncertainty related to this choice of model is determined in Sec. 11.5.

In the high statistics control channel $B^0 \to J/\psi\, K^{*0}$ there is also the decay $B_s^0 \to J/\psi\, K^{*0}$ visible (see Fig. 27). As both decays are described by very similar physics this decay is treated as signal in the control fit. The invariant mass is described by the same shape as the one of the signal decay and all mass shape parameters are shared in the fit. Additional parameters which will be left floating in this fit are the mass difference of the B_s^0 and B^0 and the event yield fraction of both. About 1% of the decays are expected to originate from a B_s^0, as previously determined [86]. In the measurement of the decay $B^0 \to K^{*0}\mu^+\mu^-$ the contribution of B_s^0 events is small enough to be neglected. A systematic uncertainty will be assigned (see Sec. 11.4).

Scaling Factor for the $K\pi\mu\mu$ invariant mass resolution

The parameters describing the mass shape are in the fit of $B^0 \to K^{*0}\mu^+\mu^-$ fixed to the values obtained in the fit of the control channel $B^0 \to J/\psi\, K^{*0}$ (see Sec. 9.3). In this channel the dimuon invariant mass is equal to the J/ψ mass, corresponding to a q^2 value of about $9\,\text{GeV}^2$. The invariant mass resolution depends on the momentum depended momentum resolution of the daughter particles and also on the decay angles. Therefore also the mass description depends on q^2. As the parameters describing the mass shape are fixed to the values obtained at $q^2 \approx 9\,\text{GeV}^2$ it is important to model any q^2 dependence of these parameters.

Using the full detector simulation (see Sec. 6.1) the q^2 dependence of the mass shape description is studied. All dependences on q^2, but the one of the mass resolution, are found to be negligible. The mass resolution gets worse by about 8% in the very high q^2 region, where the transverse momentum and thus also the opening angle of the kaon and pion is smallest.

The dependence of the mass resolution is implemented in the fit as follows. In the $B^0 \to K^{*0}\mu^+\mu^-$ full detector simulation the invariant $K\pi\mu\mu$ mass is measured in bins of q^2 and compared to the $B^0 \to J/\psi\, K^{*0}$ full detector simulation. The corresponding ratios are taken as scaling factors, which are applied in the fit to data (see Tab. 3). In the fits to data, again $B^0 \to J/\psi\, K^{*0}$ is taken as reference channel. In each q^2 bin in the fit the mass resolution is fixed to the one measured in this reference channel multiplied by the scaling factor obtained in simulation.

Table 3: Scaling Factor of the $K\pi\mu\mu$ invariant mass resolution obtained from simulation. Shown is the ratio of the measured resolution in each $B^0 \to K^{*0}\mu^+\mu^-$ q^2 bin divided by the resolution measured in the $B^0 \to J/\psi\, K^{*0}$ simulation.

$q^2[GeV^2]$	Scaling Factor
$0.1 - 0.98$	0.984 ± 0.013
$1.1 - 2.5$	1.037 ± 0.018
$2.5 - 4.0$	0.977 ± 0.017
$4.0 - 6.0$	0.992 ± 0.014
$6.0 - 8.0$	0.983 ± 0.012
$11.0 - 12.5$	1.014 ± 0.013
$15.0 - 17.0$	1.045 ± 0.013
$17.0 - 19.0$	1.075 ± 0.017

$K\pi$ invariant mass

The B^0 meson can decay via several resonances into a kaon and pion. In this analysis the decay via the $K^{*0}(892)$ (if it is clear out of the context denoted as K^{*0}) is studied. The advantage of this choice is, that because the $K^{*0}(892)$ is a vector meson there is a multitude of angular observables which can be studied.

The $K^{*0}(892)$ resonance has a rather large decay width of about 51 MeV [80] and therefore the $K\pi$ invariant mass range is chosen to be 895.9 ± 100 MeV. In this mass range there is also a significant contribution from other resonances. There is the $K_0^*(1430)$ and the $K_0^*(800)$[17] which must be also included in the description. Further resonances are found to not have a significant contribution in the studied mass range.

Both the $K_0^*(1430)$ and the $K_0^*(800)$ resonance are scalar and thus are denoted as the S-wave contribution. In contrast to decays via the $K^{*0}(892)$ resonance, the so called P-wave contribution, these decays have a different angular distribution. Therefore it is important to disentangle both components.

Due to the large decay width of the resonances, larger than the detector resolution, the invariant mass distribution is described by Breit Wigner functions. In the analysed invariant mass range only the tails of the Breit Wigner functions describing the $K_0^*(1430)$ and the $K_0^*(800)$ distributions are present. The functions describing both the S-wave and P-wave contribution are shown in Fig 30. The S/P-wave interference terms are neglected in this visualization.

In the 5D differential decay rate the $K\pi$ invariant mass distribution appears as a P-wave, S-wave and as the interference of both (see Eq. 37 in Sec. 4.3.6). The P-wave is described by a single Breit Wigner function $|BW_{\mathrm{P-wave}}(m^2_{K\pi})|^2$, where $BW_{\mathrm{R}}(m^2_{K\pi})$ is a complex factor corresponding to the resonance(s) R explained in

[17] The $K_0^*(800)$ resonance is not yet perfectly understood and needs further confirmation. However, as an effective description for the background this does not affect the current analysis.

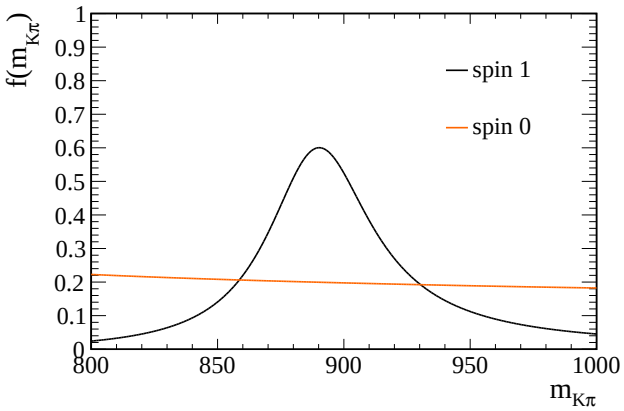

Figure 30: The function $|BW_{\mathrm{P-wave}}(m^2_{K\pi})|^2$ and $|BW_{\mathrm{S-wave}}(m^2_{K\pi})|^2$ as defined in the text for parameter values suggested by the PDG [80]. In the range of interest ($m_{K\pi} \sim 800 - 1000$) the S-wave (spin 0) contribution is nearly flat.

the following. Similarly the S-wave, $|BW_{\mathrm{S-wave}}(m^2_{K\pi})|^2$, is implemented as the sum of two Breit Wigner functions (one for the $K^*_0(1430)$ and one for the $K^*_0(800)$) scaled by a complex factor. This ansatz is the so called Isobar model. The complex scaling factor is expressed usually by two real numbers α and β as $\alpha e^{-i\beta}$:

$$BW_{\mathrm{S-wave}}(m_{K\pi}) = \alpha e^{-i\beta} BW_{800}(m_{K\pi}) + BW_{1430}(m_{K\pi}) \tag{47}$$

The complex decay amplitude BW_{R} is defined as follows:

$$BW_{\mathrm{R}}(m_{K\pi}, q^2) = N\sqrt{k}B'_{L_{K\pi}} \left(\frac{k}{k_0}\right)^{L_{K\pi}} \sqrt{p}B'_{L_B} \left(\frac{p}{p_0}\right)^{L_B} \frac{1}{m^2_{K\pi} - m^2_R - im_{K\pi}\Gamma'_R(m_{K\pi})} \tag{48}$$

where B'_i are so called Blatt-Weisskopf barrier factors accounting for spin dependent effects [87], defined as

$$B'_0(p, p_o, d) = 1 \tag{49}$$

$$B'_1(p, p_o, d) = \sqrt{\frac{1 + (p_0 d)^2}{1 + (pd)^2}} \tag{50}$$

with the commonly used value $d = 1.6\,\mathrm{GeV}^{-1}$ [88]. The other parameters are the mass of the resonance, m_{R}, the momentum of the kaon in the K^{*0} mass frame, k, the momentum of the K^{*0} in the B mass frame, p, and these values at the resonance peak k_0 and p_0. The parameter $\Gamma'_R(m_{K\pi})$ is the $K\pi$ invariant mass dependent decay width of a resonance with the decay width Γ_{R}, the so called running width:

$$\Gamma'_{\mathrm{R}}(m_{K\pi}) = \Gamma_{\mathrm{R}} B'^2_{L_{K\pi}} \left(\frac{k}{k_0}\right)^{2L_{K\pi}+1} \frac{m_{\mathrm{R}}}{m_{K\pi}} \tag{51}$$

The normalization N is defined such that the integral of $|BW_R(m^2_{K\pi})|^2$ is unity. The relative angular orbital momentum of the K^{*0} and the dimuon system, L_B, and of the kaon and pion, $L_{K\pi}$ are considered as follows. The relative angular orbital momentum, L, of mesons decaying into two pseudoscalar mesons is equal the spin of the resonance. Thus for the S-wave $L_{K\pi} = 0$ and as a consequence (the muons originate from a spin 1 boson) $L_B = 1$. In the P-wave configuration $L_{K\pi} = 1$ and $0 \leq L_B \leq 2$. The dominating decay is expected to be the one with the smallest energy, thus in a good approximation $L_B = 0$.

Systematic uncertainties related to this model are evaluated in Sec. 11.7.1 and Sec. 11.7.2.

Background Component

The background is modelled with empirical functions which describe the observed shapes in the control channel $B^0 \to J/\psi\, K^{*0}$ sufficiently well with as few parameters as possible. Due to the veto selection requirements (see Sec. 7.3) physical background can be neglected in the description. Only combinatorial background must be modelled for which all five dimensions can be treated independently. The PDF of the background is split up into three parts:

$$PDF_{\text{bkg}} = PDF_{\text{bkg\_mass}} \cdot PDF_{\text{bkg\_K}\pi} \cdot PDF_{\text{bkg\_ang}} \tag{52}$$

The functions used for $PDF_{\text{bkg\_mass}}$, which describes the shape of the $K\pi\mu\mu$ invariant mass, is a single exponential function with the parameter α_{m}. The $K\pi$ invariant mass shape, described by $PDF_{\text{bkg\_K}\pi}$, is modelled with a linear function with the slope $s_{K\pi}$. As the background shape also factorizes in the three decay angles, the angular distribution of one decay angle is modelled by a second order Chebyshev polynomial:

$$c(x) = 1 + c_1\, x + c_2\, \left(2x^2 - 1\right) \tag{53}$$

The probability density function used for the description of the angular background shape, $PDF_{\text{bkg\_ang}}$, is the product of the three different Chebyshev polynomials. In total there are eight parameters necessary for the background description. As the shape of the background depends on q^2 these parameters are left floating separately in the fit for each q^2 bin.

The systematic uncertainty related to the choice of the background models is studied in Sec. 11.

Physical Boundaries

There are different constraints on the observables, given by the relation to the amplitudes (see Sec. 4.3.4). To reduce the complexity of the fit only two physical constrains are implemented, namely that the S-wave fraction, F_S, and also the total PDF must be positive.

The later requirement is crucial for the maximum likelihood fit to work. Different options how to implement this constraint in the fit are tested. The best fit stability is achieved by testing for each event if the calculated total PDF (see Eq. 41) is negative and in that case setting it to a penalty value of 10^{-12}.

8.1.2 B^+ veto selection requirement

Due to the B^+ veto requirement (see Sec. 7.3) events with a certain combination of $\cos\theta_K$ and $K\pi\mu\mu$ invariant mass are removed. A diagonal band in these two variables is removed which is shown in a 2D plot of the control channel $B^0 \to J/\psi\, K^{*0}$ (see Fig. 31). It is checked that the effect of this selection requirement is independent of q^2.

This veto requirement does not affect the angular acceptance of the signal as only events right of the signal region are removed. Therefore the effect on the measured

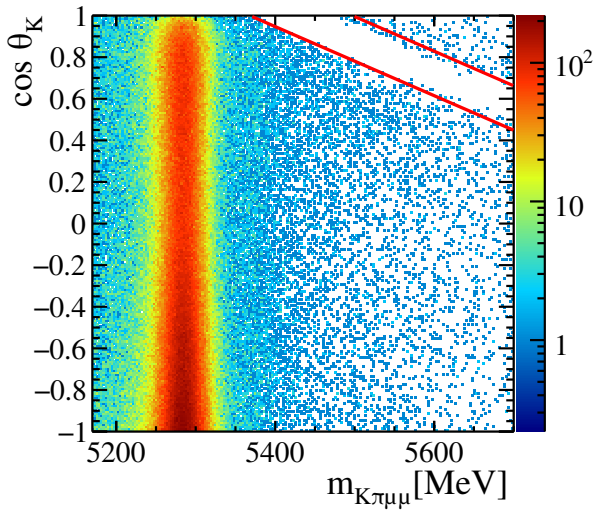

Figure 31: $K\pi\mu\mu$ invariant mass vs. $\cos\theta_K$ in the q^2 range $8 - 11\,\mathrm{GeV}^2$ after the selection and veto requirements. The effect of the B^+ veto requirement is visible as a drop of events in a diagonal band right of the signal region ($5200 - 5350\,\mathrm{MeV}$). The two red lines delimit the area which is removed from the PDF used in the maximum likelihood fit as described in the text.

observables should be small. In principle the reconstruction and selection efficiency of the background is irrelevant, as the background is only described effectively. However, the effect is highly nonlinear and also contradicts the assumption that the likelihood factorizes into mass and angles. Therefore it is explicitly corrected for this effect in the fit.

The procedure is as follows. To have a clean cut events are removed which have an invariant $K\pi\mu\mu$ mass m, with $m_{\min}(\cos\theta_K) \leq m \leq m_{\max}(\cos\theta_K)$, where

$$m_{\min}(\cos\theta_K) = 5375\,\text{MeV} + (1-\cos\theta_K)/0.55 \cdot 325\,\text{MeV} \tag{54}$$

$$m_{\max}(\cos\theta_K) = m_{\min}(\cos\theta_K) + 125\,\text{MeV}. \tag{55}$$

Afterwards in the fit it is accounted for these missing background events by modifying the PDF. Studying the effect of the B^+ veto requirement in simulation it is found that it is sufficient to modify the normalization of the background Chebyshev term describing $\cos\theta_K$ (see Sec. 8.1.1). The normalization of the function is calculated as

$$n_{\text{bkg},\cos\theta_K} = \int_{-1}^{1} \text{dcos}\theta_K \, PDF_{\cos\theta_K,\text{bkg}}(\cos\theta_K) \left(1 - \int_{m_{\min}(\cos\theta_K)}^{m_{\max}(\cos\theta_K)} \text{d}m PDF_{\text{bkg},mass}(m)\right), \tag{56}$$

where the PDF and the normalization of the PDF correspond to the one explained in Sec. 8.1.1.

The impact of the B^+ veto selection requirement is tested on the high statistics simulation (see Sec. 6.2). A simulation without this veto requirement is produced. Events are removed according to the veto and the results of the observables are compared to the nominal results. The only significant bias due to the B^+ veto requirement is a shift of the observable S_{1s} in the order of 0.004 compared to an expected statistical uncertainty of about $0.03-0.06$. This bias is completely removed by modifying the normalization of the PDF as explained above.

8.2 Method of Moments

The Method of Moments (MoM) [89] is the oldest technique for parameter estimation as already discussed by Karl Pearson in 1894 [90]. This technique is both simple and also robust.

The moments of a distribution can be related analytically to the parameters describing the distribution. For example the mean of a distribution describing the random variables with the probability density function $f(x)$ can be calculated as

$$< x >= \int x\, f(x)\, dx. \tag{57}$$

On the other hand the mean of a set of random variables can be also measured with

$$< x >= 1/N \sum x_i. \tag{58}$$

Instead of the mean also different arbitrary moments can be used. If there is the same number of independent moments as number of unknown parameters, the parameters can be calculated by solving the set of equations. If there are more conditions than parameters the system is over-constrained, which can be used to improve the sensitivity - the so called Generalized Method of Moments.

A special case occurs, if all moments are independent from each other. In this case it is not necessary to solve linear equations but each moment is related to one parameter. As all angular terms describing the P-wave contribution of the differential decay rate are orthogonal, it is convenient to use the MoM for the angular measurement. The advantage of the MoM compared to the MLE is that it is in general less biased for small data sets. However, the MoM is asymptotically less efficient than the MLE, meaning that the variance of its point estimates compared to the true value is larger. The statistical uncertainty of the measurement using the MoM can be expected to be worse than using the MLE. There exists nevertheless no strong rule how many events are necessary to be in this asymptotic regime, for which all these theorems are valid. The performance of the MLE and the MoM estimator are compared in Sec. 8.6.

8.2.1 Implementation

All P-wave observables (but $S_{1s,1c,2s,2c}$ which are again assumed to be related to each other by Eq. 36) are orthogonal to each other. Thus, in principle each observable can be calculated via one moment. However, this is no longer true if one also assumes a significant contribution from the S-wave decay. All observables scale with F_S and it is mandatory to get a precise measurement of this variable. In principle F_S could be

also calculated via the moment $\langle f_{F_S} \rangle$. However, it is more precise if F_S is determined by a 2D maximum likelihood fit of the $K\pi\mu\mu$ and $K\pi$ invariant mass (see Sec. 8.5).

The relations of the moments to the angular observables are:

$$\begin{aligned}
\langle f_1 \rangle &= 2/15(6 - (1 - F_S)(3 - 4S_{1s}) - F_S), \\
\langle f_3 \rangle &= (1 - F_S)8/25S_3, \\
\langle f_4 \rangle &= (1 - F_S)8/25S_4, \\
\langle f_5 \rangle &= (1 - F_S)2/5S_5, \\
\langle f_6 \rangle &= (1 - F_S)2/5S_{6s}, \\
\langle f_7 \rangle &= (1 - F_S)2/5S_7, \\
\langle f_8 \rangle &= (1 - F_S)8/25S_8, \\
\langle f_9 \rangle &= (1 - F_S)8/25S_9, \\
\langle f_{F_S} \rangle &= 4/5(1 - 1/3(1 - F_S)J_{1s}).
\end{aligned} \tag{59}$$

The relations are obtained by calculating:

$$\langle f_i \rangle = \int \mathrm{dcos}\theta_l \, \mathrm{dcos}\theta_K \, \mathrm{d}\phi \; f_i(\cos\theta_l, \cos\theta_K, \phi) \frac{1}{\Gamma} \frac{\mathrm{d}\Gamma}{\mathrm{dcos}\theta_l \, \mathrm{dcos}\theta_K \, \mathrm{d}\phi}, \tag{60}$$

with the angular terms defined as follows[18]:

$$\begin{aligned}
f_1(\cos\theta_l, \cos\theta_K, \phi) &= \cos^2\theta_K, \\
f_3(\cos\theta_l, \cos\theta_K, \phi) &= (1 - \cos^2\theta_K)(1 - \cos^2\theta_l)\cos 2\phi, \\
f_4(\cos\theta_l, \cos\theta_K, \phi) &= 2\cos\theta_K\sqrt{1 - \cos^2\theta_K}2\cos\theta_l\sqrt{1 - \cos^2\theta_l}\cos\phi, \\
f_5(\cos\theta_l, \cos\theta_K, \phi) &= 2\cos\theta_K\sqrt{1 - \cos^2\theta_K}\sqrt{1 - \cos^2\theta_l}\cos\phi, \\
f_6(\cos\theta_l, \cos\theta_K, \phi) &= (1 - \cos^2\theta_K)\cos\theta_l, \\
f_7(\cos\theta_l, \cos\theta_K, \phi) &= 2\cos\theta_K\sqrt{1 - \cos^2\theta_K}\sqrt{1 - cos^2\theta_l}\sin\phi, \\
f_8(\cos\theta_l, \cos\theta_K, \phi) &= 2\cos\theta_K\sqrt{1 - \cos^2\theta_K}2\cos\theta_l\sqrt{1 - \cos^2\theta_l}\sin\phi, \\
f_9(\cos\theta_l, \cos\theta_K, \phi) &= (1 - \cos^2\theta_K)(1 - \cos^2\theta_l)\sin 2\phi, \\
f_{F_S}(\cos\theta_l, \cos\theta_K, \phi) &= 1 - \cos^2\theta_l
\end{aligned} \tag{61}$$

The effect of using in addition further moments is tested in simulation but no significant gain in precision is found.

The correction of the non-uniform reconstruction and selection efficiency is implemented similarly compared to the MLE as explained in Sec. 8.3, weighting events

[18] Compare to Eq. 37 in Sec. 4.3.6. In addition the trigonometric relations are used to express the formulas by $\cos\theta_l$ and $\cos\theta_K$

with the inverse of the expected reconstruction and selection efficiency, $w_{\text{ev}} = 1/\epsilon_{\text{ev}}$. The moments explained in Eq. 59 are determined by the measurement of

$$\langle f_i \rangle = \frac{1}{\sum_{\text{ev}=1}^{N} w_{\text{ev}}} \sum_{\text{ev}=1}^{N} w_{\text{ev}} f_i(\cos\theta_{K,\text{ev}}, \cos\theta_{l,\text{ev}}, \phi_{\text{ev}}) \tag{62}$$

where $f_i(\cos\theta_{K,\text{ev}}, \cos\theta_{l,\text{ev}}, \phi_{\text{ev}})$ is the angular function evaluated for the angles measured for one event. The unfolding of the signal and background distributions is done with the so called sPlot technique, which is explained in Sec. 8.2.2. The weights w_{ev} are multiplied by the so called sWeights which are a measure for the probability of an event to be the signal decay.

Due to mixing the MLE to measure F_S and the MoM to measure the P-wave observables it is not possible to do a straight forward error calculation to obtain the uncertainties of the observables. Instead also the Bootstrap method is used for the determination of the confidence intervals, which is explained in Sec. 8.4.

8.2.2 sWeights and the sPlot technique

There are different techniques available how to unfold the signal decay from combinatorial background. Frequently the so called sideband subtraction is used. In a separating variable, m, usually an invariant mass, a signal region is defined given by the mass distribution of the signal resonance. The variable m is chosen such that it is in a good approximation uncorrelated to the variable of interest, x, which is *i.e.* a decay angle. The combinatorial background is usually flat or has an exponential shape. In m a sideband region is defined, in which the contamination from the signal is negligibly small and the distribution of x can be determined for the background. The distribution of x of the background can now be extrapolated in the signal region. A subtraction of the distribution measured in the sideband (scaled to the right amount) from the distribution measured in the signal region, results in the distribution of x of the signal only.

A novel idea of this ansatz is the so called sPlot [91, 92] technique. The observed distribution of events is written as

$$(N_s + N_b)\, f(x,m) = N_s\, s(x,m) + N_b\, b(x,m) \tag{63}$$

with N_s (N_b) signal (background) events with the PDF $s(x,m)$ ($b(x,m)$) and the PDF of the combined distribution $f(x,m)$. The desired distribution is $s(x)$, which can be obtained by weighting the observed events with weights, $w(m)$, such that

$$N_s\, s(x) = (N_s + N_b) \int \mathrm{d}m\, f(x,m)\, w(m) \tag{64}$$

The function $w(m)$ can be in principle an arbitrary function. In the case of the sideband subtraction $w(m)$ is chosen to be $+1$ in the signal region and -1 in the background region[19].

Assuming that m and x are uncorrelated

$$s(x,m) = s(x)\, s(m) \tag{65}$$
$$b(x,m) = b(x)\, b(m) \tag{66}$$

the function $w(m)$ has to be chosen such that

$$\int s(m)\, w(m)\, \mathrm{d}m = 1 \tag{67}$$
$$\int b(m)\, w(m)\, \mathrm{d}m = 0 \tag{68}$$

The smallest statistical uncertainty is achieved by minimizing the variation of the weights, which can be expressed as:

$$\int \mathrm{d}x\, \mathrm{d}m\, f(x,m)\, w(m)^2 \tag{69}$$

These conditions are enough to uniquely determine the function $w(m)$. Single weights for events with the property m are calculated according to $w(m)$. The weighted events effectively describe now the signal-only distribution of x.

[19] Assuming that the background region is chosen such that there is the same amount of background in the signal and the background region. Otherwise the weights are scaled accordingly.

8.3 Angular dependence of the reconstruction and selection efficiency

The determination of the detector geometry, trigger, reconstruction and selection efficiency as a function of the relevant decay angles, called angular acceptance, is one of the key points of this analysis. It is crucial to understand the effects of the angular acceptance on the angular distribution of the decay $B^0 \to K^{*0}\mu^+\mu^-$ to separate them from effects of New Physics.

In the track reconstruction process there are several reasons for an non-uniform efficiency. The geometrical LHCb detector acceptance is limited to a pseudorapidity range of about $2-5$. Even if a particle is inside of the detector acceptance the track reconstruction algorithms perform differently well for different pseudo-rapidity and momentum ranges. In the inner region of the detector the detector occupancy is significantly larger than in the outer region, making it more difficult to find correct tracks. Particles with a low momentum are more affected by multiple scattering. Also the trigger and offline selection efficiencies depend on the position in phase space of the particles. Especially important in the selection to efficiently suppress background are requirements on the transverse momentum of final state particles. The angle θ_K describing the angle between the flight direction of the kaon and the K^{*0} (in the K^{*0} rest frame) is related to the transverse momentum of the kaon. Similarly also the transverse momentum of the muons is related to the angle θ_l. As a consequence the selection efficiency is especially low if $\cos\theta_K$ and $\cos\theta_l$ are close to one.

There are two aspects of the angular measurement simplifying the hole procedure. As the ratio of the differential branching ratio to the total branching ratio will be measured, only the relative angular acceptance must be determined. Second, due to the small event yield the angular measurement will be statistically limited allowing also larger systematic uncertainties. Therefore it is sufficient to use a correction method based on simulation making the basic principle of the determination of the angular acceptance rather simple. In each region in phase space particles are simulated with a Monte Carlo technique (full detector simulation, see Sec. 6.1). It is checked how many of these particles are also reconstructed and selected. The ratio of the number of reconstructed and selected particles and the number of generated particles corresponds to the reconstruction and selection efficiency.

The angular acceptance is corrected with a weighting technique. In the parameter estimation each candidate is weighted with the inverse of its estimated reconstruction and selection efficiency:

$$w = \frac{1}{\varepsilon}. \tag{70}$$

In principle the most powerful and also simplest method to obtain these weights would

be a binned analysis with very fine binning. However, as the angular acceptance depends on the three decay angles and also on q^2, the study must be done in at least 4 dimensions. As the angular acceptance does not factorize completely in these dimensions a 4D histogram must be used. The amount of simulated events needed to have in each bin a reasonable number of entries is out of reach for this study.

Therefore, the determination of the angular acceptance and thus also of the correction weights is separated into two steps. In a first step the acceptance is corrected by 1D projections separately, neglecting completely any correlation between the three decay angles. Afterwards a so called Kernel Density Estimation (KDE) method [93] will be used to also include the effects of correlation. Although the KDE method is excellent in describing multidimensional distributions, it is not so powerful in describing the steep drop of efficiency at large/small $\cos\theta_{K,l}$. This two-step approach is used to take advantage of both methods.

Furthermore, the acceptance is determined in bins of q^2 (with a size of about 2 GeV, see Tab. 1 in Sec. 4.3.3), neglecting the small q^2 dependence of the acceptance within one bin. The bias related to the angular acceptance correction will be measured in simulation (found to be small) and it is corrected for (see Sec. 9.1.2).

For the calculation of the weights the $B^0 \to K^{*0}\mu^+\mu^-$ PHSP simulation of the full detector is used (see Sec. 6.1). Particles are generated with a flat distribution in the three decay angles and a q^2 distribution similar to the one observed in data. Therefore, in the Monte Carlo sample, the angular distribution of the reconstructed and selected particles in each q^2 bin is directly proportional to the angular acceptance.

Pre-Weight

At the first stage so called pre-weights are determined and applied to the events. These pre-weights are used to correct for the steep drop of efficiency at the boundaries of the distribution. Remaining differences will be corrected for in the second step. Thus, it is enough to roughly correct for the efficiency loss in this step and ignore any correlation between the variables. The pre-weights are determined in all three angles separately for each q^2 bin. The weight corresponds to the inverse of the measured normalized reconstruction and selection efficiency. The total pre-weight is

$$w_{\text{pre}}\left(q^2, \cos\theta_l, \cos\theta_K, \phi\right) = w_{\cos\theta_l}\left(q^2, \cos\theta_l\right)\, w_{\cos\theta_K}\left(q^2, \cos\theta_K\right)\, w_{\phi}\left(q^2, \phi\right) \quad (71)$$

The 1D angular distributions are fitted (binned χ^2 fit) with a polynomial function of the order, which is sufficient to describe the shape. For the variables $\cos\theta_l$, $\cos\theta_K$ and ϕ a polynomial function of respective order 6, 9 and 4 is chosen. Using this fit statistical fluctuations in the measured acceptance are smoothened.

For testing the pre-weights the total sample is split up into two parts. The larger amount of the events (80%) are used to calculate the pre-weights. The validation

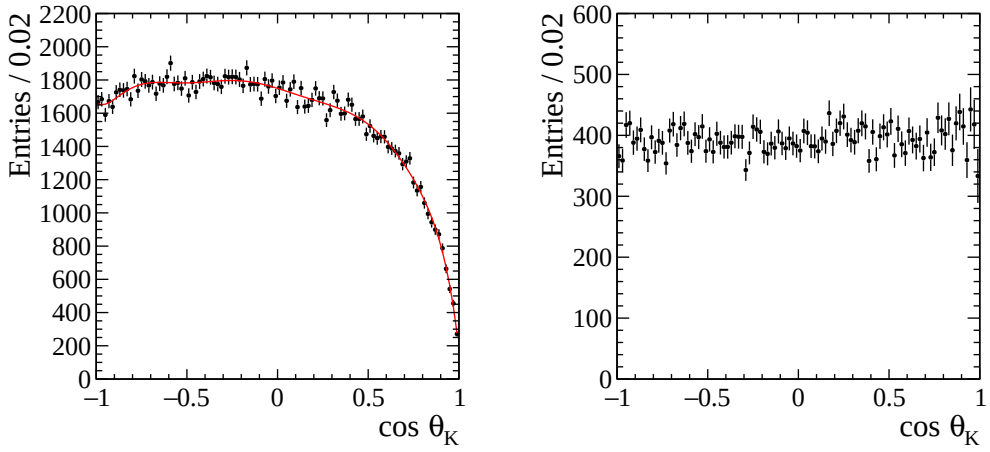

Figure 32: Polynomial χ^2 fit to the 1D projection of the $\cos\theta_K$ distribution in the q^2 bin $4.0 - 6.0\,\mathrm{GeV}^2$ of the PHSP full detector simulation to determine the pre-weights (80% of the sample, left). The distribution after applying the pre-weights (the other 20% of the sample, right).

Table 4: Test of the pre-weights in the different q^2 bins. Shown are the p values of χ^2 fits of a flat line to the corrected angular distributions.

$q^2[\,\mathrm{GeV}^2]$	$\cos\theta_K$	$\cos\theta_l$	ϕ
$0.1 - 0.98$	0.01	0.06	0.34
$1.1 - 2.5$	0.83	0.33	0.10
$2.5 - 4.0$	0.73	0.19	0.38
$4.0 - 6.0$	0.02	0.56	0.91
$6.0 - 8.0$	0.27	0.34	0.95
$8.0 - 11.0$	0.57	0.29	0.91
$11.0 - 12.5$	0.16	0.98	0.79
$12.5 - 15.0$	0.65	0.94	0.57
$15.0 - 17.0$	0.11	0.18	0.84
$17.0 - 19.0$	0.56	0.04	0.71

of the pre-weights is done on the remaining data set (20%). As in the PHSP full detector simulation the spin structure of the particles is neglected the initial angular distributions (in ϕ, $\cos\theta_K$, $\cos\theta_l$) are flat and thus also the corrected distributions are supposed to be flat. An example of this pre-weight test is shown for one of the decay angles and for one q^2 bin in Fig. 32. The results of all χ^2 fits to all corrected angular distribution in all q^2 bins are listed in Tab. 4. In all cases the agreement is

good enough that a further processing with the KDE method is possible. For the application in the real measurement the full simulation sample is used to calculate the pre-weights.

Final Weight

The final step in the calculation of the weights is done with the KDE method. A correction to the already calculated pre-weights is determined. The final weight is the pre-weight multiplied by the weight from the KDE method:

$$w_{\text{final}}\left(q^2, \cos\theta_l, \cos\theta_K, \phi\right) = w_{\text{KDE}}\left(q^2, \cos\theta_l, \cos\theta_K, \phi\right) \cdot w_{\text{pre}}\left(q^2, \cos\theta_l, \cos\theta_K, \phi\right) \tag{72}$$

The weights are normalized to represent the statistics available in the data set.

The KDE method works similar to a binned analysis, however, providing a much smoother distribution when only a small number of events is available. Instead of a discrete point in phase space, each event is replaced by a three dimensional Gaussian function. The response of the KDE method at a given point in phase space is the sum of the values of all Gaussian functions at that point. It is important that in a multidimensional environment the scales of all variables are comparable. In practice this often means that variables have to be transformed to an appropriate scale before the algorithm can be applied. The scale of $\cos\theta_{K,l}$ are the same and the angle ϕ is scaled by $1/\pi$ to also match this scale. As the acceptance is expected to be symmetric in ϕ, the acceptance is evaluated with respect to the absolute value of ϕ.

In a binned analysis it is important to choose an appropriate bin size. Similarly using the KDE method the σ_{KDE} of the Gaussian function must be tuned to achieve best results. As will be shown in the later part of this section a good value of σ_{KDE} is

$$\sigma^2_{\text{KDE}} = \sigma^2_{\phi} + \sigma^2_{\theta_l} + \sigma^2_{\theta_K} = 0.3^2 \tag{73}$$

Due to folding each point with a Gaussian the KDE method will smooth out all sharp structures, but also the edges near physical boundaries. If not correcting for this effect the response of the KDE would be significantly too small near these boundaries. Therefore the distributions are mirrored at the boundaries. As all steep structures have been removed before using the pre-weights this procedure works well.

The calculation of the multitude of Gaussian functions is very time consuming. A fast calculation of the KDE method is achieved with the program FIGTree [94], which uses the Improved Fast Gauss Transform [95] and the Approximate Nearest Neighbour search [96].

Choosing the optimal value of the parameter σ_{KDE} there are two effects which must be considered, similar to the effects present when working with binned histograms. It can be expected that by using a too large σ_{KDE} fine structures in the acceptance

Table 5: Fit to 500 simulations with each 200 signal events of the decay $B^0 \to J/\psi\, K^{*0}$. Shown is for the observable S_4 the mean of all fit results compared to the generated value, and the spread (RMS) of this distribution which corresponds to the expected statistical uncertainty.

σ	systematic shift	statistical uncertainty
0.05	-0.0063	0.0820
0.1	-0.0073	0.0817
0.2	-0.0100	0.0814
0.3	-0.0133	0.0811
0.9	-0.0278	0.0811

cannot be reproduced. On the other hand if σ_{KDE} is chosen too small for the available statistics, there will be large fluctuations present in the acceptance description. These fluctuations lead to large weight differences decreasing the statistical precision of the measurements. The effect of the choice of σ_{KDE} is tested on simulation. The full detector physics simulation of $B^0 \to J/\psi\, K^{*0}$ (see Sec. 6.1) is split up into subsets of 200 events on which the observables are measured with a maximum likelihood fit (see Sec. 8.1).

The largest effect is visible for the observable S_4. The expected bias and the expected statistical uncertainty are shown for this observable for different values of σ_{KDE} in Tab. 5. If σ_{KDE} is chosen to be smaller than 0.3 the effect of fluctuations in the acceptance description gets significant and the expected statistical precision gets worse. On the other hand due to the smeared-out description the obtained fit result gets biased with larger values of σ_{KDE}; an effect that is already starting to get visible for the smallest tested values of σ_{KDE}.

In Fig. 33 several 2D projections of the efficiency in $\cos\theta_l$ vs. $\cos\theta_K$ are shown for three different values of σ_{KDE}. The expected large statistical fluctuations can be seen for the case $\sigma_{\mathrm{KDE}} = 0.1$, whereas there is a smooth description of the efficiency for $\sigma_{\mathrm{KDE}} = 0.3$. Therefore one has to make a trade-off between the two effects. As the angular measurement is expected to be statistically limited the chosen value is $\sigma_{\mathrm{KDE}} = 0.3$ to obtain the best possible statistical uncertainty.

This method will be cross checked in simulation as well as on a control channel in data (see Sec. 9) and an adequate systematic uncertainty will be assigned (see Sec. 11.1).

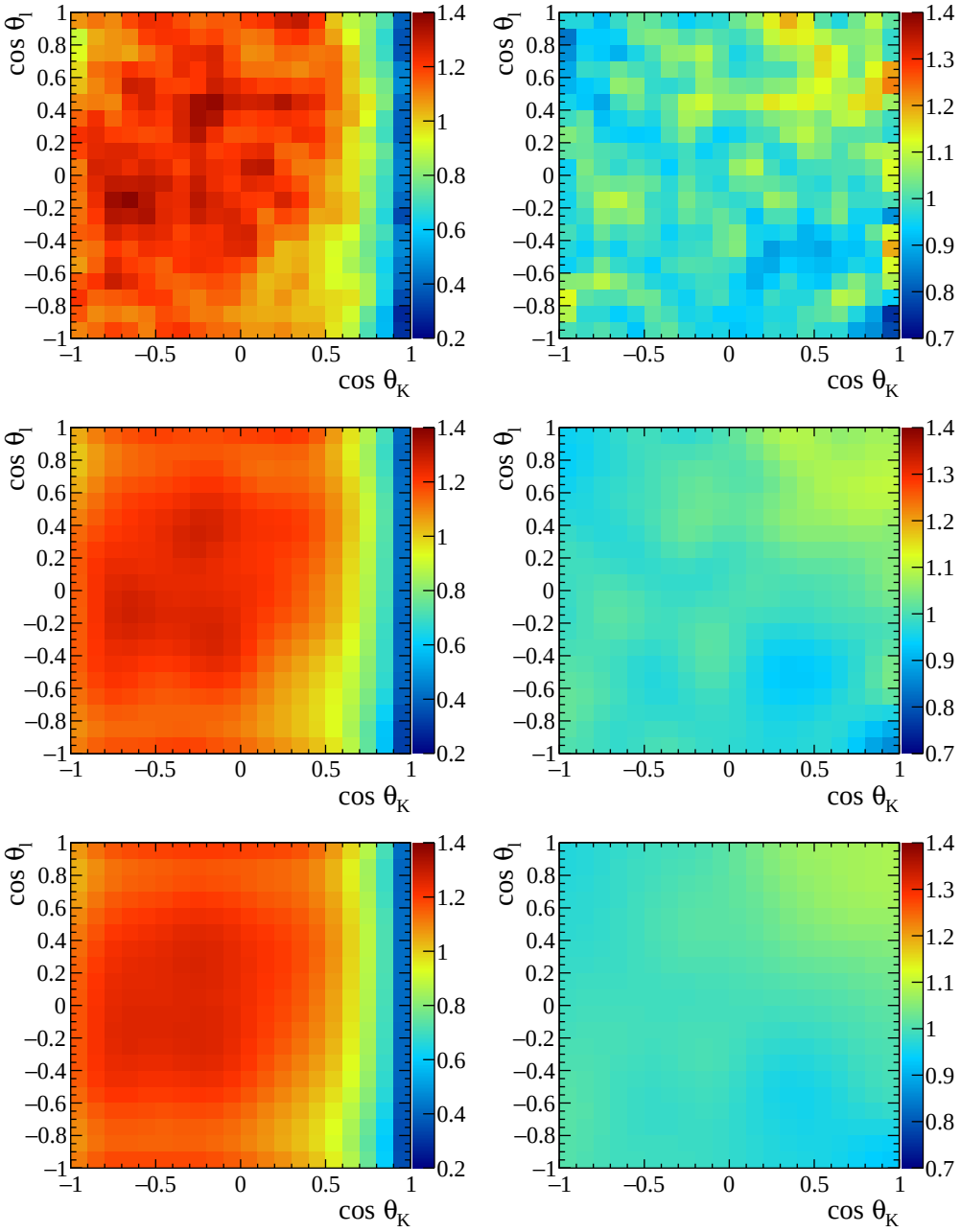

Figure 33: The normalized reconstruction and selection efficiency ($1/w_{\text{final}}$) in $\cos\theta_l$ vs. $\cos\theta_l$ for $\phi = 3$ in the q^2 bin $4 - 6\,\text{GeV}^2$ (left). Only the correction of the KDE method, $1/w_{\text{KDE}}$ (right). The plots are shown for $\sigma = 0.1$, 0.2 and 0.3 (from top to bottom).

8.4 Bootstrapping and BCa method

A powerful and also simple method to determine the confidence intervals of the measured physical parameters is the Bootstrap method and more specific the BCa (Bias Corrected and accelerated) method [97].

The idea of the Bootstrap method is that a random sample following the distribution describing the data can be drawn from the data itself. Out of the N data events one selects randomly N events with replacement. Thus some events are taken twice or even more often. In this way one obtains a random sample of the distribution, referred to in the following as bootstrap sample. For each bootstrap sample the observables are measured. The spread of the obtained results directly relates to the statistical uncertainty of an observable in the original measurement.

A 68% confidence level can be obtained by choosing an interval in which 68% of the bootstrap results are included. The naive ansatz would be to take the central and symmetric confidence interval. This choice can be improved by two adjustments implemented in the BCa method. The lower end point of the $1\,\sigma$ confidence interval is defined in the BCa method as:

$$\hat{\theta}_{\mathrm{BCa}} = \hat{G}^{-1}\left(\Phi\left(z_0 + \frac{z_0 - 1}{1 - a(z_0 + 1)}\right)\right), \tag{74}$$

where the different parameters and functions are explained in the following. The inverse cumulative distribution function $\hat{G}^{-1}(x)$, also called quantile function, returns the value of the bootstrap result, for which x% of all bootstrap results are smaller. The standard normal cumulative distribution function $\Phi(x)$ returns the integral of the standard normal distribution from $-\infty$ to x. The parameter a is the acceleration of the distribution of bootstrap results, it is related to the skewness of this distribution and is defined in Ref. [97]. If the bootstrap results are more often smaller than larger, or vice versa, compared to the original result $\hat{\theta}$ (which is in this context also called biased) this is expressed by z_0:

$$z_0 = \Phi^{-1}\left(\frac{\#(\hat{\theta}_i < \hat{\theta})}{B}\right), \tag{75}$$

with B being the number of bootstrap samples and $\Phi^{-1}(x)$ the inverse standard normal cumulative distribution function. The upper end point of the confidence interval is defined such that between the lower end point and higher end point there are 68% of all bootstrap results.

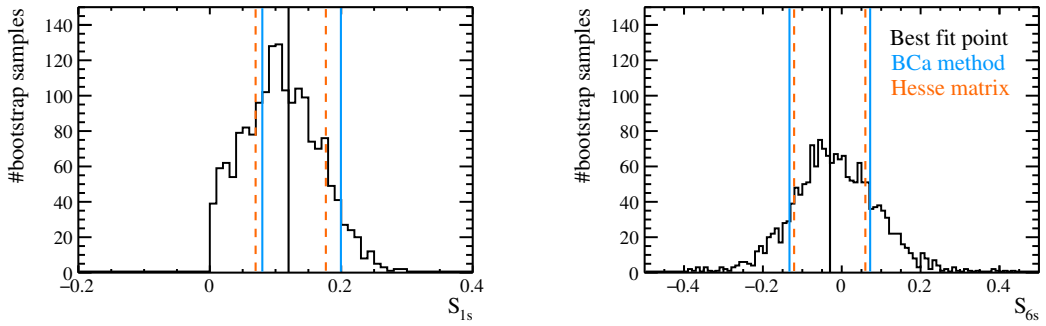

Figure 34: Measured S_{1s} (left) and S_{6s} (right) in the q^2 range from 2.5 to 4 GeV^2 for 2000 bootstrap samples of one pseudo experiment using the maximum likelihood fit. Whereas the $1\,\sigma$ confidence interval from the `Hesse` matrix (orange dashed lines) is symmetric around the best fit point (black solid line), the interval from the BCa method (blue solid lines) is asymmetric. In this case the parameters for the BCa method are for S_{1s}: $z_0 = 0.253$, $a = 0.020$ and for S_{6s}: $z_0 = -0.068$, $a = -0.015$.

In Fig. 34 the distribution of the bootstrap results and the corresponding confidence levels of the BCa method and the `Hesse` matrix from the fit are compared for two observables of one pseudo experiment. The high statistics simulation is used (see Sec. 6.2) and the event yield and signal fraction simulated are taken from a fit to data (see Tab 2). The confidence interval of the BCa method is asymmetric around the best fit point due to the effect of the physical boundaries (*i.e.* S_{1s} must not get negative). Also near physical boundaries the BCa method provides reliable estimates of the 68% uncertainty interval as is shown in Sec. 9.2. The statistical uncertainty of the estimators, which is provided for the final result, is the distance from the best fit point to the lower and higher end point of the BCa confidence interval.

As this method is rather time consuming only a limited number of bootstrap samples can be analysed. The recommended value of bootstrap repetitions is at least 2000 [97]. In this analysis 5000 bootstrap samples are evaluated. The change in the confidence interval by further increasing the number of bootstrap samples is found to be negligible.

8.5 Measurement of the S-wave fraction

In the previous measurements of the decay $B^0 \to K^{*0}\mu^+\mu^-$ the S-wave component was always only treated as systematic effect. However, in this analysis the event yield has become so large, that this is no longer a valid approach. Instead the S-wave parameters must be included in the PDF describing the differential decay rate (see Eq. 37 in Sec. 4.3.6). As can be seen in the PDF the P-wave parameters, S_i, scale with F_S. Thus a precise measurement of them is not possible without a powerful constrain of F_S. The influence of the S/P-wave interference terms, S_{Si}, on the P-wave observables is small.

The six additional parameters are unknown and there neither exists a precise theoretical prediction of the values nor measurements. Furthermore, it cannot be expected that these parameters are constant in q^2. The size of F_S in the decay $B^0 \to K^{*0}\mu^+\mu^-$ can be expected to be of similar order as in the decay $B^0 \to J/\psi\, K^{*0}$, which is about 5% (see Sec. 9.3).

There are several possibilities to measure F_S. It is one component of the differential decay rate and thus can be measured by a purely angular analysis. Furthermore, the P and S-wave have a different distribution in the $K\pi$ invariant mass. Whereas the P-wave component can be described by a Breit Wigner function the S-wave component is nearly flat in the observed mass range (see Sec. 8.1.1). The options which are studied are the 5D maximum likelihood fit (see. Sec.8.1). The same fit, however, only fitting the angular distribution and not the $K\pi$ invariant mass (4D fit). The

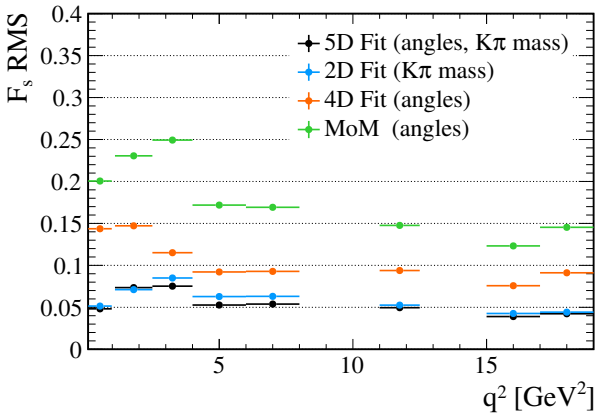

Figure 35: The total spread of the measured F_S in simulation vs. q^2 using various different methods. Shown is the Root Mean Squared (RMS) of the F_S results of multiple measurements done on different subsamples of the high statistics simulation (see Sec. 6.2). The event yield in each subsample corresponds to the one measured in data.

same fit using this time only the $K\pi$ invariant mass and not the angular distribution (2D fit). The last option is using the Method of Moments with the moment $\langle f_{F_S} \rangle$, as explained in Sec. 8.2.1).

The precision of the methods to extract F_S is shown in Fig. 35. The most precise method is the 5D fit using both information of the angular and also $K\pi$ invariant mass distribution. The 2D fit of the $K\pi$ and $K\pi\mu\mu$ invariant mass is only slightly less sensitive. The measurement using only the angular distribution is much worse using the 4D fit, and nearly unusable using the $\langle f_{F_S} \rangle$ moment.

Therefore including the S-wave component in the measurement it is utterly important to also include the $K\pi$ invariant mass information in the measurement. In the maximum likelihood fit this is done by implementing the $K\pi$ invariant mass as an additional dimension (5D fit). In the Method of Moments the value of F_S used for the calculation of the P-wave observables, is determined with the 2D fit.

8.6 Performance comparison

Two methods to estimate the true values of the observables are presented in the first two parts of this section, the maximum likelihood fit and the Method of Moments. The performance of both methods is tested on simulation. For this test the high statistics simulation is used (see Sec. 6.2). Events are generated with the conditions expected in data. A fit to the $K\pi\mu\mu$ invariant mass distribution of the decay $B^0 \to K^{*0}\mu^+\mu^-$ is used to extract the signal fraction and shape of the background in each q^2 bin observed in data (see Tab. 2). About 1500 simulation samples are generated accordingly. The spread of the results of the measured observables and thus the precision of the estimators is compared in Fig. 36. This measurement is done independently of the estimated uncertainties of the estimators and thus is unbiased.

As expected the performance of the maximum likelihood fit is better compared to the Method of Moments. In general the spread of the results is smaller by about 10%. Only in the q^2 bins $1.1 - 4.0\,\mathrm{GeV}^2$ the expected statistical uncertainty of the Method of Moments is smaller for few observables. The better performance of the Method of Moments in this region can be understood, as the number of signal candidates is small and the background level is large. Furthermore, the values lie near the physical boundary. The Method of Moments is expected to work better with very limited statistics.

Due to the in general better performance it is decided to use the maximum likelihood fit as nominal estimator. The Method of Moments is used as cross-check.

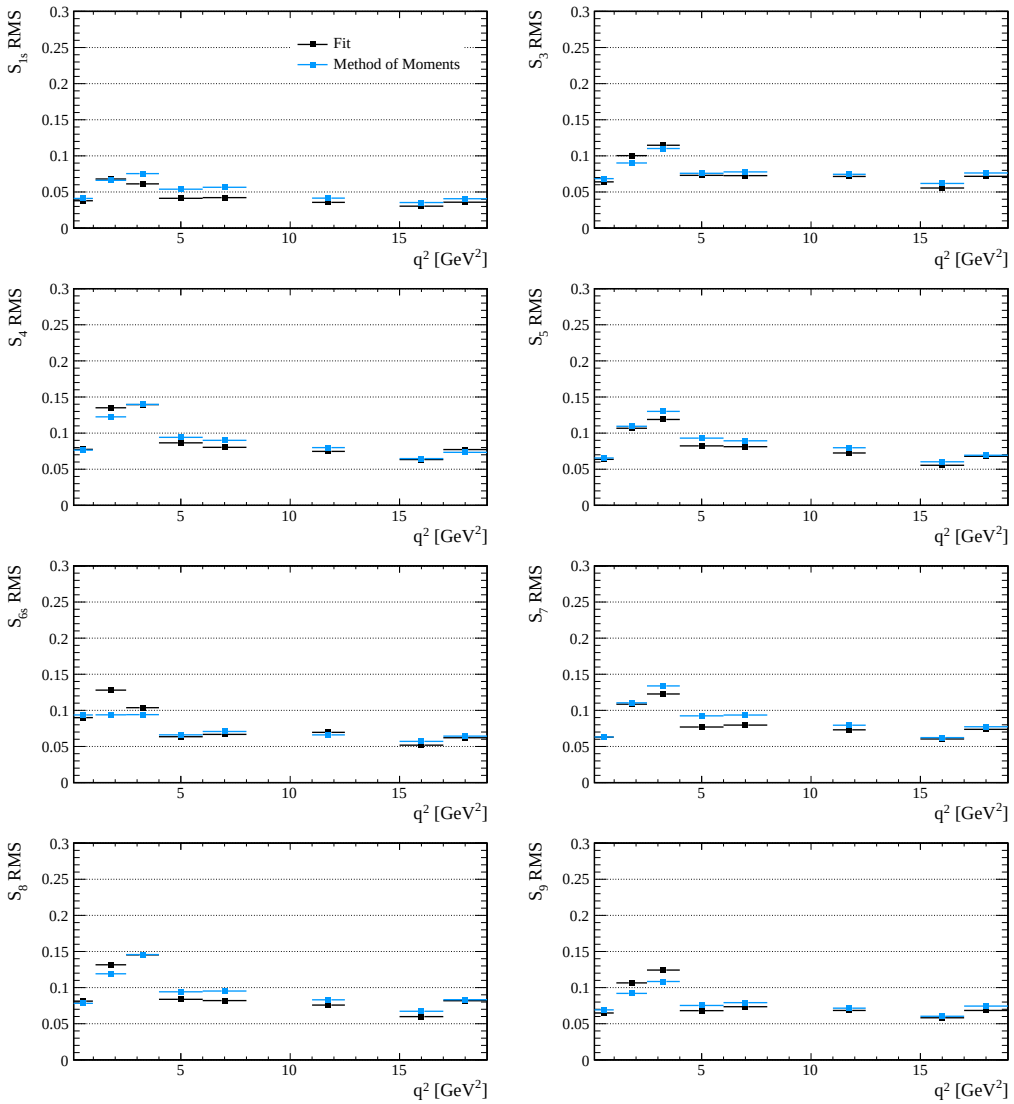

Figure 36: Spread of the results of the estimation of the observables vs. q^2 in simulation corresponding to the expected statistical uncertainty.

9 Validation of the measurement method

The measurement methods used in this analysis are due to the manifold requirements of the measurement rather complex. It is therefore all the more important to carefully validate the methods and to ensure the correctness of the results. Two different kinds of simulation are used to cross-check different aspects of the measurement: The simulation of the full detector (see Sec. 6.1) and the high statistics simulation (see Sec. 6.2).

The generation of the full detector simulation is computing time expensive and thus only the signal decay is simulated. Furthermore, as the theory knowledge about the S-wave contribution is still limited only the P-wave decay is considered. The S-wave contribution will be part of future implementations. The purpose of this simulation is the validation of the angular acceptance correction (see Sec. 8.3) and the implementation of the P-wave signal part of the PDF.

In contrast to the full detector simulation the high statistics simulation includes both signal and background events. However, the detector response is only implemented in a very simplified way, based on the acceptance correction method itself. Therefore it cannot be used to test the angular acceptance correction, but it is perfectly suited for testing the implementation of the full PDF.

A test of both the complete PDF and also a realistic detector response is performed on data using the control channel $B^0 \to J/\psi K^{*0}$. The decay in the control channel is very similar to the signal decay $B^0 \to K^{*0}\mu^+\mu^-$. The decay topology is exactly the same, and it only differs in the exact angular distribution. The result obtained in the control channel is compared to an independent angular analysis of this decay.

The validation of the fit results is done by determining the goodness of fit with the Point-to-Point dissimilarity method, as explained in the last part of this section.

9.1 Validation of the angular acceptance and signal decay description

9.1.1 ... using $B^0 \to J/\psi K^{*0}$ full detector simulation

The correction of the non-uniform reconstruction and selection efficiency is tested with the simulation of the full detector response (see Sec. 6.1).

A three dimensional maximum likelihood fit (implemented as described in Sec. 8.1) to the three decay angles is done using the differential decay rate described in Sec. 4.3.2. As only the signal decay is simulated the $K\pi\mu\mu$ invariant mass description used for separating signal and background is not considered for this test. The correction of the angular acceptance as explained in Sec. 8.3 is applied. The fit used in that section to find the optimal configuration of the correction procedure is identical to the one used in this section.

The three 1D projections of the fit model can be seen in Fig. 37. The projections are compared to the distribution of simulated events, which are weighted according to the angular acceptance correction. The largest correction weights are in the region $\cos\theta_K \approx 1$ (an example of the q^2 bin $6 - 8\,\mathrm{GeV}^2$ is shown in Fig. 32 in Sec. 8.3). Therefore the uncertainties and fluctuations are larger in this region. Within the statistical uncertainties the angular distributions are well described by the maximum likelihood fit including the acceptance correction as shown by the pulls in Fig. 37.

The results of the fit are compared to the theoretical values used in the simulation, the corresponding values are listed in Tab. 6. Besides a small, but due to the high statistics significant, deviation in S_4 the generated values are estimated within the statistical uncertainties. The bias in S_4 is related to the acceptance correction as discussed in Sec. 8.3. It is a small effect compared to the expected statistical uncertainty of the measurement of $B^0 \to K^{*0}\mu^+\mu^-$. Due to the q^2 dependence of the angular acceptance this bias is expected to be also q^2 depended. Therefore it is corrected using the full detector simulation of the $B^0 \to K^{*0}\mu^+\mu^-$ decay as explained in the next subsection.

9.1.2 ... using $B^0 \to K^{*0}\mu^+\mu^-$ full detector simulation

The signal PDF and angular acceptance correction are not only tested on the decay $B^0 \to J/\psi\, K^{*0}$ as detailed in the last subsection, but also on the signal decay $B^0 \to K^{*0}\mu^+\mu^-$. The fit is done in bins of q^2. The configuration of the maximum likelihood fit and the acceptance correction is the same as the one used in the last subsection. The total yield of simulated events in each q^2 bin is about a factor 20 less than the event yield available for the $B^0 \to J/\psi\, K^{*0}$ full detector simulation. Therefore, the fit done in this subsection is less precise. The fitted and generated values of the observables are shown in Fig. 38.

In the last subsection small deviations from generated and measured values are observed because of the acceptance correction. Thus also in the fit done in this subsection smaller deviations are expected. The bias is expected to be like the angular acceptance correction itself q^2 depended. The deviations observed in the different q^2 bins for the decay $B^0 \to K^{*0}\mu^+\mu^-$ are listed in Tab. 7. All deviations are small compared to the expected statistical uncertainties of the measurement. Due to neglecting lepton masses in the PDF (see Sec. 4.3.2) an effect is expected in the first q^2 bin ($0.1 - 0.98\,\mathrm{GeV}^2$). A bias of about 0.02 in S_{1s} is observed which is in agreement with the one measured in the high statistics simulation (see Sec. 9.2).

The results of the final maximum likelihood fit are corrected for these shifts, the related uncertainties are included in the systematic uncertainty (see Sec. 11.1.2). As already known from previous analyses the measured values are expected to be close to the Standard Model predictions. Therefore, the simulation provides realistic estimates of effects related to the angular acceptance correction.

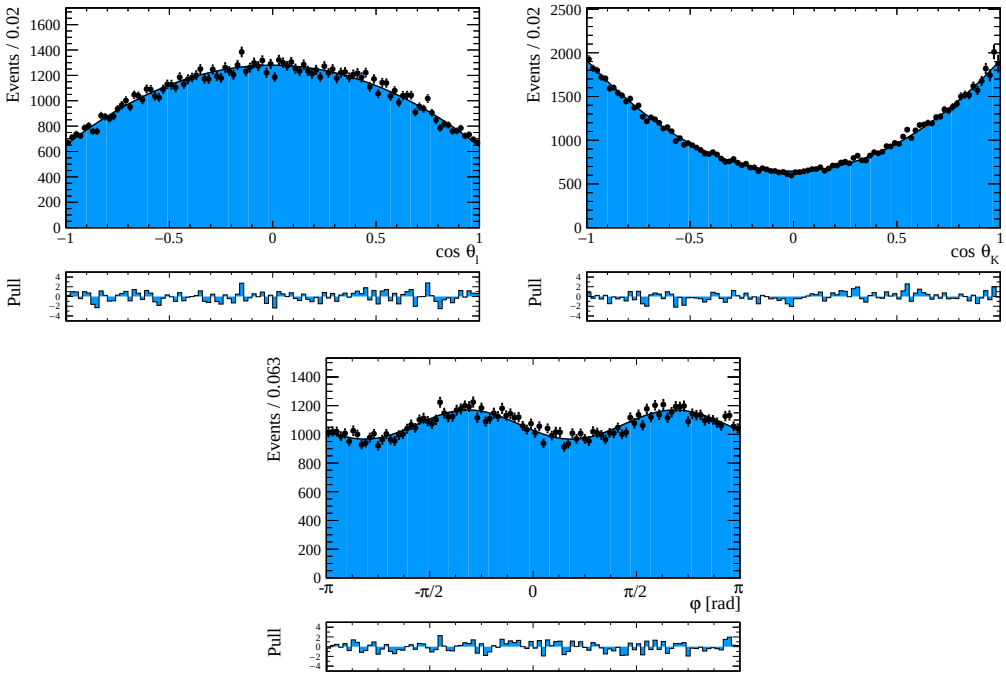

Figure 37: Projections of the 3D fit to the $B^0 \to J/\psi K^{*0}$ simulation. The projections of the fit function (blue) are compared to the weighted data point (black).

Table 6: Results of the fit to the $B^0 \to J/\psi K^{*0}$ simulation compared to the values used for the generation of this simulation. The given uncertainties are statistical only.

Parameter	Fitted	Generated
S_{1s}	0.302 ± 0.002	0.300
S_3	-0.037 ± 0.003	-0.040
S_4	-0.223 ± 0.003	-0.215
S_5	0.001 ± 0.003	0
S_{6s}	-0.003 ± 0.003	0
S_7	-0.004 ± 0.003	0
S_8	0.035 ± 0.003	0.037
S_9	-0.087 ± 0.003	-0.089

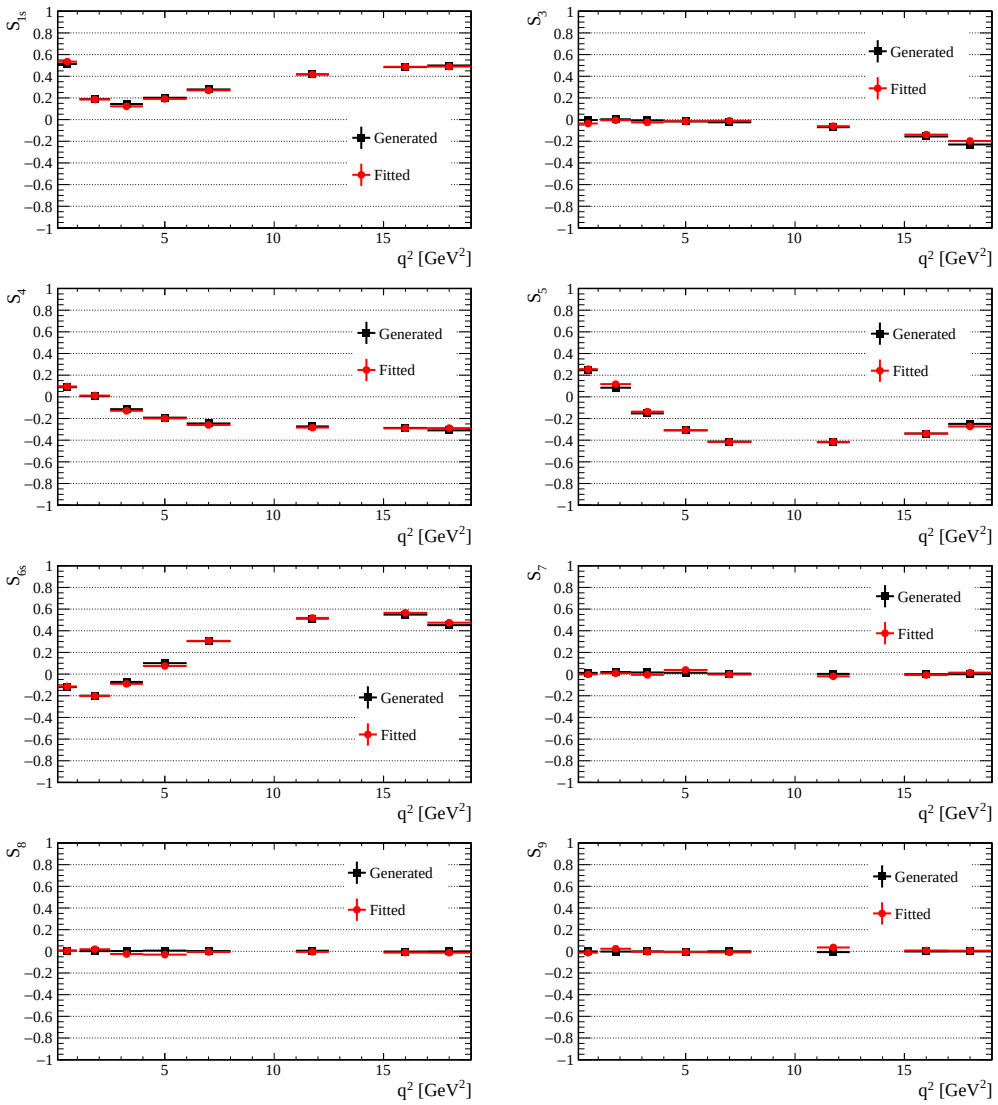

Figure 38: The angular observables of the $B^0 \to K^{*0}\mu^+\mu^-$ simulation measured with the maximum likelihood fit (red) compared to the values used in the generation (black).

Table 7: Differences between the generated and fitted values of the $B^0 \to K^{*0}\mu^+\mu^-$ simulation (see Fig. 38).

$q^2[GeV^2]$	S_{1s}	S_3	S_4	S_5
$0.1 - 1.0$	0.022 ± 0.009	-0.033 ± 0.020	0.007 ± 0.018	0.009 ± 0.017
$1.1 - 2.5$	-0.003 ± 0.010	-0.009 ± 0.016	0.005 ± 0.020	0.033 ± 0.022
$2.5 - 4.0$	-0.022 ± 0.009	-0.019 ± 0.015	-0.015 ± 0.020	0.016 ± 0.021
$4.0 - 6.0$	-0.011 ± 0.008	0.000 ± 0.013	-0.007 ± 0.015	0.002 ± 0.016
$6.0 - 8.0$	-0.009 ± 0.007	0.011 ± 0.012	-0.004 ± 0.010	0.001 ± 0.012
$11.0 - 12.5$	0.000 ± 0.007	0.005 ± 0.012	-0.020 ± 0.010	-0.007 ± 0.011
$15.0 - 17.0$	0.002 ± 0.007	0.017 ± 0.012	-0.004 ± 0.010	-0.001 ± 0.010
$17.0 - 19.0$	-0.012 ± 0.009	0.033 ± 0.016	0.020 ± 0.015	-0.024 ± 0.014

$q^2[GeV^2]$	S_{6s}	S_7	S_8	S_9
$0.1 - 1.0$	0.006 ± 0.019	-0.014 ± 0.018	-0.003 ± 0.019	-0.012 ± 0.020
$1.1 - 2.5$	-0.003 ± 0.015	-0.011 ± 0.023	0.017 ± 0.021	0.026 ± 0.016
$2.5 - 4.0$	-0.018 ± 0.013	-0.022 ± 0.022	-0.027 ± 0.020	-0.004 ± 0.013
$4.0 - 6.0$	-0.026 ± 0.013	0.026 ± 0.018	-0.037 ± 0.017	-0.003 ± 0.013
$6.0 - 8.0$	0.003 ± 0.011	-0.007 ± 0.016	-0.006 ± 0.015	-0.010 ± 0.013
$11.0 - 12.5$	0.006 ± 0.010	-0.023 ± 0.015	-0.011 ± 0.014	0.043 ± 0.012
$15.0 - 17.0$	0.015 ± 0.010	-0.008 ± 0.014	-0.009 ± 0.014	0.009 ± 0.012
$17.0 - 19.0$	0.021 ± 0.016	0.015 ± 0.019	-0.012 ± 0.019	0.006 ± 0.017

9.2 Validating of the PDFs using the high statistics simulation

In the high statistics simulation (see Sec. 6.2) the signal decay and also background is simulated. The complete implementation of the PDF (as explained in Sec. 8.1) is tested for correctness. The total data set of about 10^7 simulated events is divided into smaller samples, each containing the number of events as observed in data (see Tab. 2 in Sec. 8.6). Each sample is analysed with the maximum likelihood fit and a so called pull study is done:

For each sample j the differences between the fit results of the observables and the generated values divided by the expected uncertainty are calculated:

$$\Delta S_{i,j} = \frac{S_{i,j} - S_{i,\mathrm{gen}}}{\sigma_{S_{i,j}}} \tag{76}$$

The ΔS_i distribution, also called pull distribution, describes the fluctuations of the measured results with respect to the estimated uncertainty of each measurement. If the uncertainties of the measured values are Gaussian distributed and the fit implementation is correct the pull distribution will have a Gaussian shape with mean zero and σ one. Therefore this pull study is a good test of both the best fit point and also the estimated uncertainty of the measurement. However, the test method fails if observables are near to the physical boundaries and the confidence intervals have a non-Gaussian shape.

The mean and the σ of the Gaussian functions fitted to the pull distributions are shown in Figs. 39 and 40. The Method of Moments, which is the more robust of both methods, behaves very well in all observables and all q^2 bins. Only in the first q^2 bin ($0.1 - 0.98\,\mathrm{GeV}^2$) the observable S_{1s} shows a bias of about 45% of the size of the statistical uncertainty. The absolute size of this bias is about 0.02. The bias is related to the neglect of photon masses in the calculations. A bias of very similar size is observed with the maximum likelihood fit. The effect is also observed with the full detector simulation (see Sec. 9.1.2).

The maximum likelihood fit shows in general a good performance. However, in contrast to the Method of Moments its results seem to be slightly biased. Especially in the very first and very last q^2 bins the fitted values of some observables are shifted by up to 20% of the size of the statistical uncertainty. The width of the pulls is often smaller than one, thus it seems that the statistical uncertainty is slightly overestimated.

To test if the observed bias is related to limited statistics the full high statistics simulation sample is measured at once. By increasing the number of fitted events

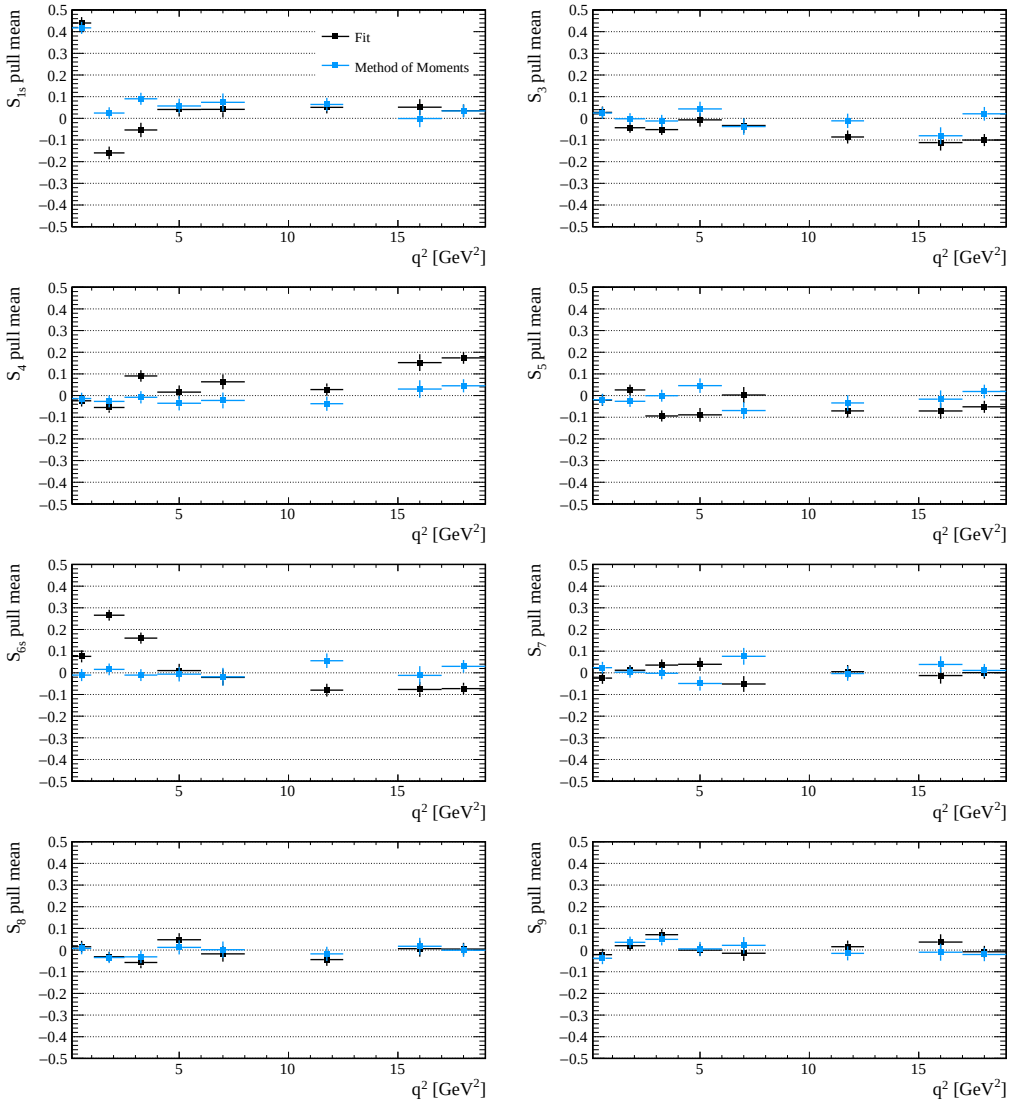

Figure 39: Mean of the pull vs. q^2 for the observables $S_{1s} - S_9$. In general there is a good agreement with zero. In S_{1s} in the first q^2 bin the effect of neglecting the lepton mass is visible. Smaller shifts are related to physical boundaries and low statistics (see text).

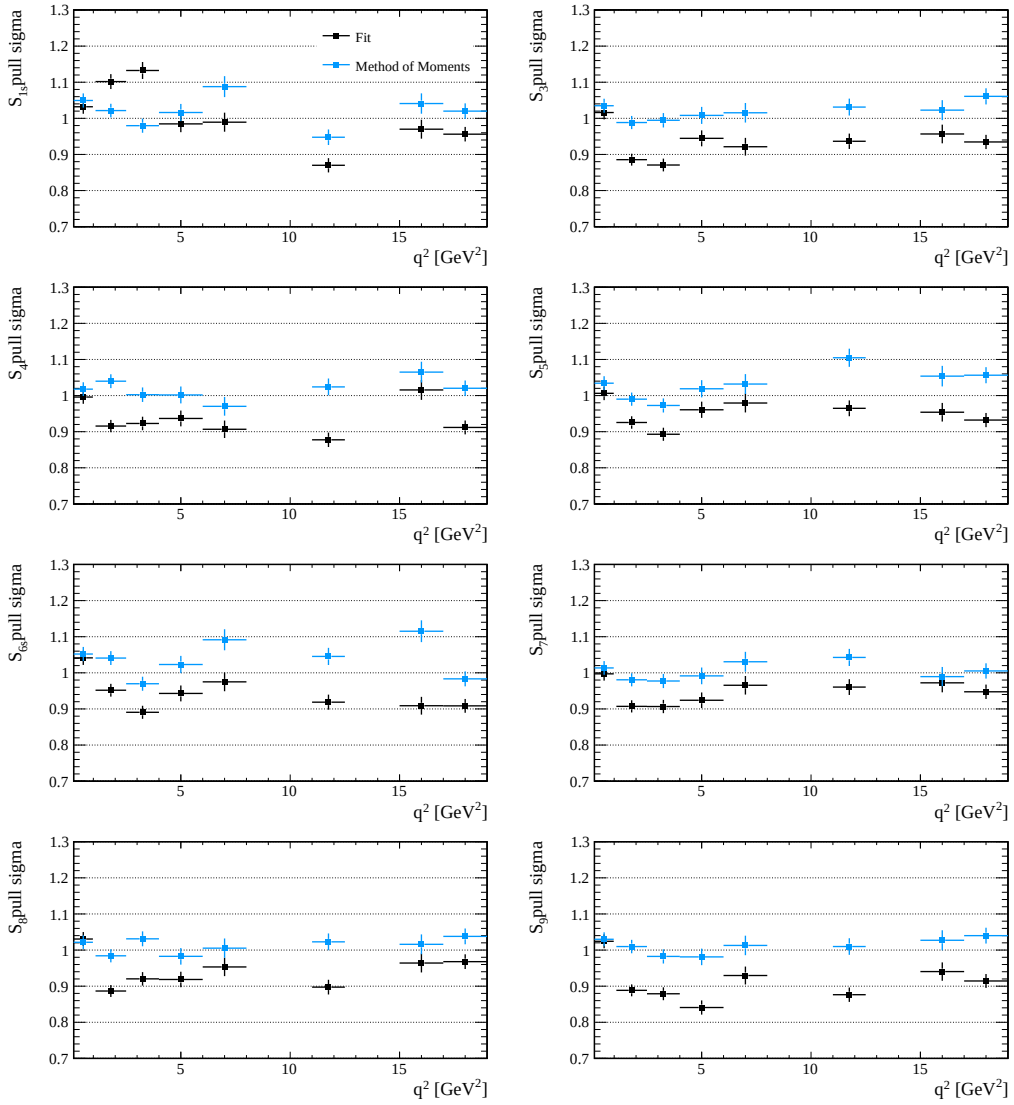

Figure 40: The width of the pulls, σ, vs. q^2 (compare Fig. 39). Only in the case of Gaussian distributed uncertainties σ can be expected to be unity. The 68% ($1\,\sigma$) confidence interval of the fit is also provided in cases where the width of the pull deviates from unity (see text).

the fit results agree very well with the generated values with respect to the expected uncertainties. Therefore it is concluded that the PDF is implemented correctly and the observed bias is due to the non-Gaussian uncertainties of the measurements.

The effect of physical boundaries can also be seen by looking at the pull distributions. Two of these distributions from the maximum likelihood fit which are not perfectly normal are shown in Fig. 41. It can be seen that there is a large step in the distribution at zero. Due to the physical boundaries the measured values of the observables can sometimes only fluctuate into one direction. In these cases the mean of the pull is also shifted into the corresponding direction. As the physical boundaries are not strict but also depend on the best fit values of the other observables[20] the effect is smeared out. Nevertheless, the confidence interval provided by the bootstrap method for the maximum likelihood fit is still correct. The two pull plots show that $67.3 \pm 1.2\%$ of the fit-results for S_{1s} ($70.4 \pm 1.2\%$ for S_{6s}) are within a range of ± 1, which is in good agreement with the expected 68%.

In conclusion, it can be said that both estimators work fine and can be used for this angular analysis.

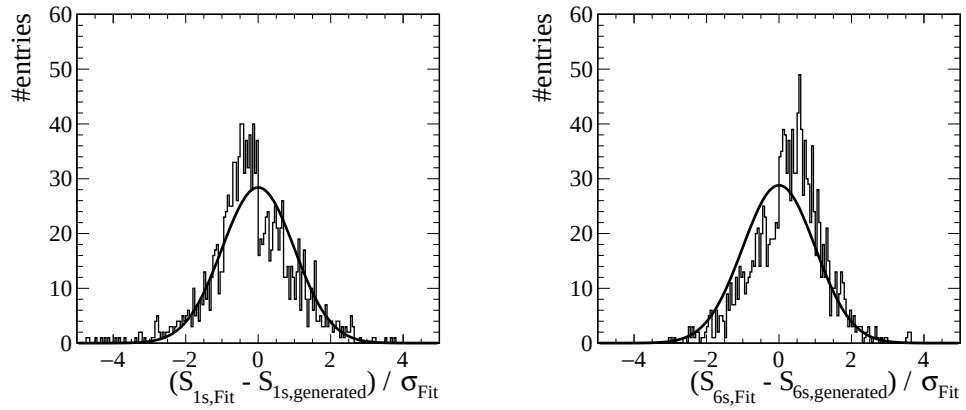

Figure 41: Pull distributions of the observables S_{1s} and S_{6s} from the maximum likelihood fit in the q^2 bin $1.1 - 2.5\,\mathrm{GeV}^2$ compared to a normal Gaussian (mean=0, σ=1). Due to boundary effects the distributions do not have a perfect Gaussian shape, however, the 68% coverage is still provided (see text).

[20]The main physical boundary is that the decay rate is not allowed to become negative.

9.3 Full validation using the $B^0 \to J/\psi K^{*0}$ control channel

A test of the complete PDF combined with a real detector response is done with the control channel $B^0 \to J/\psi K^{*0}$ from data. In this control channel about 300 000 signal candidates are available, which is about a factor of 1000 more than in each q^2 bin of $B^0 \to K^{*0}\mu^+\mu^-$. Besides validating the measurement methods, the fit of the control channel is also used to obtain values of the parameters describing the signal mass shapes. In the fits to $B^0 \to K^{*0}\mu^+\mu^-$ these parameters are fixed to the values obtained in the fit detailed in this subsection.

A 5D maximum likelihood fit is performed, similar to the one used in the final measurement. Since the mass shape parameters are determined in this fit, there are 36 floating parameters in total. Furthermore, in the fit of the control channel also the decay $B_s^0 \to J/\psi K^{*0}$ is described.

In the $K\pi\mu\mu$ invariant mass distribution besides the signal decay also the decay $B_s^0 \to J/\psi K^{*0}$ is visible (see Fig. 27). Due to the larger mass of the B_s^0 resonance the invariant mass of these decays is shifted by about 90 MeV to larger values with respect to the $B^0 \to J/\psi K^{*0}$ decays. Compared to the signal decay an event yield of about 1% is expected [86]. The decay of the B_s^0 meson can be expected to be similar to the decay of the B^0 meson, as the physics describing both decays is also very similar. Therefore the B_s^0 component is treated as signal for the test done in this section. In the final measurement of $B^0 \to K^{*0}\mu^+\mu^-$ the B_s^0 component is much less pronounced due to larger background and a smaller total event yield. The B_s^0 decay is ignored and an adequate systematic uncertainty is evaluated (see Sec. 11.4).

The results of the fit to this control channel are compared to the results of a published LHCb analysis on the same decay using the 2011 data set [98]. About one third of the data is shared with the data used for this cross check. Also the angular acceptance correction is based on the same strategy. A comparison with additionally also a measurement of another experiment, when available, will be favourable. Nevertheless, this cross checks especially allows to test the full implementation of the maximum likelihood fit, which was developed completely independent in both analyses. Comparing both results it must be noted that the previous analysis is systematically limited. For this comparison no systematic study of the control channel is done, as systematic effects are expected to be highly correlated between both studies.

The measured observables obtained with the 5D maximum likelihood fit and the results from Ref. [98] are listed in Tab. 8. The results from [98] are translated to the description used in this analysis using the equations 25 to 35 in Sec. 4.3.4. Within

these uncertainties both results agree very well. Considering the about a factor 1000 less signal candidates in each q^2 bin of the $B^0 \to K^{*0}\mu^+\mu^-$ decay the performance of the maximum likelihood fit is fully sufficient.

In Tab. 9 the complete set of the remaining floating parameters of the fit is shown. In the fit to $B^0 \to K^{*0}\mu^+\mu^-$ the parameters describing the signal mass shapes are fixed to these values. The parameter σ_{m} describing the invariant mass resolution will be scaled according to simulation (see Sec. 8.1.1).

The five 1D projections of the fit are presented in Fig. 42. Additional also 2D pull histograms are shown in Fig. 43 and Fig. 44. The given value in each bin of such a pull histogram refers to the measured number of events minus the event yield predicted by the fit divided by the estimated uncertainty. In general a good agreement of the fit projections with the data points is observed. Two effects, which will be discussed in the following, are visible.

First, due to the very low level of combinatorial background in the decay $B^0 \to J/\psi\, K^{*0}$ the background shape in the higher invariant $K\pi\mu\mu$ mass region is not optimally described. In the decay $B^0 \to K^{*0}\mu^+\mu^-$ there is a much larger background-to-signal ratio and a lower event yield. Thus, for the fit to $B^0 \to K^{*0}\mu^+\mu^-$, this background description is sufficient. The parameter α_m describing the slope of the background is left floating in each $B^0 \to K^{*0}\mu^+\mu^-$ fit.

Second, there are several artefacts visible in the $\cos\theta_K$ projection. These artefacts are related to the angular acceptance and physical background. A better understanding of these artefacts is obtained from the 2D pull histograms of the $K\pi\mu\mu$ invariant mass and $\cos\theta_K$ (see Fig. 44) .The effect of the B^+ veto selection requirement is visible for a $K\pi\mu\mu$ invariant mass larger than 5400 MeV and $\cos\theta_K > 0.5$. This effect is accounted for as described in Sec. 8.1.2. As all events from this region are excluded from the analysis no pull can be calculated there. The artefact in $\cos\theta_K < 0$ is located primarily in the signal mass region (≈ 5300 MeV). Compared to $B^0 \to K^{*0}\mu^+\mu^-$ there is additional background expected in the fit of the control channel $B^0 \to J/\psi\, K^{*0}$. For example the decay $B^0 \to Z(4430)^-(\to J/\psi\,\pi^-)K^+$, whose existence was recently confirmed [99]. Due to its angular distribution a peaking structure at $\cos\theta_K \approx -0.5$ is expected. Since this decay has the same initial and final state as the signal decay it has a similar invariant mass distribution. Therefore this decay is located in phase space where the artefact is visible. The $K\pi$ invariant mass distribution of this decay is flat comparable to the one of the S-wave contribution of the signal decay. Because the Z(4430) decays via a J/ψ resonance, which is explicitly removed in the selection of $B^0 \to K^{*0}\mu^+\mu^-$, this background is only visible in this control channel. Another source of background with peaking structures is the decay

$B_s^0 \to \phi\mu^+\mu^-$. This decay is expected to have a contribution of about 0.3% compared to the signal yield (see Sec. 7.3). Due to the mis-identification of one of the kaons in the final state as a pion the reconstructed mass of the B meson is smaller. Therefore this background is present especially below the signal mass region ($< 5300\,\text{MeV}$). Due to the mis-reconstruction in the angular distribution this background is present mainly at $\cos\theta_K < 0$. Therefore this would explain why below the signal mass ($< 5300\,\text{MeV}$) there is less background at $\cos\theta_K > 0$ compared to $\cos\theta_K < 0$. The effect of mis-reconstructed background on the final result is studied in Sec. 11.3.

Compared to the expected statistical uncertainty of the measurement of $B^0 \to K^{*0}\mu^+\mu^-$ all observed effects are small. It is shown that the description of the PDF and the angular correction is fully sufficient.

Table 8: Comparison of the results obtained with the maximum likelihood fit to $B^0 \to J/\psi K^{*0}$ and a published LHCb analysis of this decay [98].

Parameter	Fitted value	Ref. [98]
S_{1s}	0.333 ± 0.001	0.321 ± 0.011
S_3	0.003 ± 0.002	-0.013 ± 0.007
S_4	-0.257 ± 0.002	-0.250 ± 0.005
S_5	-0.005 ± 0.002	0
S_{6s}	0.004 ± 0.002	0
S_7	0.002 ± 0.002	0
S_8	-0.053 ± 0.002	-0.048 ± 0.009
S_9	-0.085 ± 0.002	-0.084 ± 0.009

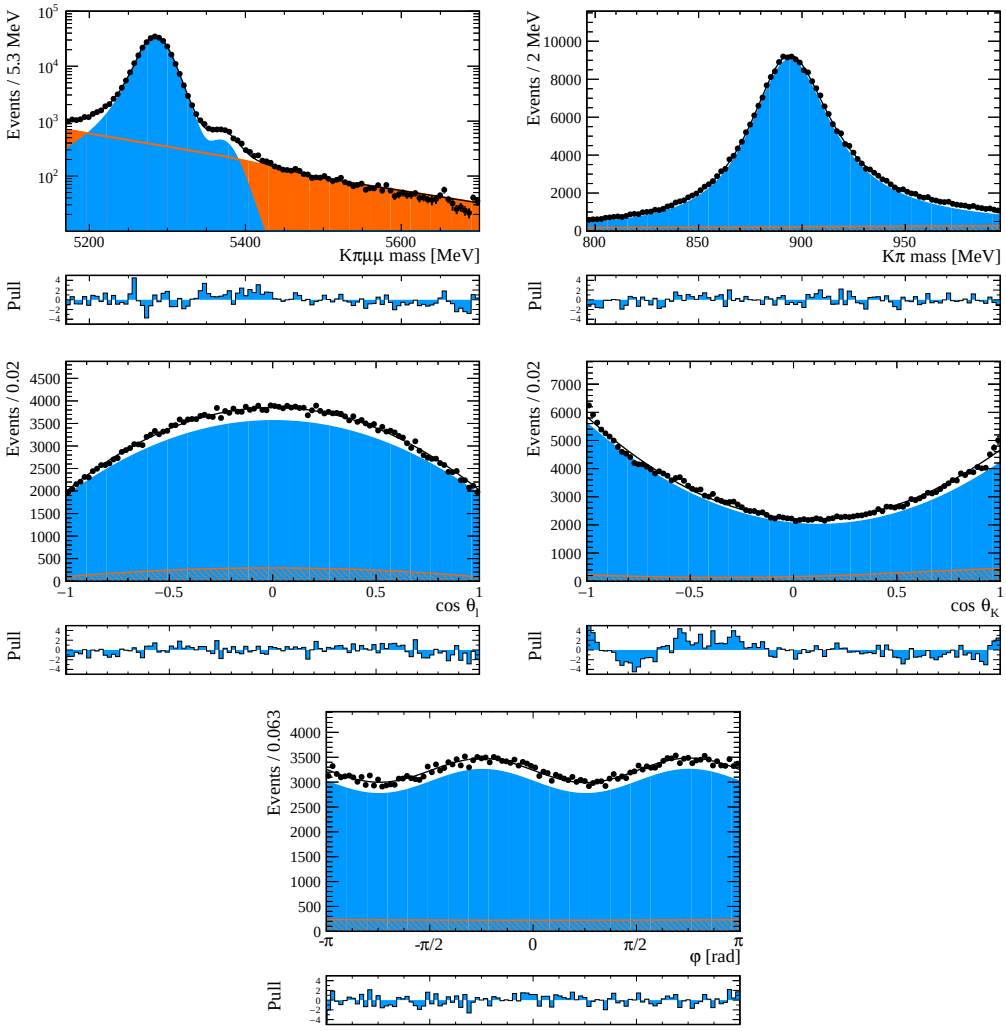

Figure 42: Projections of the 5D maximum likelihood fit to $B^0 \to J/\psi K^{*0}$. The signal component (blue) and the combinatorial background (orange) are compared to the weighted data points (black).

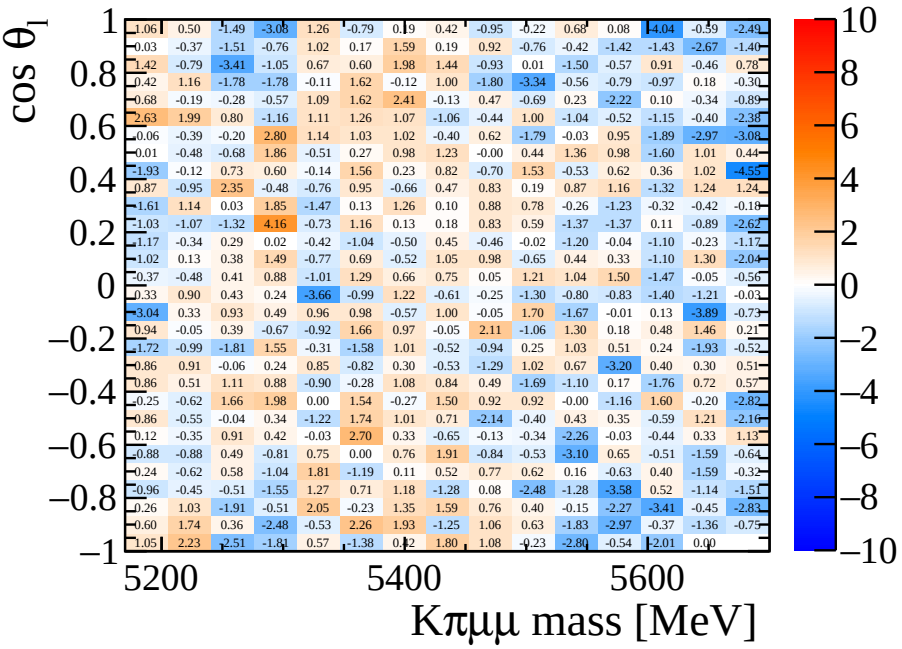

Figure 43: 2D Pulls of the maximum likelihood fit to $B^0 \to J/\psi K^{*0}$ for the $K\pi\mu\mu$ invariant mass vs. $\cos\theta_l$. In each bin it is shown: $(N_{\rm ev} - N_{\rm fit})/\sigma_{fit}$.

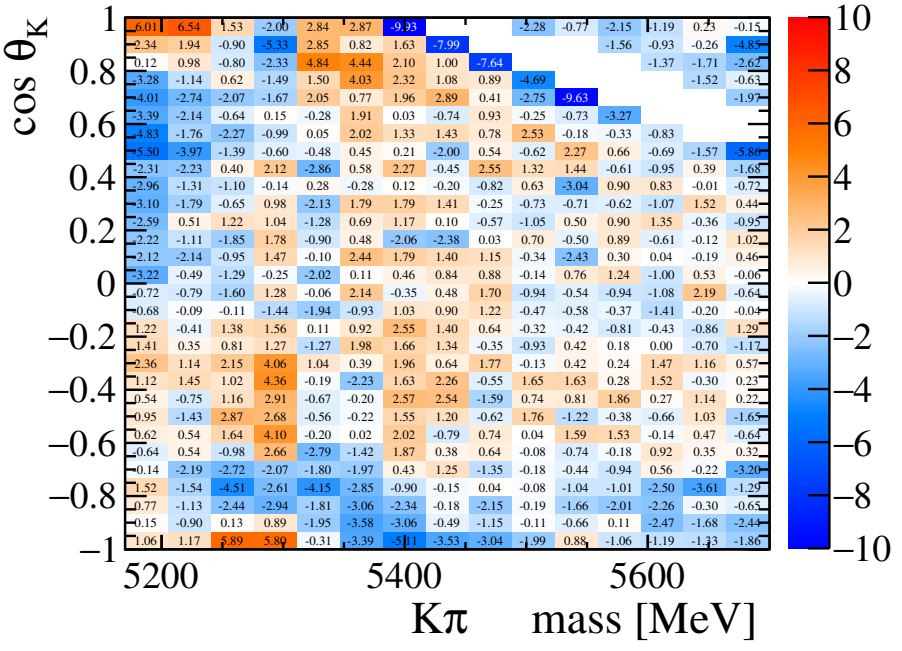

Figure 44: 2D Pulls of the maximum likelihood fit to $B^0 \to J/\psi K^{*0}$ for the $K\pi\mu\mu$ invariant mass vs. $\cos\theta_K$. In each bin it is shown: $(N_{\rm ev} - N_{\rm fit})/\sigma_{fit}$.

Table 9: A complete list of nuisance and mass shape parameters in the maximum likelihood fit to $B^0 \to J/\psi K^{*0}$ data. The parameters describing the signal $K\pi\mu\mu$ and $K\pi$ invariant mass shape are fixed in the $B^0 \to K^{*0}\mu^+\mu^-$ fits to these values (plus scaling factor).

Parameter	Description	Fitted value
	Invariant $K\pi\mu\mu$ mass shape	
$f_{\rm sig}$	signal fraction	0.9298 ± 0.0009
$m_{B^0}[\,\mathrm{MeV}]$	B^0 mass	5284.282 ± 0.047
$f_{CB1/CB2}$	fraction 1st CB	0.644 ± 0.037
$\sigma_{\rm m,1}[\,\mathrm{MeV}]$	width of 1st CB	15.12 ± 0.24
$\sigma_{\rm m,2}/\sigma_{\rm m,1}$	ratio width 2nd/1st CB	1.628 ± 0.026
n_{CB}	n of both CB	2.12 ± 0.17
α_{CB1}	α of 1st CB	1.653 ± 0.026
α_{CB2}	α of 2nd CB	1.746 ± 0.081
$f_{B_s^0/B^0}$	fraction of B_s^0 events	0.01238 ± 0.00044
$\Delta m_{B_s^0}[\,\mathrm{MeV}]$	mass difference B_s^0 and B^0	85.70 ± 0.84
$\alpha_{\rm m}[\,\mathrm{MeV}^{-1}]$	background exponent	0.00552 ± 0.00007
	Invariant $K\pi$ mass shape	
$m_{K^{*0}}[\,\mathrm{MeV}]$	K^{*0} mass	895.380 ± 0.064
$\Gamma_{K^{*0}}[\,\mathrm{MeV}]$	K^{*0} decay width	48.92 ± 0.19
$\alpha_{\rm S-wave}$	relative amplitude size	3.8 ± 1.1
$\beta_{\rm S-wave}$	relative phase difference	-1.148 ± 0.031
$s_{K\pi}[\,\mathrm{MeV}^{-1}]$	background slope of $K\pi$	0.438 ± 0.030
	Angular background	
$c_{\cos\theta_l,1}$	Chebyshev $cos\theta_l$	-0.008 ± 0.016
$c_{\cos\theta_l,2}$	Chebyshev $cos\theta_l$	-0.508 ± 0.019
$c_{\cos\theta_K,1}$	Chebyshev $cos\theta_K$	0.573 ± 0.016
$c_{\cos\theta_K,2}$	Chebyshev $cos\theta_K$	0.412 ± 0.012
$c_{\phi,1}[rad^{-1}]$	Chebyshev ϕ	-0.002 ± 0.005
$c_{\phi,2}[rad^{-2}]$	Chebyshev ϕ	0.003 ± 0.002
	S-wave fraction and interference	
F_S	S-wave fraction	0.0490 ± 0.0016
$s1$	S/P-wave interference	-0.5758 ± 0.0056
$s2$	S/P-wave interference	0.1157 ± 0.0052
$s3$	S/P-wave interference	-0.0006 ± 0.0035
$s4$	S/P-wave interference	0.0009 ± 0.0033
$s5$	S/P-wave interference	-0.1484 ± 0.0038

9.4 Goodness of fit - Point-to-Point Dissimilarity Method

The best fit point and the confidence intervals are based on the maximum likelihood method. Nevertheless, it is not wise to also compute the goodness of fit from the likelihood [100]. There is a multitude of methods available which can be used, performing best in different scenarios [101].

Due to low statistics the typically used binned χ^2 method would give only rather poor results for the present case. The challenge are bins with only a few or zero entries, which will especially occur if not only the 1D projections but the multidimensional distribution is analysed. The bins would have to be chosen rather large removing local structures.

Instead the unbinned point-to-point dissimilarity (P2PD) method is used to test the agreement of the fits with data. The point-to-point dissimilarity method [101] is designed to test if two data sets originate from the same PDF. A comparison of the fitted PDF with data can be performed by generating pseudo-experiments according to the PDF proposed by the fit. It is then tested if the data events agree with the simulated events. To reduce statistical fluctuations, the number of generated events should be significantly larger than the number of data events. As a result this method is rather time consuming and impractical for a large data set.

The test statistic T for the point-to-point dissimilarity method is calculated according to the following formula:

$$T = \frac{1}{n_{\mathrm{d}}^2} \sum_{i=1}^{n_{\mathrm{d}}} \sum_{j>i}^{n_{\mathrm{d}}} \Psi(|x_i^{\mathrm{d}} - x_j^{\mathrm{d}}|) - \frac{1}{n_{\mathrm{d}} n_{\mathrm{sim}}} \sum_{i=1}^{n_{\mathrm{d}}} \sum_{j=1}^{n_{\mathrm{sim}}} \Psi(|x_i^{\mathrm{d}} - x_j^{\mathrm{sim}}|) \tag{77}$$

with n_{d} data and n_{sim} simulated events and a function Ψ of the distance between two points. The first term means that Ψ is evaluated and summed up for all data points. In the second term Ψ is calculated for all data and simulated events. A third term which calculates the distance between all simulated events is usually neglected.

There is no strict rule for the choice of the distance function Ψ. As discussed in Ref. [101] a Gaussian function provides good results. The width of the Gaussian must be tuned with simulation to give best results. The maximum likelihood fit used for measuring the angular observables is done in five dimensions. The $K\pi\mu\mu$ invariant mass is assumed to be completely orthogonal to the other four variables. Besides a slight influence of the S/P-wave interference terms the $K\pi$ invariant mass is also independent of the three angular dimensions. Therefore it is convenient to also test these contributions separately. Three goodness-of-fit p-value are calculated for each fit. Tested are the projections of the $K\pi\mu\mu$ invariant mass, the $K\pi$ invariant mass

and the decay angles. In the first two cases a one-dimensional Gaussian function is used for Ψ, in the last case a three-dimensional Gaussian function is used. The distance functions are defined as follows:

$$\Psi_{m,B^0}(m_i, m_j) = \exp\left(-\frac{(m_i - m_j)^2}{2\,\sigma^2_{\mathrm{m},B^0}}\right) \tag{78}$$

$$\Psi_{m,K^{*0}}(m_i, m_j) = \exp\left(-\frac{(m_i - m_j)^2}{2\,\sigma^2_{\mathrm{m},K^{*0}}}\right) \tag{79}$$

$$\Psi_{\mathrm{angles}}(x_i, x_j) = \exp\left(-\frac{(\phi_i - \phi_j)^2}{2\,\sigma^2_{\phi}} - \frac{(\theta_{l,i} - \theta_{l,i})^2}{2\,\sigma^2_{\theta_l}} - \frac{(\theta_{k,i} - \theta_{k,i})^2}{2\,\sigma^2_{\theta_k}}\right) \tag{80}$$

with the width of the Gaussian functions tuned to be $\sigma_{\mathrm{m},K^{*0}} = 100\,\mathrm{MeV}$, $\sigma_{\mathrm{m},B^0} = 40\,\mathrm{MeV}$, $\sigma_{\phi} = \pi/2$, $\sigma_{\theta_l} = 1$, $\sigma_{\theta_K} = 1$.

The p-value is determined in the following way. First, the test statistic of the data sample is calculated. Afterwards, the data events are replaced by the same number of simulated events and T is calculated again[21]. This is repeated until the T distribution of the specific case is known with a sufficient precision. The p-value is the fraction of calculations, for which the obtained value of T is lower than the one calculated on data. The Goodness of fit p-values of the P2PD method for the final maximum likelihood fits are given in Sec. 12.

[21] Actually due to timing reasons no new simulated events are generated, but a subsample is randomly drawn from the combined pool of simulated and data events.

10 Correlations

The correlation between two random variables X,Y is defined as

$$\mathrm{cor}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle}{\sigma_X \sigma_Y} \tag{81}$$

where cov is the covariance between the two random variables, σ the standard deviation and $\langle\rangle$ the expectation value operator. Although most of the angular terms are orthogonal this does not mean that they are also uncorrelated. The correlation of the different angular terms can be calculated analytically using their moments:

$$\mathrm{cor}(S_i, S_k) = \frac{\langle f_i f_k \rangle - \langle f_i \rangle \langle f_k \rangle}{\sqrt{\langle f_i^2 \rangle - \langle f_i \rangle^2} \sqrt{\langle f_k^2 \rangle - \langle f_k \rangle^2}} \tag{82}$$

where f_i are the angular terms, as defined in Eq. 61 in Sec. 8.2.1. The formulas to calculate the correlations can be found in appendix A.1. The corresponding correlation matrices in the q^2 bin are shown in appendix. A.2.

The mean correlation of the observables of an estimator can also be obtained from simulation. Doing multiple pseudo experiments the correlation can be measured according to:

$$\mathrm{cor}(S_i, S_k) = \frac{1}{N_{\mathrm{sim}}} \sum_{n=1}^{N_{\mathrm{sim}}} \frac{(S_{i,n} - \langle S_i \rangle)(S_{k,n} - \langle S_j \rangle)}{\sigma_{S_{i,n}} \sigma_{S_{k,n}}} \tag{83}$$

where N_{sim} is the number of pseudo experiments and S_i are the observables measured in each experiment with the uncertainty σ_{S_i}.

In Fig. 45 the correlation for two sets of observables which have a significant correlation is shown. The high statistics simulation is used (see Sec. 6.2) and about 1500 pseudo experiment per q^2 bin are done. Only the signal decay is simulated as the combinatorial background has a different correlation. The maximum likelihood fit is affected by physical boundaries. Therefore the mean correlation of the observables determined with this estimator is slightly different compared to the correlation expected due to the PDF. In contrast, the correlation of the Method of Moment Estimator agrees very well with the analytical calculation according to the PDF.

In each data sample there will be a specific correlation of the observables due to fluctuations that are present in this specific data set. These correlations must be determined on a data driven method. The correlation of the PDF determined previously can be taken as a good indication for what correlation can be expected.

The default option is to use the correlation matrix which is provided by `Hesse` in `Minuit` (see Sec. 8.1.1). This matrix is the inverse of the second derivative matrix at the best fit point. However, if this point is near a physical boundary the likelihood becomes highly non-Gaussian and the calculation by `Minuit` via the second derivatives no longer accurately reflects the correlations.

A possibility to take the boundaries into account is to scan the 2D likelihood profile. For each set of two parameters the 2D likelihood profile is sampled and the correlation is calculated via the following formula:

$$cor(S_i, S_k) = \frac{1}{\sum p_{\text{point}}} \sum\nolimits_{\text{point}} p_{\text{point}} \frac{(S_{i,\text{point}} - S_i)(S_{k,\text{point}} - S_k)}{\sigma_{S_i}\sigma_{S_k}} \tag{84}$$

where p_{point} is the relative likelihood (without log) at the sampled point compared to the maximum likelihood value at the best fit point.

In Fig. 46 the different methods are compared in a simulation (about 1500 pseudo experiments per q^2 bin). This time background is included in the simulation. As the sampling of the likelihood is very time consuming in this comparison the likelihood is sampled at 1000 random points, which is sufficient for a qualitative comparison. In the real measurement the correlation will be calculated based on a grid of 100×100 points.

The mean of the correlation of the observables determined with `Hesse` and the likelihood scan method are compared to the measured correlation of the fit. Whereas in some regions the mean of the matrix from `Hesse` shows deviations from the measured correlation of the fit, the likelihood scan method performs better. The observed deviations of the likelihood scan method and the measured correlation of the fit are of the order 5%. When using the provided correlation matrices, the coefficients should be varied by this quantity in a systematic study. Due to the better performance the correlation matrix provided for the real measurement is based on the likelihood scan method.

As discussed in Sec. 9.2 the uncertainties of the observables are expected to be in some cases slightly non-Gaussian. Therefore additionally also 2D profile likelihood scans are provided for the real measurement. With these scans it is possible to also take non-linear correlation of observables in a theoretical study into account.

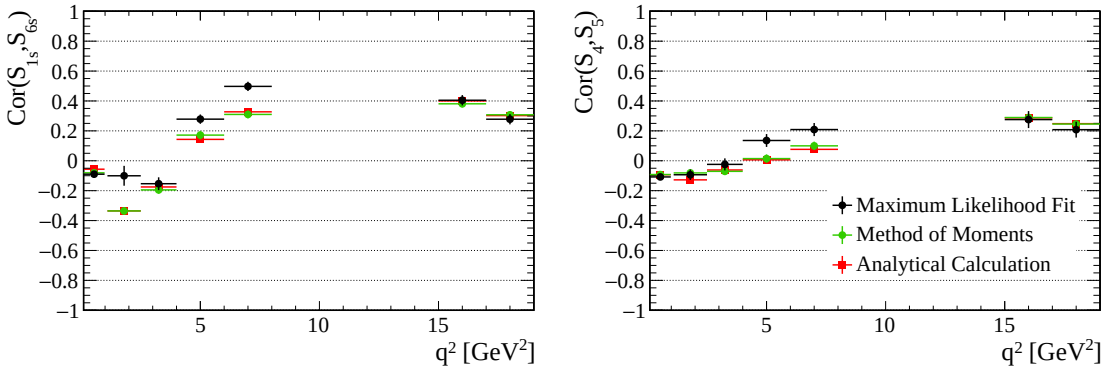

Figure 45: Mean Correlation of two sets of observables which have a significant correlation in bins of q^2. The error bars show the RMS of the measured distribution. The analytical calculation of the PDF (red, see Eq. 82) is in good agreement with the mean correlation measured with the Method of Moments (green, see Eq. 83). In contrast further structures are visible in the mean correlation of the results of the maximum likelihood fit (black, see Eq. 83), which are constrained by the physical boundaries.

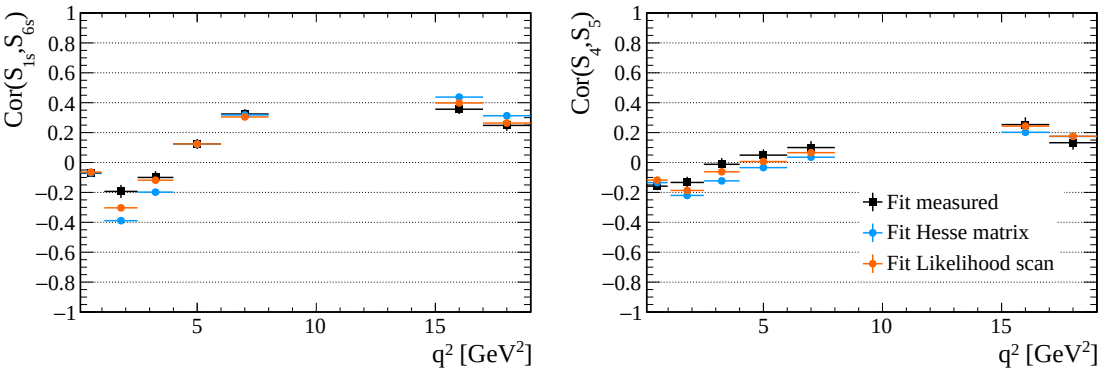

Figure 46: Mean Correlation of two sets of observables which have a significant correlation in bins of q^2. The error bars show the RMS of the measured distribution. Shown is the measured mean correlation of the results of the maximum likelihood fit (black, see Eq. 83), the mean of the correlation values provided by `Hesse` (blue) and the mean of the values obtained with the likelihood scan method (orange, see Eq. 84).

11 Systematic uncertainties

The systematic uncertainties of the P-wave observables are estimated on different simulations. Effects related to the reconstruction and selection efficiency are studied using the full detector simulation (physics simulation, see Sec. 6.1), for which there is, however, only limited statistics available ($\sim$ 4k events/q^2 bin). The empirical models for the invariant mass and background description are tested with the high statistics simulation (see Sec. 6.2), which is not suited for testing the detector response but is ideal for testing the maximum likelihood fit. Although a difference between the Standard Model prediction and the angular measurement has been observed in an earlier measurement, deviations are small enough to get reliable estimation of the systematic uncertainties based on simulation.

In general the determination of the systematic uncertainty is done by comparing two different models or two different methods. The nominal option is the maximum likelihood fit as explained in Sec. 8.1. All systematic studies will be explained in detail in the following subsections.

11.1 Angular acceptance

11.1.1 Statistical uncertainty

Due to the limited size of the simulation sample used to determine the angular acceptance (PHSP simulation, see Sec. 6.1) there is a statistical uncertainty related to the acceptance correction.

The effect on the observables is measured with the Bootstrapping technique [97]. A random subset of events is drawn from the simulation sample (with replacement) which has in total the size of the original sample. The acceptance correction is calculated based on this subset and the observables are measured on the physics full detector simulation using this correction.

The procedure is repeated 100 times and the RMS of the distribution of measured observables is taken as systematic uncertainty. The largest effect is on the observables S_{1s}, S_4, S_{6s} and S_8 which is a shift[22] of the order of $0.002 - 0.005$.

11.1.2 Systematic shift

When measuring the observables using the $B^0 \to J/\psi K^{*0}$ detector simulation it is observed that the observable S_4 is systematically shifted by about 0.01 compared to the generated value (see Sec. 9.1.1) . This bias is related to the acceptance correction, where a trade-off between the statistical uncertainty of the measurement and the ability to describe steep changes in the efficiency has to be made (see Sec. 8.3).

The differences between generated and measured values are measured with the $B^0 \to K^{*0}\mu^+\mu^-$ full detector simulation and it is corrected for these shifts. As a

[22] If not explicitly stated otherwise all numbers quoted in this chapter correspond to the absolute (not relative) effect.

consequence the statistical uncertainty of this measurement is taken as a systematic uncertainty which is of the size $0.01 - 0.02$. This is the leading systematic uncertainty in the measurement. The uncertainty scales with the precision of the fit to measure a certain observable and has in general a size of about 20% compared of the expected statistical uncertainty.

The systematic shift which is corrected for is in half of the cases below 0.01. The largest observed shifts are 0.04 in S_9 (q^2 bin $11.0 - 12.5\,\mathrm{GeV}^2$), 0.03 in S_5 (q^2 bin $1.1 - 2.5\,\mathrm{GeV}^2$) and 0.03 in S_3 (q^2 bin $0.1 - 0.98\,\mathrm{GeV}^2$, $17.0 - 19.0\,\mathrm{GeV}^2$). The complete list of shifts can be seen in Tab. 7 in Sec. 9.1.2. All effects are small compared to the expected statistical uncertainty of the angular measurement. Also the effect of neglecting the lepton masses in the description of the differential decay rate is included in this correction. This effect is a systematic shift in S_{1s} of about 0.02 in the first q^2 bin ($0.1 - 0.98\,\mathrm{GeV}^2$).

11.1.3 Differences in data and simulation

Although there has been a lot of improvement in the last years the simulation of the detector response and the physical description of the proton-proton interaction is still not perfect. Comparing kinematic distributions of the particles smaller differences are observed. Especially the transverse momentum of the pion is too soft in simulation compared to data. The track multiplicity of the underlying event is in general too low.

The reconstruction and selection efficiency depends on kinematic observables and it is crucial that these distributions are described correctly in simulation. The effect of the differences is measured by weighting events in simulation to match the distributions in data. The observables which are investigated are the transverse momentum of the pion, kaon and muons, the track multiplicity in the event, the vertex quality of the B^0, and q^2.

The comparison of data and simulation is done in the high statistics control channel $B^0 \to J/\psi\, K^{*0}$. As in simulation the S-wave is not described, a tight selection requirement of $\pm 20\,\mathrm{MeV}$ around the nominal K^{*0} mass is applied to select mainly the resonant P-wave decay. To extract the signal distribution in data the sPlot technique (see Sec. 8.2.2) is used. As the q^2 distribution cannot be measured with the control channel the comparison of this variable is done on the real $B^0 \to K^{*0}\mu^+\mu^-$ decay and the corresponding $B^0 \to K^{*0}\mu^+\mu^-$ physics simulation.

The acceptance correction (see Sec. 8.3) is calculated based on the weighted events and the fit is done based on this correction. The differences to the nominal result are taken as systematic uncertainties. The largest effect is visible in S_{1s} and S_{6s} when weighting according to the differences observed in the pion transverse momentum, which is of the size of up to 0.008. Therefore in these observable the systematic uncertainty is occasionally of a similar size as the leading systematic uncertainty discussed in Sec 11.1.2.

11.1.4 Track reconstruction efficiency

The track reconstruction efficiency has been measured and compared to simulation [102]. The analysis is done using $J/\psi \to \mu^+\mu^-$ decays and exploiting the tag and probe technique. Differences are found to be small.

The effect on the angular acceptance correction is studied using the full detector simulation. Events of the PHSP simulation are weighted according to the weights obtained in the measurement of tracking efficiencies (binned in p, η). Observables are measured in the $B^0 \to K^{*0}\mu^+\mu^-$ full detector simulation one time using the nominal correction and one time using the weighted correction. The difference between both results is taken as systematic uncertainty. The difference is at most 0.008.

11.1.5 $K\pi$ invariant mass dependence

The angular acceptance is measured in three dimensions (three decay angles) and is binned in q^2. Due to the limited size of the simulation sample it is not feasible to also include a possible dependence on the $K\pi$ invariant mass in the acceptance correction. In the 1D $K\pi$ invariant mass projection the acceptance is constant within about 5%.

The effect of ignoring the $K\pi$ invariant mass in the acceptance is studied on the high statistics simulation. In the measurement of the simulation an artificial dependence to the $K\pi$ invariant mass is introduced. The weights related to the acceptance correction are increased/decreased linearly with the $K\pi$ invariant mass by 5%. The change in the observables compared to the nominal case is taken as systematic uncertainty. Always the larger deviation of both cases $(+/-)$ is taken as uncertainty.

This affects mostly the interference parameters which, however, are only nuisances parameters. The effect on the P-wave observables is very small.

11.2 Angular resolution

The momentum resolution of final state particles in the LHCb detector is of the order of $\mathrm{d}p/p = 0.4\% - 1\%$ (see Sec. 3.1.2). As a result also the decay angles in the Eigensystem can only be determined up to about 10 mrad.

A poor resolution of the decay angles would smear out structures in the angular distribution leading to a biased measurement. Also the correction of the angular acceptance could suffer from resolution effects.

Nevertheless, the LHCb detector is precise enough that these effects do not become significant. The effect of angular resolution is tested with the complete detector simulation. The fit of the observables is done one time using the reconstructed quantities and one time using the generated ones. In both cases the acceptance weights are calculated again. As a systematic uncertainty the difference between both results is taken. The difference is for all observables and all q^2 bins smaller than 0.003.

11.3 Physics background

There is not only combinatorial background but also background due to misreconstructed decays. The $K\pi\mu\mu$ invariant mass distribution of these decays normally has a peaking structure. Due to the peaking structure this background usually causes systematic effects if neglected. After the selection and the explicit veto requirements there are three major physics background components remaining, which have a relative contribution compared to the signal of up to 1% (see Sec. 7.3):

- $\Lambda_b^0 \to pK^-\mu^+\mu^-$ $(1.0 \pm 0.4\%)$
- $B^0 \to K^{*0}\mu^+\mu^-$ $(0.64 \pm 0.06\%)$
- $B_s^0 \to \phi\mu^+\mu^-$ $(0.33 \pm 0.12\%)$

In the nominal high statistics simulation no physics background is included. For a systematic study artificially physics background is introduced according to the expected amount and the expected phase space distributions from simulation. Due to the unknown exact angular distribution of the $\Lambda_b^0 \to \Lambda^*(1520)^0\mu^+\mu^-$ decay a phase space simulation is used. The main structure in the angular distribution is related to the wrong mass hypothesis therefore this procedure is sufficient for a systematic study.

The differences of the results obtained with the nominal maximum likelihood fit using the nominal simulation and the simulation including physics background is taken as systematic uncertainty. The differences are always smaller than 0.003.

11.4 Neglecting of $B_s^0 \to K^{*0}\mu^+\mu^-$

In the fit of the control channel $B^0 \to J/\psi\, K^{*0}$ also the channel $B_s^0 \to J/\psi\, K^{*0}$ is visible with a fraction relative to the signal of about 1%.

In the fit of $B^0 \to K^{*0}\mu^+\mu^-$ the event yield is much lower and also the expected background level is significantly larger. Therefore, in these fits the decay of the B_s^0 meson is much less pronounced and for simplicity neglected in the fit. For a systematic study a high statistics simulation with a 1.2% B_s^0 fraction is generated. The angular distribution of the B_s^0 component is assumed to be identical to the one of the signal $B^0 \to J/\psi\, K^{*0}$ decay.

The fit is done once including the B_s^0 component in the PDF and once setting the B_s^0 fraction in the PDF to zero. The difference between both results is taken as systematic uncertainty. The effect is small and at most of the size 0.0007.

11.5 Signal $K\pi\mu\mu$ invariant mass model

In the nominal fit the signal distribution of the $K\pi\mu\mu$ invariant mass is described by two Crystal Ball functions (see Sec. 8.1.1).

For the systematic study in the control channel $B^0 \to J/\psi\, K^{*0}$ the $K\pi\mu\mu$ invariant mass is fitted with two Gaussian functions. A high statistics simulation is produced according to the measured mass shape. This simulation is measured one time with two Gaussian functions and one time with the nominal description of two Crystal Balls functions. The difference between both models, which is used as an estimation for the systematic uncertainty, is always smaller than 0.001.

11.6 Background angular model

The nominal description of the angular distribution of the combinatorial background is based on three second order Chebyshev polynomials (see Sec. 8.1.1).

The control channel $B^0 \to J/\psi\, K^{*0}$ is fitted with three Chebyshev polynomials of fourth order. A high statistic simulation is generated accordingly. The systematic uncertainty is estimated as the difference between fitting this simulation with three second order Chebyshev polynomials and three fourth order Chebyshev polynomials. The difference between both models is always smaller than 0.002

11.7 $K\pi$ invariant mass

11.7.1 S-wave model

The nominal description of the $K\pi$ invariant mass distribution of the S-wave is the Isobar model (see Sec. 8.1.1). An alternative description, the LASS parametrisation [103], is tested as a systematic study. This description is similar to the Isobar model, however, instead of the $K_0^*(800)$ an effective component is used in the description of the S-wave decay amplitude.

The nominal high statistics simulation, which was generated using the Isobar model, is fitted using the LASS parametrisation and the difference to the nominal result is taken as systematic uncertainty. Both models agree very well and the systematic uncertainty accounted for is everywhere below 0.001.

11.7.2 Background model

The background in the $K\pi$ invariant mass distribution is described by a linear model (see Sec. 8.1.1). In the control channel $B^0 \to J/\psi\, K^{*0}$ the background distribution is fitted with a fourth order Chebyshev polynomial.

A high statistics simulation is produced according to the measured distributions. This simulation is fitted once with a fourth order Chebyshev polynomial and once with the nominal linear model. The difference between both results is taken as systematic uncertainty. The difference between both models is below 0.002.

11.8 Summary

All systematic uncertainties are small compared to the statistical uncertainty of the measurement. The total expected systematic uncertainty is calculated as the squared sum of the single components. For the q^2 bin $4 - 6\,\mathrm{GeV}^2$ a complete overview of the systematic effects is shown in Tab. 10. The detailed list of systematic uncertainties for all q^2 bins is shown in Appendix C. In all cases the total systematic uncertainty is significant smaller than the expected statistical uncertainty therefore the measurement is statistically limited.

The dominating systematic uncertainty is due to the correction of smaller biases introduced by the acceptance correction which is based on the full detector simulation. As the production of this simulation is rather time consuming only a limited amount of statistics is available. The size of this uncertainty is about 0.01 to 0.02 which corresponds to about 20% of the expected statistical uncertainty.

The second largest uncertainty is also related to the simulation used for the acceptance correction. The kinematics of the pion are not perfectly described in simulation affecting mainly the observables S_{1s} and S_{6s}. The uncertainties related to this have a size of up to 0.008. All remaining uncertainties are small.

In the previous angular analysis of $B^0 \to K^{*0}\mu^+\mu^-$ at LHCb the dominating systematic uncertainties are related first also to the angular acceptance correction and second to the S-wave, which is neglected in the fit [43]. Similarly also in the angular analysis of $B^0 \to J/\psi\, K^{*0}$ at LHCb the dominating systematic is due to the acceptance determination and especially the simulation of the kaon and pion kinematics [98].

In future measurements it will be crucial to develop a more precise technique for the determination of the reconstruction and selection efficiency. It would be advisable to use a data driven technique to not rely anymore on the simulation. Such a technique would require the precise knowledge of an angular distribution of a process to which the distributions after reconstructing and selecting the particles could be normalized.

Table 10: Systematic uncertainties of the maximum likelihood fit in the q^2 range $4.0 - 6.0\,\mathrm{GeV}^2$. Shown is the statistical uncertainty of angular acceptance (stat. ang. acc., Sec. 11.1.1), statistical uncertainty of the simulation (stat. sim., Sec. 11.1.2), weighting according to differences in simulation (weight x, Sec. 11.1.3), track reconstruction efficiency (tracking eff., Sec 11.1.4), $K\pi$ dependence of the reconstruction efficiency (rec. eff. $K\pi$, Sec. 11.1.5), angular resolution (ang. resolution, Sec.11.2), peaking background (peaking bkg., Sec. 11.3), neglecting of B_s^0 component (neglecting B_s^0, Sec. 11.4), $K\pi\mu\mu$ signal mass model (sig. mass model, Sec. 11.5), background angular model (bkg. ang. model, Sec. 11.6), S-wave model of $K\pi$ invariant mass (S-wave model, Sec. 11.7.1), background model of $K\pi$ invariant mass (bkg. $K\pi$ model, Sec. 11.7.2). The values are coloured according to: < 0.0001 $0.0001 - 0.0002$ $0.0002 - 0.0005$ $0.0005 - 0.003$ $0.003 - 0.005$ $0.005 - 0.008$ $0.008 - 0.01$ $0.01 - 0.015$ > 0.015 .

Syst. Uncer.	S_{1s}	S_3	S_4	S_5	S_{6s}	S_7	S_8	S_9
stat. ang. acc.	0.0009	0.0000	0.0010	0.0001	0.0012	0.0001	0.0015	0.0001
stat. sim.	0.0072	0.0124	0.0140	0.0148	0.0117	0.0164	0.0159	0.0124
weight π p_T	0.0061	0.0014	0.0012	0.0004	0.0006	0.0002	0.0012	0.0001
weight K p_T	0.0005	0.0002	0.0004	0.0004	0.0001	0.0001	0.0004	0.0000
weight μ p_T	0.0002	0.0000	0.0002	0.0001	0.0001	0.0000	0.0000	0.0000
weight #tracks	0.0000	0.0001	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000
weight B^0 vtx	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
weight q^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
tracking eff.	0.0005	0.0001	0.0002	0.0001	0.0000	0.0001	0.0001	0.0000
rec. eff. $K\pi$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ang. resolution	0.0000	0.0011	0.0019	0.0015	0.0000	0.0023	0.0007	0.0002
peaking bkg.	0.0005	0.0002	0.0003	0.0008	0.0004	0.0001	0.0001	0.0000
neglecting B_s^0	0.0005	0.0001	0.0003	0.0004	0.0000	0.0001	0.0000	0.0001
sig. mass model	0.0004	0.0003	0.0007	0.0010	0.0001	0.0002	0.0001	0.0003
bkg. ang. model	0.0006	0.0001	0.0008	0.0015	0.0003	0.0003	0.0000	0.0010
S-wave model	0.0001	0.0000	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000
bkg. $K\pi$ model	0.0008	0.0000	0.0006	0.0012	0.0003	0.0001	0.0000	0.0000
quadratic sum	0.0096	0.0125	0.0142	0.0151	0.0118	0.0166	0.0161	0.0125
exp. stat. uncer.	0.0412	0.0730	0.0865	0.0823	0.0635	0.0769	0.0838	0.0681

12 Results

12.1 Results of the angular observables

The angular observables related to the normalized differential branching fraction of the decay $B^0 \to K^{*0}\mu^+\mu^-$ (see Eq. 37 in Sec. 4.3.6) are determined. In the measurement of the *CP*-averaged observables, S_i, (see Eq. 12) it is assumed that the *CP*-asymmetric observables, A_i, (see Eq. 13) are negligibly small. The nominal result is obtained with a 5D maximum likelihood fit as discussed in Sec. 8.1. In comparison to the Method of Moments (see Sec. 8.2) it has in general a better significance of about 10% (see Sec. 8.6). The dominating systematic uncertainty is due to the correction for the non-uniform reconstruction and selection efficiency (see Sec. 8.3). This uncertainty is, however, small compared to the statistical uncertainty (see Sec. 11).

The 1D projections of the maximum likelihood fit are shown in Fig. 47 to Fig. 54. The results of the angular observables are listed in Tab. 11. The results are also visualized in Fig. 55, where they are compared to the results of the Method of Moments and the predictions from the Standard Model taken from Ref. [50]. The 68% confidence intervals of the maximum likelihood fit include the systematic uncertainties, whereas for the Method of Moments only the statistical uncertainties are shown. The systematic uncertainties are expected to be mostly independent of the estimators. Within the expected deviations due to statistical fluctuations the result of the maximum likelihood fit and the Method of Moments agree very well.

The p-values of the goodness of fit test of the maximum likelihood fit, based on the strategy explained in Sec. 9.4, are listed in Tab. 12. None of the p-values indicates a problem in one of the fits. In the ideal case the p-values are evenly distributed between zero and one. There is a slight tendency to larger p-values (i.e. the smallest of the 24 p-values is 0.19) indicating a small overfitting of the 24 floating parameters in the fit. However, it is shown that using the Bootstrap method the resulting confidence levels are correct (see Sec. 9.2).

The correlation matrices of the observables and the 2D profile log likelihood scans from which these correlations are calculated (see Sec. 10) are shown in Appendix B. In the 2D likelihood plots it can be seen that all fits have converged, although in certain cases the best fit point is closer than 2σ to the physical boundaries. Comparing the measured correlation with the ones intrinsically in the PDF (see appendix. A.2) it can be seen that most of the correlations which are expected are also

present in the measurement. Additional there are also larger correlations which are related to physical boundaries and to the statistical fluctuations in the data sample. The correlations are up to 0.5, thus when comparing the full set of observables to predictions these correlations must be included in the analysis.

The measured statistical uncertainty agrees in general very well with the expected statistical uncertainty (compare Tab. 11 with Fig. 36 in Sec. 8.6). In the second q^2 bin the measured uncertainty of several observables is slightly worse than the mean expected uncertainty, whereas in the third q^2 bin it is slightly better. It is checked with pseudo-experiments of the high statistics simulation that in at least 5% of the cases such deviations are obtained in simulation. Therefore the maximum likelihood fit behaves well within expectations.

In this analysis a fit to the complete set of eight *CP*-averaged observables is done for the first time, providing also the correlation between the different observables. The quantities F_L and A_{FB}, which are related to S_{1s} and S_{6s} respectively (see Eq. 14), were measured by several experiments. The analyses done by BaBar [104], Belle [105] and CDF [106] are using the one dimensional differential decay rates in $\cos\theta_K$ and/or $\cos\theta_l$. In the analysis done by CMS [107] the 2D differential decay rate in $\cos\theta_K$ and $\cos\theta_l$ is fitted. Whereas CMS and CDF are only looking at the decay $B^0 \to K^{*0}\mu^+\mu^-$, BaBar and Belle were combining several leptonic decays $B^0 \to K^*\ell^+\ell^-$. In a previous analysis by LHCb, using $1\,\text{fb}^{-1}$ of data from 2011 [43], the three dimensional differential decay rate is folded, such that simultaneously the observables S_{1s}, S_3, S_{6s} and S_9 are determined. The data set used in that analysis is included also in the current analysis, providing about one third of the total statistics. A direct comparison of the results is due different q^2 binning schemes not possible. Nevertheless, the general shapes agree between all measurements. The different results are shown in Fig. 56.

Comparing the results of this analysis to the Standard Model expectation in general there is also a good agreement. Smaller deviations are visible in the observables S_5 and S_{6s}. In the q^2 region $1.1 - 6.0\,\text{GeV}^2$ the Standard Model prediction is lower in S_5 and larger in S_{6s} by $1 - 2\,\sigma$ in three consecutive q^2 bins. In a previous measurement at LHCb a deviation compared to the Standard Model prediction of an observable called P_5', which is related to S_5, is observed. A detailed discussion and comparison will be done in the next section. The deviations in S_{6s} are consistent with the previous measurement at LHCb.

12.2 Result of an angular observable in a second angular basis

The full information of the *CP*-averaged normalized angular distribution is included inside the measured angular observables S_i and their correlation as described in the last section. Nevertheless, for a comparison to the result obtained in Ref. [5], where a significant deviation to the Standard Model prediction is observed, also a measurement of P_5' is performed. The observable P_5' relates to S_5 and S_{1s} in the following way:

$$P_5' = S_5 / \sqrt{4/3 S_{1s}(1 - 4/3 S_{1s})} \tag{85}$$

This observable could be calculated from the results of the last section. Instead the $P^{(\prime)}$ observables are directly implemented in the PDF of the 5D maximum likelihood fit and the measurement is redone. Besides a better handle of the correlations also it is possible to re-evaluate precisely the systematic uncertainties. Systematic uncertainties are determined in a similar way as explained in Sec.11.

The results compared to the binned Standard Model prediction from [53] are shown in Fig. 57 and listed in Tab. 13. In the table also the deviations of the measurement to the Standard Model expectation are listed for each q^2 bin.

In the figure the result is compared to the previous analysis done at LHCb [5]. One third of the data set ($1\,\text{fb}^{-1}$ data from 2011 data taking at LHCb) is in common with the current analysis. In the old analysis due to historical reasons a different q^2 binning is chosen. The results between both measurements agree and a deviation compared to the Standard Model prediction in the q^2 range $4.3 - 8.68\,\text{GeV}^2$ is visible. Nevertheless, the absolute value of the difference got smaller. In the last analysis in this bin a deviation from the Standard Model prediction of $3.7\,\sigma$ is seen, which is of similar size if combining the observed deviations of the current measurement.

The agreement of the Standard Model prediction and the measurement is quantified by combining the squared observed deviations of the first five q^2 bins in a χ^2 test. The p-value that the observed deviations are a statistical fluctuation is $8.2 \cdot 10^{-4}$. Thus it is highly unlikely that the deviations in the lower q^2 bins are purely statistical in nature, however, to claim a discovery a new more precise measurement with more statistics is necessary. The observed deviations are compatible with predictions of several New Physics models (see Sec. 4.4).

Table 11: The results of the maximum likelihood fit. The first uncertainties are statistical and the second systematic. Due to low statistics and the physical boundaries the 68% confidence interval is in few cases highly asymmetric around the best fit point.

$q^2[GeV^2]$	S_{1s}	S_3	S_4
$0.1-1.0$	$0.521^{+0.035}_{-0.044}\pm0.011$	$0.011^{+0.062}_{-0.063}\pm0.017$	$-0.014^{+0.070}_{-0.083}\pm0.016$
$1.1-2.5$	$0.176^{+0.098}_{-0.056}\pm0.012$	$-0.021^{+0.171}_{-0.152}\pm0.014$	$-0.158^{+0.120}_{-0.176}\pm0.019$
$2.5-4.0$	$0.178^{+0.078}_{-0.046}\pm0.010$	$0.033^{+0.079}_{-0.094}\pm0.014$	$-0.116^{+0.139}_{-0.126}\pm0.018$
$4.0-6.0$	$0.281^{+0.041}_{-0.040}\pm0.010$	$0.058^{+0.070}_{-0.084}\pm0.013$	$-0.229^{+0.099}_{-0.074}\pm0.014$
$6.0-8.0$	$0.346^{+0.037}_{-0.036}\pm0.010$	$-0.054^{+0.061}_{-0.062}\pm0.011$	$-0.288^{+0.081}_{-0.074}\pm0.011$
$11.0-12.5$	$0.384^{+0.047}_{-0.033}\pm0.009$	$-0.151^{+0.093}_{-0.057}\pm0.012$	$-0.271^{+0.067}_{-0.090}\pm0.009$
$15.0-17.0$	$0.483^{+0.026}_{-0.030}\pm0.007$	$-0.189^{+0.055}_{-0.039}\pm0.012$	$-0.297^{+0.054}_{-0.047}\pm0.010$
$17.0-19.0$	$0.497^{+0.034}_{-0.039}\pm0.010$	$-0.225^{+0.077}_{-0.067}\pm0.015$	$-0.282^{+0.073}_{-0.060}\pm0.015$

$q^2[GeV^2]$	S_5	S_{6s}	S_7
$0.1-1.0$	$0.148^{+0.061}_{-0.066}\pm0.015$	$-0.098^{+0.085}_{-0.078}\pm0.017$	$0.005^{+0.056}_{-0.065}\pm0.016$
$1.1-2.5$	$0.192^{+0.129}_{-0.125}\pm0.021$	$-0.392^{+0.171}_{-0.081}\pm0.015$	$-0.340^{+0.148}_{-0.147}\pm0.021$
$2.5-4.0$	$-0.023^{+0.114}_{-0.108}\pm0.020$	$-0.164^{+0.130}_{-0.092}\pm0.013$	$0.050^{+0.114}_{-0.116}\pm0.020$
$4.0-6.0$	$-0.156^{+0.082}_{-0.082}\pm0.015$	$0.062^{+0.063}_{-0.066}\pm0.012$	$-0.082^{+0.082}_{-0.091}\pm0.017$
$6.0-8.0$	$-0.288^{+0.080}_{-0.060}\pm0.012$	$0.164^{+0.059}_{-0.054}\pm0.013$	$-0.032^{+0.073}_{-0.067}\pm0.015$
$11.0-12.5$	$-0.334^{+0.079}_{-0.059}\pm0.010$	$0.409^{+0.059}_{-0.062}\pm0.012$	$-0.112^{+0.075}_{-0.078}\pm0.014$
$15.0-17.0$	$-0.305^{+0.053}_{-0.047}\pm0.010$	$0.511^{+0.040}_{-0.053}\pm0.011$	$0.064^{+0.053}_{-0.061}\pm0.014$
$17.0-19.0$	$-0.329^{+0.068}_{-0.060}\pm0.014$	$0.420^{+0.057}_{-0.063}\pm0.016$	$-0.013^{+0.078}_{-0.081}\pm0.019$

$q^2[GeV^2]$	S_8	S_9
$0.1-1.0$	$0.107^{+0.079}_{-0.079}\pm0.017$	$-0.103^{+0.062}_{-0.062}\pm0.017$
$1.1-2.5$	$-0.210^{+0.182}_{-0.141}\pm0.019$	$-0.292^{+0.166}_{-0.125}\pm0.015$
$2.5-4.0$	$0.060^{+0.156}_{-0.116}\pm0.019$	$-0.057^{+0.106}_{-0.104}\pm0.012$
$4.0-6.0$	$0.181^{+0.092}_{-0.095}\pm0.016$	$0.030^{+0.079}_{-0.073}\pm0.012$
$6.0-8.0$	$-0.103^{+0.074}_{-0.068}\pm0.014$	$0.001^{+0.072}_{-0.069}\pm0.012$
$11.0-12.5$	$-0.002^{+0.069}_{-0.082}\pm0.014$	$-0.065^{+0.070}_{-0.070}\pm0.011$
$15.0-17.0$	$0.038^{+0.058}_{-0.063}\pm0.014$	$-0.026^{+0.055}_{-0.053}\pm0.012$
$17.0-19.0$	$0.103^{+0.084}_{-0.076}\pm0.019$	$-0.114^{+0.073}_{-0.060}\pm0.017$

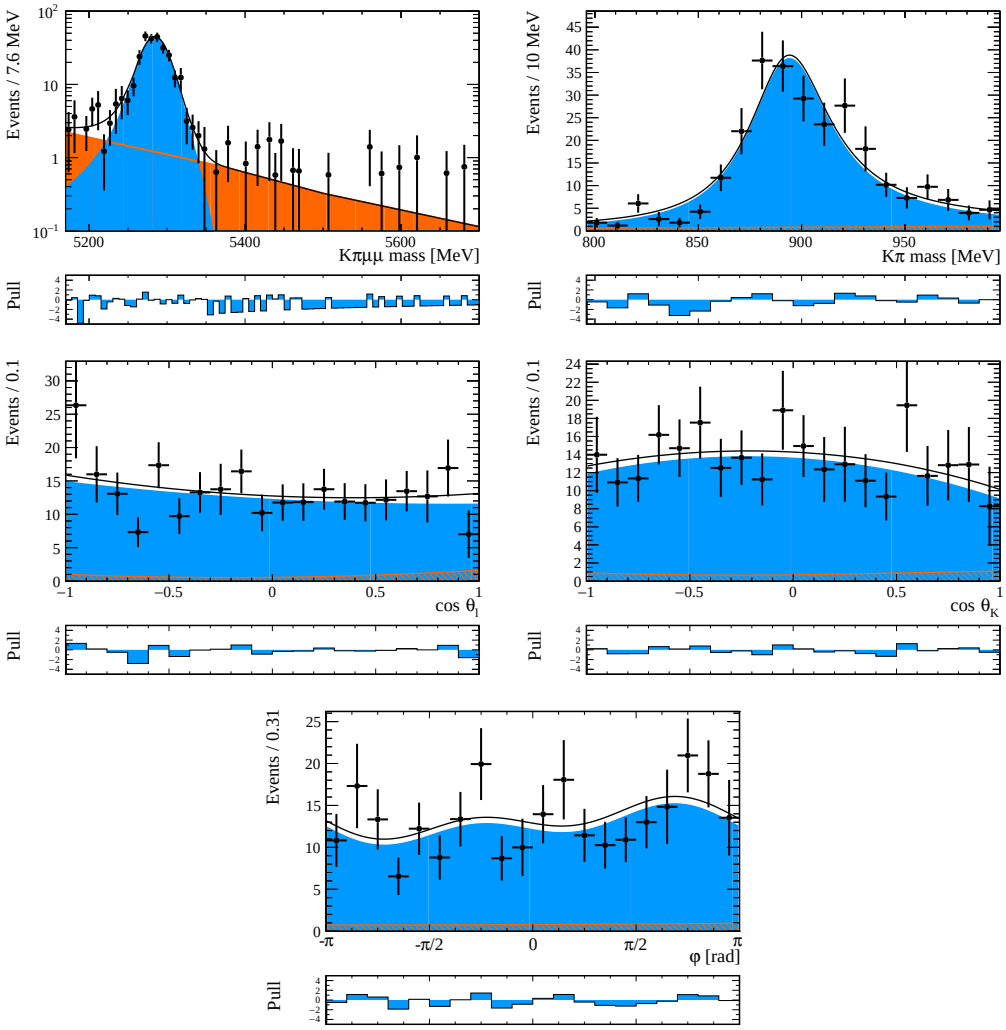

Figure 47: Projections of the 5D maximum likelihood fit of $B^0 \to K^{*0}\mu^+\mu^-$ in the q^2 bin $0.1 - 0.98\,\mathrm{GeV}^2$. Besides for the $K\pi\mu\mu$ invariant mass, projections are shown for the signal region $\pm 50\,\mathrm{MeV}$ around the signal peak.

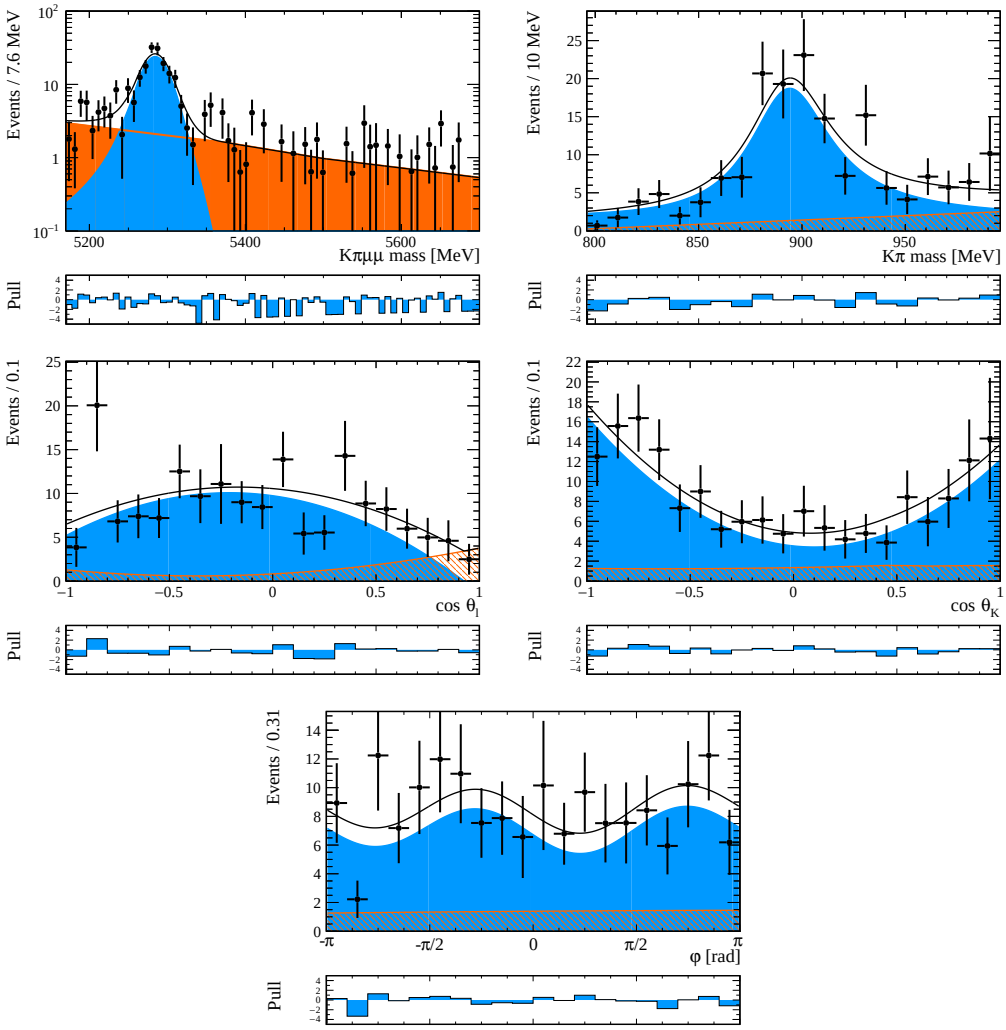

Figure 48: Projections of the 5D maximum likelihood fit of $B^0 \to K^{*0}\mu^+\mu^-$ in the q^2 bin $1.1 - 2.5\,\mathrm{GeV}^2$. Besides for the $K\pi\mu\mu$ invariant mass, projections are shown for the signal region $\pm 50\,\mathrm{MeV}$ around the signal peak.

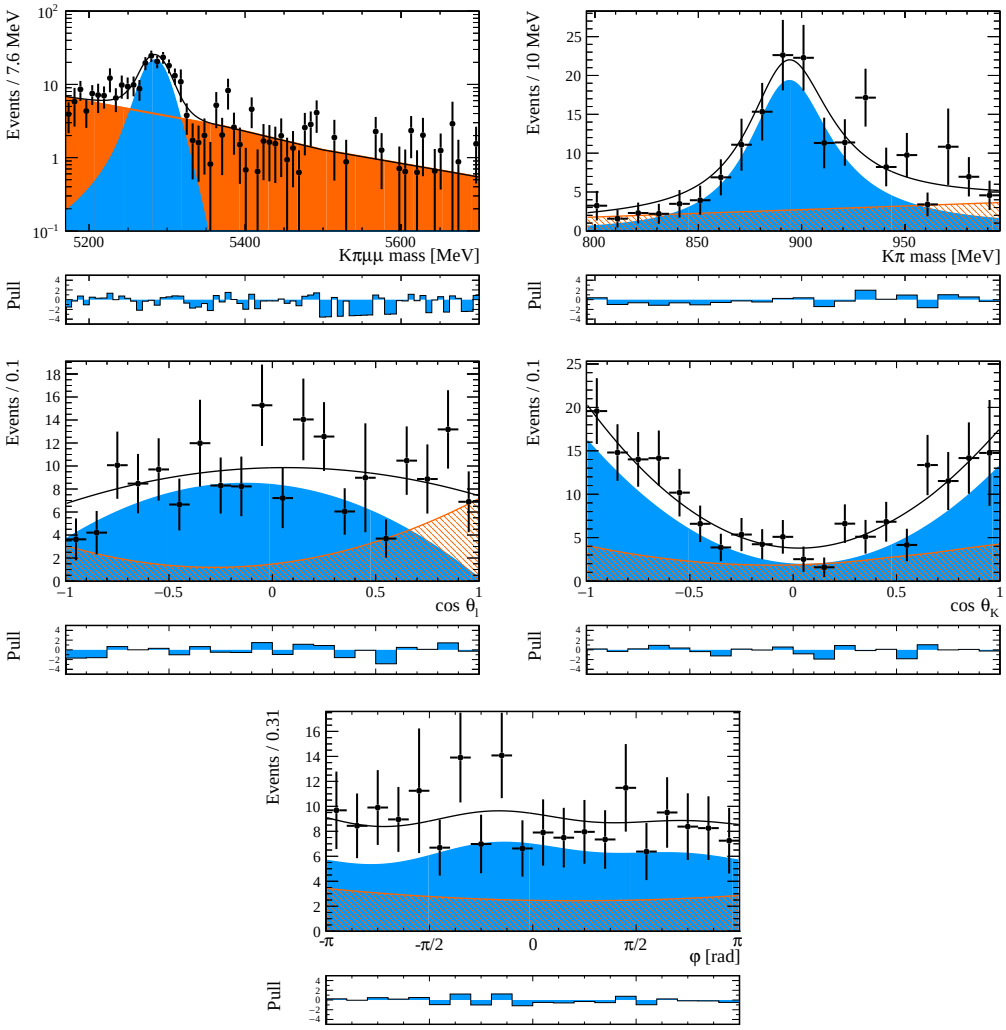

Figure 49: Projections of the 5D maximum likelihood fit of $B^0 \to K^{*0}\mu^+\mu^-$ in the q^2 bin $2.0 - 4.0\,\mathrm{GeV}^2$. Besides for the $K\pi\mu\mu$ invariant mass, projections are shown for the signal region $\pm 50\,\mathrm{MeV}$ around the signal peak.

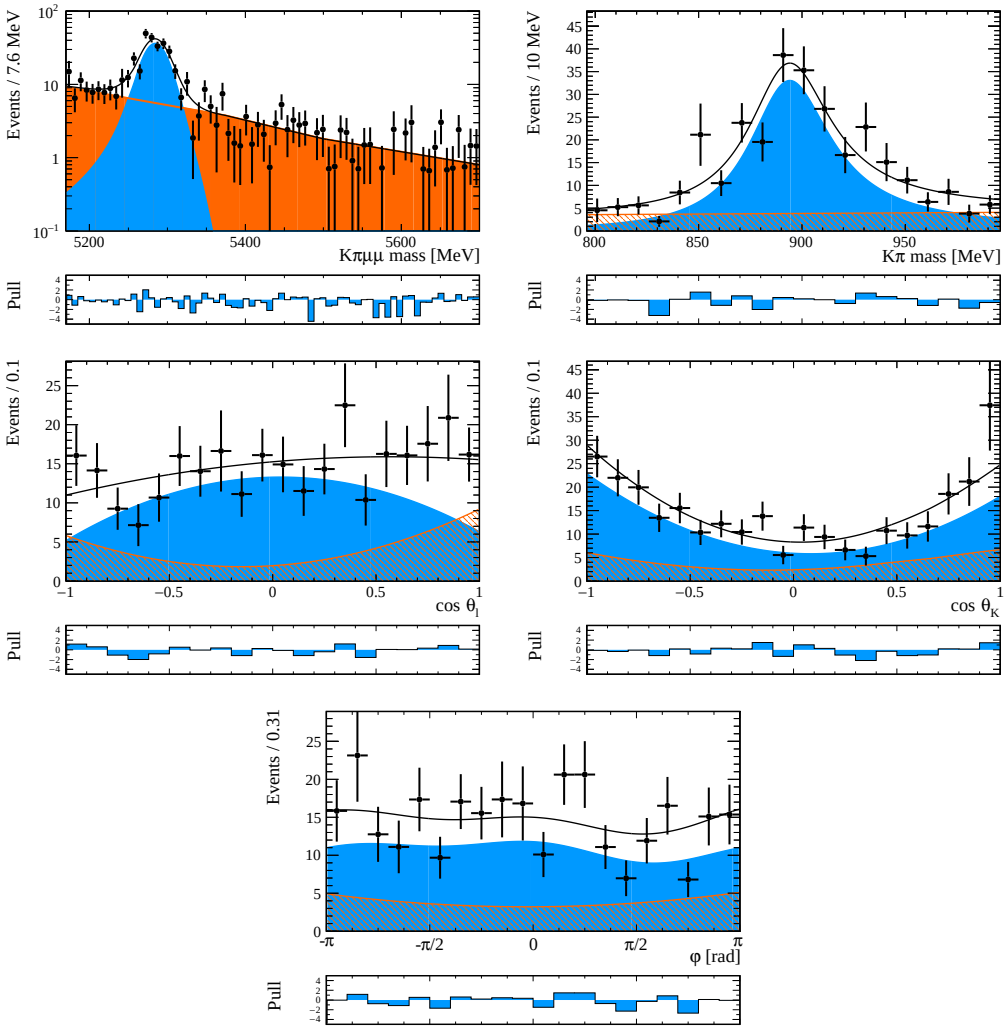

Figure 50: Projections of the 5D maximum likelihood fit of $B^0 \to K^{*0}\mu^+\mu^-$ in the q^2 bin $4.0 - 6.0\,\mathrm{GeV}^2$. Besides for the $K\pi\mu\mu$ invariant mass, projections are shown for the signal region $\pm 50\,\mathrm{MeV}$ around the signal peak.

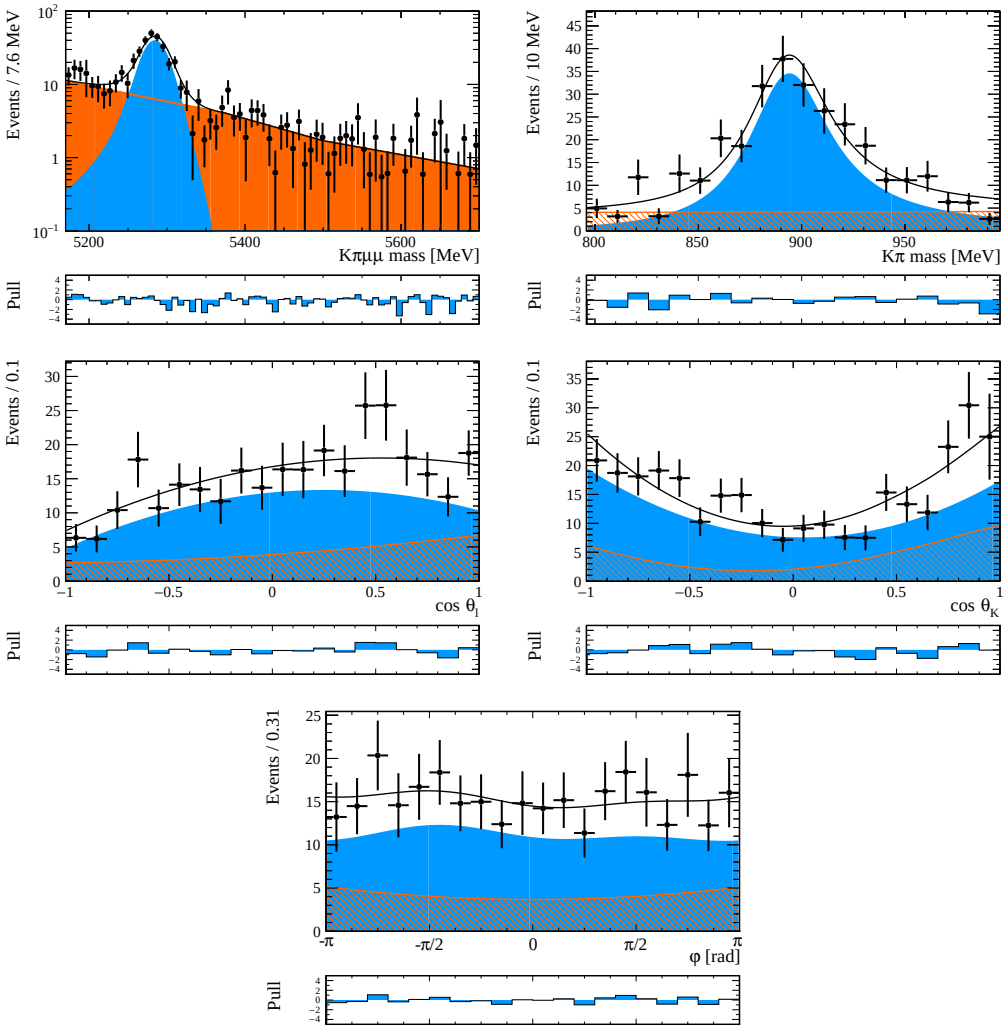

Figure 51: Projections of the 5D maximum likelihood fit of $B^0 \to K^{*0}\mu^+\mu^-$ in the q^2 bin $6.0 - 8.0\,\mathrm{GeV}^2$. Besides for the $K\pi\mu\mu$ invariant mass, projections are shown for the signal region $\pm 50\,\mathrm{MeV}$ around the signal peak.

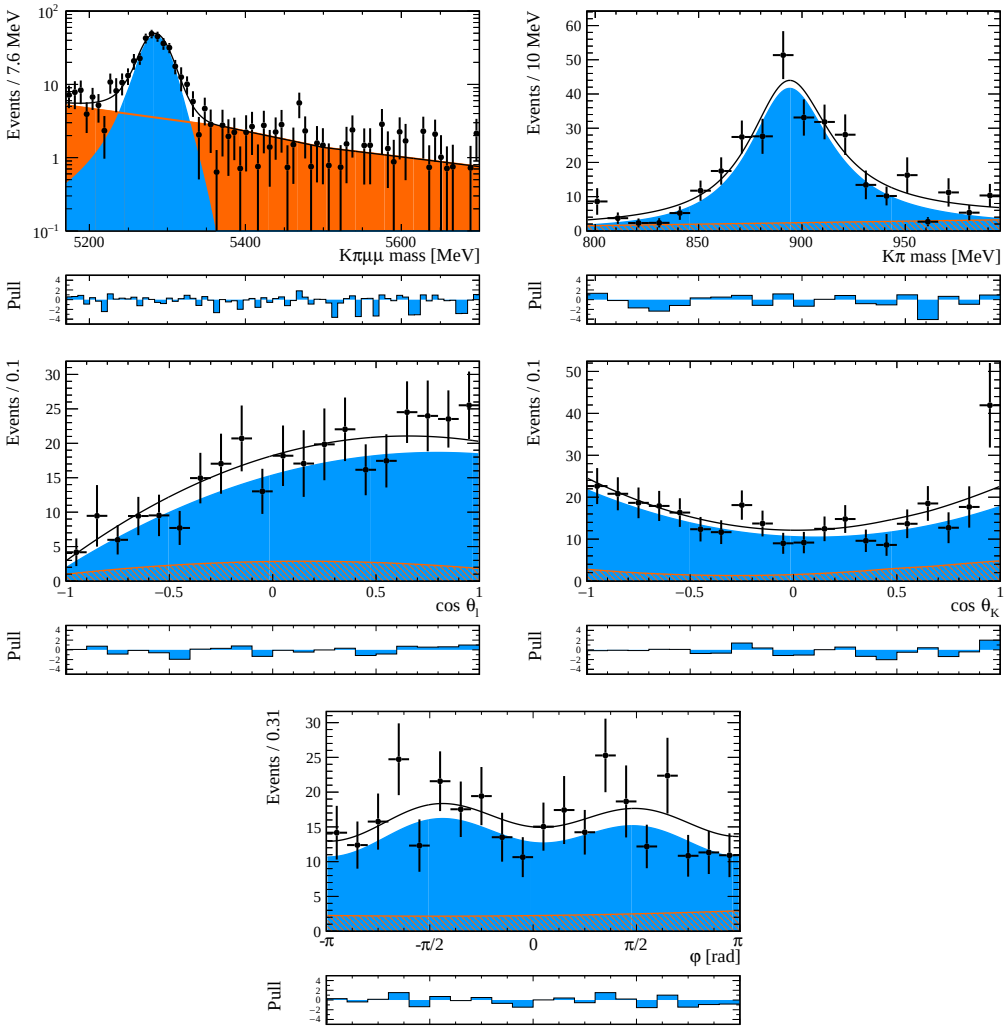

Figure 52: Projections of the 5D maximum likelihood fit of $B^0 \to K^{*0}\mu^+\mu^-$ in the q^2 bin $11.0 - 12.5\,\mathrm{GeV}^2$. Besides for the $K\pi\mu\mu$ invariant mass, projections are shown for the signal region $\pm 50\,\mathrm{MeV}$ around the signal peak.

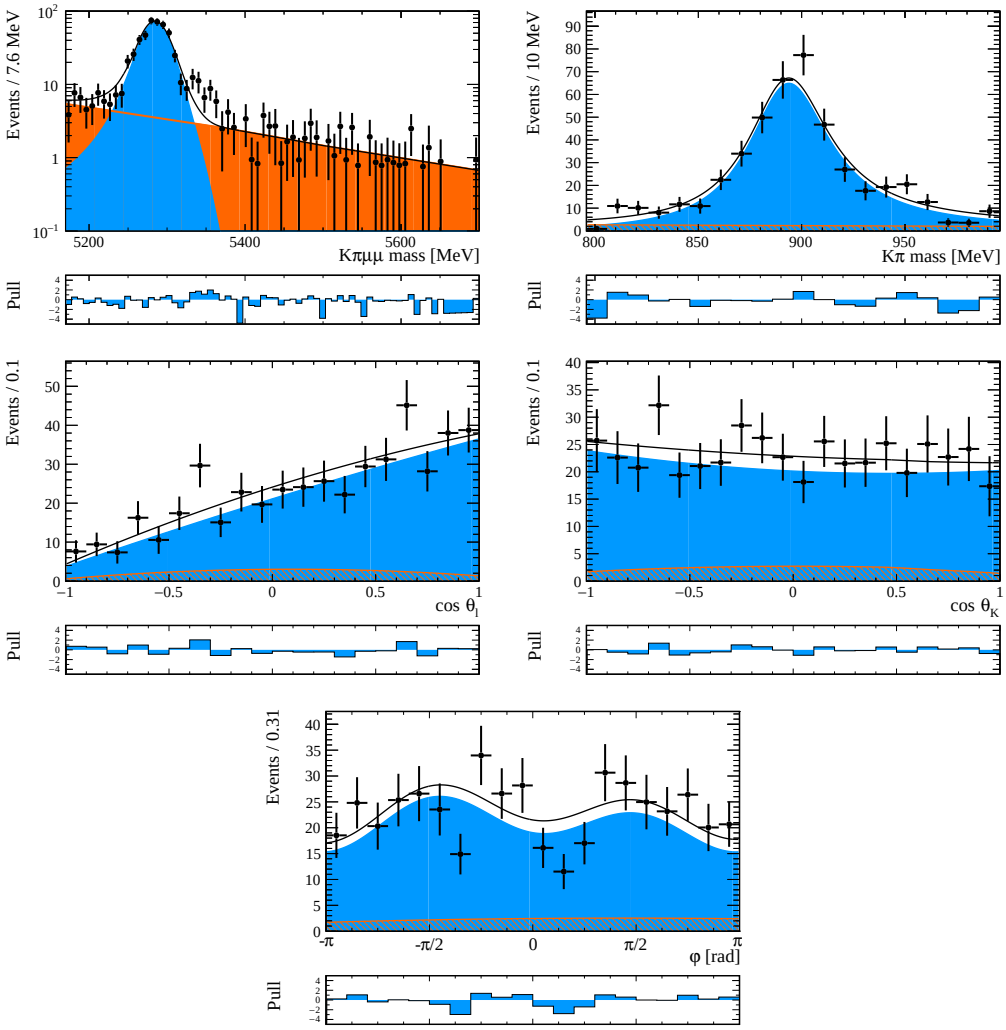

Figure 53: Projections of the 5D maximum likelihood fit of $B^0 \to K^{*0}\mu^+\mu^-$ in the q^2 bin $15.0 - 17.0\,\mathrm{GeV}^2$. Besides for the $K\pi\mu\mu$ invariant mass, projections are shown for the signal region $\pm 50\,\mathrm{MeV}$ around the signal peak.

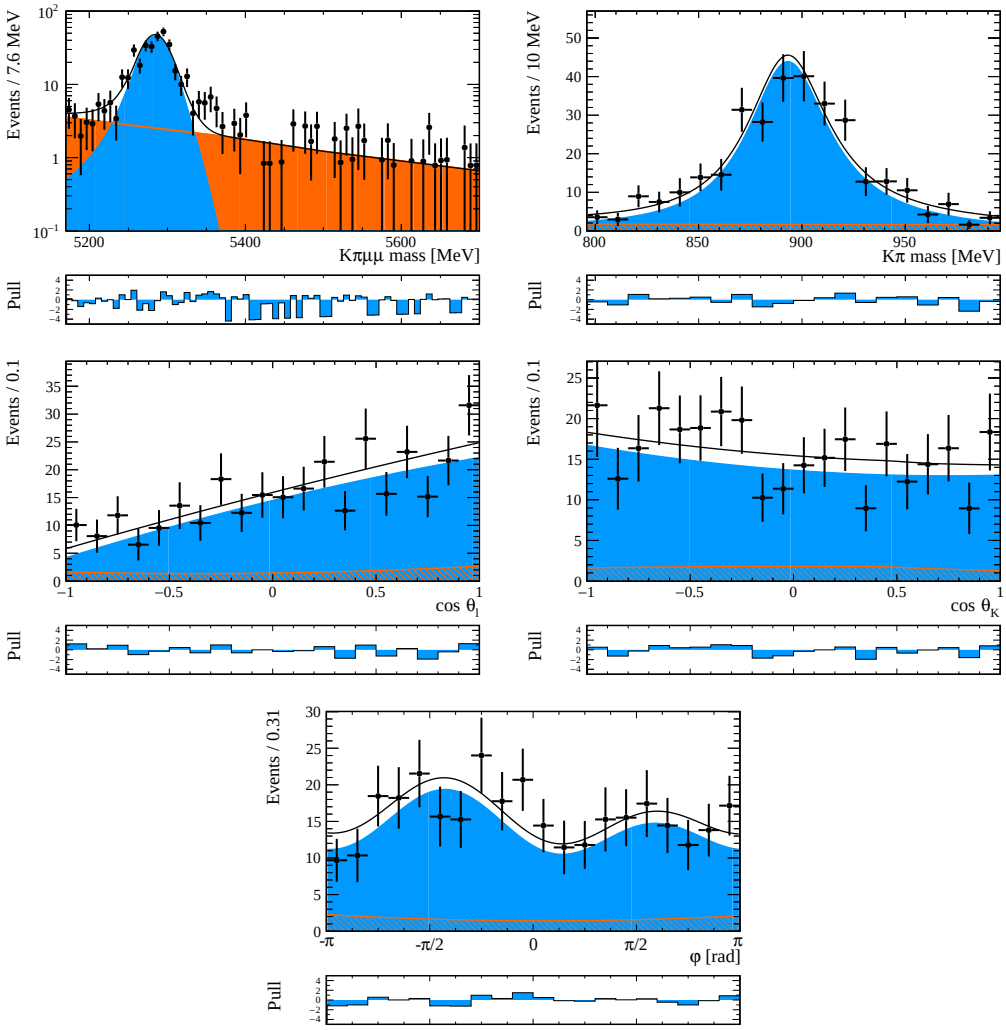

Figure 54: Projections of the 5D maximum likelihood fit of $B^0 \to K^{*0}\mu^+\mu^-$ in the q^2 bin $17.0 - 19.0\,\mathrm{GeV}^2$. Besides for the $K\pi\mu\mu$ invariant mass, projections are shown for the signal region $\pm 50\,\mathrm{MeV}$ around the signal peak.

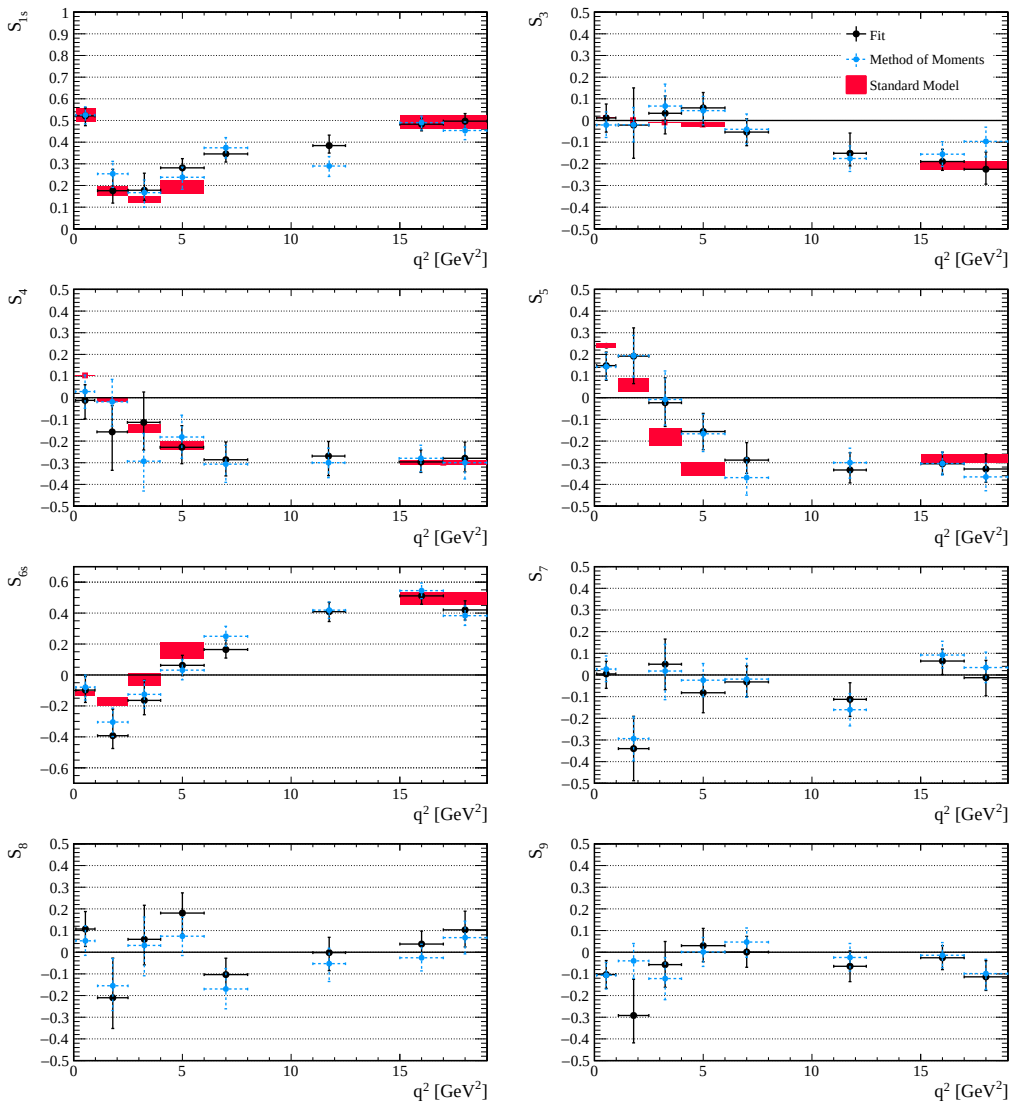

Figure 55: The results of the maximum likelihood fit (black), the Method of Moments (blue) compared to the Standard Model prediction [50] (red). The observables $S_7 - S_9$ are expected to be zero in the Standard Model.

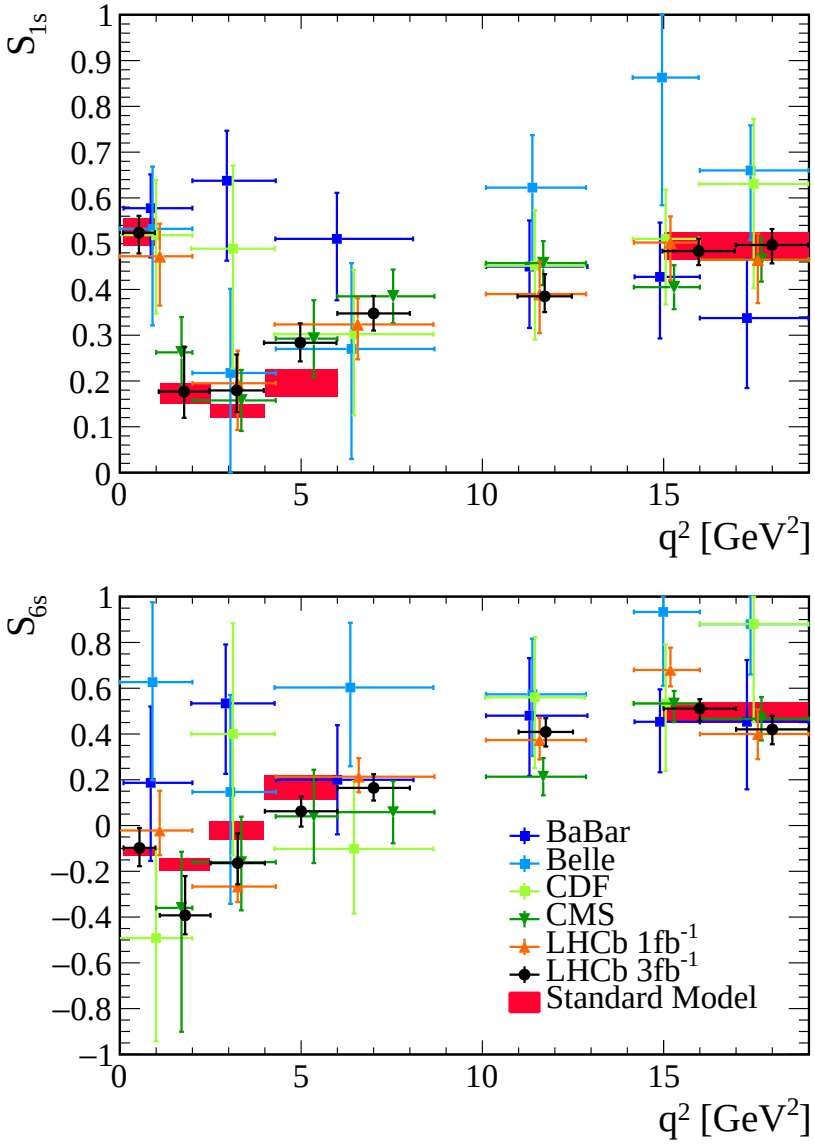

Figure 56: The results of the maximum likelihood fit (black circles) compared to previous analyses of S_{1s} and S_{6s}.

Table 12: The p-values for determining the goodness of fit as explained in Sec. 9.4.

$q^2[GeV^2]$	$K\pi\mu\mu$ mass	$K\pi$ mass	angles
$0.1 - 0.98$	0.68	0.38	0.92
$1.1 - 2.5$	0.74	0.68	0.41
$2.5 - 4.0$	0.89	0.24	0.46
$4.0 - 6.0$	0.85	0.39	0.48
$6.0 - 8.0$	0.36	0.21	0.21
$11.0 - 12.5$	0.59	0.44	0.91
$15.0 - 17.0$	0.21	0.51	0.78
$17.0 - 19.0$	0.19	0.70	0.56

Table 13: The results of the measurement of P_5' (first column). The first uncertainties are statistical and the second systematic. The result is compared to the Standard Model prediction from Ref. [53] and the deviation with respect to the estimated uncertainties is calculated.

$q^2[GeV^2]$	fit	Standard Model	Deviation $[\sigma]$
$0.1 - 1.0$	$0.315^{+0.132}_{-0.157} \pm 0.037$	$0.678^{+0.033}_{-0.040}$	-2.5
$1.1 - 2.5$	$0.460^{+0.318}_{-0.261} \pm 0.052$	$0.170^{+0.096}_{-0.117}$	1.0
$2.5 - 4.0$	$-0.033^{+0.303}_{-0.287} \pm 0.058$	$-0.492^{+0.104}_{-0.118}$	1.5
$4.0 - 6.0$	$-0.319^{+0.171}_{-0.177} \pm 0.037$	$-0.789^{+0.066}_{-0.081}$	2.4
$6.0 - 8.0$	$-0.576^{+0.162}_{-0.120} \pm 0.026$	$-0.882^{+0.049}_{-0.059}$	2.3
$11.0 - 12.5$	$-0.666^{+0.163}_{-0.121} \pm 0.023$	-	-
$15.0 - 17.0$	$-0.637^{+0.118}_{-0.093} \pm 0.020$	-	-
$17.0 - 19.0$	$-0.692^{+0.147}_{-0.127} \pm 0.030$	-	-

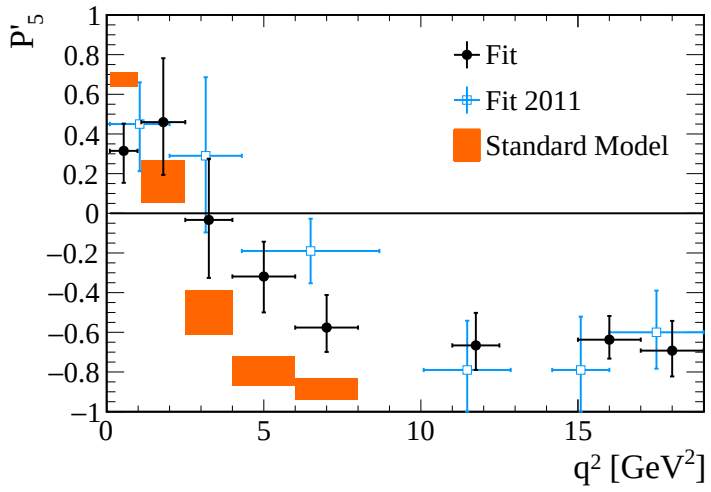

Figure 57: The results of the maximum likelihood fit (black), the result of a previous measurements (blue) compared to the Standard Model prediction [53] (orange).

13 Conclusion

Using the full Run 1 data set with an integrated luminosity of $3\,\text{fb}^{-1}$ from the LHCb experiment, angular observables of the decay $B^0 \to K^{*0}\mu^+\mu^-$ are determined in bins of the dimuon invariant mass squared, q^2. For the first time the complete set of *CP*-averaged angular observables is determined simultaneously in a 5D maximum likelihood fit. The variables which are used are the $K\pi\mu\mu$ invariant mass, the $K\pi$ invariant mass and the three decay angles θ_K, θ_l and ϕ. They fully constrain the decay $B^0 \to K^{*0}\mu^+\mu^-$ for a given value of q^2. The simultaneous determination of the observables allows to provide the complete correlation matrix of the observables. The measurement presented in this thesis is up to date the world best one. For the observables which can be compared, results are in good agreement with the measurements done at different experiments.

Mostly, results are in good agreement with the Standard Model predictions. For the observables S_5 and S_{6s} the measured values in the q^2 region $1-6\,\text{GeV}^2$ are slightly above and below the predictions, respectively. Additionally in a second angular basis the observable P_5' is measured. A local deviation from the Standard Model prediction observed in P_5' in a previous measurement is confirmed. Although the absolute value of the deviation got smaller the significance of the deviation stayed similar. In the q^2 region $0.1 - 8\,\text{GeV}^2$ the measurement of P_5' is compatible with the Standard Model prediction with only a p-value of $8.2 \cdot 10^{-4}$.

The observed deviations from the Standard Model prediction are consistent with different New Physics models. If all experimental and theoretical uncertainties are estimated correctly it seems very likely that a sign for New Physics has been observed. However, there is still a large debate in the theoretical community especially about the size of hadronic effects which, if underestimated, could result in a similar deviation.

Thus, in the future the validation of the theoretical uncertainties will be of outermost important. Due to the small event yield, the measurement of the angular observables of $B^0 \to K^{*0}\mu^+\mu^-$ is statistically limited. In the following years, further data will be taken at the LHCb experiment. Until the year 2030 the integrated luminosity of the recorded data is expected to increase by at least a factor of 10. Therefore, the precision of future measurements will significantly improve. Also additional measurements of different decays and different observables will be performed. Especially interesting is the comparison of angular observables of the decay $B^0 \to K^{*0}e^+e^-$ with the observables measured in this thesis. If lepton universality would be broken in $b - s$ transitions, as indicated by the measurement of R_K, this measurement could be highly sensitive to related effects.

Although at present time it cannot be said if New Physics has been observed, and if yes what exactly it would be, first signs have emerged that the understanding of the Standard Model is not complete. Already in the near future with further measurements and more precise theory calculations the reason for the observed deviations can be expected to be revealed.

References

[1] R. D. Stutzke *et al.*, *Search for the Decay $K_S{}^0 \to \mu^+\mu^-$*, Phys. Rev. **177** (1969) 2009.

[2] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, Phys. Rev. **D2** (1970) 1285.

[3] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B716 (2012) 1-29 (2012) , arXiv:1207.7214.

[4] CMS collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B 716 (2012) 30 (2013) , arXiv:1207.7235.

[5] LHCb collaboration *et al.*, *Measurement of form-factor independent observables in the decay $B^0 \to K^{*0}\mu^+\mu^-$*, Phys. Rev. Lett. 111, 191801 (2013) (2013) , arXiv:1308.1707.

[6] LHCb collaboration, *Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay*, LHCb-CONF-2015-002 (2015), https://cds.cern.ch/record/2002772.

[7] M. Thomson, *Modern Particle Physics*, Cambridge University Press, Oct., 2013.

[8] N. Tuning and P. Kooijman, *CP Violation*, www.nikhef.nl/~h71/Lectures/2012/cp-080212.pdf, Feb., 2012.

[9] C. Quigg, *Unanswered Questions in the Electroweak Theory*, Ann. Rev. Nucl. Part. Sci. **59** (2009) 505, arXiv:0905.3187.

[10] *Standard Model Plot*, https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg.

[11] *CKM Fitter*, http://ckmfitter.in2p3.fr/www/results/plots_eps15/ckm_res_eps15.html, 2015.

[12] (LHCb Collaboration), LHCb collaboration *et al.*, *Precision Measurement of CP Violation in $B_s^0 \to J/\psi K^+K^-$ decays*, Phys. Rev. Lett. **114** (2015) 041801.

[13] *HFAG 2014*, http://www.slac.stanford.edu/xorg/hfag/osc/fall_2014/.

[14] G. 't Hooft, *Naturalness, chiral symmetry, and spontaneous chiral symmetry breaking*, NATO Sci. Ser. B **59** (1980) 135, https://inspirehep.net/record/144074.

[15] J. J. Balmer, *Notiz über die Spectrallinien des Wasserstoffs*, Ann. Phys. , 261: 80–87 (1885), http://gallica.bnf.fr/ark:/12148/bpt6k15268j.image.f88.pagination.langEN.

[16] LHCb collaboration, *LHCb outreach material* , http://lhcb-public.web.cern.ch/lhcb-public/en/lhcb-outreach/multimedia/.

[17] ATLAS, G. Aad *et al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[18] CMS, R. Adolphi *et al.*, *The CMS experiment at the CERN LHC*, JINST **0803** (2008) S08004.

[19] ALICE, K. Aamodt *et al.*, *The ALICE experiment at the CERN LHC*, JINST **0803** (2008) S08002.

[20] E. Norrbin and T. Sjöstrand, *Production and Hadronization of Heavy Quarks*, Eur. Phys. J. C17:137-161,2000 (2000) , arXiv:hep-ph/0005110v1.

[21] LHCb, A. Alves *et al.*, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.

[22] LHCb collaboration *et al.*, *LHCb Detector Performance*, Int. J. Mod. Phys. A 30, 1530022 (2015) (2015) , arXiv:1412.6352.

[23] S. Borghi *et al.*, *First spatial alignment of the LHCb VELO and analysis of beam absorber collision data*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **618** (2010), no. 1–3 108.

[24] L. V. Group *et al.*, *Performance of the LHCb Vertex Locator*, 2014 JINST 9 P09007 (2014) , arXiv:1405.7808.

[25] L. O. T. group *et al.*, *Performance of the LHCb Outer Tracker*, JINST 9 (2014) P01002 (2014) , arXiv:1311.3893.

[26] J. Luisier, *Performance of LHCb Silicon Tracker Detector in the LHC*, Physics Procedia **37** (2012) 851, http://dx.doi.org/10.1016/j.phpro.2012.04.097.

[27] A. A. A. Jr *et al.*, *Performance of the LHCb muon system*, JINST 015P 1112 (2013) arXiv:1211.1346.

[28] M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. C 73 (2013) 2431 (2013) , arXiv:1211.6759.

[29] R. Aaij *et al.*, *The LHCb trigger and its performance in 2011*, Journal of Instrumentation **8** (2013), no. 04 P04022, http://arxiv.org/abs/1211.3055.

[30] J. Albrecht, V. V. Gligorov, G. Raven, and S. Tolk, *Performance of the LHCb High Level Trigger in 2012*, J. Phys. : Conf. Ser. 513 (2014) 012001 (2013) , arXiv:1310.8544.

[31] LHCb collaboration, *LHCb Event Display*, https://lbgroups.cern.ch/EventDisplay/index.html.

[32] F. Rademakers and R. Brun, *ROOT: An Object-oriented Data Analysis Framework*, Linux J. **1998** (1998), http://dl.acm.org/citation.cfm?id=327422.362112.

[33] LHCb collaboration, *The Moore Project*, http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/moore/.

[34] LHCb collaboration, *The Brunel Project*, http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/brunel/.

[35] LHCb collaboration, *The DaVinci Project*, http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/davinci/.

[36] W. D. Hulsbergen, *Decay Chain Fitting with a Kalman Filter*, Nucl. Instrum. Meth. A552 (2005) 566-575 (2005) , arXiv:physics/0503191v1.

[37] G. Buchalla, A. J. Buras, and M. E. Lautenbacher, *Weak Decays Beyond Leading Logarithms*, Rev. Mod. Phys. 68:1125-1144,1996 (1995) , arXiv:hep-ph/9512380v1.

[38] LHCb collaboration *et al.*, *Implications of LHCb measurements and future prospects*, EPJ C 73 (2013) 2373 (2013) , arXiv:1208.3355.

[39] CMS collaboration and LHCb collaboration, *Observation of the rare $B^0_s \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data*, Nature **522** (2015) 68.

[40] C. Elsasser, *Feynman Diagram Library*, http://www.physik.uzh.ch/~che/FeynDiag/.

[41] J. Drobnak, *Constraints on new physics from top quark decays at high precision*, PhD thesis, Oct., 2012, arXiv:1210.5051.

[42] W. Altmannshofer *et al.*, *Symmetries and Asymmetries of $B^0 \to K^{*0}\mu^+\mu^-$ Decays in the Standard Model and Beyond*, JHEP 0901:019,2009 (2011) , arXiv:0811.1214.

[43] LHCb collaboration *et al.*, *Differential branching fraction and angular analysis of the decay $B^0 \to K^{*0}\mu^+\mu^-$*, JHEP 1308 (2013) 131 (2013) , arXiv:1304.6325.

[44] T. Blake, T. Gershon, and G. Hiller, *Rare b hadron decays at the LHC*, Invited review for Annual Reviews of Nuclear and Particle Physics (2015) arXiv:1501.0330.

[45] M. Beneke, T. Feldmann, and D. Seidel, *Systematic approach to exclusive B V gamma decays*, Nucl. Phys. B612:25-58,2001 (2001) , arXiv:hep-ph/0106067v2.

[46] B. Grinstein and D. Pirjol, *Exclusive rare B K*e+e- decays at low recoil: controlling the long-distance effects*, Phys. Rev. D70 (2004) 114005 (2004) , arXiv:hep-ph/0404250v3.

[47] D. Straub, *Implications of $b \to s$ measurements for model-building*, 50th Rencontres de Moriond EW (2015), http://moriond.in2p3.fr.

[48] R. R. Horgan, Z. Liu, S. Meinel, and M. Wingate, *Rare B decays using lattice QCD form factors*, Proceedings of Science (LATTICE2014) 372 (2015) arXiv:1501.0036.

[49] A. Khodjamirian, T. Mannel, A. A. Pivovarov, and Y. M. Wang, *Charm-loop effect in $B \to K^{(*)}l^+l^-$ and $B \to K^*\gamma$*, JHEP 1009:089,2010 (2010) , arXiv:1006.4945.

[50] W. Altmannshofer and D. M. Straub, *State of new physics in $b-s$ transitions*, (2015) arXiv:1411.3161, http://arxiv.org/abs/1411.3161v2.

[51] F. Beaujean, C. Bobeth, and D. van Dyk, *Comprehensive Bayesian Analysis of Rare (Semi)leptonic and Radiative B Decays*, Eur. Phys. J. C 74 (2014) 2897 (2014) , arXiv:1310.2478.

[52] F. Mahmoudi, S. Neshatpour, and J. Virto, *$B^0 \to K^{*0}\mu^+\mu^-$ optimised observables in the MSSM*, Eur. Phys. J. C74 (2014) 2927 (2014) arXiv:1401.2145.

[53] S. Descotes-Genon, L. Hofer, J. Matias, and J. Virto, *On the impact of power corrections in the prediction of $B^0 \to K^{*0}\mu^+\mu^-$ observables*, JHEP12(2014)125 (2014) arXiv:1407.8526.

[54] S. Jäger and J. M. Camalich, *Reassessing the discovery potential of the $B \to K^*\ell^+\ell^-$ decays in the large-recoil region: SM challenges and BSM opportunities*, (2014) arXiv:1412.3183.

[55] D. Becirevic and A. Tayduganov, *Impact of $B \to K_0^*\ell^+\ell^-$ on the New Physics search in $B \to K^*\ell^+\ell^-$ decay*, Nuclear Physics B **868** (2013) 368, arXiv:1207.4004.

[56] BaBar collaboration, J. L. Ritchie *et al.*, *BABAR Results on $B \to X_s\gamma$*, Proceedings of CKM 2012 (2013) arXiv:1301.0836.

[57] LHCb collaboration, *Measurement of the direct CP asymmetry in the $B^0_d \to K^{*0}\gamma$ decay*, LHCb-CONF-2012-004 (2012), https://cds.cern.ch/record/1424352.

[58] BaBar collaboration, *Measurement of Branching Fractions and CP and Isospin Asymmetries in $B \to K^{*0}\gamma$ Decays*, Phys. Rev. Lett. 103:211802,2009 (2009) , arXiv:0906.2177.

[59] R. Aaij *et al.*, *Differential branching fractions and isospin asymmetries of $B \to K^{(*)}\mu^+\mu^-$ decays*, JHEP 06 (2014) 133 (2014) , arXiv:1403.8044.

[60] LHCb collaboration *et al.*, *Differential branching fraction and angular analysis of the decay $B^0_s \to \phi\mu^+\mu^-$*, JHEP 1307 (2013) 084 (2013) , arXiv:1305.2168.

[61] LHCb collaboration *et al.*, *Test of lepton universality using $B^+ \to K^+\ell^+\ell^-$ decays*, Phys. Rev. Lett. 113, 151601 (2014) (2014) , arXiv:1406.6482.

[62] CMS collaboration, *Search for anomalous t t-bar production in the highly-boosted all-hadronic final state*, JHEP 09 (2012) 029 (2013) , arXiv:1204.2488.

[63] ATLAS collaboration, *A search for high-mass resonances decaying to $\tau^+\tau^-$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, JHEP **07** (2015) arXiv:1502.0717.

[64] CMS collaboration, *Search for high-mass resonances decaying into tau-lepton pairs in pp collisions at sqrt(s) = 7 TeV*, Phys. Lett. B 716 (2012) 82 (2013) , arXiv:1206.1725.

[65] CMS collaboration, *Search for physics beyond the standard model in dilepton mass spectra in proton-proton collisions at sqrt(s) = 8 TeV*, JHEP 04 (2015) 025 (2015) arXiv:1412.6302.

[66] ATLAS Collaboration, *Search for high-mass dilepton resonances in pp collisions at sqrt(s) = 8 TeV with the ATLAS detector*, Phys. Rev. D. 90, 052005 (2014) (2014) , arXiv:1405.4123.

[67] CERN, *Worldwide LHC Computing Grid* , http://wlcg.web.cern.ch/.

[68] LHCb collaboration, *The Gauss project*, http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/gauss/.

[69] LHCb collaboration, *The Boole project*, http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/boole/.

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

[71] T. Sjöstrand, S. Mrenna, and P. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. 178:852-867,2008 (2007) , arXiv:0710.3820.

[72] GEANT4, S. Agostinelli *et al.*, *GEANT4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

[73] J. Allison *et al.*, *Geant4 developments and applications*, Nuclear Science, IEEE Transactions on (2006) 270.

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

[75] D. van Dyk, I. Toijala, and F. Beaujean, *EOS - A HEP Program for Flavour Observables*, http://project.het.physik.tu-dortmund.de/source/eos.

[76] C. Bobeth, G. Hiller, and D. van Dyk, *The Benefits of $B^0 \to K^* \ell^+ \ell^-$ Decays at Low Recoil*, JHEP 1007:098,2010 (2010) , arXiv:1006.5013.

[77] Y. Freund and R. E. Schapire, *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting.*, J. Comput. Syst. Sci. **55** (1997), no. 1 119, http://dx.doi.org/10.1006/jcss.1997.1504.

[78] A. Hoecker *et al.*, *TMVA - Toolkit for Multivariate Data Analysis*, PoS ACAT:040,2007 (2009) arXiv:physics/0703039v5.

[79] A. Blum, A. Kalai, and J. Langford, *Beating the Hold-out: Bounds for K-fold and Progressive Cross-validation*, in *Proceedings of the Twelfth Annual Conference on Computational Learning Theory*, COLT '99, (New York, NY, USA), pp. 203–208, ACM, 1999. doi: 10.1145/307400.307439.

[80] Particle Data Group, K. Olive *et al.*, *Review of Particle Physics 2014*, Chin. Phys. **C38** (2014) 090001.

[81] S. Kotz and N. L. Johnson, *Encyclopedia of Statistical Sciences Volume 5*, pp. 340–350. Wiley, New York, NY, USA, 1985.

[82] NIST/SEMATECH, *e-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook/, 2015.

[83] S. M. Stigler, *The Epic Story of Maximum Likelihood*, Statistical Science 2007, Vol. 22, No. 4, 598-620 (2008) , arXiv:0804.2996.

[84] F. James and M. Winkler, *MINUIT User's Guide*, www.cern.ch/minuit, 2004.

[85] J. E. Gaiser, *Appendix-F Charmonium Spectroscopy from Radiative Decays of the J/Psi and Psi-Prime*, PhD thesis, Stanford, California, Aug., 1982, http://inspirehep.net/record/183554.

[86] LHCb collaboration *et al.*, *Measurement of the $B^0 \to J/\psi K^{*0}$ branching fraction and angular amplitudes*, Phys. Rev. D 86, 071102(R) (2012) (2012) , arXiv:1208.0738.

[87] J. M. Blatt and V. F. Weisskopf, *Theoretical Nuclear Physics*, Springer-Verlag, 1952.

[88] Belle collaboration *et al.*, *Experimental constraints on the spin and parity of the Z(4430)+*, Phys. Rev. D 88, 074026 (2013) (2013) , arXiv:1306.4894.

[89] S. Kotz and N. L. Johnson, *Encyclopedia of Statistical Sciences Volume 5*, pp. 467–472. Wiley, New York, NY, USA, 1985.

[90] K. Pearson, *Contributions to the Mathematical Theory of Evolution*, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **185** (1894) 71.

[91] M. Pivk and F. R. L. Diberder, *sPlot: a statistical tool to unfold data distributions*, Nucl. Instrum. Meth. A555:356-369,2005 (2005) , arXiv:physics/0402083v3.

[92] M. Pivk, *sPlot: A Quick Introduction*, Contributed to PHYSTAT 05 - Oxford 12th (2006) arXiv:physics/0602023v1.

[93] M. P. Wand and M. C. Jones, *Kernel smoothing*, vol. 60, Crc Press, 1994.

[94] V. I. Morariu *et al.*, *Automatic online tuning for fast Gaussian summation*, in *Advances in Neural Information Processing Systems 21* (D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, eds.), pp. 1113–1120. Curran Associates, Inc., 2009. http://papers.nips.cc/paper/3420-automatic-online-tuning-for-fast-gaussian-summation.pdf.

[95] L. Greengard and J. Strain, *The Fast Gauss Transform.*, SIAM J. Scientific Computing **12** (1991), no. 1 79, http://dx.doi.org/10.1137/0912004.

[96] E. Kushilevitz, R. Ostrovsky, and Y. Rabani, *Efficient Search for Approximate Nearest Neighbor in High Dimensional Spaces*, in *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, STOC '98, (New York, NY, USA), pp. 614–623, ACM, 1998. doi: 10.1145/276698.276877.

[97] T. J. DiCiccio and B. Efron, *Bootstrap Confidence Intervals*, Statistical Science **11** (1996), no. 3 189, http://www.jstor.org/stable/2246110.

[98] LHCb collaboration *et al.*, *Measurement of the polarisation amplitudes in $B^0 \to J/\psi K^{*0}$ decays*, Phys. Rev. D 88, 052002 (2013) (2013) , arXiv:1307.2782.

[99] LHCb collaboration *et al.*, *Observation of the resonant character of the $Z(4430)^-$ state*, Phys. Rev. Lett. 112, 222002 (2014) (2014) , arXiv:1404.1903.

[100] J. Heinrich, *Pitfalls of Goodness-of-Fit from Likelihood*, ECONF C030908:MOCT001,2003 (2003) arXiv:physics/0310167v1.

[101] M. Williams, *How good are your fits? Unbinned multivariate goodness-of-fit tests in high energy physics*, JINST 5:P09004,2010 (2010) , arXiv:1006.3019.

[102] LHCb collaboration *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, JINST 10 (2015) P02007 (2015) , arXiv:1408.1251.

[103] D. Aston *et al.*, *A study of $K^-\pi^+$ scattering in the reaction $K^-p \to K^+\pi^+n$ at 11 GeV/c*, Nuclear Physics B **296** (1988), no. 3 493.

[104] BaBar collaboration, J. L. Ritchie *et al.*, *Angular Analysis of $B^0 \to K^*\ell^+\ell^-$ in BaBar*, (2013) arXiv:1301.1700.

[105] J. T. Wei and P. Chang, *Measurement of the Differential Branching Fraction and Forward-Backword Asymmetry for $B^0 \to K^*\ell^+\ell^-$*, Phys. Rev. Lett. 103:171801,2009 (2009) , arXiv:0904.0770.

[106] CDF collaboration, *Measurements of the Angular Distributions in the Decays $B^0 \to K^{*0}\mu^+\mu^-$ at CDF*, Phys. Rev. Lett. **108** (2012) , arXiv:1108.0695.

[107] CMS collaboration, *Angular analysis of the decay $B^0 \to K^{*0}\mu^+\mu^-$ from pp collisions at sqrt(s) = 8 TeV*, (2015) arXiv:1507.0812.

Appendix

A Theoretical correlation

A.1 Moments of angular terms

$$
\begin{aligned}
\langle f_1\rangle &= 2/15(6+3F_l(F_S-1)-F_S),\\
\langle f_3\rangle &= -8/25(-1+F_S)S_3,\\
\langle f_4\rangle &= -8/25(-1+F_S)S_4,\\
\langle f_5\rangle &= -2/5(-1+F_S)S_5,\\
\langle f_6\rangle &= -2/5(-1+F_S)S_{6s},\\
\langle f_7\rangle &= -2/5(-1+F_S)S_7,\\
\langle f_8\rangle &= -8/25(-1+F_S)S_8,\\
\langle f_9\rangle &= -8/25(-1+F_S)S_9,\\
\langle f_1f_1\rangle &= 8/105(9+6F_l(-1+F_S)-2F_S),\\
\langle f_1f_3\rangle &= -(48/175)(-1+F_S)S_3,\\
\langle f_1f_4\rangle &= -(32/175)(-1+F_S)S_4,\\
\langle f_1f_5\rangle &= -(8/35)(-1+F_S)S_5,\\
\langle f_1f_6\rangle &= -(12/35)(-1+F_S)S_{6s},\\
\langle f_1f_7\rangle &= -(8/35)(-1+F_S)S_7,\\
\langle f_1f_8\rangle &= -(32/175)(-1+F_S)S_8,\\
\langle f_1f_9\rangle &= -(48/175)(-1+F_S)S_9,\\
\langle f_3f_3\rangle &= (32(6+3F_l(-1+F_S)+F_S))/1225,\\
\langle f_3f_4\rangle &= -((64(-1+F_S)S_4)/1225),\\
\langle f_3f_5\rangle &= -(16/175)(-1+F_S)S_5,\\
\langle f_3f_6\rangle &= 0,\\
\langle f_3f_7\rangle &= (16/175)(-1+F_S)S_7,\\
\langle f_3f_8\rangle &= (64(-1+F_S)S_8)/1225,\\
\langle f_3f_9\rangle &= 0,\\
\langle f_4f_4\rangle &= -32/3675(-15+3F_l(-1+F_S)+F_S-6S_3+6F_SS_3),\\
\langle f_4f_5\rangle &= -(16/175)(-1+F_S)S_{6s},\\
\langle f_4f_6\rangle &= -(16/175)(-1+F_S)S_5,\\
\langle f_4f_7\rangle &= 0,\\
\langle f_4f_8\rangle &= -(64(-1+F_S)S_9)/1225,\\
\langle f_4f_9\rangle &= -(64(-1+F_S)S_8)/1225,
\end{aligned}
\tag{86}
$$

$$
\begin{aligned}
\langle f_5 f_5 \rangle &= -(8/525)(-9 + 9F_l(-1 + F_S) - 6S_3 + F_S(-5 + 6S_3)),\\
\langle f_5 f_6 \rangle &= -(16/175)(-1 + F_S)S_4,\\
\langle f_5 f_7 \rangle &= -(16/175)(-1 + F_S)S_9,\\
\langle f_5 f_8 \rangle &= 0,\\
\langle f_5 f_9 \rangle &= -(16/175)(-1 + F_S)S_7,\\
\langle f_6 f_6 \rangle &= 8/525(18 + 15F_l(-1 + F_S) - 11F_S),\\
\langle f_6 f_7 \rangle &= -(16/175)(-1 + F_S)S_8,\\
\langle f_6 f_8 \rangle &= -(16/175)(-1 + F_S)S_7,\\
\langle f_6 f_9 \rangle &= 0,\\
\langle f_7 f_7 \rangle &= -(8/525)(-9 + 9F_l(-1 + F_S) + 6S_3 - F_S(5 + 6S_3)),\\
\langle f_7 f_8 \rangle &= -(16/175)(-1 + F_S)S_{6s},\\
\langle f_7 f_9 \rangle &= -(16/175)(-1 + F_S)S_5,\\
\langle f_8 f_8 \rangle &= -32/3675(-15 + 3F_l(-1 + F_S) + F_S + 6S_3 - 6F_S S_3),\\
\langle f_8 f_9 \rangle &= -(64(-1 + F_S)S_4)/1225,\\
\langle f_9 f_9 \rangle &= (32(6 + 3F_l(-1 + F_S) + F_S))/1225.
\end{aligned}
\tag{87}
$$

A.2 Correlation matrices of PDF

Table 14: The correlation matrix of the PDF obtained from an analytical calcualtion in the q^2 range $0.1 - 0.98\,\mathrm{GeV}^2$

	S_{1s}	S_3	S_4	S_5	S_{6s}	S_7	S_8	S_9
S_{1s}	1.00	−0.00	−0.03	0.10	−0.06	0.01	−0.00	−0.00
S_3		1.00	−0.03	0.13	−0.00	−0.01	0.00	−0.00
S_4			1.00	0.09	0.13	0.00	−0.00	−0.00
S_5				1.00	−0.07	−0.00	0.00	0.01
S_{6s}					1.00	−0.00	0.01	−0.00
S_7						1.00	0.08	0.13
S_8							1.00	−0.03
S_9								1.00

Table 15: The correlation matrix of the PDF obtained from an analytical calcualtion in the q^2 range $1.1 - 2.5\,\mathrm{GeV}^2$

	S_{1s}	S_3	S_4	S_5	S_{6s}	S_7	S_8	S_9
S_{1s}	1.00	0.00	0.00	−0.01	−0.34	−0.01	0.00	−0.00
S_3		1.00	−0.01	0.06	0.00	−0.02	0.01	0.00
S_4			1.00	0.13	0.08	0.00	−0.00	−0.01
S_5				1.00	−0.04	−0.00	0.00	0.02
S_{6s}					1.00	−0.02	0.03	−0.00
S_7						1.00	0.13	0.06
S_8							1.00	−0.01
S_9								1.00

Table 16: The correlation matrix of the PDF obtained from an analytical calcualtion in the q^2 range $2.5 - 4.0\,\text{GeV}^2$

	S_{1s}	S_3	S_4	S_5	S_{6s}	S_7	S_8	S_9
S_{1s}	1.00	0.00	−0.03	0.04	−0.15	−0.01	0.00	−0.00
S_3		1.00	0.05	−0.10	0.00	−0.02	0.01	0.00
S_4			1.00	0.06	−0.14	−0.00	0.00	−0.01
S_5				1.00	0.09	0.00	−0.00	0.02
S_{6s}					1.00	−0.01	0.03	−0.00
S_7						1.00	0.05	−0.10
S_8							1.00	0.05
S_9								1.00

Table 17: The correlation matrix of the PDF obtained from an analytical calcualtion in the q^2 range $4.0 - 6.0\,\text{GeV}^2$

	S_{1s}	S_3	S_4	S_5	S_{6s}	S_7	S_8	S_9
S_{1s}	1.00	0.01	−0.03	0.06	0.17	−0.00	0.00	−0.00
S_3		1.00	0.09	−0.22	−0.00	−0.02	0.00	0.00
S_4			1.00	−0.01	−0.23	−0.00	0.00	−0.00
S_5				1.00	0.08	0.01	−0.00	0.02
S_{6s}					1.00	−0.00	0.02	0.00
S_7						1.00	−0.06	−0.20
S_8							1.00	0.08
S_9								1.00

Table 18: The correlation matrix of the PDF obtained from an analytical calcualtion in the q^2 range $6.0 - 8.0\,\text{GeV}^2$

	S_{1s}	S_3	S_4	S_5	S_{6s}	S_7	S_8	S_9
S_{1s}	1.00	0.01	−0.02	0.03	0.34	−0.00	0.00	−0.00
S_3		1.00	0.10	−0.28	−0.00	−0.02	0.00	0.00
S_4			1.00	−0.08	−0.24	−0.00	0.00	−0.00
S_5				1.00	0.01	0.01	−0.00	0.02
S_{6s}					1.00	0.00	0.02	0.00
S_7						1.00	−0.15	−0.25
S_8							1.00	0.10
S_9								1.00

Table 19: The correlation matrix of the PDF obtained from an analytical calcualtion in the q^2 range $11.0 - 12.5\,\text{GeV}^2$

	S_{1s}	S_3	S_4	S_5	S_{6s}	S_7	S_8	S_9
S_{1s}	1.00	0.03	0.04	−0.08	0.45	−0.00	−0.00	−0.00
S_3		1.00	0.12	−0.30	−0.02	0.00	0.00	0.00
S_4			1.00	−0.24	−0.17	0.00	0.00	−0.00
S_5				1.00	−0.10	0.00	−0.00	0.00
S_{6s}					1.00	−0.00	−0.00	0.00
S_7						1.00	−0.29	−0.26
S_8							1.00	0.10
S_9								1.00

Table 20: The correlation matrix of the PDF obtained from an analytical calcualtion in the q^2 range $15.0 - 17.0\,\text{GeV}^2$

	S_{1s}	S_3	S_4	S_5	S_{6s}	S_7	S_8	S_9
S_{1s}	1.00	0.08	0.07	−0.10	0.41	0.00	−0.00	−0.00
S_3		1.00	0.15	−0.28	−0.07	−0.00	0.00	0.00
S_4			1.00	−0.28	−0.09	−0.00	0.00	−0.00
S_5				1.00	−0.05	0.00	−0.00	0.00
S_{6s}					1.00	−0.00	−0.00	0.00
S_7						1.00	−0.30	−0.21
S_8							1.00	0.11
S_9								1.00

Table 21: The correlation matrix of the PDF obtained from an analytical calcualtion in the q^2 range $17.0 - 19.0\,\text{GeV}^2$

	S_{1s}	S_3	S_4	S_5	S_{6s}	S_7	S_8	S_9
S_{1s}	1.00	0.13	0.09	−0.08	0.30	−0.00	−0.00	−0.00
S_3		1.00	0.19	−0.23	−0.09	0.00	0.00	0.00
S_4			1.00	−0.24	−0.04	0.00	0.00	−0.00
S_5				1.00	0.05	0.00	−0.00	0.00
S_{6s}					1.00	−0.00	−0.00	0.00
S_7						1.00	−0.24	−0.15
S_8							1.00	0.11
S_9								1.00

B Correlation matrices and 2D profile log likelihood scans

The correlation matrices of the maximum likelihood fit are shown in Tab. 22 to 29. The corresponding 2D profile log likelihood scans can be seen in Fig. 58 to 89. In these plots every single pixel corresponds to a fit, in which the two parameters are fixed to the corresponding values. Shown is the difference in the likelihood $2\Delta lnL$ compared to the nominal result. The 68.3% ($1\,\sigma$) confidence region is in 2D where $2\Delta lnL = 2.30$; the 95.5% ($2\,\sigma$) confidence region where $2\Delta lnL = 6.18$.

Table 22: The correlation matrix for the maximum likelihood fit in the q^2 range $0.1 - 0.98\,\mathrm{GeV}^2$

	S_{1s}	S_3	S_4	S_5	S_{6s}	S_7	S_8	S_9
S_{1s}	1.00	−0.02	0.01	−0.03	−0.09	0.02	−0.00	−0.09
S_3		1.00	−0.03	0.09	0.00	−0.05	−0.04	−0.01
S_4			1.00	0.10	0.11	−0.02	0.09	0.07
S_5				1.00	−0.01	−0.02	−0.03	0.05
S_{6s}					1.00	0.05	−0.03	−0.04
S_7						1.00	−0.02	0.13
S_8							1.00	0.03
S_9								1.00

Table 23: The correlation matrix for the maximum likelihood fit in the q^2 range $1.1 - 2.5\,\mathrm{GeV}^2$

	S_{1s}	S_3	S_4	S_5	S_{6s}	S_7	S_8	S_9
S_{1s}	1.00	−0.03	0.04	0.01	−0.26	0.14	0.20	−0.07
S_3		1.00	−0.26	−0.20	−0.16	0.28	0.04	−0.08
S_4			1.00	−0.49	−0.02	−0.01	−0.00	0.11
S_5				1.00	−0.13	−0.10	−0.03	0.08
S_{6s}					1.00	0.12	0.15	−0.17
S_7						1.00	−0.19	0.38
S_8							1.00	0.24
S_9								1.00

Table 24: The correlation matrix for the maximum likelihood fit in the q^2 range $2.5 - 4.0\,\mathrm{GeV}^2$

	S_{1s}	S_3	S_4	S_5	S_{6s}	S_7	S_8	S_9
S_{1s}	1.00	0.25	−0.04	−0.05	−0.19	−0.03	−0.04	−0.05
S_3		1.00	−0.09	−0.02	0.34	0.05	0.02	0.01
S_4			1.00	−0.27	−0.22	−0.04	−0.02	0.14
S_5				1.00	−0.09	0.07	−0.02	0.28
S_{6s}					1.00	0.03	−0.09	0.06
S_7						1.00	−0.04	−0.04
S_8							1.00	−0.04
S_9								1.00

Table 25: The correlation matrix for the maximum likelihood fit in the q^2 range $4.0 - 6.0\,\mathrm{GeV}^2$

	S_{1s}	S_3	S_4	S_5	S_{6s}	S_7	S_8	S_9
S_{1s}	1.00	0.01	−0.03	−0.02	0.04	0.07	−0.03	0.03
S_3		1.00	−0.04	−0.03	0.11	−0.11	−0.02	−0.08
S_4			1.00	0.08	−0.08	−0.01	−0.06	0.02
S_5				1.00	−0.06	−0.04	−0.00	−0.01
S_{6s}					1.00	0.02	0.03	−0.01
S_7						1.00	0.07	−0.07
S_8							1.00	0.00
S_9								1.00

Table 26: The correlation matrix for the maximum likelihood fit in the q^2 range $6.0 - 8.0\,\mathrm{GeV}^2$

	S_{1s}	S_3	S_4	S_5	S_{6s}	S_7	S_8	S_9
S_{1s}	1.00	−0.05	−0.04	−0.03	0.23	0.07	0.01	0.04
S_3		1.00	−0.25	−0.29	−0.12	−0.02	0.03	−0.03
S_4			1.00	−0.28	−0.28	−0.02	−0.05	−0.05
S_5				1.00	−0.29	−0.07	−0.03	−0.04
S_{6s}					1.00	−0.02	0.05	−0.05
S_7						1.00	0.08	−0.07
S_8							1.00	−0.10
S_9								1.00

Table 27: The correlation matrix for the maximum likelihood fit in the q^2 range $11.0 - 12.5\,\mathrm{GeV}^2$

	S_{1s}	S_3	S_4	S_5	S_{6s}	S_7	S_8	S_9
S_{1s}	1.00	−0.31	0.05	−0.05	0.37	−0.00	−0.04	0.09
S_3		1.00	0.01	−0.39	−0.23	0.08	−0.00	0.20
S_4			1.00	0.01	0.05	−0.04	−0.12	−0.06
S_5				1.00	−0.07	−0.12	−0.04	−0.07
S_{6s}					1.00	−0.05	0.02	0.11
S_7						1.00	0.23	−0.16
S_8							1.00	−0.08
S_9								1.00

Table 28: The correlation matrix for the maximum likelihood fit in the q^2 range $15.0 - 17.0\,\mathrm{GeV}^2$

	S_{1s}	S_3	S_4	S_5	S_{6s}	S_7	S_8	S_9
S_{1s}	1.00	−0.14	0.12	0.00	0.39	0.02	0.06	−0.11
S_3		1.00	−0.08	−0.08	0.01	−0.08	−0.02	−0.02
S_4			1.00	0.21	−0.09	−0.07	0.11	0.01
S_5				1.00	−0.09	0.06	−0.06	0.01
S_{6s}					1.00	0.06	−0.03	−0.05
S_7						1.00	0.25	−0.20
S_8							1.00	−0.10
S_9								1.00

Table 29: The correlation matrix for the maximum likelihood fit in the q^2 range $17.0 - 19.0\,\mathrm{GeV}^2$

	S_{1s}	S_3	S_4	S_5	S_{6s}	S_7	S_8	S_9
S_{1s}	1.00	−0.01	−0.15	−0.02	0.35	−0.04	−0.09	−0.13
S_3		1.00	−0.08	−0.23	−0.08	−0.04	−0.07	0.06
S_4			1.00	0.11	−0.30	0.02	0.11	0.03
S_5				1.00	−0.13	0.00	0.05	−0.01
S_{6s}					1.00	0.01	−0.06	−0.12
S_7						1.00	0.24	−0.17
S_8							1.00	−0.13
S_9								1.00

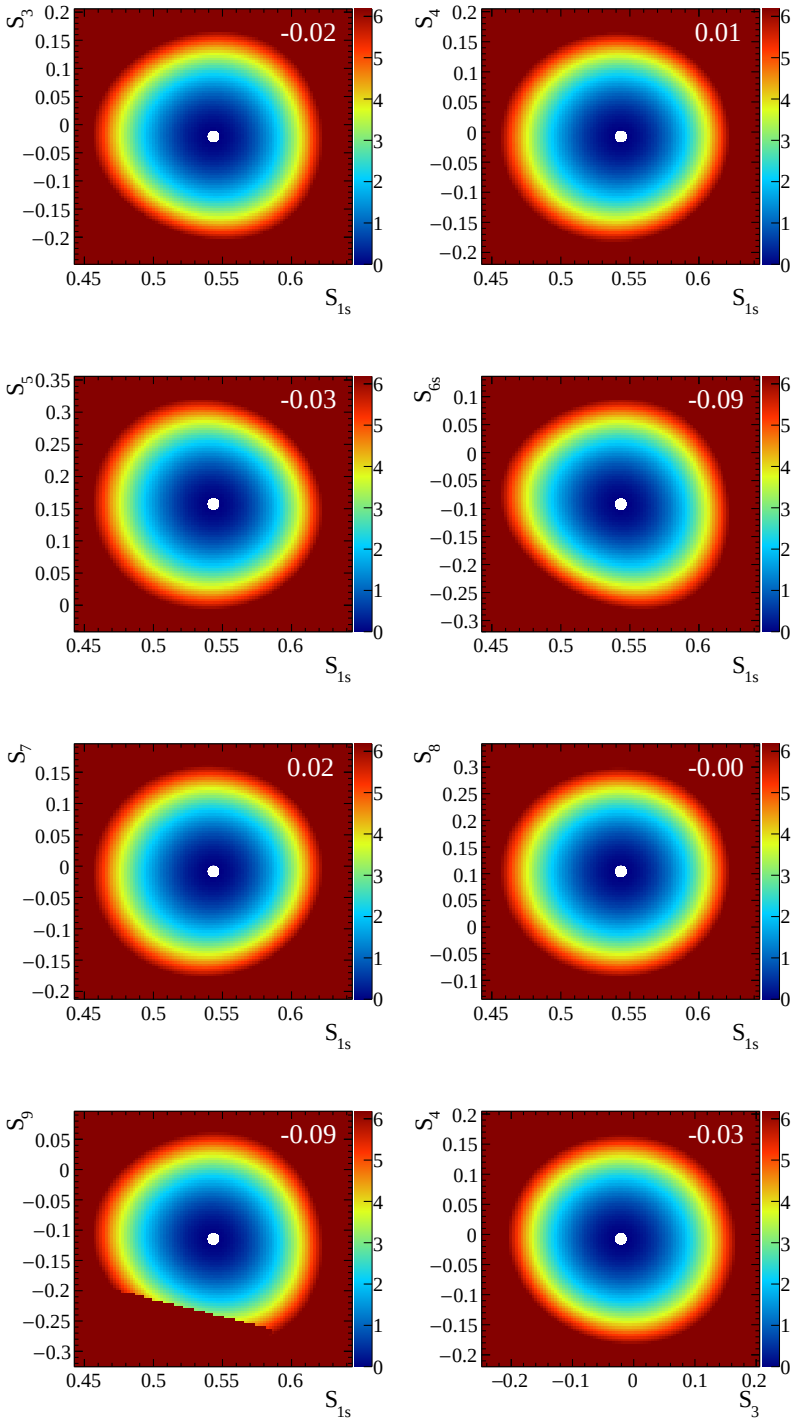

Figure 58: 2D profile log likelihood scans in the q^2 bin 0.1-0.98 GeV^2

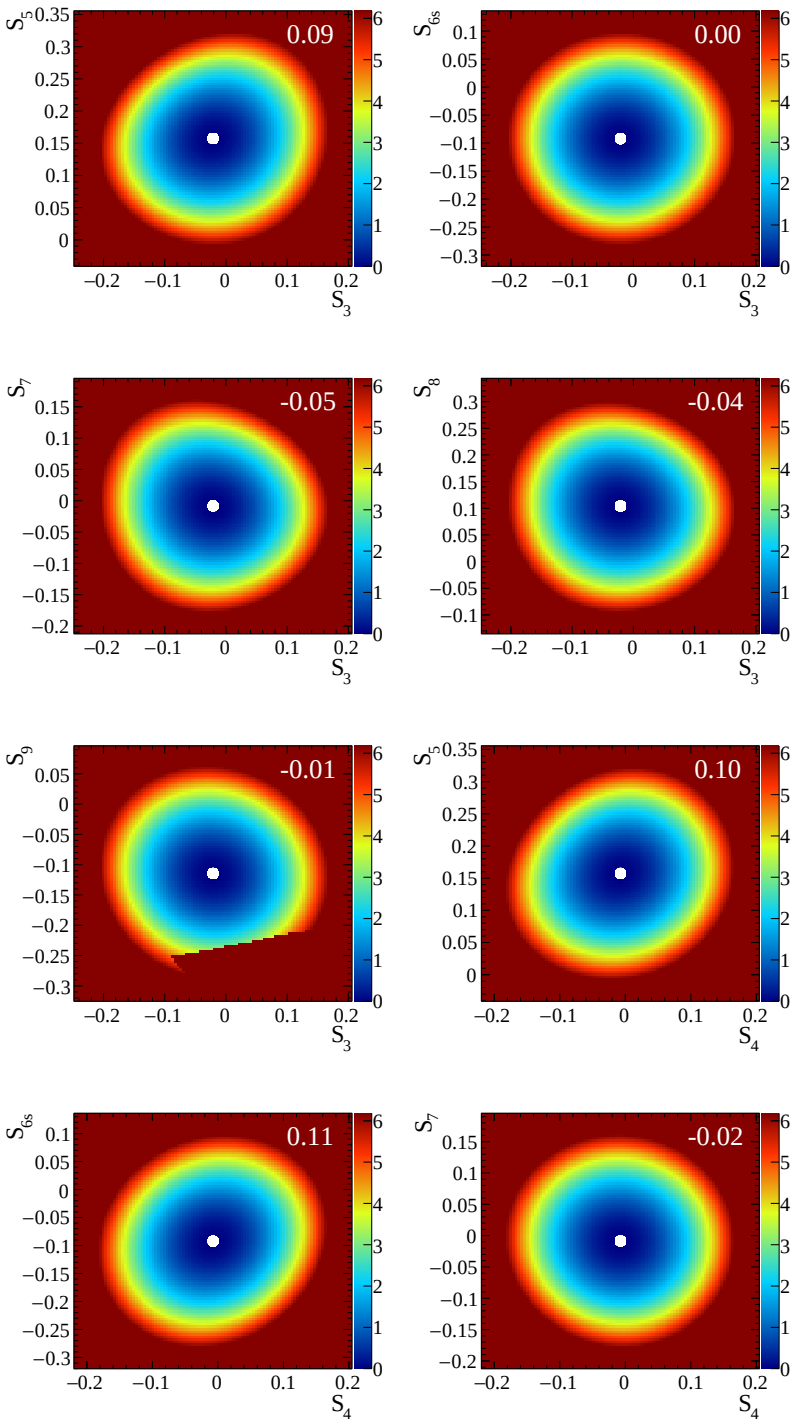

Figure 59: 2D profile log likelihood scans in the q^2 bin 0.1-0.98 GeV^2

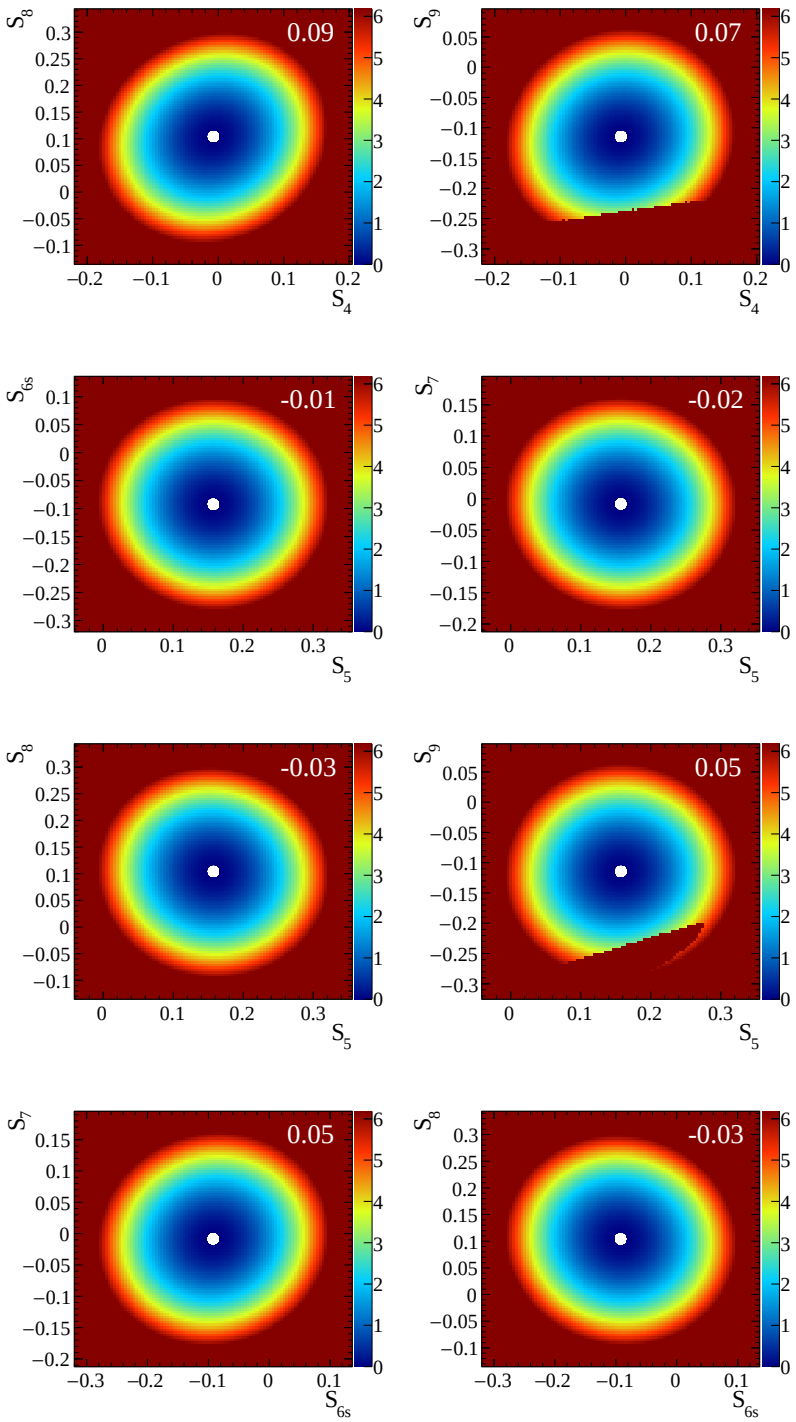

Figure 60: 2D profile log likelihood scans in the q^2 bin 0.1-0.98 GeV^2

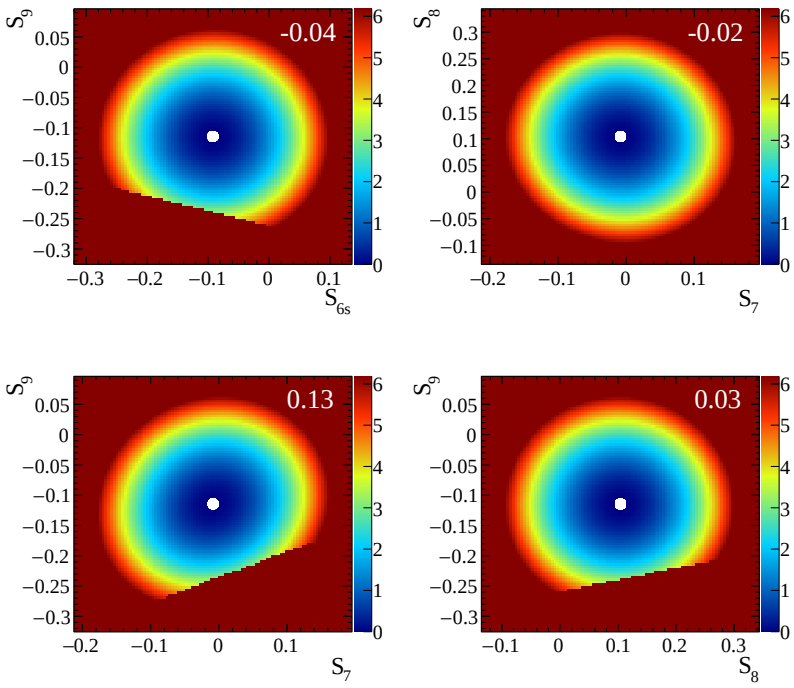

Figure 61: 2D profile log likelihood scans in the q^2 bin 0.1-0.98 GeV^2

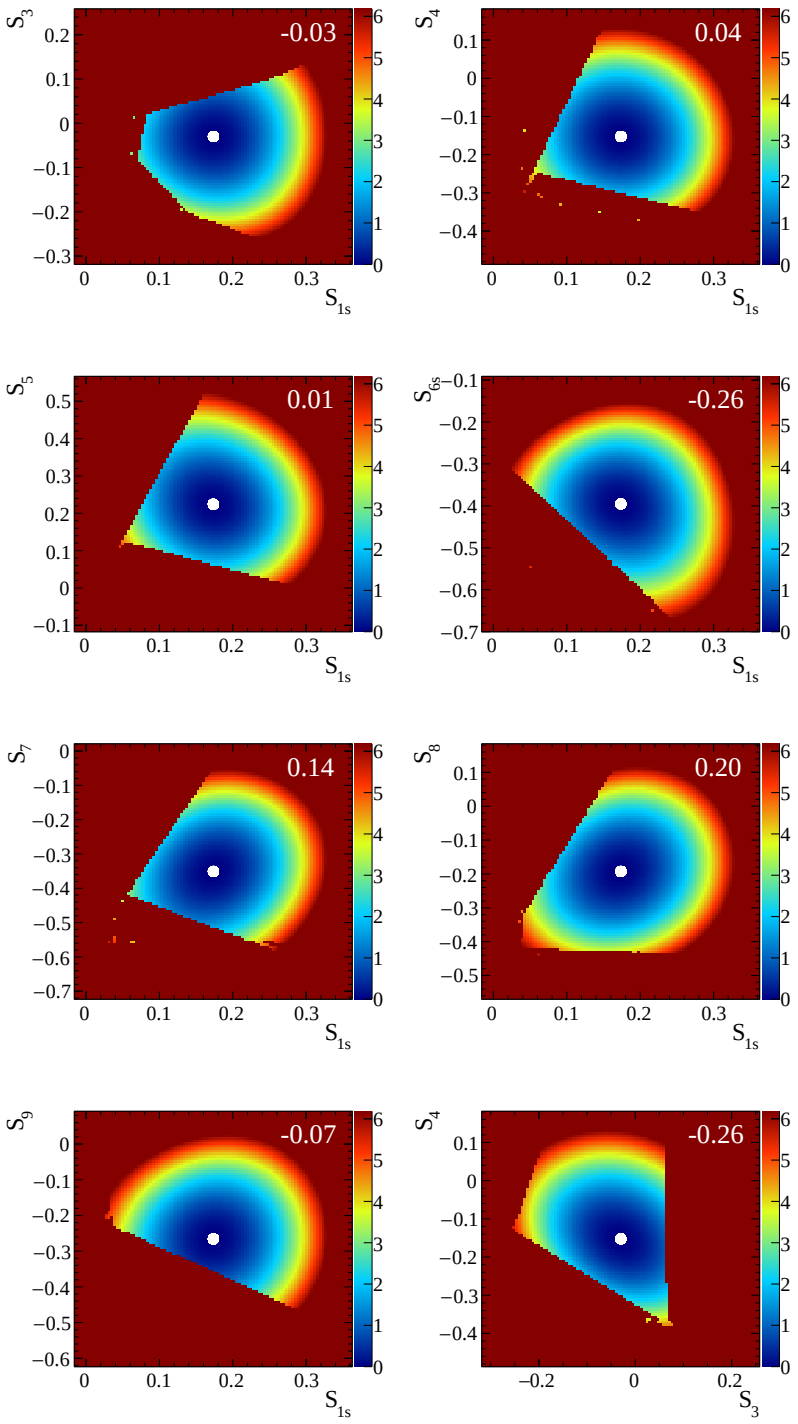

Figure 62: 2D profile log likelihood scans in the q^2 bin 1.1-2.5 GeV^2

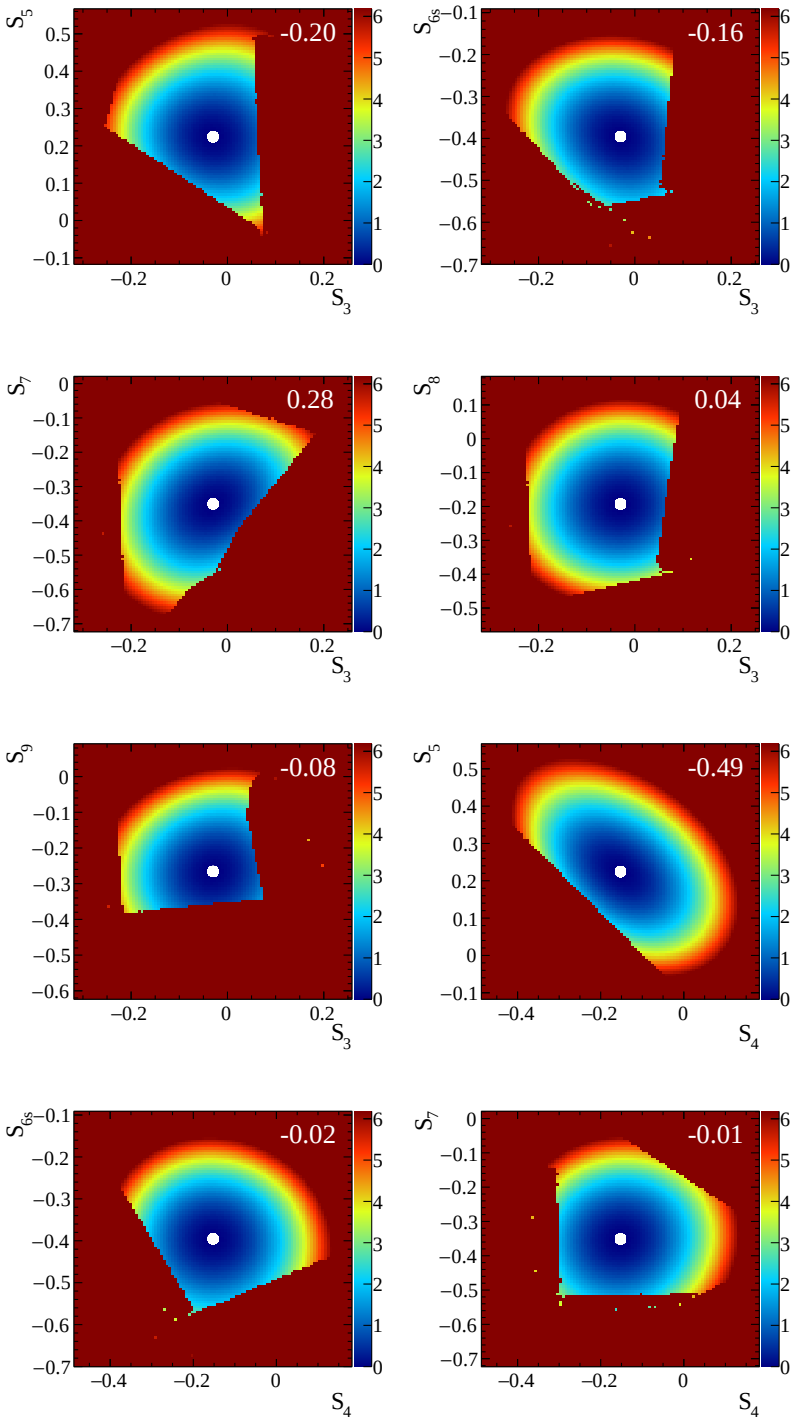

Figure 63: 2D profile log likelihood scans in the q^2 bin 1.1-2.5 GeV^2

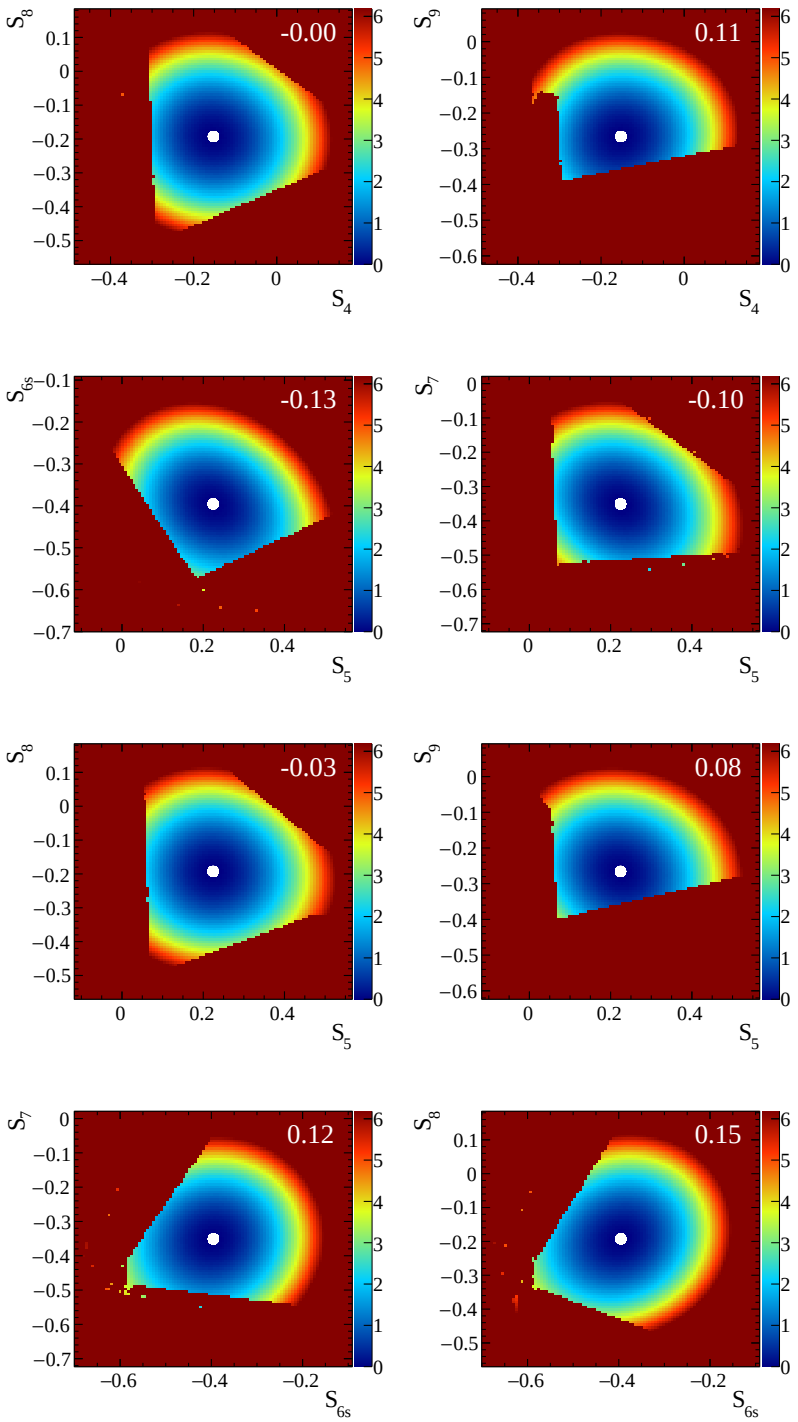

Figure 64: 2D profile log likelihood scans in the q^2 bin 1.1-2.5 GeV^2

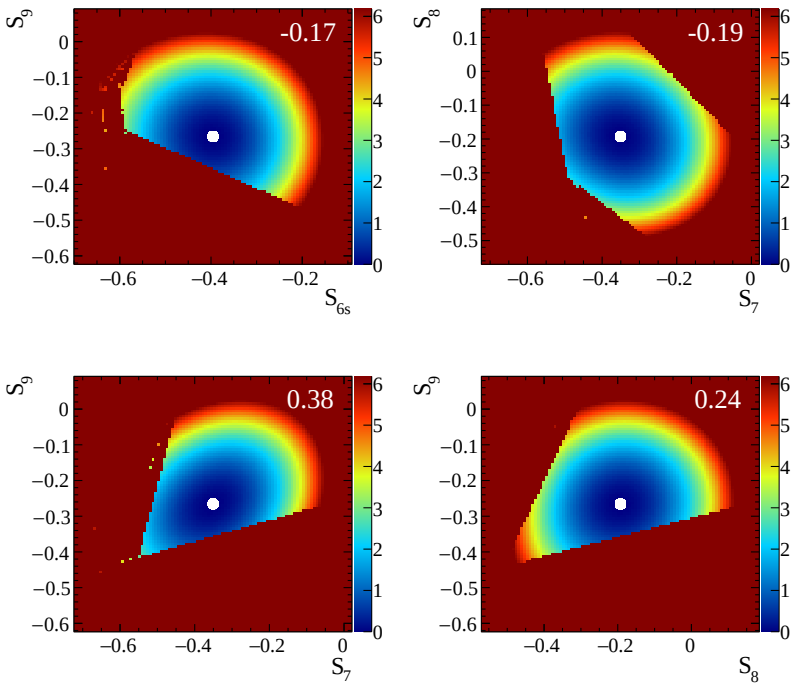

Figure 65: 2D profile log likelihood scans in the q^2 bin 1.1-2.5 GeV^2

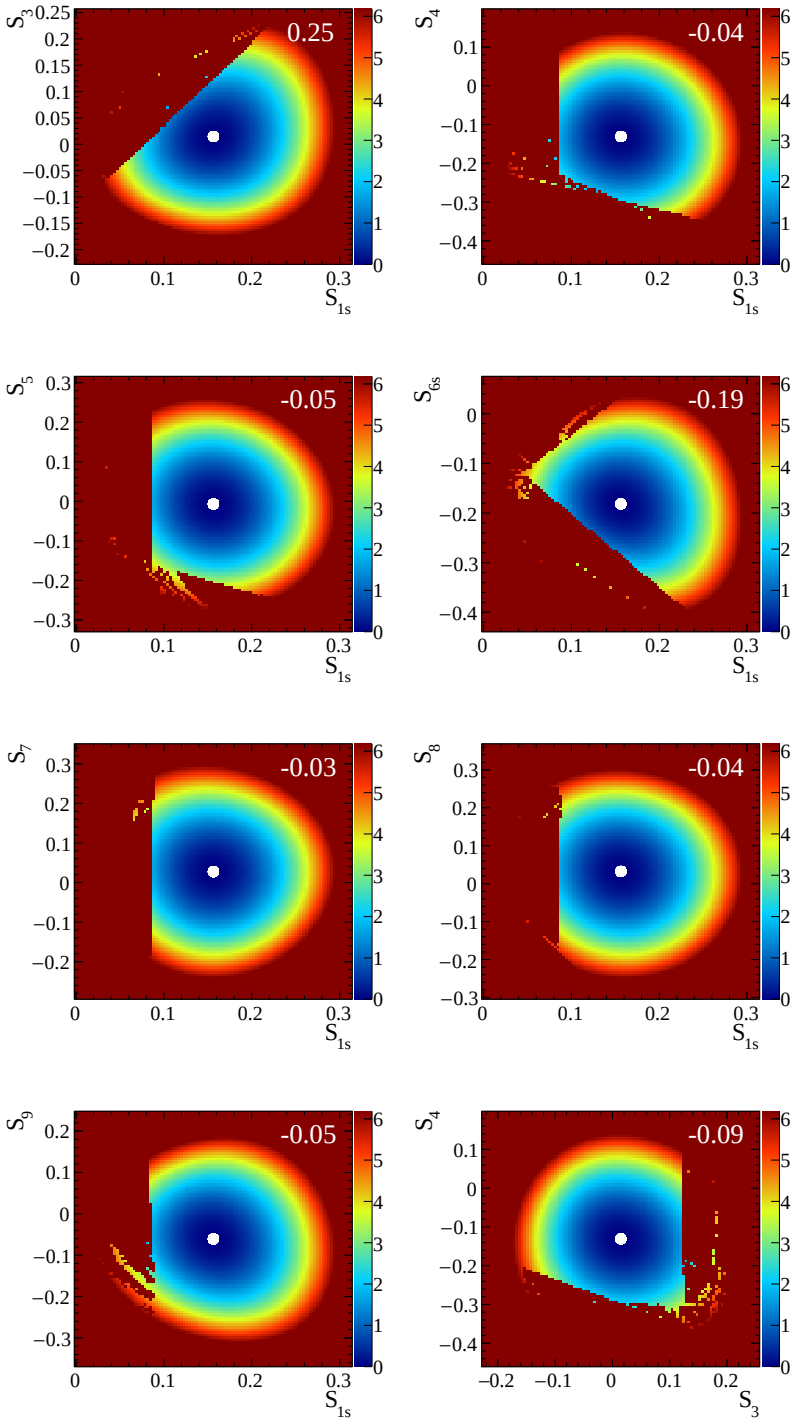

Figure 66: 2D profile log likelihood scans in the q^2 bin 2.5-4.0 GeV^2

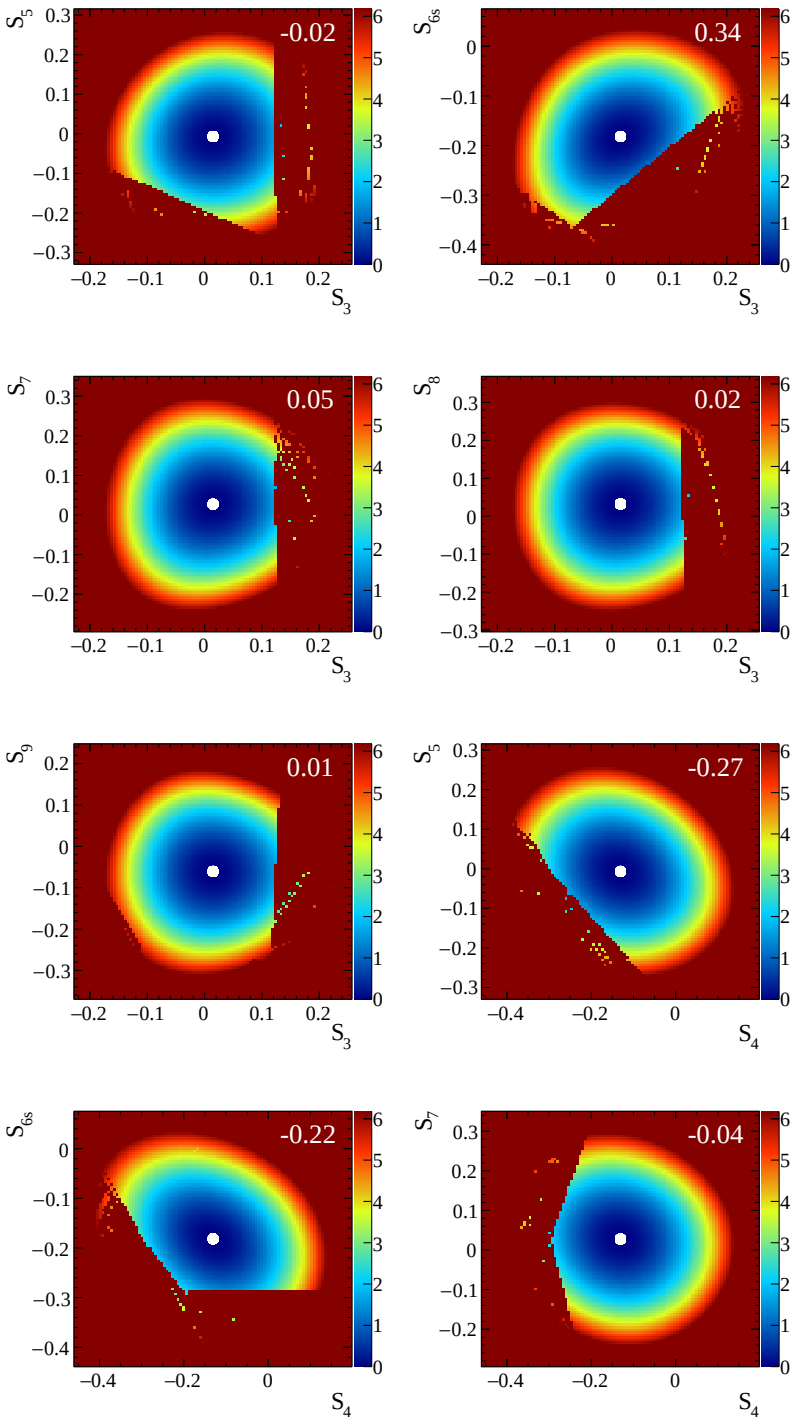

Figure 67: 2D profile log likelihood scans in the q^2 bin 2.5-4.0 GeV^2

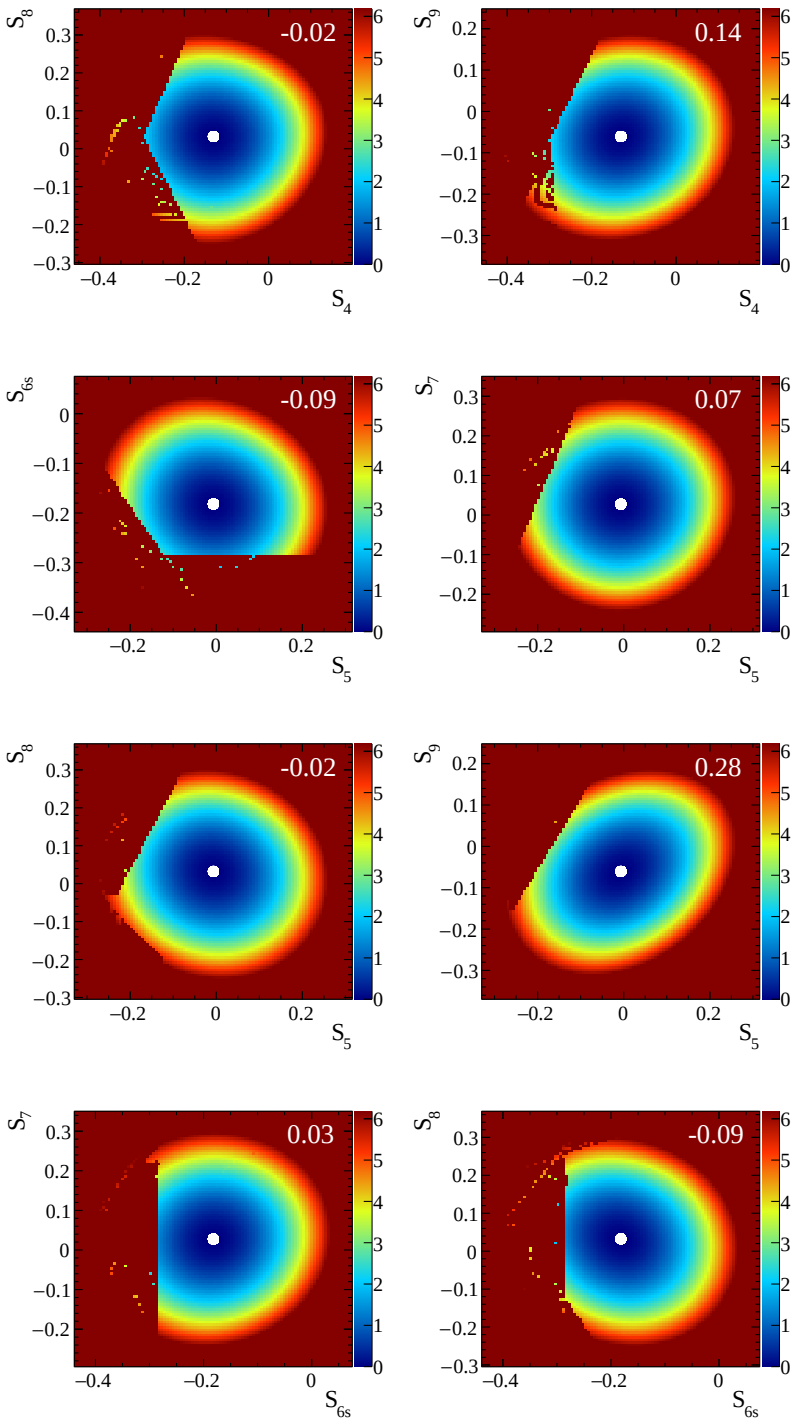

Figure 68: 2D profile log likelihood scans in the q^2 bin 2.5-4.0 GeV^2

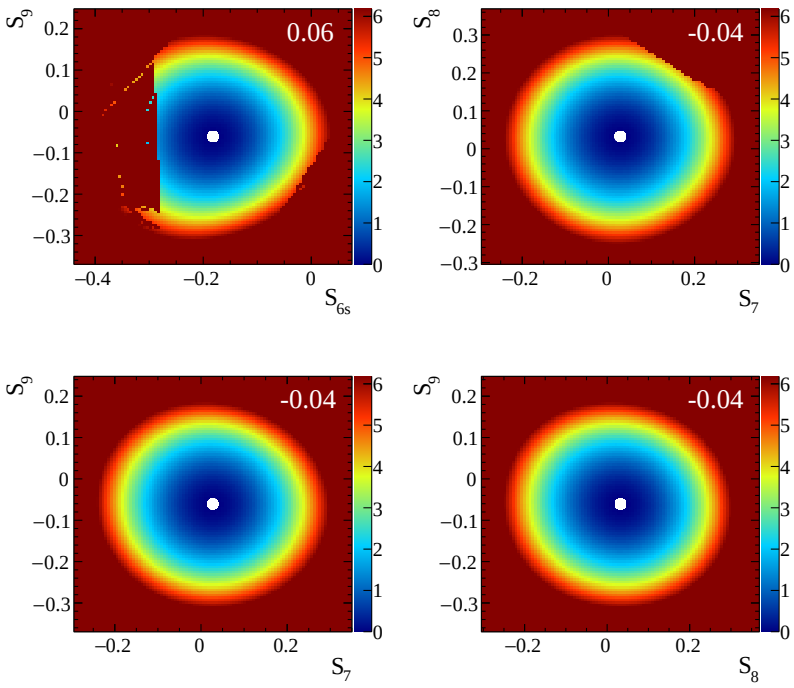

Figure 69: 2D profile log likelihood scans in the q^2 bin 2.5-4.0 GeV^2

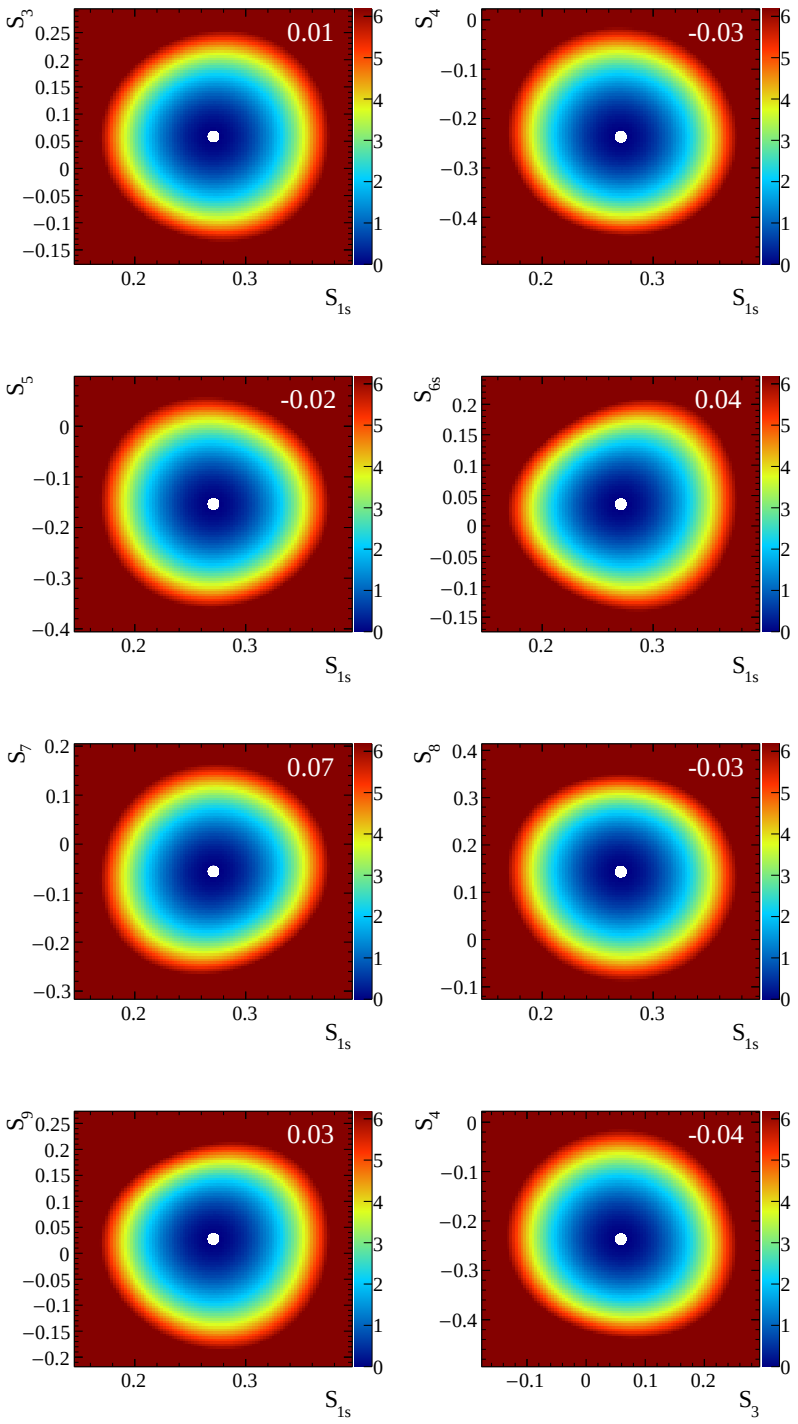

Figure 70: 2D profile log likelihood scans in the q^2 bin 4.0-6.0 GeV^2

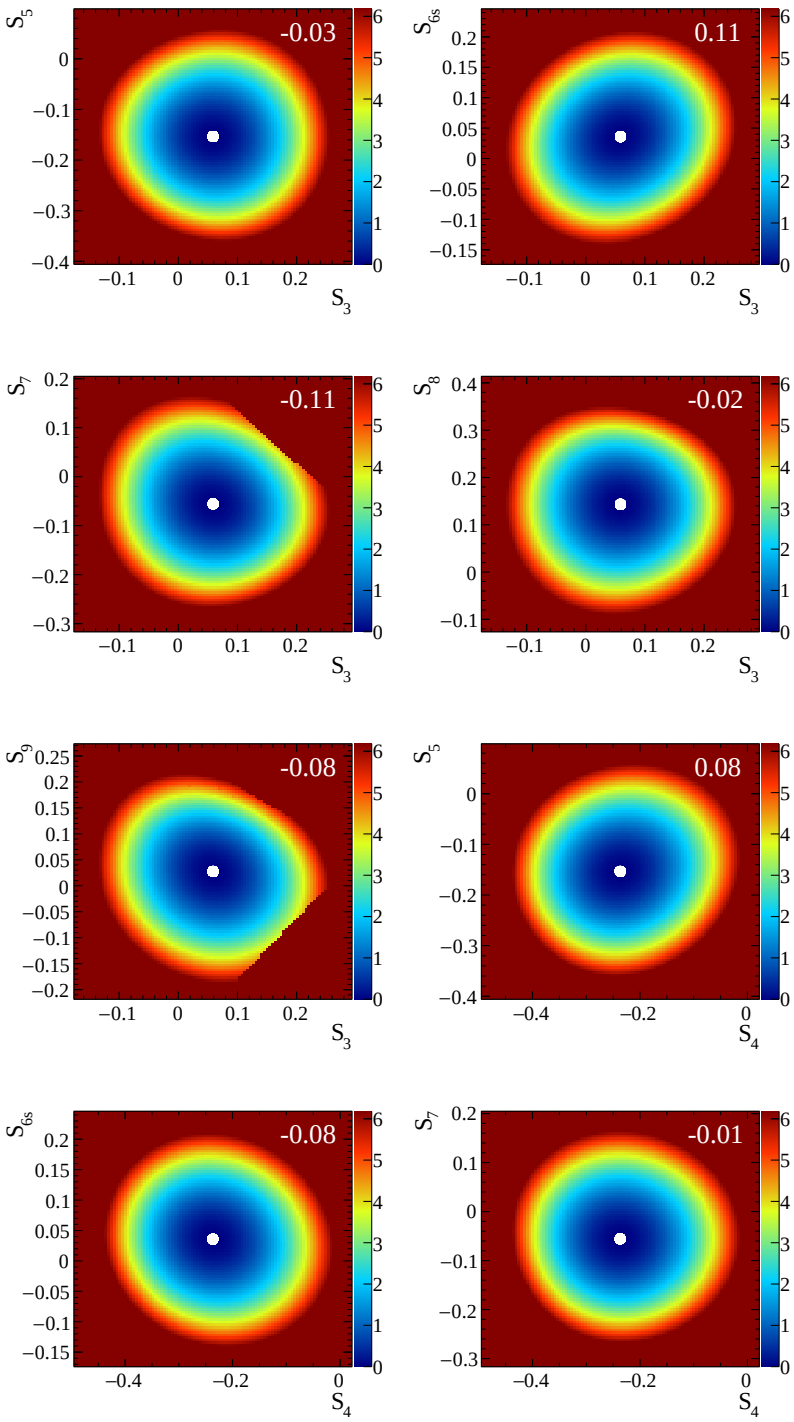

Figure 71: 2D profile log likelihood scans in the q^2 bin 4.0-6.0 GeV^2

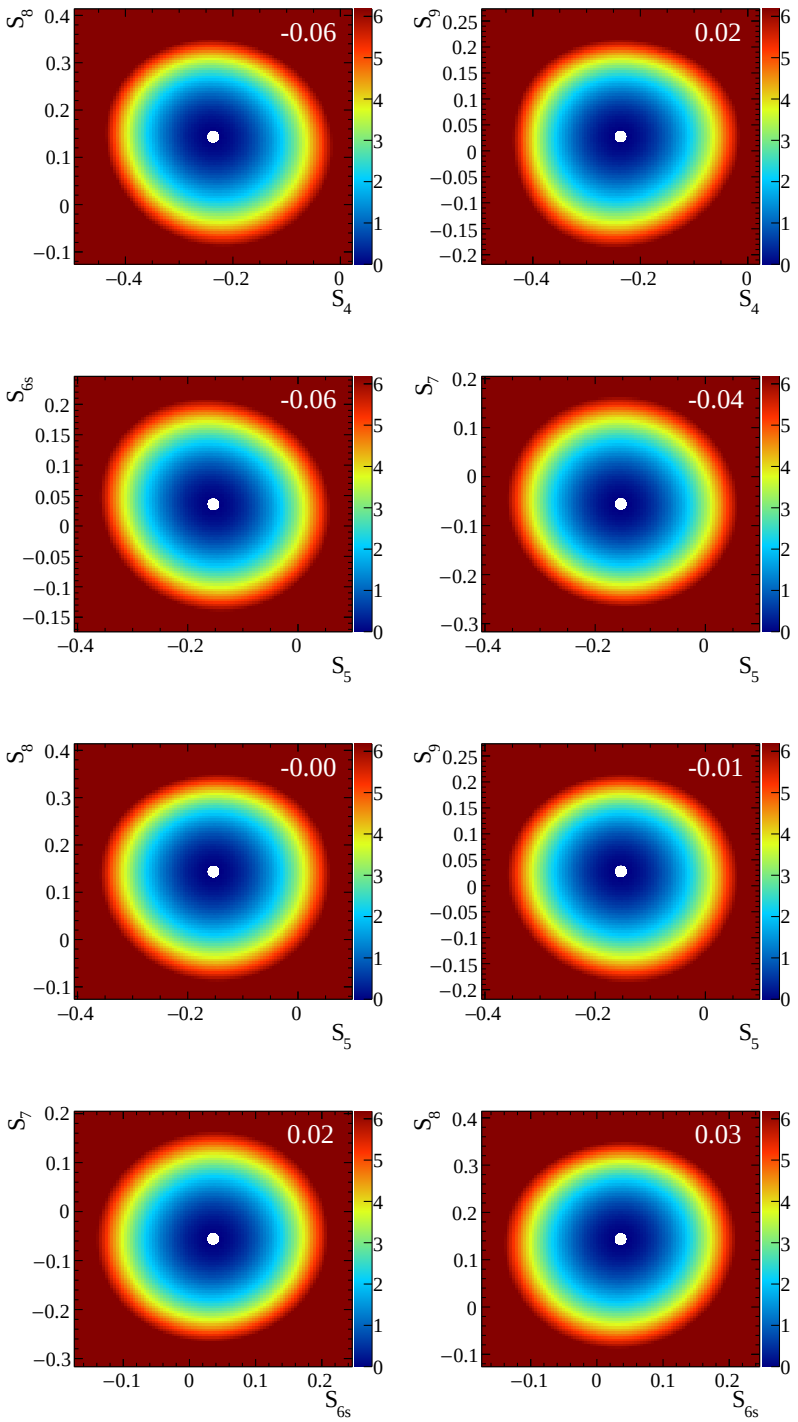

Figure 72: 2D profile log likelihood scans in the q^2 bin 4.0-6.0 GeV^2

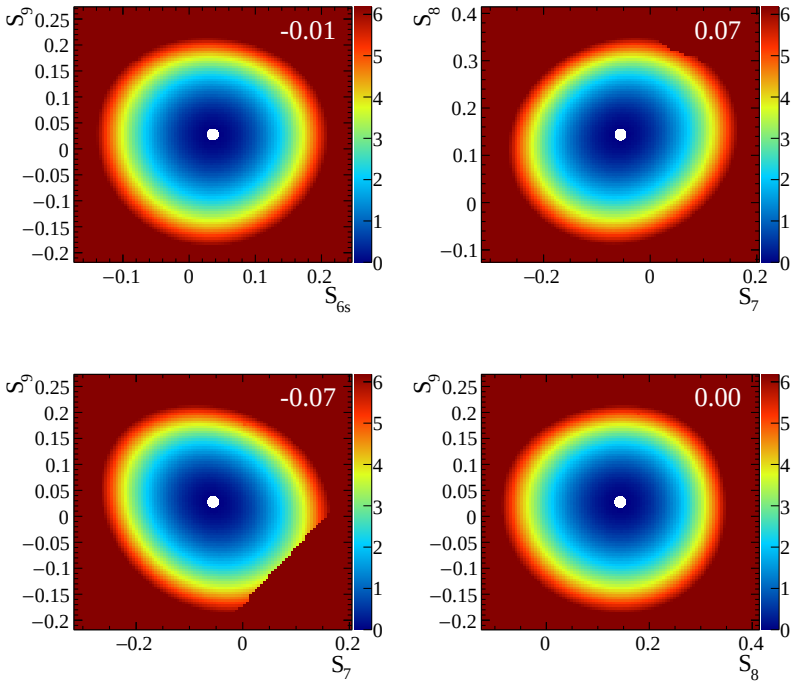

Figure 73: 2D profile log likelihood scans in the q^2 bin 4.0-6.0 GeV^2

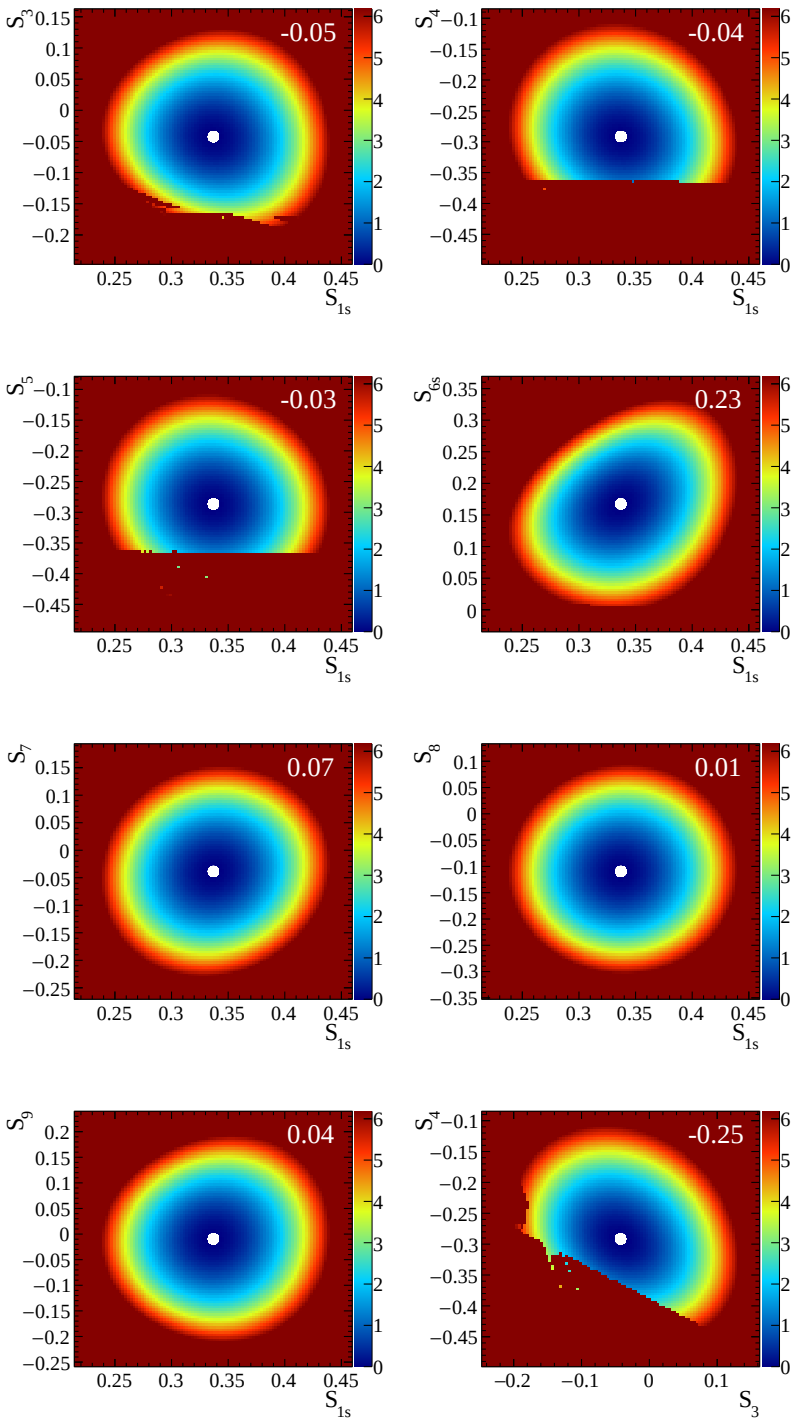

Figure 74: 2D profile log likelihood scans in the q^2 bin 6.0-8.0 GeV^2

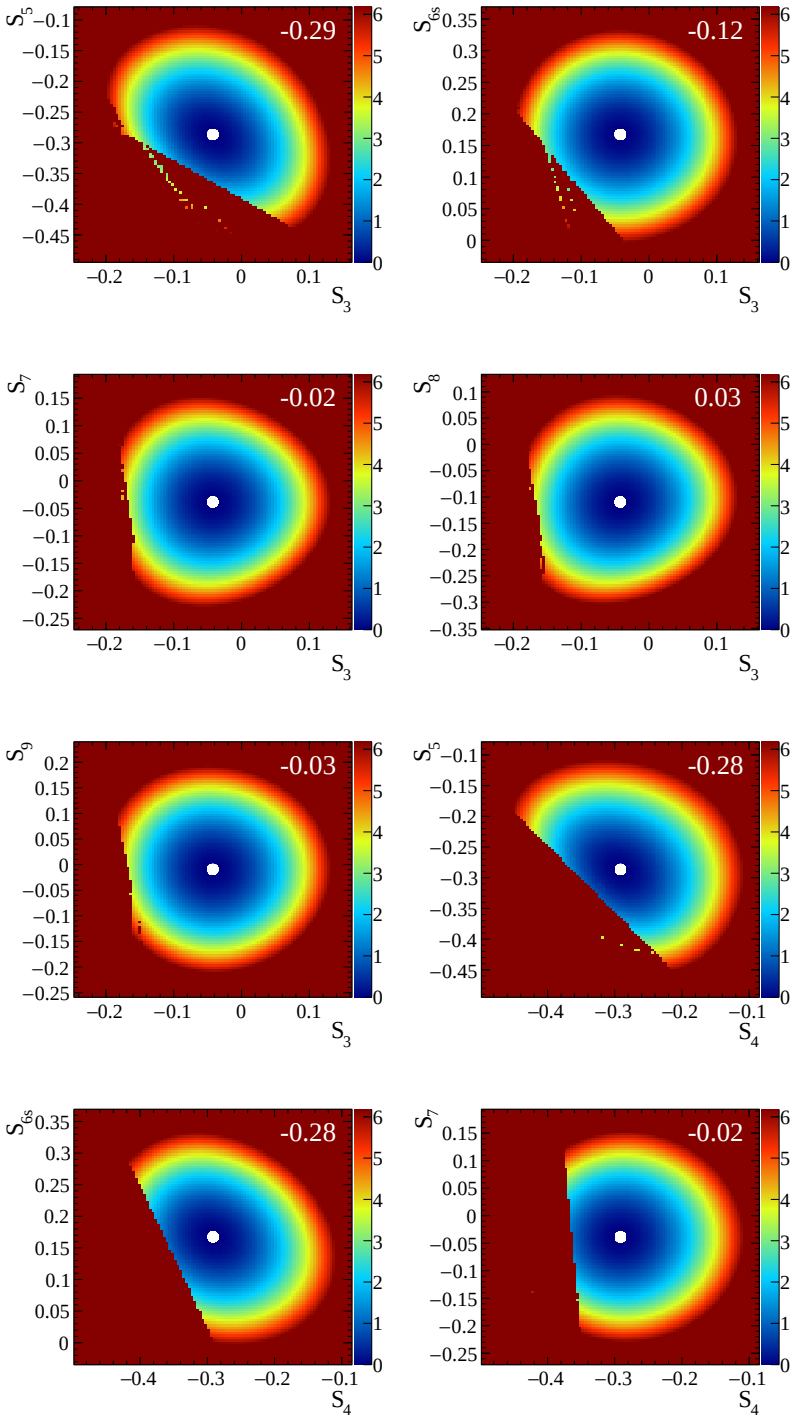

Figure 75: 2D profile log likelihood scans in the q^2 bin 6.0-8.0 GeV^2

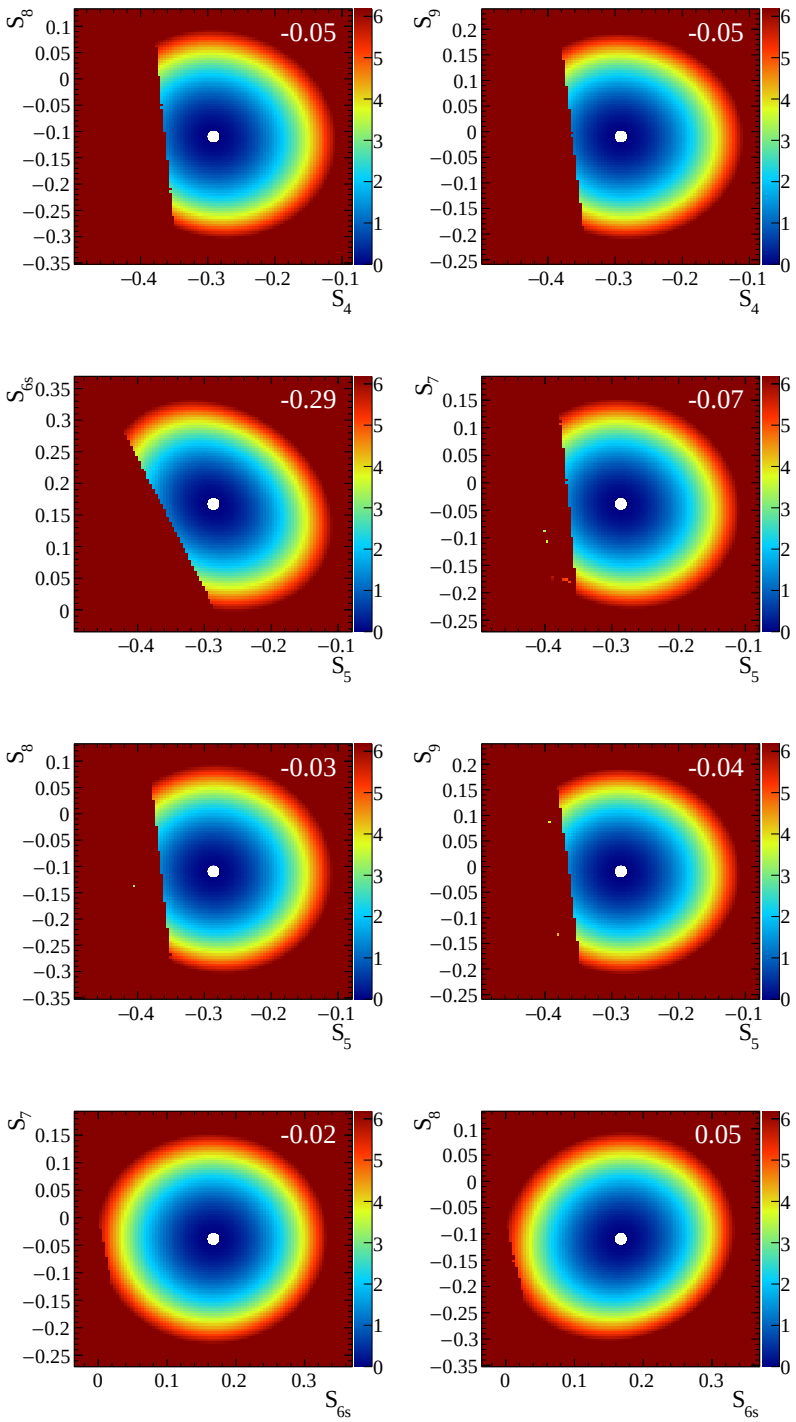

Figure 76: 2D profile log likelihood scans in the q^2 bin 6.0-8.0 GeV^2

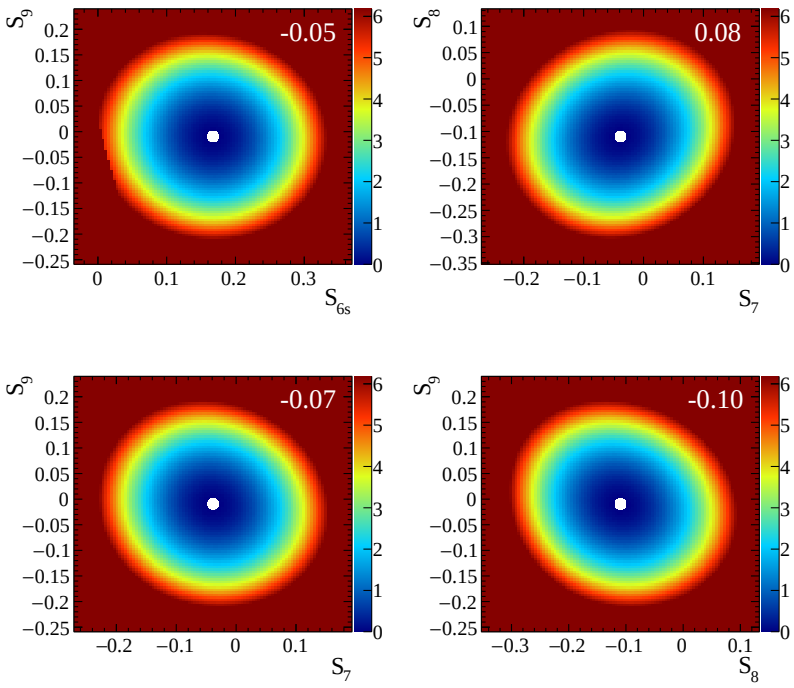

Figure 77: 2D profile log likelihood scans in the q^2 bin 6.0-8.0 GeV^2

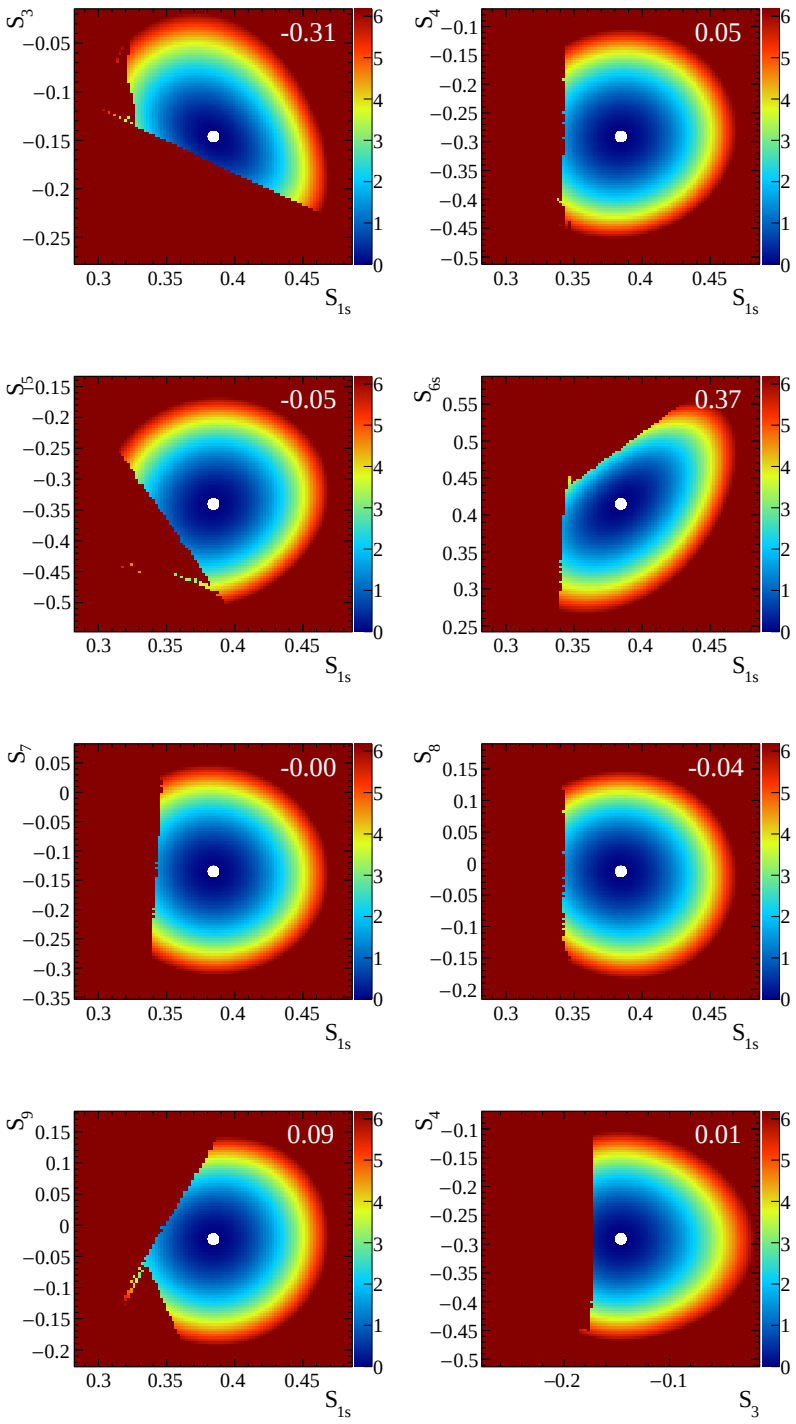

Figure 78: 2D profile log likelihood scans in the q^2 bin 11.0-12.5 GeV^2

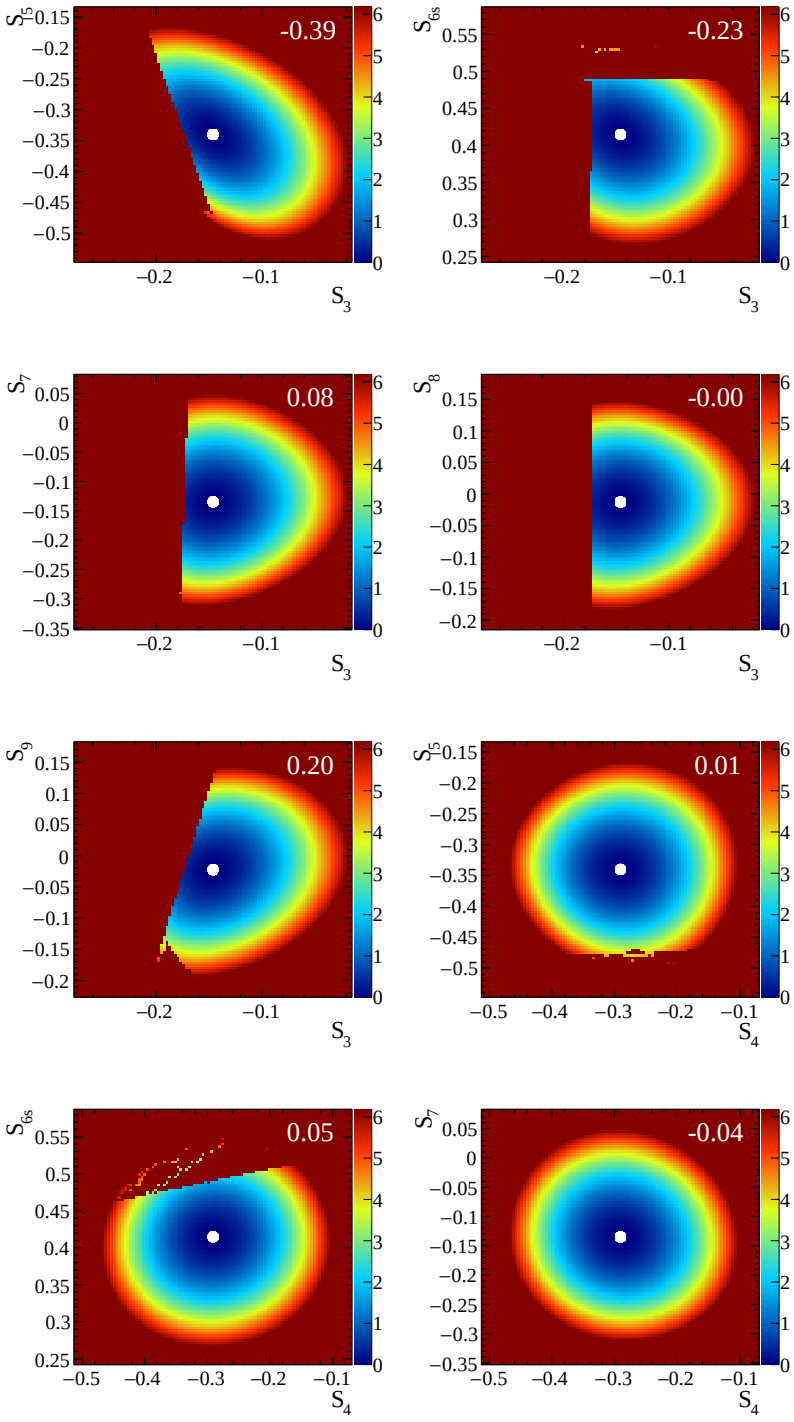

Figure 79: 2D profile log likelihood scans in the q^2 bin 11.0-12.5 GeV^2

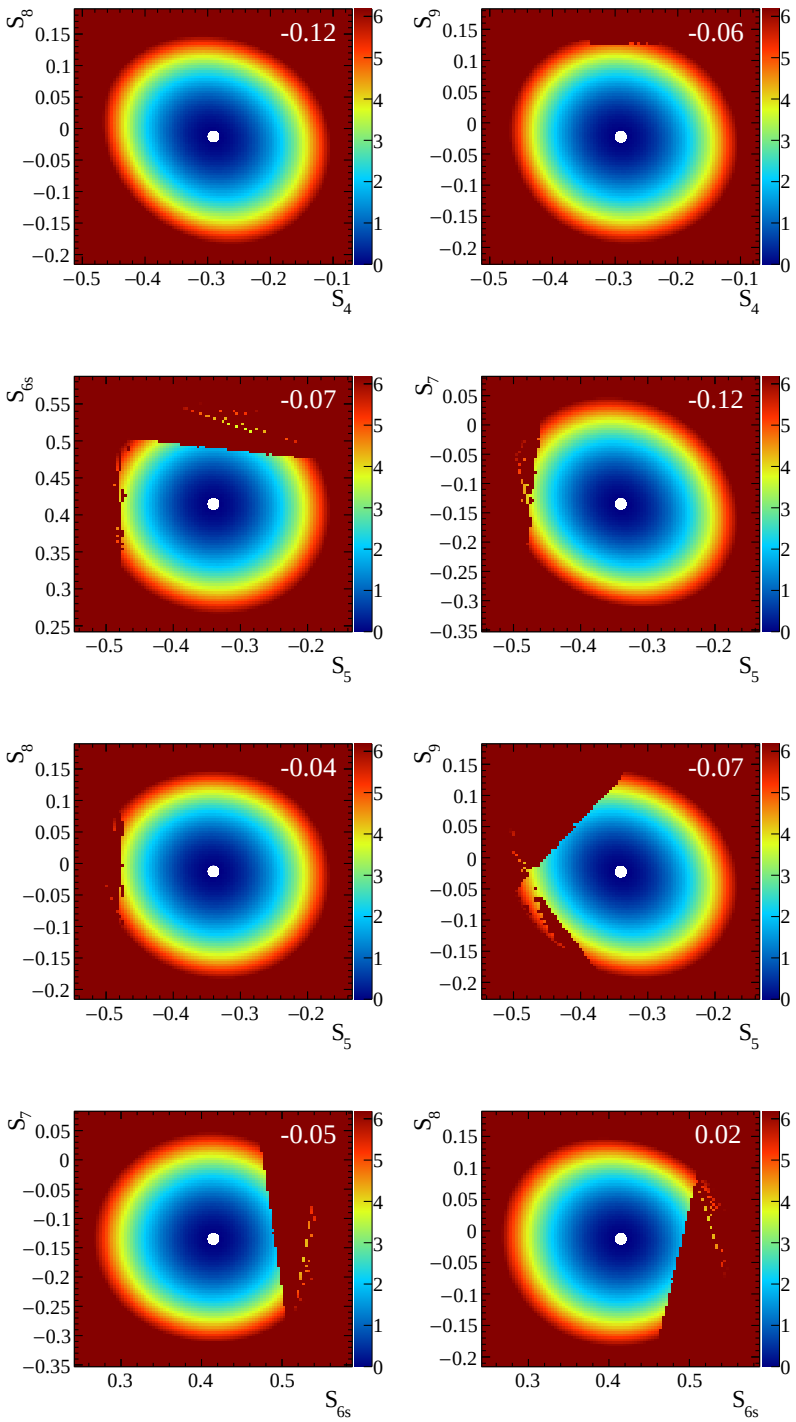

Figure 80: 2D profile log likelihood scans in the q^2 bin 11.0-12.5 GeV^2

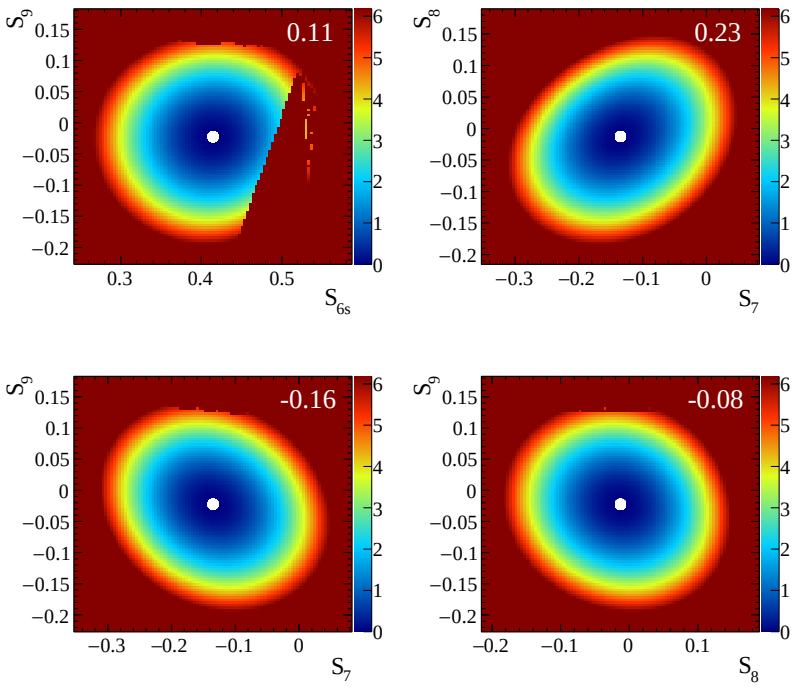

Figure 81: 2D profile log likelihood scans in the q^2 bin 11.0-12.5 GeV^2

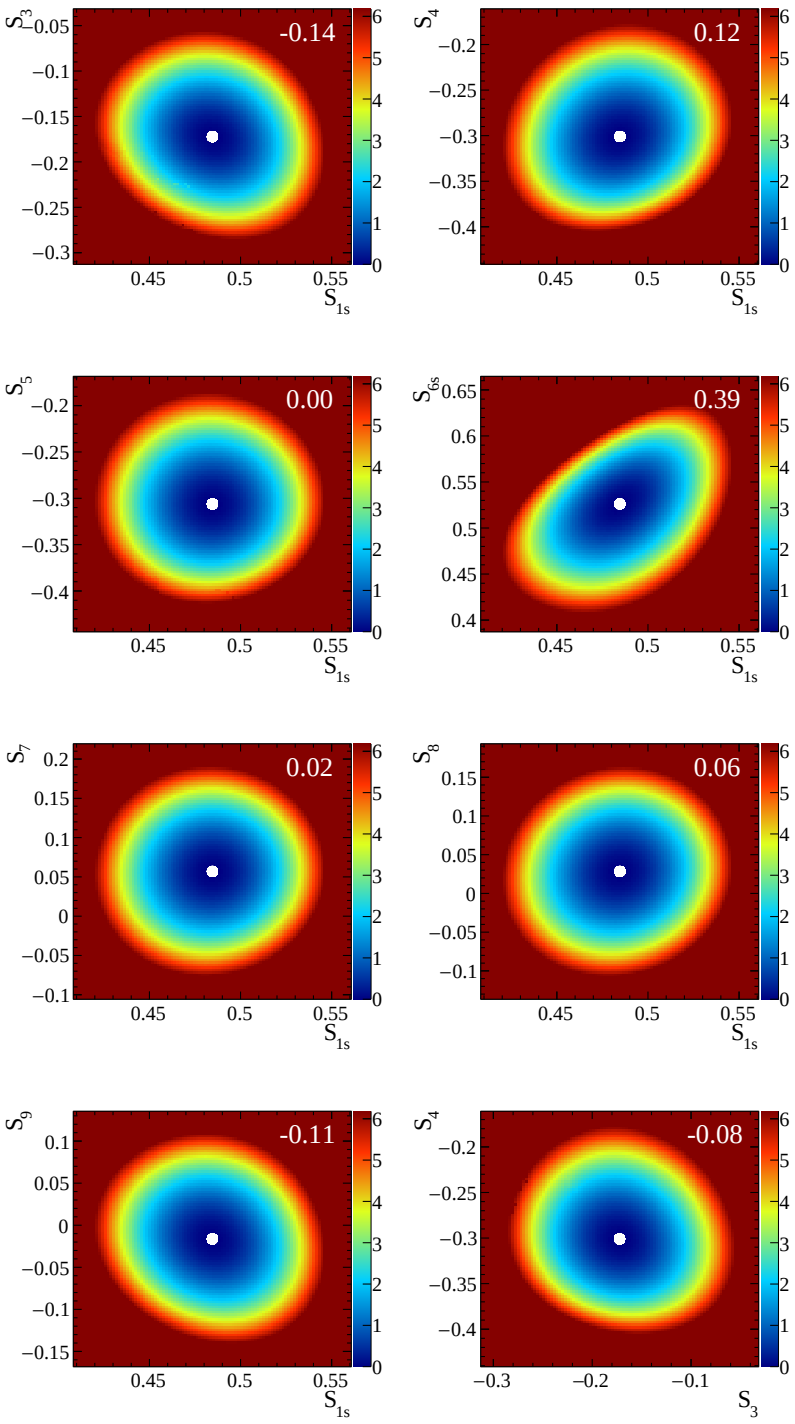

Figure 82: 2D profile log likelihood scans in the q^2 bin 15.0-17.0 GeV^2

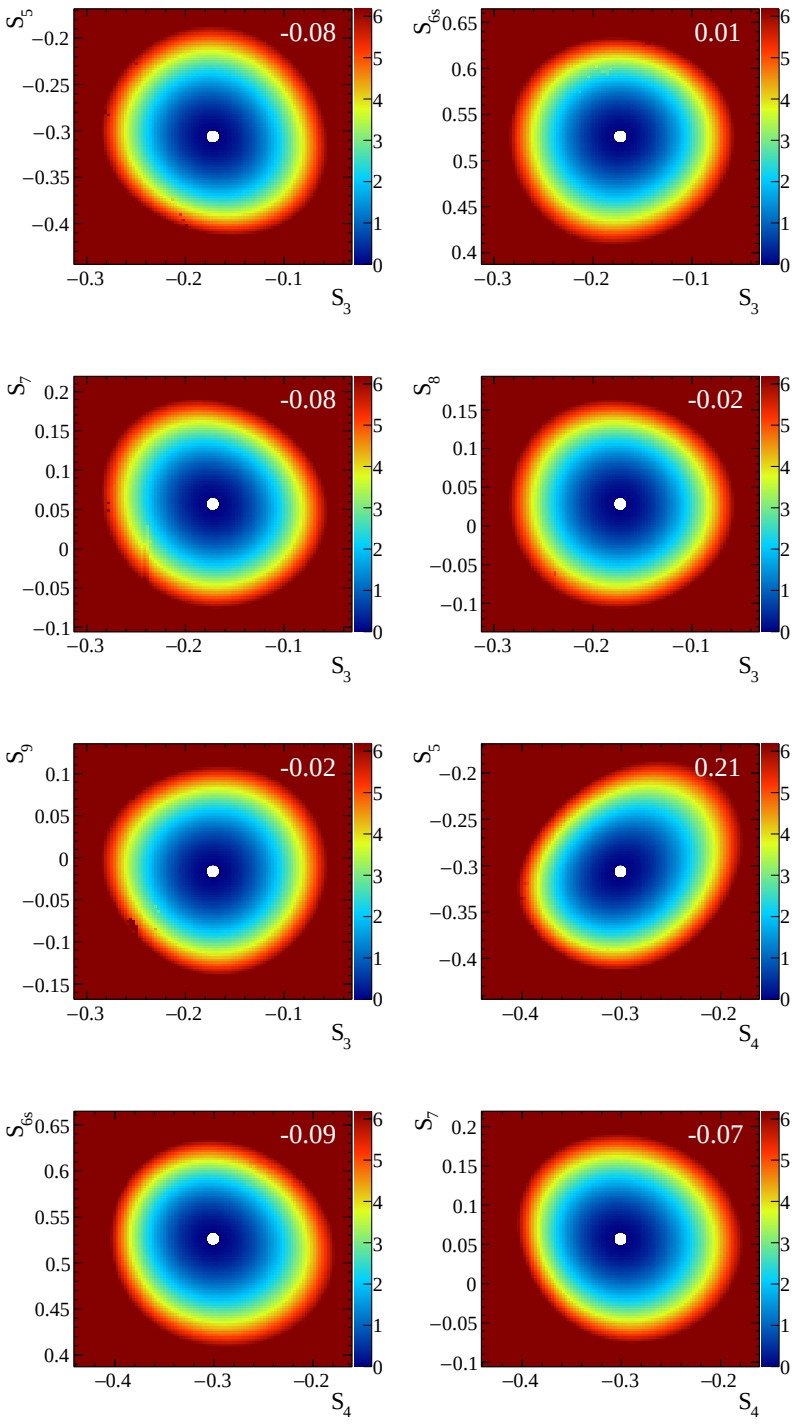

Figure 83: 2D profile log likelihood scans in the q^2 bin 15.0-17.0 GeV^2

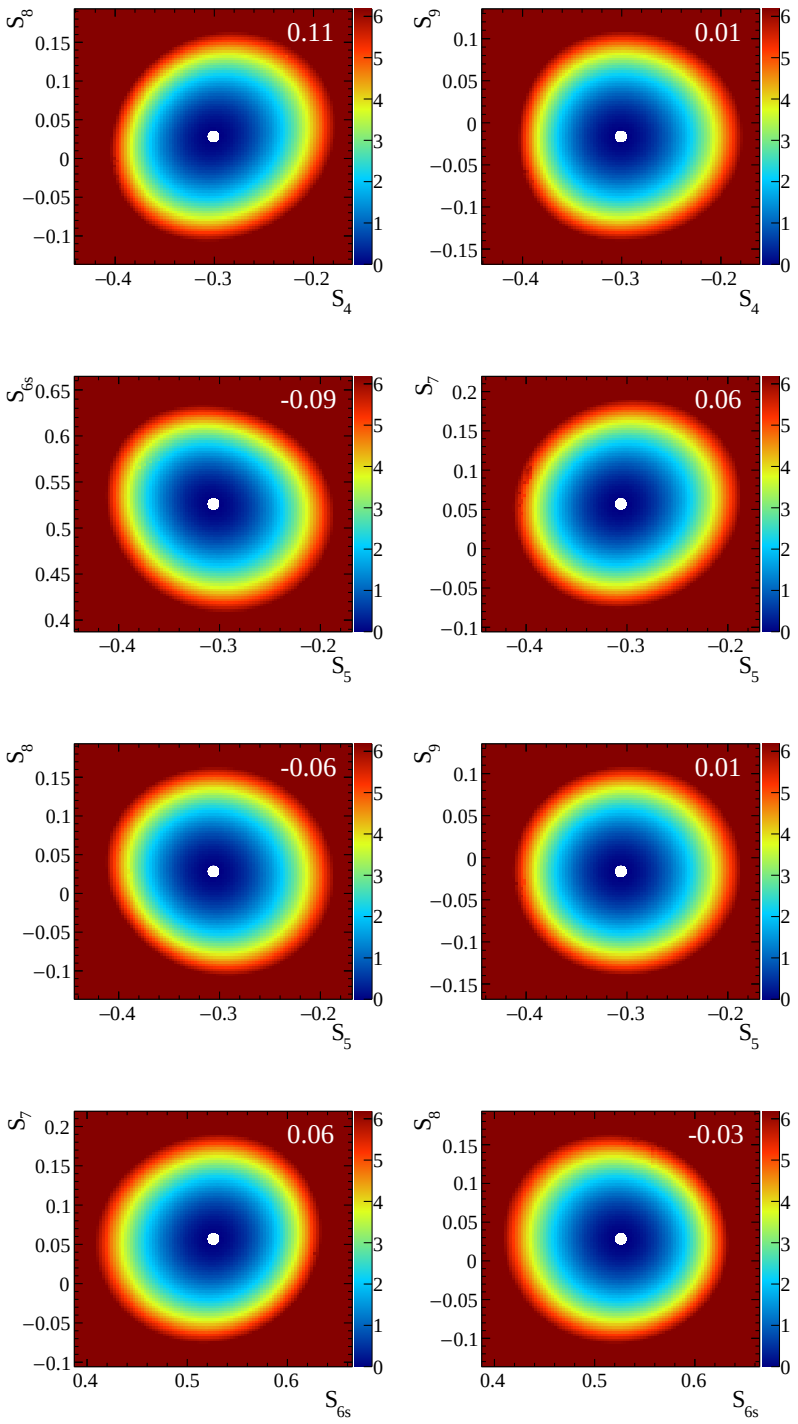

Figure 84: 2D profile log likelihood scans in the q^2 bin 15.0-17.0 GeV^2

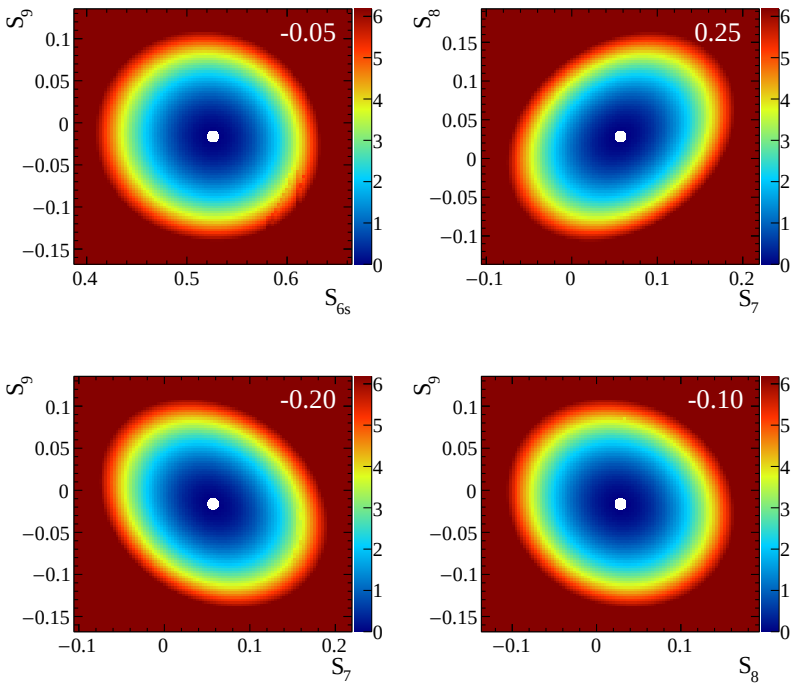

Figure 85: 2D profile log likelihood scans in the q^2 bin 15.0-17.0 GeV^2

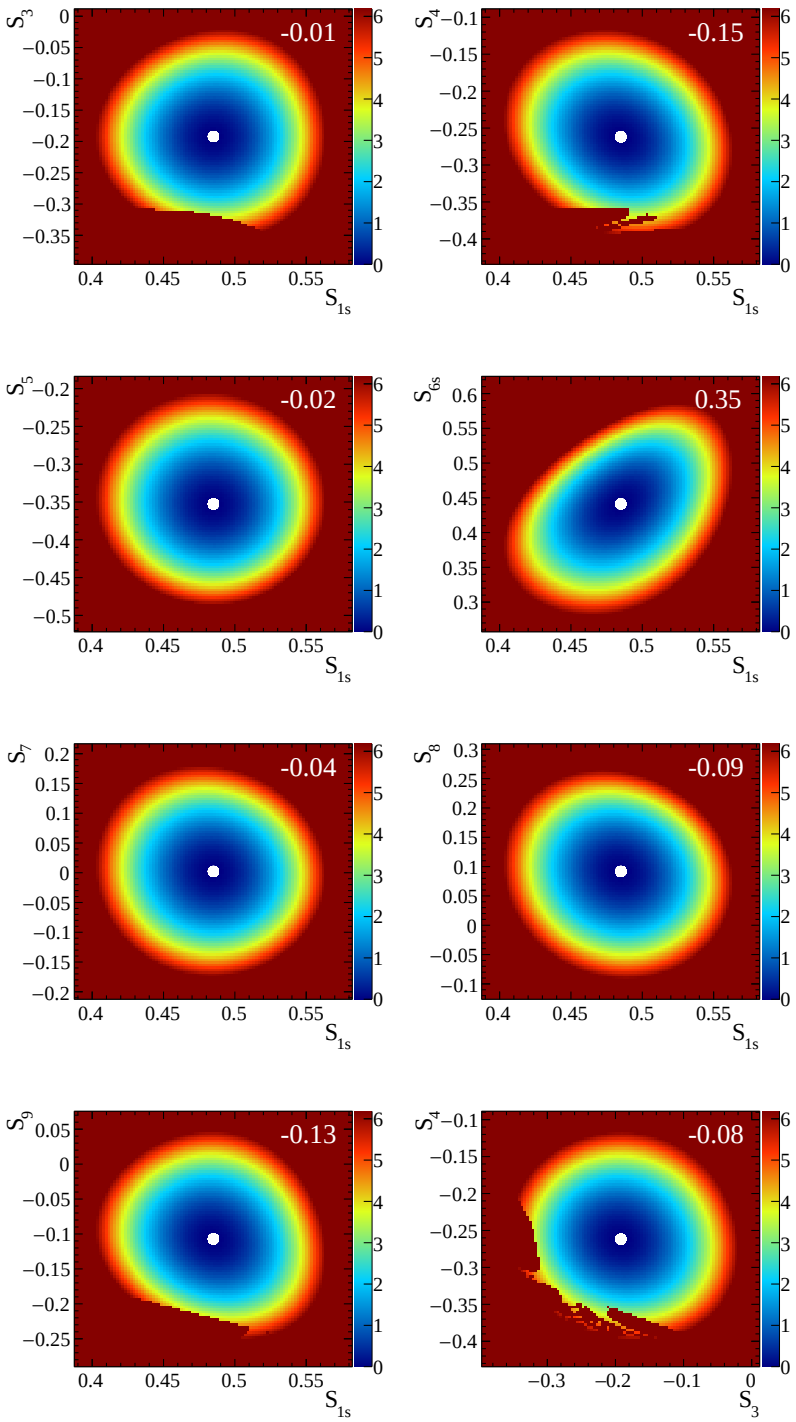

Figure 86: 2D profile log likelihood scans in the q^2 bin 17.0-19.0 GeV^2

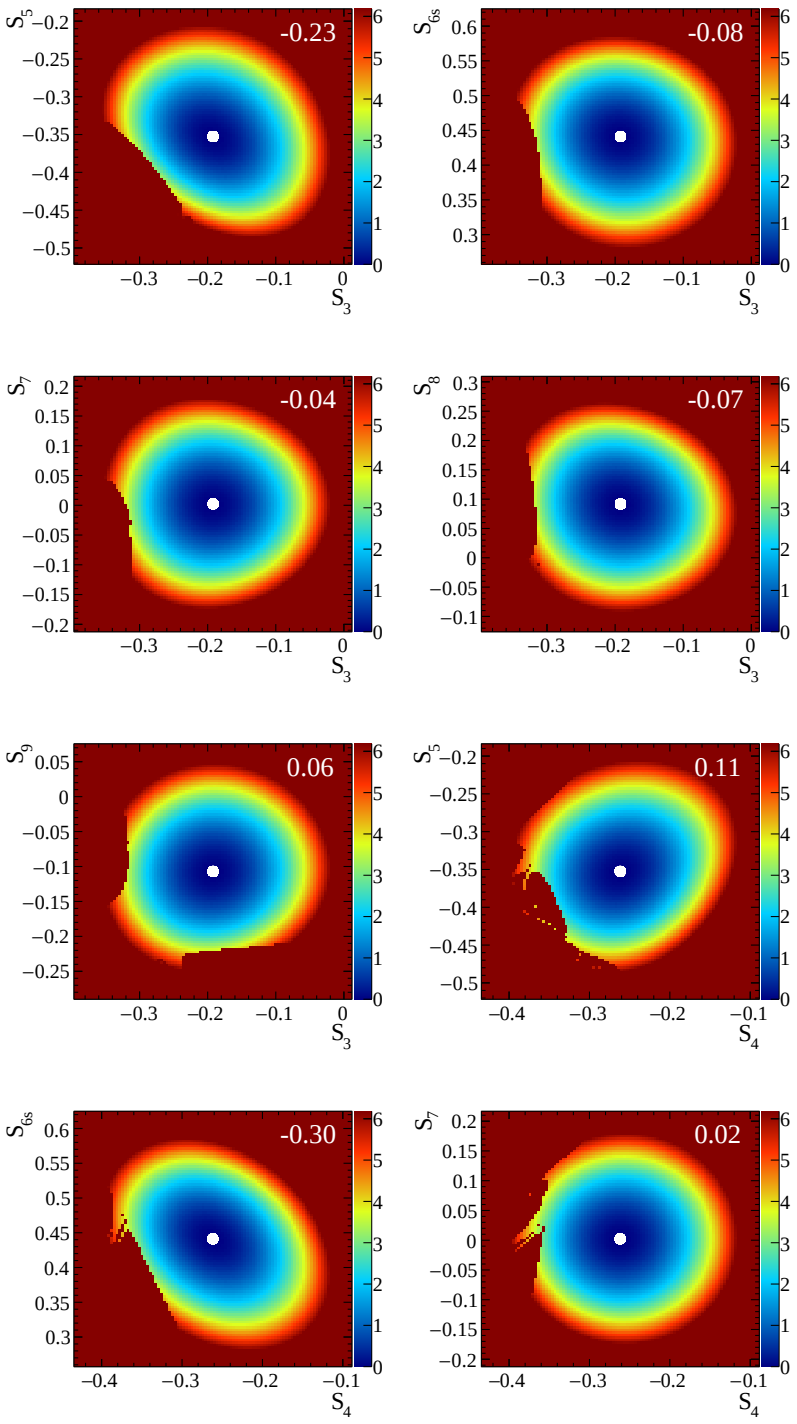

Figure 87: 2D profile log likelihood scans in the q^2 bin 17.0-19.0 GeV^2

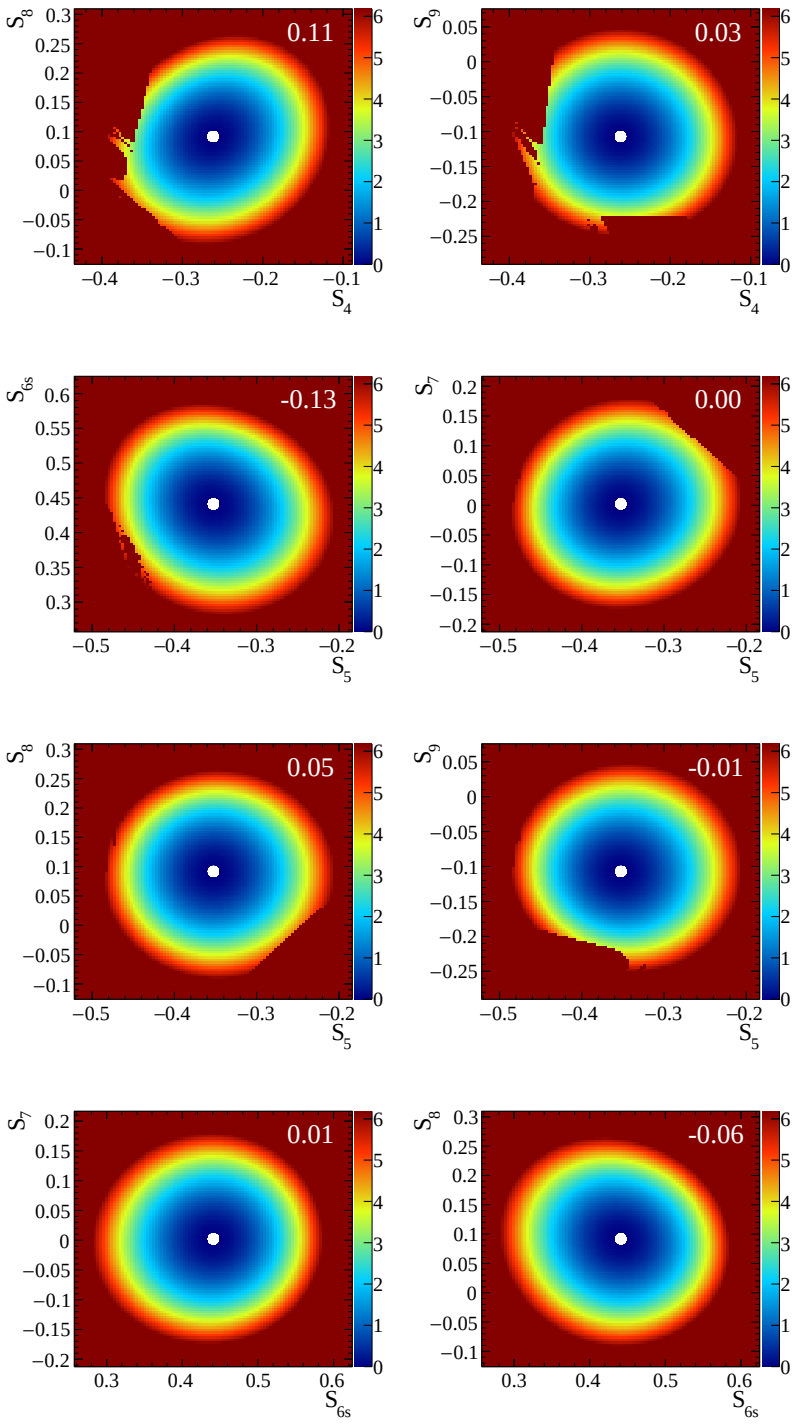

Figure 88: 2D profile log likelihood scans in the q^2 bin 17.0-19.0 GeV^2

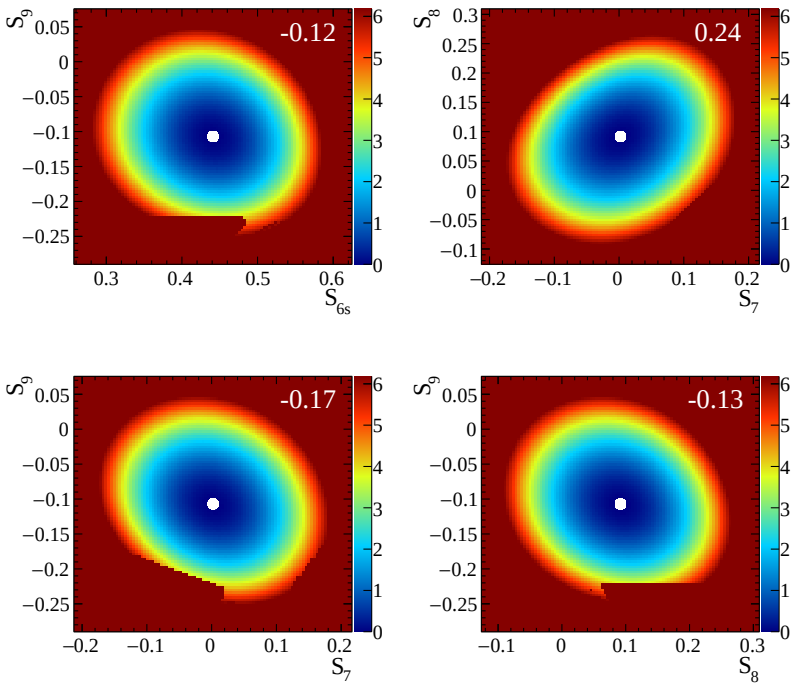

Figure 89: 2D profile log likelihood scans in the q^2 bin 17.0-19.0 GeV^2

C Systematic uncertainty tables

Tables of the estimated systematic uncertainty of the maximum likelihood fit in the different q^2 bins. The definition of the names and colour coding is the same as explained in Sec. 11.

Table 30: Systematic uncertainties of the maximum likelihood fit in the q^2 range $0.1 - 1.0\,\mathrm{GeV}^2$

Syst. Uncer.	S_{1s}	S_3	S_4	S_5	S_{6s}	S_7	S_8	S_9
stat. ang. acc.	0.0023	0.0001	0.0032	0.0002	0.0026	0.0002	0.0027	0.0002
stat. sim.	0.0076	0.0171	0.0154	0.0145	0.0168	0.0156	0.0166	0.0169
weight π p_T	0.0081	0.0007	0.0022	0.0031	0.0003	0.0003	0.0001	0.0002
weight K p_T	0.0002	0.0002	0.0007	0.0006	0.0008	0.0003	0.0001	0.0001
weight μ p_T	0.0002	0.0001	0.0002	0.0002	0.0002	0.0001	0.0001	0.0000
weight #tracks	0.0002	0.0001	0.0000	0.0001	0.0002	0.0000	0.0000	0.0000
weight B^0 vtx	0.0000	0.0001	0.0000	0.0001	0.0001	0.0000	0.0000	0.0000
weight q^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
tracking eff.	0.0007	0.0001	0.0001	0.0001	0.0007	0.0000	0.0000	0.0000
rec. eff. $K\pi$	0.0001	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000
ang. resolution	0.0012	0.0002	0.0015	0.0015	0.0001	0.0010	0.0019	0.0001
peaking bkg.	0.0003	0.0002	0.0001	0.0001	0.0012	0.0001	0.0001	0.0002
neglecting B_s^0	0.0004	0.0000	0.0001	0.0004	0.0002	0.0002	0.0000	0.0001
sig. mass model	0.0006	0.0002	0.0001	0.0007	0.0005	0.0000	0.0001	0.0001
bkg. ang. model	0.0008	0.0003	0.0001	0.0006	0.0000	0.0001	0.0000	0.0004
S-wave model	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000
bkg. $K\pi$ model	0.0001	0.0000	0.0001	0.0003	0.0001	0.0000	0.0000	0.0000
quadratic sum	0.0115	0.0171	0.0159	0.0150	0.0171	0.0156	0.0169	0.0169
exp. stat. uncer.	0.0379	0.0640	0.0775	0.0637	0.0899	0.0630	0.0813	0.0649

Table 31: Systematic uncertainties of the maximum likelihood fit in the q^2 range $1.1 - 2.5\,\mathrm{GeV}^2$

Syst. Uncer.	S_{1s}	S_3	S_4	S_5	S_{6s}	S_7	S_8	S_9
stat. ang. acc.	0.0014	0.0000	0.0010	0.0001	0.0018	0.0001	0.0023	0.0001
stat. sim.	0.0089	0.0144	0.0187	0.0203	0.0134	0.0208	0.0189	0.0148
weight π p_T	0.0075	0.0012	0.0004	0.0008	0.0066	0.0006	0.0005	0.0003
weight K p_T	0.0005	0.0002	0.0001	0.0005	0.0002	0.0001	0.0003	0.0001
weight μ p_T	0.0001	0.0001	0.0003	0.0002	0.0001	0.0001	0.0001	0.0000
weight #tracks	0.0000	0.0000	0.0001	0.0002	0.0000	0.0000	0.0000	0.0000
weight B^0 vtx	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000
weight q^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
tracking eff.	0.0008	0.0001	0.0001	0.0001	0.0007	0.0000	0.0002	0.0001
rec. eff. $K\pi$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ang. resolution	0.0005	0.0016	0.0022	0.0026	0.0002	0.0016	0.0018	0.0005
peaking bkg.	0.0001	0.0002	0.0000	0.0001	0.0006	0.0001	0.0000	0.0002
neglecting B_s^0	0.0003	0.0000	0.0001	0.0002	0.0001	0.0001	0.0000	0.0000
sig. mass model	0.0008	0.0002	0.0000	0.0004	0.0002	0.0001	0.0001	0.0001
bkg. ang. model	0.0005	0.0000	0.0000	0.0005	0.0006	0.0003	0.0001	0.0010
S-wave model	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
bkg. $K\pi$ model	0.0003	0.0000	0.0000	0.0002	0.0002	0.0001	0.0000	0.0000
quadratic sum	0.0118	0.0145	0.0189	0.0205	0.0151	0.0208	0.0191	0.0148
exp. stat. uncer.	0.0680	0.1003	0.1351	0.1067	0.1280	0.1086	0.1316	0.1065

Table 32: Systematic uncertainties of the maximum likelihood fit in the q^2 range $2.5 - 4.0\,\mathrm{GeV}^2$

Syst. Uncer.	S_{1s}	S_3	S_4	S_5	S_{6s}	S_7	S_8	S_9
stat. ang. acc.	0.0010	0.0000	0.0010	0.0001	0.0017	0.0001	0.0021	0.0001
stat. sim.	0.0079	0.0142	0.0183	0.0194	0.0119	0.0202	0.0183	0.0121
weight π p_T	0.0050	0.0010	0.0013	0.0014	0.0037	0.0001	0.0004	0.0006
weight K p_T	0.0002	0.0001	0.0003	0.0002	0.0001	0.0000	0.0001	0.0001
weight μ p_T	0.0000	0.0001	0.0001	0.0003	0.0000	0.0000	0.0000	0.0000
weight #tracks	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
weight B^0 vtx	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
weight q^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
tracking eff.	0.0005	0.0001	0.0002	0.0000	0.0004	0.0001	0.0000	0.0000
rec. eff. $K\pi$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ang. resolution	0.0016	0.0021	0.0002	0.0015	0.0005	0.0010	0.0020	0.0012
peaking bkg.	0.0002	0.0000	0.0001	0.0000	0.0008	0.0001	0.0000	0.0002
neglecting B_s^0	0.0006	0.0001	0.0002	0.0002	0.0001	0.0001	0.0001	0.0000
sig. mass model	0.0002	0.0001	0.0004	0.0004	0.0003	0.0001	0.0001	0.0001
bkg. ang. model	0.0005	0.0003	0.0005	0.0008	0.0004	0.0002	0.0001	0.0013
S-wave model	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
bkg. $K\pi$ model	0.0013	0.0000	0.0005	0.0009	0.0002	0.0002	0.0000	0.0000
quadratic sum	0.0096	0.0144	0.0184	0.0196	0.0126	0.0203	0.0185	0.0123
exp. stat. uncer.	0.0612	0.1147	0.1395	0.1189	0.1037	0.1227	0.1453	0.1243

Table 33: Systematic uncertainties of the maximum likelihood fit in the q^2 range $4.0 - 6.0\,\mathrm{GeV}^2$

Syst. Uncer.	S_{1s}	S_3	S_4	S_5	S_{6s}	S_7	S_8	S_9
stat. ang. acc.	0.0009	0.0000	0.0010	0.0001	0.0012	0.0001	0.0015	0.0001
stat. sim.	0.0072	0.0124	0.0140	0.0148	0.0117	0.0164	0.0159	0.0124
weight π p_T	0.0061	0.0014	0.0012	0.0004	0.0006	0.0002	0.0012	0.0001
weight K p_T	0.0005	0.0002	0.0004	0.0004	0.0001	0.0001	0.0004	0.0000
weight μ p_T	0.0002	0.0000	0.0002	0.0001	0.0001	0.0000	0.0000	0.0000
weight #tracks	0.0000	0.0001	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000
weight B^0 vtx	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
weight q^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
tracking eff.	0.0005	0.0001	0.0002	0.0001	0.0000	0.0001	0.0001	0.0000
rec. eff. $K\pi$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ang. resolution	0.0000	0.0011	0.0019	0.0015	0.0000	0.0023	0.0007	0.0002
peaking bkg.	0.0005	0.0002	0.0003	0.0008	0.0004	0.0001	0.0001	0.0000
neglecting B_s^0	0.0005	0.0001	0.0003	0.0004	0.0000	0.0001	0.0000	0.0001
sig. mass model	0.0004	0.0003	0.0007	0.0010	0.0001	0.0002	0.0001	0.0003
bkg. ang. model	0.0006	0.0001	0.0008	0.0015	0.0003	0.0003	0.0000	0.0010
S-wave model	0.0001	0.0000	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000
bkg. $K\pi$ model	0.0008	0.0000	0.0006	0.0012	0.0003	0.0001	0.0000	0.0000
quadratic sum	0.0096	0.0125	0.0142	0.0151	0.0118	0.0166	0.0161	0.0125
exp. stat. uncer.	0.0412	0.0730	0.0865	0.0823	0.0635	0.0769	0.0838	0.0681

Table 34: Systematic uncertainties of the maximum likelihood fit in the q^2 range $6.0 - 8.0\,\mathrm{GeV}^2$

Syst. Uncer.	S_{1s}	S_3	S_4	S_5	S_{6s}	S_7	S_8	S_9
stat. ang. acc.	0.0013	0.0000	0.0013	0.0001	0.0043	0.0009	0.0024	0.0001
stat. sim.	0.0067	0.0107	0.0112	0.0114	0.0101	0.0144	0.0138	0.0116
weight π p_T	0.0065	0.0015	0.0008	0.0005	0.0058	0.0013	0.0000	0.0005
weight K p_T	0.0004	0.0001	0.0000	0.0003	0.0003	0.0004	0.0000	0.0002
weight μ p_T	0.0004	0.0001	0.0001	0.0000	0.0004	0.0000	0.0000	0.0000
weight #tracks	0.0001	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
weight B^0 vtx	0.0001	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
weight q^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
tracking eff.	0.0003	0.0000	0.0001	0.0000	0.0003	0.0001	0.0000	0.0001
rec. eff. $K\pi$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ang. resolution	0.0000	0.0002	0.0001	0.0008	0.0005	0.0008	0.0009	0.0016
peaking bkg.	0.0004	0.0000	0.0010	0.0025	0.0013	0.0002	0.0001	0.0000
neglecting B_s^0	0.0003	0.0002	0.0003	0.0003	0.0003	0.0000	0.0001	0.0000
sig. mass model	0.0003	0.0004	0.0007	0.0009	0.0003	0.0001	0.0002	0.0000
bkg. ang. model	0.0010	0.0002	0.0008	0.0020	0.0011	0.0004	0.0001	0.0016
S-wave model	0.0001	0.0000	0.0001	0.0002	0.0001	0.0000	0.0000	0.0000
bkg. $K\pi$ model	0.0005	0.0000	0.0006	0.0013	0.0006	0.0001	0.0000	0.0000
quadratic sum	0.0095	0.0108	0.0114	0.0120	0.0126	0.0146	0.0140	0.0118
exp. stat. uncer.	0.0422	0.0726	0.0803	0.0812	0.0667	0.0796	0.0820	0.0735

Table 35: Systematic uncertainties of the maximum likelihood fit in the q^2 range $11.0 - 12.5\,\mathrm{GeV}^2$

Syst. Uncer.	S_{1s}	S_3	S_4	S_5	S_{6s}	S_7	S_8	S_9
stat. ang. acc.	0.0019	0.0001	0.0024	0.0001	0.0044	0.0003	0.0053	0.0006
stat. sim.	0.0068	0.0117	0.0085	0.0094	0.0100	0.0142	0.0132	0.0113
weight π p_T	0.0052	0.0014	0.0000	0.0014	0.0055	0.0000	0.0008	0.0007
weight K p_T	0.0003	0.0001	0.0000	0.0001	0.0003	0.0001	0.0002	0.0001
weight μ p_T	0.0005	0.0001	0.0000	0.0001	0.0006	0.0000	0.0001	0.0001
weight #tracks	0.0001	0.0000	0.0001	0.0001	0.0002	0.0000	0.0000	0.0000
weight B^0 vtx	0.0001	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
weight q^2	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
tracking eff.	0.0003	0.0000	0.0001	0.0001	0.0004	0.0000	0.0000	0.0000
rec. eff. $K\pi$	0.0000	0.0000	0.0000	0.0001	0.0001	0.0000	0.0000	0.0000
ang. resolution	0.0009	0.0014	0.0012	0.0030	0.0013	0.0008	0.0005	0.0005
peaking bkg.	0.0002	0.0001	0.0003	0.0001	0.0006	0.0002	0.0002	0.0000
neglecting B_s^0	0.0002	0.0001	0.0005	0.0006	0.0007	0.0001	0.0001	0.0000
sig. mass model	0.0003	0.0005	0.0012	0.0011	0.0003	0.0000	0.0000	0.0000
bkg. ang. model	0.0010	0.0001	0.0008	0.0017	0.0011	0.0003	0.0001	0.0007
S-wave model	0.0000	0.0001	0.0002	0.0005	0.0003	0.0000	0.0000	0.0000
bkg. $K\pi$ model	0.0000	0.0001	0.0004	0.0008	0.0007	0.0000	0.0000	0.0000
quadratic sum	0.0089	0.0119	0.0091	0.0102	0.0125	0.0143	0.0143	0.0114
exp. stat. uncer.	0.0356	0.0715	0.0747	0.0725	0.0695	0.0730	0.0758	0.0684

Table 36: Systematic uncertainties of the maximum likelihood fit in the q^2 range $15.0 - 17.0\,\mathrm{GeV}^2$

Syst. Uncer.	S_{1s}	S_3	S_4	S_5	S_{6s}	S_7	S_8	S_9
stat. ang. acc.	0.0017	0.0000	0.0020	0.0001	0.0012	0.0001	0.0015	0.0001
stat. sim.	0.0065	0.0116	0.0094	0.0097	0.0103	0.0137	0.0137	0.0118
weight π p_T	0.0026	0.0012	0.0000	0.0004	0.0022	0.0003	0.0003	0.0003
weight K p_T	0.0001	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0001
weight μ p_T	0.0004	0.0001	0.0000	0.0001	0.0003	0.0000	0.0000	0.0000
weight #tracks	0.0001	0.0001	0.0000	0.0000	0.0002	0.0000	0.0000	0.0000
weight B^0 vtx	0.0001	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
weight q^2	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
tracking eff.	0.0003	0.0000	0.0000	0.0001	0.0004	0.0000	0.0000	0.0000
rec. eff. $K\pi$	0.0001	0.0002	0.0001	0.0002	0.0002	0.0000	0.0000	0.0000
ang. resolution	0.0011	0.0002	0.0008	0.0015	0.0008	0.0009	0.0011	0.0007
peaking bkg.	0.0001	0.0003	0.0001	0.0002	0.0007	0.0001	0.0001	0.0000
neglecting B_s^0	0.0004	0.0001	0.0006	0.0006	0.0007	0.0000	0.0000	0.0001
sig. mass model	0.0003	0.0006	0.0010	0.0012	0.0005	0.0000	0.0000	0.0001
bkg. ang. model	0.0010	0.0004	0.0005	0.0010	0.0010	0.0003	0.0001	0.0007
S-wave model	0.0001	0.0004	0.0004	0.0007	0.0007	0.0000	0.0000	0.0000
bkg. $K\pi$ model	0.0000	0.0003	0.0004	0.0006	0.0007	0.0000	0.0000	0.0000
quadratic sum	0.0074	0.0117	0.0097	0.0100	0.0108	0.0137	0.0138	0.0118
exp. stat. uncer.	0.0303	0.0555	0.0634	0.0555	0.0518	0.0604	0.0598	0.0582

Table 37: Systematic uncertainties of the maximum likelihood fit in the q^2 range $17.0 - 19.0\,\mathrm{GeV}^2$

Syst. Uncer.	S_{1s}	S_3	S_4	S_5	S_{6s}	S_7	S_8	S_9
stat. ang. acc.	0.0028	0.0000	0.0027	0.0002	0.0021	0.0002	0.0023	0.0001
stat. sim.	0.0090	0.0153	0.0148	0.0141	0.0161	0.0187	0.0185	0.0172
weight π p_T	0.0006	0.0008	0.0003	0.0005	0.0003	0.0005	0.0002	0.0004
weight K p_T	0.0000	0.0002	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
weight μ p_T	0.0002	0.0002	0.0001	0.0000	0.0002	0.0000	0.0000	0.0000
weight #tracks	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
weight B^0 vtx	0.0001	0.0001	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000
weight q^2	0.0003	0.0001	0.0001	0.0001	0.0001	0.0000	0.0000	0.0000
tracking eff.	0.0001	0.0003	0.0000	0.0002	0.0001	0.0001	0.0000	0.0000
rec. eff. $K\pi$	0.0001	0.0005	0.0003	0.0004	0.0005	0.0000	0.0000	0.0000
ang. resolution	0.0013	0.0002	0.0019	0.0001	0.0014	0.0002	0.0044	0.0013
peaking bkg.	0.0003	0.0005	0.0002	0.0002	0.0003	0.0002	0.0001	0.0000
neglecting B_s^0	0.0003	0.0000	0.0007	0.0006	0.0003	0.0002	0.0001	0.0001
sig. mass model	0.0004	0.0002	0.0014	0.0014	0.0013	0.0002	0.0001	0.0000
bkg. ang. model	0.0010	0.0004	0.0006	0.0008	0.0008	0.0003	0.0001	0.0008
S-wave model	0.0000	0.0003	0.0001	0.0002	0.0001	0.0000	0.0000	0.0000
bkg. $K\pi$ model	0.0000	0.0003	0.0003	0.0003	0.0004	0.0000	0.0000	0.0000
quadratic sum	0.0097	0.0154	0.0153	0.0142	0.0164	0.0187	0.0192	0.0173
exp. stat. uncer.	0.0359	0.0717	0.0772	0.0679	0.0621	0.0737	0.0818	0.0684