Search for a long-lived spin-0 particle in $b \to s$ quark transitions at the Belle II experiment

Dissertation
zur Erlangung des Doktorgrades
an der Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Physik
der Universität Hamburg

vorgelegt von
Sascha Simon Dreyer
aus Göttingen

Hamburg
2023

Gutachter/innen der Dissertation:	Prof. Dr. Johannes Haller Prof. Dr. Torben Ferber
Zusammensetzung der Prüfungskommission:	Prof. Dr. Johannes Haller Prof. Dr. Torben Ferber Prof. Dr. Jochen Liske Dr. Krisztian Peters Dr. Kai Schmidt-Hoberg
Vorsitzender der Prüfungskommission:	Prof. Dr. Jochen Liske
Datum der Disputation:	21.12.2023
Vorsitzender Fach-Promotionsausschusses PHYSIK:	Prof. Dr. Günter H. W. Sigl
Leiter des Fachbereichs PHYSIK:	Prof. Dr. Wolfgang J. Parak
Dekan der Fakultät MIN:	Prof. Dr.-Ing. Norbert Ritter

Abstract

Phenomena such as dark matter can be connected to the standard model via additional spin-0 mediators. This doctoral thesis describes a search for a long-lived spin-0 particle S in rare B-meson decays mediated by $b \to s$ quark transitions.

A dataset corresponding to an integrated luminosity of $189\,\text{fb}^{-1}$ of e^+e^- collisions collected at the $\Upsilon(4S)$ resonance energy by the Belle II experiment is analysed. The search is carried out in eight exclusive channels, the two modes of production $B^+ \to K^+S$ and $B^0 \to K^{*0}\,S$ with decays via $S \to e^+e^-/\,\mu^+\mu^-/\,\pi^+\pi^-/\,K^+K^-$. Decays of the S to standard model particles are motivated if the decay to dark matter is kinematically not possible. Mediator masses between $(0.025 - 4.78)\,\text{GeV}/c^2$ and lifetimes between $(0.001 - 400)\,\text{cm}$ are probed. Long-lived particle reconstruction is studied and validated with K^0_S-mesons. Maximum likelihood fits to the reconstructed S mass distribution are used to determine the signal yield.

No evidence for the signal process is found. Model-independent upper limits are derived on the branching fractions of the signal processes. The upper limits extend down to the order of $\mathcal{O}(10^{-7})$. These are the first constraints on hadronic final states of the S produced in $B \to KS$ and the most stringent bounds from a direct search for $S \to e^+e^-$ at e^+e^- colliders.

The results are interpreted in models that predict a dark Higgs-like scalar and an axionlike particle with fermion couplings. The model-dependent bounds are competitive with existing experimental constraints.

Zusammenfassung

Phänomene wie dunkle Materie können durch zusätzliche Mediatorteilchen mit Spin-0 mit dem Standardmodell verbunden werden. Diese Doktorarbeit beschreibt eine Suche nach einem langlebigen Teilchen mit Spin-0, S, in seltenen B-Meson Zerfällen via $b \to s$ Quarkübergänge.

Ein Datensatz enstprechend einer integrierten Luminosität von $189\,\text{fb}^{-1}$ von e^+e^- Kollisionen bei der $\Upsilon(4S)$ Resonanzenergie, aufgezeichnet vom Belle II Experiment, wird analysiert. Die Suche wird in acht exklusiven Kanälen durchgeführt, bestehend aus den beiden Produktionskanälen $B^+ \to K^+S$ und $B^0 \to K^{*0}\,S$, und Zerfällen via $S \to e^+e^-\,/\,\mu^+\mu^-\,/\,\pi^+\pi^-\,/\,K^+K^-$. Zerfälle vom S in Standardmodellteilchen sind motiviert, wenn der direkte Zerfall in dunkle Materie kinematisch nicht möglich ist. Mediatormassen zwischen $(0.025 - 4.78)\,\text{GeV}/c^2$ und Lebenszeiten zwischen $(0.001 - 400)\,\text{cm}$ werden untersucht. Die Rekonstruktion von langlebigen Teilchen wird mithilfe von K^0_S-Mesonen studiert und validiert. Die Anzahl an Signalkandidaten wird durch Fits mit der maximum-likelihood-Methode an die rekonstruierte S Massenverteilung bestimmt.

Kein statistisch signifikanter Hinweis auf die Signalprozesse wird gefunden. Modellunabhängige obere Ausschlussgrenzen werden für die Verzweigungsverhältnisse der Signalprozesse bestimmt. Die Ausschlussgrenzen erstecken sich bis runter auf die Größenordnung $\mathcal{O}(10^{-7})$. Diese sind die ersten Beschränkungen für hadronische Endzustände des S produziert in $B \to KS$, und die strengsten einer direkten Suche nach $S \to e^+e^-$ an e^+e^- Beschleunigern.

Die Ergebnisse werden in Modellen interpretiert, welche ein Higgs-ähnliches skalares Teilchen und ein Axion-ähnliches Teilchen mit Kopplungen an Fermionen vorhersagen. Die modellabhängigen Beschränkungen sind konkurrenzfähig mit existierenden experimentellen Ausschlussgrenzen.

Contents

1. Introduction

How do we find clues of long-lived particles buried within hundreds of millions of electron-positron collisions? Could these particles aid our attempts to understand dark matter and other new physics phenomena? These are the type of questions that the search presented in this thesis aims to answer.

Searches for new physics are motivated by the fact that our theory of elementary particles and their interactions, the Standard Model [1] (SM), is very successful, albeit not complete. One of the most important pieces of evidence for this incompleteness is dark matter [2] which is found to contribute by 85% to the total matter content in the universe [3] but is not described by the SM. Dark matter is an elusive form of matter which is only observed to interact via gravitation with SM particles. However, the exact nature of dark matter is not yet known. Particle dark matter [4] with new non-gravitational interactions extending those of the SM is a compelling option. The same quantum field theory approaches used for SM physics can be used for its description. Furthermore, a search at existing particle collider experiments is possible.

This thesis presents a search for a long-lived spin-0 particle S. The S can act as a mediator between the SM and dark matter and is not a candidate for dark matter itself. The S signal is searched for via its couplings to SM particles in rare decays of B-mesons that are produced in e^+e^- collisions at the Belle II experiment in Tsukuba, Japan. Many extensions of the SM that predict the existence of additional spin-0 particles assume Yukawa-like couplings, similar to the Higgs boson, which is the only elementary spin-0 particle in the SM. This coupling structure motivates the production of the S through radiation from heavy virtual particles in the $b \to s$ process.

The dataset analysed in this work was collected in 2019 and 2020. The search is car-

ried out in eight exclusive channels with the production of the S through $B^+ \to K^+S$ and $B^0 \to K^{*0}(\to K^+\pi^-)S$. Charge conjugated decays are implied throughout this thesis, and K^{*0} refers to the neutral $K^*(892)$ meson. Visible decays of the S via $S \to e^+e^-/\,\mu^+\mu^-/\,\pi^+\pi^-/\,K^+K^-$ are considered in the search. Decays of the S to dark matter would manifest as an invisible signature in Belle II and are assumed to be absent, e.g. due to kinematically inaccessible decays to dark matter. The couplings of the S to the SM are small enough for the S to be long-lived on a scale such that its decay vertex is sufficiently displaced from the e^+e^- collision region to be resolved. The decay products of the S are fully reconstructed using the tracking subdetectors of Belle II. Very large lifetimes cannot be probed since too large of a fraction of decays take place outside the detector. The displaced S vertex is a particular experimental signature. Is is used as a powerful tool to suppress backgrounds and thereby increase the sensitivity to the signal process. However, displaced vertices were relatively unexplored in Belle II at the time of this search. Therefore, a significant part of this work is dedicated to understand and validate long-lived particle (LLP) reconstruction performance.

This thesis is organised as follows:

- Chapter 2 provides the context of the presented work.
- Chapter 3 describes the experimental setup.
- Chapter 4 gives an overview of the search strategy and details of the datasets.
- Chapter 5 presents the signal selections that suppress background.
- Chapter 6 introduces the templates used to describe the signal and background shapes.
- Chapter 7 describes the performance of LLP and additional correction factors.
- Chapter 8 presents the validation of the simulation of backgrounds and of a process close to the S.
- Chapter 9 gives an overview of the systematic uncertainties.

- Chapter 10 describes the statistical treatment of the results.
- Chapter 11 presents the results and their interpretation.
- Chapter 12 concludes this thesis with a summary and gives an outlook for future searches.

2. Physics overview

This chapter motivates the search presented in this thesis and introduces the theoretical models for physics beyond the SM that are used to interpret the results.

2.1. Physics motivation

The currently observed abundance of dark matter in the universe is an important constraint for extensions of the SM. Particle dark matter models often use a thermal freeze-out mechanism [5] to set the relic abundance, although alternatives are studied as well [6]. In these theories, the candidate for dark matter is often a Dirac fermion χ. The freeze-out is described by thermal equilibrium between dark matter and SM particles in the early universe. This is characterised by equal rates of dark matter annihilating into SM particles and vice versa. The expansion of the universe leads to cooling. At temperatures below the dark matter mass, it cannot be produced by the SM thermal bath. From this point on, the annihilation of dark matter reduces its abundance. The annihilation slows down with the expansion of the universe leading to constant observable abundance which depends on the annihilation cross-section. However, simple dark matter models used for benchmarking of experimental results do not necessarily need to predict the correct relic abundance. These models can be part of more complex dark matter scenarios with additional processes altering the relic abundance [7].

Searches for dark matter at particle colliders aim to measure the direct production of dark matter. Mediator particles that correspond to a new interaction connect the SM to dark matter. Dark matter particles are invisible to particle detectors and manifest as an imbalance of energy in the final state. Other measurements focus on

the mediator and search for its decay to SM particles, usually via the same coupling that governs its production.

In addition to collider searches there are two more classes of experimental techniques to search for dark matter, direct and indirect detection. Most galaxies seem to reside in larger concentrations of dark matter, termed halo. Direct detection experiments aim to measure the scattering of dark matter particles from the galactic halo with detector material. Indirect detection experiments are used to search for the products of dark matter annihilation in the spectra of cosmic particles.

Apart from dark matter, beyond the SM (BSM) mediators can provide explanations to further observed deviations from the SM such as the anomalous magnetic moment of the muon [8]. Dark sectors are introduced as a generic term for BSM physics that is coupled to the SM via not yet discovered interactions, usually with strengths well below the electroweak coupling.

Earlier collider searches focused on weakly interacting massive particle (WIMP) [9]–[11] dark matter. This was motivated by WIMPs naturally predicting the correct relic density for dark matter masses around the weak scale, which can be tested well at the Large Hadron Collider (LHC).

Light dark sectors [12]–[17] feature couplings weak enough to have evaded detection and receive interest in the community due to the nonobservation of WIMPs [18].

In particular, mediators with masses of order $\mathcal{O}(\,\mathrm{MeV}/c^2 - \mathrm{GeV}/c^2)$ can be searched for in e^+e^- collisions at the Belle II experiment. Corresponding efforts are categorised by the type of mediator. These include the vector portal [19]–[21] with additional mediator particles such as dark photons and Z'-bosons, the neutrino portal [22]–[24] with heavy neutral leptons, and the scalar and pseudoscalar portals as discussed in Sec. 2.2.

Searches for light dark sectors at Belle II are facilitated by the clean e^+e^- collision environment, full knowledge of the initial state, and dedicated event trigger categories.

The mediators searched for at Belle II can be directly produced in the e^+e^- collision, with visible decays to SM particles [25], invisible decays to dark matter or SM neutrinos [26], or with final states combining visible and invisible signatures [27]. Further options are the production in τ-lepton decays [28], or in meson decays, as

presented in the following.

2.1.1. Long-lived particles

Theories of BSM physics often predict new particles. An interesting class of such particles are LLPs [29], [30].

Long-lived refers to lifetimes large enough to be resolved experimentally. Faster decays are generally termed prompt. Long lifetimes are an effect of small decay rates which can arise through decays via heavy off-shell mediators, a reduced available phase space for the decay, or small values of coupling parameters. Muons acquire a long lifetime through their decay mediated by the exchange of a heavy W-boson. The difference between the neutron and proton masses is small. The decay of neutrons is hence suppressed by the small available phase space, leading to a long lifetime. Underlying effects that lead to small couplings in a BSM scenario can also be approximate symmetries or additional particles at larger energy scales. The BSM models that are used in this work for interpretation of the results provide examples of long lifetimes due to small couplings.

SM particles such as electrons and photons are stable, and muons, pions, and kaons exhibit a long lifetime. Particle detectors in collider experiments are used to measure the energy depositions of these particles. Shorter-lived particles like B- and D-mesons or τ-leptons are detected by their decays to more long-lived particles. Highly granular detectors close to the interaction point can resolve their decay vertices. This allows for efficient identification of these particles and strong rejection of prompt backgrounds. Flavour physics experiments such as Belle II use this to measure lifetimes directly [31]–[34] or measure the violation of charge-parity (CP) [35], [36]. At the Large Hadron Collider (LHC), sprays of particles, reconstructed as jets, can be identified to originate from b- or c-quarks [37], [38].

SM particles with larger masses, such as the electroweak W/Z-bosons, the top quark, and the Higgs boson, decay promptly. Early searches for new heavy resonances at the LHC hence focused on prompt decays, but found no evidence [1]. LLPs, however, can easily be missed if their particular experimental signatures are not explicitly taken into account. Possible signatures include tracks and decay vertices displaced

from the interaction point, late timings, tracks with kinks, and emergent showers in the calorimeter or muon systems [30], [39], [40]. Hence, there is growing interest for LLPs in the community and a rise in corresponding searches.

Recent searches for LLPs were performed at the LHC [41]–[46], at Belle [47], and BaBar [48], [49]. The analysis presented in this thesis is the first search for LLPs at Belle II. Dedicated experiments such as FASER [50] are already taking data or are proposed and currently discussed within the community [29], [51]–[53]. In the context of this work, a LLP extension to Belle II, the GAZELLE [54] detector, was studied. However, LLP searches were also part of the physics programme of previous experiments. Examples include efforts at LEP [55], [56], at the Tevatron [57], [58] and at beam-dump experiments [59].

2.1.2. $b \to s$ quark transitions

Flavour-changing neutral current (FCNC) $b \to s$ quark transitions occur within the SM only through suppressed $W - t$ loops [60]. Therefore, the corresponding meson-level $B \to KX$ decays are rare and present an opportunity to search for small deviations from the SM with small background levels. The kaon arises from the hadronisation of the s-quark and can be the lighter pseudoscalar mesons K or heavier mesons with potentially different structures, such as the vector K^*. The LHCb collaboration observed anomalies in $b \to s$ transitions accompanied by two charged leptons $\ell^+\ell^-$ produced by an off-shell photon or Z-boson radiated from the loop. Branching fractions for $B \to K\ell^+\ell^-$ and $B \to K^*\ell^+\ell^-$ are of the order of 10^{-7} and 10^{-6}, respectively [1]. Discrepancies were found between the data and the SM predictions for the branching fractions $\mathcal{B}(B^{+(0)} \to K^{+(*0)}\mu^+\mu^-)$ [61] and in angular observables [62]. The theoretically cleanest observables are ratios of branching fractions into pairs of charged leptons from different generations. These correspond to tests of lepton flavour universality [63]. The LHCb collaboration also found significant discrepancies in these ratios [64]–[67]. However, a simultaneous reanalysis of some of the ratios was performed recently, which yielded perfect agreement between SM and observations [68].

This thesis presents a search for LLPs produced on-shell in $b \to s$ transitions. Due to their small predicted couplings, they cannot explain the size of any of the observed

discrepancies. However, the potential contributions of new physics in $b \to s$ transitions and the clean experimental signature provide strong motivation for the search for the production of BSM mediators such as LLPs.

A related search involves the SM process $b \to s\nu\overline{\nu}$, which has not been observed [69], and could be enhanced by new physics [70]. In the SM the neutrinos are produced by a Z-boson radiated from the loop. A similar signature arises if a LLP is produced in $b \to s$ transitions and decays to dark matter or outside of the detector acceptance. Detector imperfections, such as inefficiencies in the reconstruction, can have the same effect. However, a difference arises due to the two-body kinematics of the on-shell LLP production compared to the three-body kinematics of $\ell^+\ell^-/\nu\overline{\nu}$ production in SM $b \to s$ processes via off-shell gauge bosons. Other physics processes close to the signal searched for in this work are $s \to d$ transitions with a FCNC between lighter quark flavours.

2.2. Phenomenological models

The main results of this search are model-independent upper limits on the signal branching fractions for an electrically neutral LLP with spin-0 produced in $b \to s$ transitions. The upper limits are derived as a function of the S mass and for different S lifetimes. They can be interpreted in BSM models that predict the signal process. The limits on the branching fraction then translate into excluded regions in the plane of model parameters.

The kinematics of the spin-0 signal allow for interpretations of the results in scalar and pseudo-scalar portal models. These interactions are simple extensions of the SM that have natural connections to the SM Higgs sector. The SM Higgs boson is the only elementary particle with spin-0. However, not all its couplings have been measured yet, and the precision of the measured couplings leaves room for BSM effects to enter [71], [72]. BSM models with additional spin-0 particles are therefore very interesting. Two models that could give rise to the signal process, a dark Higgs-like scalar and an axionlike particle with fermion couplings, are used for the interpretation of the results of this search and described later in this section.

Results in such specific BSM models should not be considered as definite exclusions of additional scalar or pseudo-scalar particles. More complex coupling structures, additional symmetries, or larger numbers of additional particles can extend the models and modify the numerical values of the results. Examples of these include:

- Inelastic dark matter models with two mediators, a scalar LLP and a dark photon, and two dark matter states [40], [73]–[75],
- two Higgs-doublet models and their variations, introducing a larger number of additional scalar and pseudo-scalar particles, including observable charged states [76]–[80],
- supersymmetrical extensions with partners of opposite spin for every SM particle [81]–[83], which can also incorporate two Higgs-doublet models.

The BSM models investigated in this thesis are chosen as rather simple representatives that provide an opportunity to benchmark existing experimental bounds. They can furthermore be used to guide future experimental efforts for searches for BSM physics.

Future reinterpretations in any specific model that predicts the signal process can be performed with the model-independent limits derived in this search. Model-independent limits on the signal branching fractions are provided separately for different lifetimes. This effectively decouples the production from the decay of the LLP [84], and thus can help distinguish between different dark sectors [85]. The inclusion of two different production modes $B^+ \to K^+ S$ and $B^0 \to K^{*0} S$ increases the number of predicted signal candidates, and hence the sensitivity. A comparison of the relative rates in both production modes allows one to distinguish between spin-0 and spin-1 mediators [86].

2.2.1. Dark Higgs-like scalar

A simple model that gives rise to a LLP that can be produced in $b \to s$ transitions is a real scalar field ϕ_S that mixes with the SM Higgs boson through the scalar portal.

The additional terms in the Lagrangian $\mathcal{L}$ are:

$$\mathcal{L}_{\mathrm{BSM}} = -\frac{1}{2}m_{\phi_S}^2\phi_S^2 - \lambda_3|H|^2\phi_S - y_\chi\bar{\chi}\chi\phi_S - \frac{1}{2}m_\chi\bar{\chi}\chi, \tag{2.1}$$

where H is the SM Higgs doublet, and χ describes a Dirac fermion field. χ is a singlet under the SM gauge group and serves as a stable dark matter candidate. Both χ and ϕ_S are electroweak and strong singlets. The dark matter candidate χ is charged under a discrete $\mathbb{Z}_2$ symmetry to avoid mixing with neutrinos. After electroweak symmetry breaking, this yields the observed SM Higgs boson h and a dark scalar S. The model features two free parameters that are both experimentally accessible: the mass of the physical dark scalar m_S and the mixing angle between the SM Higgs boson and the dark scalar θ_S. The mixing angle θ_S is defined by:

$$\begin{aligned} \sin^2\theta_S &= \frac{1}{2}\left(1 + \frac{m_{\phi_S}^2 - m_{h_0}^2}{\Delta m^2}\right), \\ \left(\Delta m^2\right)^2 &= 4\nu^2\lambda_3^2 + \left(m_{\phi_S}^2 - m_{h_0}^2\right)^2, \end{aligned} \tag{2.2}$$

with the neutral component of the SM Higgs doublet h_0 and its vacuum expectation value ν. The masses of the physical states are then given by:

$$\begin{aligned} m_S^2 &= \frac{1}{2}\left(m_{\phi_S}^2 - m_{h_0}^2 - \Delta m^2\right), \\ m_h^2 &= \frac{1}{2}\left(m_{\phi_S}^2 - m_{h_0}^2 + \Delta m^2\right) \approx \left(125\,\mathrm{GeV}/c^2\right)^2. \end{aligned} \tag{2.3}$$

The sensitivity of Belle II to this model with the expected full dataset was studied in Refs. [86], [87].

The correct thermal relic abundance can be achieved for dark matter masses $2m_\chi > m_S$ via the $\chi\chi \to SS$ and $\chi\chi \to S \to$ SM annihilation processes in the early universe [88], [89].

The dark scalar can also be used to break an additional dark gauge symmetry like $U'(1)$ to generate mass for the corresponding dark photon and dark matter states similar to electroweak symmetry breaking in the SM [40].

The couplings of S to SM particles resemble the couplings of h weighted by the sine

of the mixing angle $\sin\theta_S$. The production of the dark scalar in $b \to s$ transitions works via radiation of the dark scalar from the top quark or W-boson in the loop. The coupling parameter θ_S affects both the production rate and the decay rate and therefore the lifetime. For a given mass of the dark scalar, a specific value of the production rate fixes its lifetime, and vice versa.

The predicted values for the dark scalar production and decay rates used here are taken from Ref. [87]. In this study, results from Ref. [90] are used to calculate the hadronic decay rates of the dark scalar. The decay rate of S to dark matter particles is assumed to be zero. This is achieved i.e. by heavy dark matter candidates that are kinematically not accessible. There are no theoretical predictions for the decay rate to pairs of pions and kaons above masses of $m_S > 2\,\mathrm{GeV}/c^2$ because the method used at lower masses ceases to be valid, leading to large uncertainties [90]. The predictions for the lifetime above this point take into account the dark scalar decay width to hadronic final states through the decay to light quark and gluon pairs. The dark scalar can also mix with the scalar f_{980} resonance [90]. This results in an increased decay rate into pairs of pions and hence a reduced lifetime for dark scalar masses close to $m_{f_{980}}$. The calculation of the production rate of the dark scalar in $B^+ \to K^+ S$ uses form factors from lattice QCD [91]. Form factors from light-cone sum rules [92] are used in $B^0 \to K^{*0} S$.

The dark scalar production and decay branching fractions for the channels searched for in this work as well as their lifetimes for representative mixing angles are shown in Fig. 2.1. Production, individual decay, and total decay rates scale with $\sin^2\theta$. Therefore, the lifetime is proportional to $\sin^{-2}\theta$. The values of the decay branching fractions of the S are independent of the mixing angle, since they are ratios of decay rates. The depicted branching fractions do not add up to one as the rates in photons, light quarks, and gluons above $2\,\mathrm{GeV}/c^2$ are not shown. Decays into two photons are possible via a top-quark loop similar to the SM Higgs boson [93]. The LLP $\to \tau^+\tau^-$ channel is experimentally more challenging due to the additional invisible neutrinos in the final state and is not included in this search.

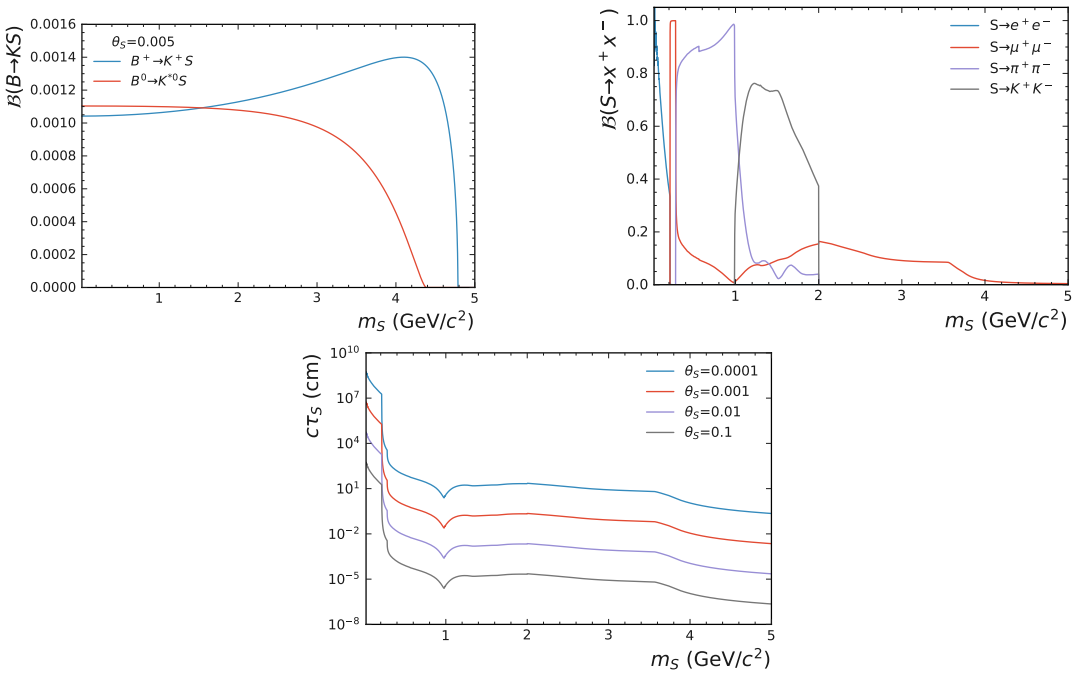

Figure 2.1.: Model predictions for the dark scalar S (upper left) production branching fraction in B-meson decays, (upper right) decay branching fractions into final states searched for here, and (lower) lifetime for different values of the mixing angle θ_S.

2.2.2. Axionlike particle model

The second model considered here uses the pseudoscalar portal in the form of an axionlike particle (ALP) to couple new physics to the SM. It is identified by the Physics Beyond Collider (PBC) working group as the BC10 benchmark model [94]. The BC10 model assumes couplings to fermions to dominate, couplings to gauge bosons are hence not included. Sensitivities to this model from different experiments, especially in the flavour sector, were studied [84], [95]. The corresponding additional terms in the Lagrangian with the ALP field a read:

$$\mathcal{L}_{\text{BSM}} = -\frac{1}{2}m_a^2 a^2 + \frac{\delta_\mu a}{f_l}\sum_\alpha \bar{l_\alpha}\gamma_\mu\gamma_5 l_\alpha + \frac{\delta_\mu a}{f_q}\sum_\beta \bar{q_\beta}\gamma_\mu\gamma_5 q_\beta. \tag{2.4}$$

The relevant parameters of the model are m_a, f_q and f_l. Furthermore, the couplings to leptons and quarks are set to be equal $f_l = f_q$. The dimensionless coupling parameter $g_Y = 2v/f$ normalised by the SM Higgs vacuum expectation value $v = 246.2\,\text{GeV}$ is used to present the results of this search. For small coupling values of g_Y the ALP becomes long-lived because of its small decay rate. The structure of the couplings to fermions leads to Yukawa-like couplings proportional to the fermion mass. This appears naturally if the ALP arises from an extended Higgs sector [96]. Couplings to fermionic dark matter candidates would take a similar form as the couplings to SM fermions.

The production of the ALP in $b \to s$ transitions has to be calculated in an effective field theory approach due to divergences in the vertex at one-loop level [96]. Integrating out the $W - t$ loop results in the effective $b - s - a$-vertex [94], [97] of form:

$$L \supset \frac{a}{f}\bar{s_L}b_R \frac{\sqrt{2}G_F m_t^2 m_b V_{ts}^* V_{tb}}{8\pi^2}\log\frac{\Lambda^2}{m_t^2}, \tag{2.5}$$

with the Fermi constant G_F, the CKM [98] mixing matrix elements for quarks a, b V_{ab}, left-handed strange and right handed bottom quark fields s_L and b_R, quark masses $m_{t,b}$, and the ultraviolet UV scale Λ. The UV scale corresponds to the energy at which further BSM particles are expected to appear in a more complete model. A class of more complete BSM models that do not feature the divergence are two Higgs-doublet

models [99]. The predicted ALP production rates used here are calculated with a fixed UV scale of $\Lambda = 1\,\text{TeV}$ in accordance with the PBC benchmark scenario [94].

It should be noted that a variation of Λ independently of the coupling g_Y effectively changes the ALP production rate without affecting its decay rate and hence lifetime. Furthermore, the ratio between the production rates in $B \to Ka$ and $B \to K^*a$ is independent of Λ [84]. The ratio is used in this search to fix the relative fractions in the model-dependent determination of the signal in Sec. 10.2.1. Similarly to the dark scalar model it is possible to obtain the correct thermal relic density by $\chi\chi \to aa$ and $\chi\chi \to a \to \text{SM}$ processes which are mostly dominant for small and, respectively, large fermion couplings [96].

The ALP decays to pairs of charged leptons and photons as well as to hadronic final states with three mesons. Decays to two mesons are forbidden by CP conservation. Therefore, only $a \to e^+e^-$ and $a \to \mu^+\mu^-$ are used here in both production modes to constrain the ALP model. Again, invisible decays to dark matter are not considered and the $a \to \tau^+\tau^-$ final state is not included.

To compute the lifetime of the ALP and the branching fraction of the leptonic modes, the leptonic and total decay widths are required. The ALP decay rates are taken from the PBC benchmark [94]. The leptonic widths use results from Ref. [96] and the hadronic widths are computed with inputs from Ref. [100]. The hadronic decays are modelled using chiral perturbation theory for lower ALP masses. A perturbative spectator model is used for masses $m_a > 1.2\,\text{GeV}/c^2$. Furthermore, the mixing between the ALP and the η' is taken into account. The mixing manifests itself in an enlarged decay rate into two photons and hadrons for ALP masses close to $m_{\eta'}$. This increase in the total ALP decay rate leads to a shorter lifetime and a smaller branching fraction into muons. The ALP could also mix with the η meson, although the effect is not visible due to the small width of the η. For masses below the di-muon threshold, there is no clear consensus in the theory community. Predictions for the ALP decay rates in electrons and photons are taken from Ref. [96]. These do not include possible mixing of the pseudo-scalar LLP with the π^0. The branching fractions of the ALP production and decay for the channels included in this search, as well as its lifetime for representative values for g_Y are shown in Fig. 2.2. The scaling with

the coupling is similar to the dark scalar case. The production as well as individual decay and total decay rates scale with g_Y^2, hence the lifetime scales with g_Y^{-2}. The decay branching fractions are independent of g_Y.

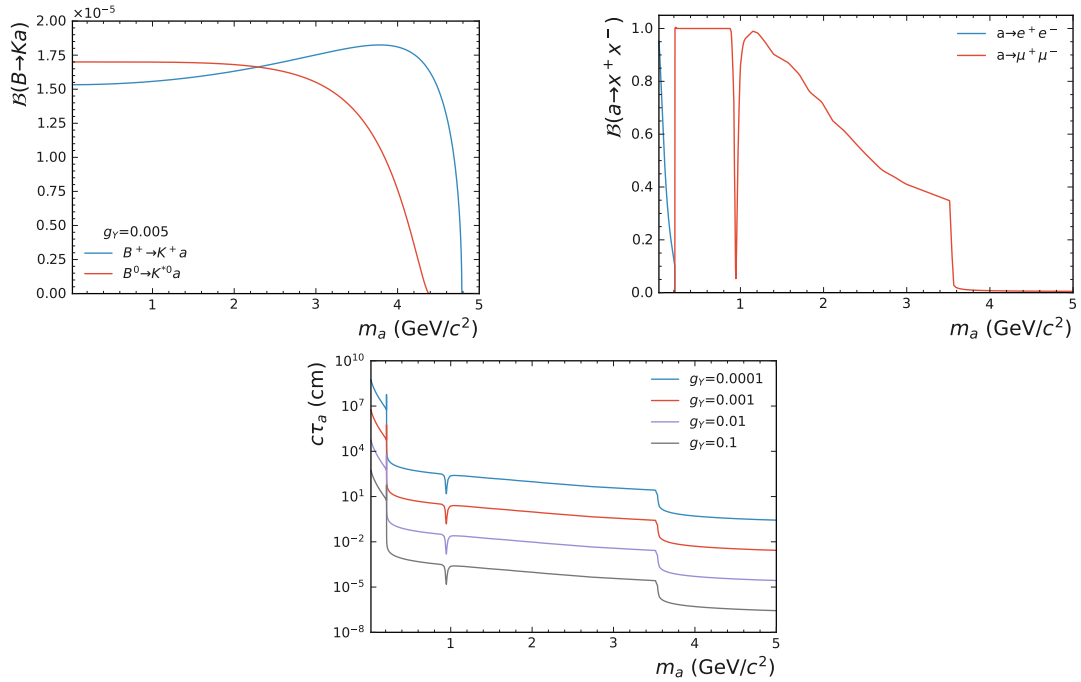

Figure 2.2.: Model predictions for the ALP a (upper left) production branching fraction in B-meson decays, (upper right) decay branching fractions into final states searched for here, and (lower) lifetime for different values of the coupling g_Y.

2.2.3. Current bounds

A summary of the most stringent experimental bounds on light dark spin-0 particles in the BSM models discussed before is given in the following.

- Searches by the LHCb collaboration in the $B^+ \to K^+ S(\to \mu^+\mu^-)$ [101] and $B^0 \to K^{*0}(\to K^+\pi^-)S(\to \mu^+\mu^-)$ [102] channels with a displaced muon pair, reported in Ref. [90].

- The L3 collaboration searched for Higgs-like scalars in $e^+e^- \to Z^*S$ [103] with hadronic decays of the S reconstructed as jets.

- The KTEV collaboration searched for the decay $K_L^0 \to \pi^0\mu^+\mu^-$ [104] reinterpreted in Ref. [90]. In the $s \to d$ transition, the production of the long-lived dark scalar or pseudo-scalar works via the same mechanism as in the described $b \to s$ case. Due to the smaller mass of the kaon, the available range of mediator masses is narrower.

- The E949 collaboration searched for the $K^+ \to \pi^+\nu\overline{\nu}$ process [105]. The underlying quark-level process $s \to d\nu\overline{\nu}$ is similar to $b \to s\nu\overline{\nu}$, discussed in Sec. 2.1.2. The missing energy of the neutrinos can be interpreted as an undetected LLP taking into account the three- versus two-body decay kinematics.

- The PS191 beam-dump experiment searched for sterile neutrinos in meson decays. The nonobservation of the signal was interpreted as bounds on the dark scalar model [106] using the $K_L^0 \to \pi^0 S$ and $K^+ \to \pi^+ S$ channels with decays of the dark scalar to electron-positron or muon pairs.

- Another beam-dump experiment, CHARM searched for displaced lepton pairs [59] but found no signal. The results were interpreted in both BSM models by the authors of Ref. [94].

- The BaBar collaboration performed an inclusive search for dark scalar production in $B \to X_s S$ with any strange meson X_s [48]. The results were interpreted in the dark scalar model by the authors of Ref. [87].

- The NA62 collaboration performed two direct searches for $K^+ \to \pi^+\nu\overline{\nu}$ [107] and invisible decays of the π^0 [108]. The latter is interpreted as bounds on $K^+ \to \pi^+ X$ with an undetected particle X with a mass close to the π^0.

- The MicroBooNE neutrino-scattering experiment searched for dark scalars produced in kaon decays with decays to electron-positron pairs [109].

- The CMS collaboration searched for LLPs in the decay to two muons at small masses using a special trigger stream for muons with low momenta [110]. The correspondence of the model assumptions underlying the CMS result with the benchmark model considered in this work is not yet clear at the time of writing. Therefore, they are not included in the figures showing the model-dependent results.

The experimental results take the form of upper bounds on the coupling parameters of the BSM models as functions of mediator mass. Lower bounds on the LLP couplings arise from cosmological observations. The authors of Ref. [87] performed model-dependent studies of constraints from primordial nucleosynthesis (BBN) [111] and supernova SN1987A [112], [113] on the dark scalar model. The resulting lower bounds for the mixing angle are at the level of 10^{-5} for $m_S \lesssim 0.2\,\mathrm{GeV}/c^2$ and 10^{-7} for $0.2 < m_S \lesssim 4\,\mathrm{GeV}/c^2$.

BBN provides a bound by considering the mediator lifetime in a freeze-out scenario as introduced in Sec. 2.1. The abundance of the mediator after freeze-out is high, because of its small coupling strength with the SM. If it does not decay before BBN it would have an observable effect on the abundances of the light elements. This leads to an upper bound on the lifetime and hence lower bound on the coupling. The supernova bound arises from the observation of thermal neutrinos from SN1987A. These neutrinos could not have been detected if the supernova had cooled too fast. This can happen by emission of a high rate of light spin-0 mediators. One way to avoid this is to impose couplings of the mediators large enough for them to get trapped.

3. Experimental setup

This chapter introduces the two main experimental apparatuses used in the presented search, the SuperKEKB particle accelerator and the Belle II detector. Fundamental objects that are derived from the detector signals and used in the analysis of the data are defined.

3.1. SuperKEKB collider

The SuperKEKB particle accelerator [114] is used to accelerate and collide electrons with positrons at the $\Upsilon(4S)$ resonance energy. The $\Upsilon(4S)$ decays to B-mesons, which are studied at the Belle II detector. It is located at The High Energy Accelerator Research Organization (KEK) in Tsukuba, Japan. A schematic view of the accelerator complex is shown in Fig. 3.1.

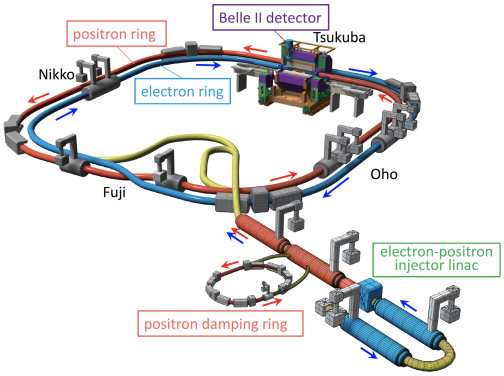

Figure 3.1.: Schematic view of the SuperKEKB accelerator complex, taken from Ref. [114]. Straight sections are named after Japanese cities.

The source of electrons is a photocathode, while positrons are produced in collisions of an electron beam with a tungsten target. A linear accelerator is used to accelerate electron and positron bunches to energies of 7 GeV and 4 GeV, respectively. The emittance of the positron beam is reduced by a damping ring before injection. Both bunched beams are injected into the main storage ring complex, which consists of a high-energy ring for the electrons and a low-energy ring for the positrons. The collision of both beams takes place in an interaction region within the Belle II detector. During data taking, electron and positron bunches are continuously injected into the storage rings.

The design luminosity of the SuperKEKB collider is $L = 8 \times 10^{35}\,\mathrm{cm}^{-2}s^{-1}$ which is around 40 times higher than its predecessor, the KEKB collider [115], [116]. This large increase is planned to be achieved by a twofold increase in beam intensity and another factor of 20 by squeezing the beams in a nano-beam scheme [117]. The highest luminosity achieved by the accelerator by Summer 2022 is $L = 4.65 \times 10^{34}\,\mathrm{cm}^{-2}s^{-1}$, a new world record [118].

The cms energy of the electron-positron collision $\sqrt{s} \approx 10.58\,\mathrm{GeV}$ allows the resonant production of the $\Upsilon(4S)$-meson which subsequently decays to pairs of charged (B^+B^-) and neutral ($B^0\overline{B}^0$) mesons. Furthermore, a large number of u, d, s, c quark pairs and τ-lepton pairs are produced in the e^+e^- collision, see Ref. [119] for a summary of relevant cross sections. The beam background sources are described in detail in Ref. [120].

3.2. Belle II detector

The Belle II detector [121] is used to detect the particles that are produced in the e^+e^- collision at the SuperKEKB accelerator. An overview of the diverse physics programme can be found in Ref. [119]. It consists of several subdetectors which are arranged cylindrically around the collision region. A schematic view of the Belle II detector with its subdetectors is shown in Fig. 3.2.

The origin of the Belle II coordinate system is the nominal collision point. The x-axis is defined in the horizontal direction towards the outside of the accelerator

tunnel. The y-axis is vertical with the positive direction pointing upward. The z-axis coincides with the symmetry axis of the solenoid. Its positive, or forward, direction approximately follows the electron beam. The azimuthal angle ϕ is defined in the transverse $x - y$ plane with $\phi = 0$ for $y = 0$ in the positive x direction. The polar angle θ is defined with respect to the z-axis.

The asymmetric electron and positron beam energies lead to a Lorentz boost of $\beta\gamma \approx 0.28$ for the B-mesons. The boost corresponds to an average flight distance of the B-mesons of around 130 μm. This leads to an experimentally resolvable separation between the decay vertices of the two B-mesons. The size of the interaction region is around 250 μm in z-direction, 10 μm in x-direction and 0.3 μm in y-direction, the size and position are regularly measured using $e^+e^- \to \mu^+\mu^-$ events [31]. The interaction point (IP) is defined in the centre of the interaction region.

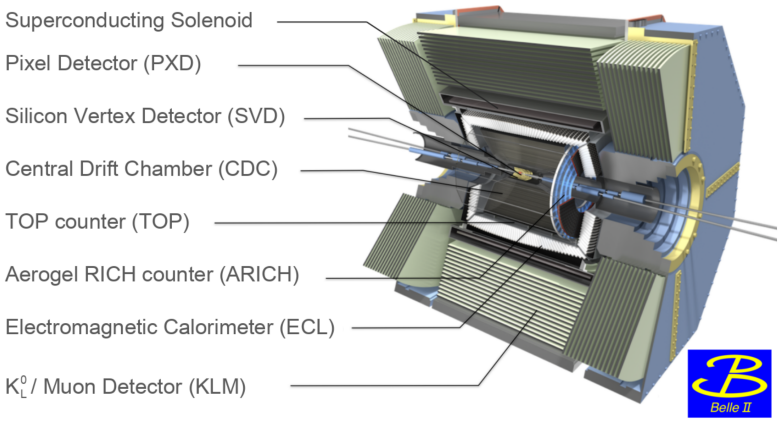

Figure 3.2.: Schematic view of the Belle II detector. Adapted from Ref. [121].

The following sections briefly describe the subdetectors starting from the innermost system, closest to the interaction region, and gradually moving further out.

3.2.1. Tracking and vertexing

Electromagnetically charged particles ionise detector materials, recorded as hits. The detector hits are used to reconstruct the particle trajectories, called tracks. Tracks are associated with the decay of a specific mother particle by determining its decay vertex. The decay vertex is found as the common geometrical origin of the tracks.

The tracking and vertexing system consists of three different detectors. Two detectors with high spatial resolution are used close to the interaction region. This allows to distinguish decay vertices of particles that decay close to interaction region, such as B-mesons, with high precision. A larger tracking detector is placed further outside with less precise spatial information at multiple points. This provides a large lever arm for a precise determination of the momenta of charged particles. The momentum of a particle is measured from the curvature of its track, which is bent in the magnetic field.

Pixel Vertex Detector

The Pixel Vertex Detector (PXD) [122] uses two layers of silicon sensors made up of depleted p-type field effect transistors [123]. The first PXD layer is installed at a radial distance of 1.4 cm from the IP and the second layer at 2.2 cm. They are installed along the z-direction around the beryllium beam pipe of 1 cm radius, and are arranged to overlap in ϕ. The pixel sizes range between $(50 \times 55)\,\mu\text{m}^2$ and $(50 \times 85)\,\mu\text{m}^2$. Charged particles which traverse the sensor produce electron-hole pairs. The electrons are accelerated by an electric field towards the gate leading to a current signal across the transistor. The PXD achieves a spatial resolution of around $12\,\mu$m [122], [124].

The data used here were collected with one sixth of the second layer installed. The PXD will be replaced with a complete configuration by the end of 2023.

Silicon Vertex Detector

The Silicon Vertex Detector (SVD) [125] is constructed of four layers of double-sided silicon strip sensors arranged around the beam pipe at radii of 3.9 cm, 8.0 cm, 10.4 cm and 13.5 cm from the IP. The modules are arranged in the z-direction in the barrel

region. The most forward modules of the outer three layers are slanted towards the beam pipe. The two sides of the sensors are used to determine three-dimensional information of electron-hole pairs produced by particles traversing the SVD. The first side uses silicon strips along $r - \phi$ collecting electrons, the second uses strips parallel to the z-direction collecting holes. The SVD achieves a spatial resolution ranging from 18 μm to 35 μm [126].

Central Drift Chamber

The Central Drift Chamber (CDC) [127] is the main tracking detector of the experiment. The chamber uses wires arranged in 56 layers between radii of 16.8 cm and 113.0 cm from the IP. Three-dimensional information is retrieved using two wire configurations. Axial wires are installed along the z-direction, whereas stereo wires are skewed with respect to the former. The chamber is filled with a gas mixture composed of 50% helium and 50% ethane. Charged particles ionise the gas while traversing the CDC. The primary ionisation electrons are accelerated towards the wires by an electric field generated by a high wire voltage. This leads to an avalanche of electrons, which is detected at the wires. The CDC achieves a spatial resolution of around 120 μm [128] and a relative momentum resolution of around 0.1% [129].

3.2.2. Dedicated particle identification detectors

The identification of the type of charged particles is an important part of physics analyses to differentiate between final states. Two dedicated detectors mainly aimed at distinguishing charged pions from kaons are installed in Belle II. The time-of-propagation (TOP) detector provides identification in the barrel region. The aerogel ring-imaging Cherenkov detector (ARICH) covers the forward region. Due to the forward boost of the center-of-mass (cms) system, there is no dedicated particle identification system in the backward region. Information from other subdetectors is also used for particle identification, see Sec. 3.3.3

TOP

The TOP [130] detector is installed in the outer cover of the CDC and uses the Cherenkov effect to identify particles. It consists of 16 quartz bars along the z-direction arranged in ϕ around the interaction region. The read-out uses photomultiplier tubes [131] coupled to the bars at the backward end. A mirror is placed at the forward end. Charged particles which traverse a TOP quartz emit Cherenkov photons at a specific angle, depending on their mass and momentum. The angle is determined from the measured time of the photons. Combined with momentum information, the particle type can be inferred. The TOP achieves a kaon identification efficiency of around 85% at a pion misidentification probability of 10% [132].

ARICH

The ARICH [133] detector also uses the Cherenkov effect for particle identification and is placed in the forward region of Belle II. It uses proximity-focussing ring-imaging Cherenkov detector technology. Charged particles entering the ARICH emit Cherenkov photons in two layers of silica aerogel tiles. The emitted photons are distributed in a ring whose radius depends on the Cherenkov angle. Hybrid Avalanche Photo Detectors [134] collect photons. The ARICH detector achieves a kaon identification efficiency of 93% with a pion misidentification probability of 10% [135].

3.2.3. Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECL) [136] is used to measure the energy of particles. It is required to detect neutral particles, such as photons or π^0-mesons that leave no signal in the tracking detectors. The ECL is a homogeneous calorimeter made up of scintillating caesium iodide crystals doped with thallium, corresponding to approximately 16 radiation lengths. It consists of a barrel part and two endcaps in the forward and backward regions. Energies are measured by collecting the photons of the electromagnetic showers initiated by particles entering the ECL. Photodiodes at the rear crystal surface collect the photons. The energy resolution of the ECL ranges between around 4% at 100 MeV and 2% at 8 GeV, leading to a π^0 mass resolution of

5 MeV [119], [137].

3.2.4. Solenoid magnet

A solenoid magnet [121] placed outside the ECL provides a 1.5 T magnetic field to bend the trajectories of charged particles. The magnetic field direction is parallel to the z-direction leading to a curvature of the trajectories in the transverse plane. The solenoid is made up of a niobium-titanium superconductor cooled with liquid helium. The homogeneous magnetic field in the inner part of the detector is measured to vary at a level of $\mathcal{O}(1\%)$ [138].

3.2.5. K_L^0 and muon system

The K_L^0 and muon system (KLM) [121], [139] is the outermost part of the Belle II detector. It is used to measure muons that pass through all previous detector systems and long-lived neutral K_L^0-mesons. The KLM consists of a barrel part and two endcaps. It is made up of alternating layers of active material and iron for passive deceleration. The active layers in the endcaps and two innermost barrel layers are scintillating strips. Resistive plate chambers are used in the outer layers of the barrel. The muon identification efficiency is around 89% for muons with a momentum of 1 GeV/c at a hadronic misidentification rate of 1.3% [121].

3.2.6. Trigger

The Belle II trigger system [140] decides whether an e^+e^- collision event is recorded and consists of a hardware and a software stage. The hardware trigger is based on field-programmable gate arrays that use information primarily from the CDC and the ECL. A low-level reconstruction of the TOP and KLM information complements the decision. The maximum output rates of the hardware and software stages are 30 kHz and 10 kHz, respectively. The software trigger has access to full detector information except for the PXD and is based on CPUs. Regions of interest are identified in the PXD for offline reconstruction. The trigger is optimised to provide near 100%

efficiency for $B\overline{B}$ events. Further important categories are events with low track multiplicity such as $\tau^+\tau^-$ or in dark sector searches [26], [28].

3.3. Object definitions

This section briefly describes how the main objects used in this search are derived from the detector information. These are tracks of charged particles, particle identification likelihoods, and ECL clusters.

3.3.1. Track reconstruction and vertexing

Charged particles are reconstructed as tracks [141], [142]. Track finding associates detector hits to a given particle trajectory. Tracks are first found in the CDC only [143], after which SVD hits are attached. Then, unused SVD hits are grouped into tracks [144]. These can originate from particles with low momentum that do not reach the CDC. In a further step, the SVD tracks are attached to the CDC tracks. However, SVD tracks for which no corresponding CDC tracks are found are used as well. PXD hits from regions of interest are attached to the tracks in a last step. A helix function is fitted to the hits [145] to extract the track parameters. A track is defined by the set of five helix parameters estimated by the fit:

- d_0: the distance between the coordinate system origin and the track helix at its point of closest approach projected on the transverse plane,
- z_0: the same distance projected on the z-direction,
- ω: the curvature of the track in the $x-y$ plane, signed with the electric charge,
- θ: the polar angle of the track at the point of closest approach.

The transverse momentum p_T of a track is determined from its curvature, its electric charge q, and the magnetic field as $p_\mathrm{T} = |q|B|\omega|$.

Common production vertices of multiple tracks are found using a kinematic fit algorithm [146]. This vertex fit improves the momentum estimation of the involved

tracks which leads to a higher resolution for combined quantities such as the invariant mass of the mother particle.

3.3.2. ECL cluster reconstruction

ECL clusters are reconstructed from energy deposits in the ECL crystals [147]. Connected regions of adjacent crystals above a threshold energy are formed. Clusters are identified around local energy maxima within the connected regions. Tracks are associated to ECL clusters by geometric matching. ECL clusters not associated with a track from a charged particle are treated as photon candidates.

3.3.3. Particle identification

Particle identification (PID) for charged particles uses information from all subdetectors except PXD and SVD. The CDC provides PID information using specific energy loss along the trajectory of the particles dE/dx. The TOP and ARICH contribute to the PID using the Cherenkov effect combined with the momentum measurement. For tracks with an associated ECL cluster, the ratio of measured energy to momentum E/p provides further PID information. Minimum ionising particles, such as muons, leave less energy in the ECL compared to, e.g., electrons, leading to a small E/p. Muon identification is improved by matching tracks to hit patterns in the KLM.

Each subdetector provides likelihoods $\mathcal{L}$ for the electron, muon, pion, kaon, proton, and deuteron hypotheses. The likelihoods are determined by comparing the observed pattern of information from the subdetector with expected distributions for each hypothesis. If a subdetector can provide only little information on a specific particle hypothesis, the corresponding likelihood is weak. PID information from all subdetectors is combined by summation over the likelihoods. The two main PID definitions employed in Belle II analyses are the global and binary scores. The global PID score is computed by comparing the likelihood of a specific particle hypothesis with the sum of all hypotheses. The binary PID score compares two specific hypotheses, dividing the likelihood of one of the two hypotheses under study by the sum of both. This gives a measure on whether a certain hypothesis is found more likely than another,

even if both have a small global score. A typical example is a binary PID score with respect to the pion hypothesis, as pions are often the largest background source due to their abundant production.

4. Strategy and data samples

This chapter gives a brief overview of the data analysis strategy and the experimental and simulated datasets that are used to perform the presented search.

4.1. Analysis strategy

The search for a LLP with spin-0, S, produced in $b \to s$ transitions is performed in eight exclusive channels. These consist of the production of the S in association with two different kaon states $B^+ \to K^+ S$ and $B^0 \to K^{*0} S$, and decays $S \to e^+e^- / \mu^+\mu^- / \pi^+\pi^- / K^+K^-$, referred to in this work as S production mode and final state channel, respectively. Only the fully charged decay channel of the K^{*0} is considered: $K^{*0} \to K^+\pi^-$. This choice is made due to the large branching fraction of $\mathcal{B}(K^{*0} \to K^+\pi^-) \approx 2/3$ and to minimise differences between production modes in the kinematic features used for the signal selection. The latter can arise because of differences in the energy and momentum resolutions between charged-particle tracks and neutral final-state particles. In the following, the charged decay channel of the K^* is implied.

Important background processes to this search are generic decays of B-mesons that can pass the specific kinematic selections, as well as light-quark pairs $e^+e^- \to q\bar{q}$, where $q \in u, d, s, c$, which give rise to a combinatorial background component. The $e^+e^- \to \tau^+\tau^-$ background is already suppressed to negligible levels with loose requirements detailed in Sec. 5.2. This search relies on two key methods to separate backgrounds from the signal process. The S candidate is reconstructed from a pair of oppositely charged tracks with a common production vertex. A minimum distance (displacement) of the S decay vertex from the IP is required. This selection heavily

suppresses most prompt and short-lived backgrounds. The (almost) prompt signal case cannot be studied with this dedicated LLP search. This is motivated by strong experimental constraints on the prompt production of additional spin-0 particles, see the experimental status in Sec. 2.2.3.

A kaon candidate that originates from close to the IP is identified depending on the production mode. The signal B-meson candidate is then constructed by combining the S with the kaon candidate. This allows to place highly effective selections based on the expected kinematics of a B-meson to further suppress combinatorial backgrounds from light-quark-pair production.

The search for the signal is carried out by scanning the reconstructed mass distribution of the S candidate separately for different lifetime hypotheses and analysis channels. A signal is expected to yield a peaking excess of candidates over the locally smooth remaining background.

Monte Carlo simulation of signal and background processes are used to:

- optimise a set of rectangular selection requirements for sensitivity to the signal process,
- compute the signal efficiency at various channel, mass and lifetime combinations,
- define signal templates for the extraction of the signal yield via fits to the experimental data,
- determine the functional form of the background templates for the extraction.

The quantities derived from signal simulation are validated in data using a control sample enriched with K_S^0, the closest SM proxy to the S. Differences between data and simulation that are found for K_S^0 are used to correct the signal simulation.

Systematic uncertainties arise due to the finite knowledge of the size of the K_S^0 correction factors. Further systematic uncertainties taken into account affect the number of B-mesons in the dataset, as well as the tracking efficiency and momenta of charged particle that originate from near the IP in simulation.

The signal yield in data is estimated using maximum likelihood fits of the signal plus background templates to the reconstructed S mass distribution. The significance

of a signal excess is quantified by performing a second fit of only the background template and comparing the maximised likelihoods of both hypotheses. Furthermore, a simultaneous signal extraction is performed with a combined fit in all accessible analysis channels for different lifetime hypotheses using predictions of the two benchmark models.

Differences in the lifetime hypothesis manifest mainly in the signal efficiency, signal template parameters, and some selection requirements.

The signal extraction method is validated by repeating the fit multiple times on datasets sampled from the templates. This tests the stability of the procedure against statistical fluctuations.

The signal selection as well as the signal extraction method are studied and decided on using simulated events. They are validated in data using background-enriched control regions, before opening the region in data where the signal is expected to avoid experimenters bias.

The main result of the search are upper limits on the product of branching fractions $\mathcal{B}(B \to K^{(*)}S) \times \mathcal{B}(S \to x^+x^-)$ as a function of the S mass for a range of different lifetime hypotheses in all eight exclusive channels. The upper limits for different lifetime hypotheses are translated into a two-dimensional excluded region for the two benchmark models. These are the plane of dark scalar mass m_S versus mixing angle θ in the scalar model and ALP mass m_a versus coupling g_Y in the pseudoscalar model.

4.2. Experimental dataset

The data used for the presented search were collected by the Belle II experiment at the $\Upsilon(4S)$ resonance energy in 2019 and 2020 and corresponds to an integrated luminosity of $189\,\text{fb}^{-1}$. With a $e^+e^- \to \Upsilon(4S) \to B\overline{B}$ cross section of 1.1 nb [119] at the nominal collision energy, this yields approximately 198×10^6 of B-meson pairs in the dataset.

4.3. Monte Carlo simulation

This section summarises the generators and settings used to model the background and signal processes required for this search. The Belle II detector response to the final-state particles is simulated using GEANT4 [148] with a dedicated Belle II implementation [149], [150].

4.3.1. Background simulation

Background processes to the search are studied using official Belle II simulation from the 14th run-independent simulation campaign. Run-independent simulation uses general detector settings and simulated beam-backgrounds that do not take into account possible variations between runs of data-taking.

The simulated background corresponds to an integrated luminosity of $1\,\text{ab}^{-1}$, roughly five times the statistics of the data used for this search. It is often scaled down to $250\,\text{fb}^{-1}$ to study the sensitivity with statistical precision close to the experimental dataset. Scaling reduces the statistical uncertainty compared to sampling the same fraction of the total simulated sample. The background processes considered for this search and the event generators used for their modelling are summarised in Tab. 4.1.

Table 4.1.: Background processes considered to this search with the event generators used. The (γ) denotes additional electromagnetic radiation.

Process	Generator
$e^+e^- \to \Upsilon(4S) \to B^+B^-$ $e^+e^- \to \Upsilon(4S) \to B^0\overline{B}^0$	EVTGEN [151]
$e^+e^- \to u\bar{u}(\gamma)$ $e^+e^- \to d\bar{d}(\gamma)$ $e^+e^- \to s\bar{s}(\gamma)$ $e^+e^- \to c\bar{c}(\gamma)$	KKMC[152] for the production interfaced with PYTHIA8 [153] for hadronisation and decay and EVTGEN for decay
$e^+e^- \to \tau^+\tau^-(\gamma)$	KKMC for generation interfaced with TAUOLA [154] for decay

Electromagnetic final-state radiation from charged particles generated by EVTGEN

is modelled using PHOTOS [155]. The B-mesons are set to generically decay to a variety of final states with branching fractions according to the official Belle II simulation [150].

4.3.2. Signal simulation

The signal processes are simulated with the same detector settings and beam-background conditions used for the background simulation described before.

Signal events are generated using EVTGEN. The production and decay of the S follow the kinematics of a spin-0 particle. The $\Upsilon(4S)$ resonance is set to decay to a pair of oppositely flavoured B-mesons, B^+B^- or $B^0\overline{B}^0$, depending on the production mode. One of the two B-mesons decays via the signal process. The second B-meson decays generically via the same processes and probabilities as used for the $B\overline{B}$ background simulation. The EVTGEN decay models used for the simulation of the signal processes are listed below:

- $\Upsilon(4S) \to B\overline{B}$ via VSS,
- $B^+ \to K^+ S$ via PHSP,
- $B^0 \to K^{*0}\, S$ via SVS,
- $K^{*0} \to K^+\pi^-$ via VSS,
- $S \to x^+x^-$ via PHSP.

Here, VSS and SVS describe the kinematics of a vector particle that decays to two scalar particles and a vector particle that decays to two scalars, respectively. PHSP describes generic decays according to the two-body phase space, where all spins in the initial and final states are averaged.

The S is initially generated with zero lifetime. Finite lifetimes are considered by displacing the S decay vertices and daughter particle production vertices according to exponential decay distributions with different lifetimes. All lifetime values used in this work are given in units of cm for $c\tau$ where τ denotes the lifetime in the S rest frame. The propagation of the S through the detector is simulated with GEANT4

using a dedicated LLP definition that automatically detects non-SM neutral LLPs and treats them correctly without further user action. Both parts were developed and implemented in the open source Belle II software framework [150] in the course of this work.

Generated model parameters

Simulated signal events are produced for various combinations of channels, mass, and lifetime. The list of simulated S lifetimes is given in Tab. 4.2, and the list of masses is given in Tab. 4.3. At each S mass value, events are simulated for all kinematically accessible production modes and final states. The lifetime points are chosen to cover a region that is experimentally accessible for the analysis. Too small lifetimes cannot be tested because of the minimum vertex displacement selection. Lifetimes that are too large have reduced efficiency due to a decrease in reconstruction efficiency as a function of displacement and at some point a large fraction of S decay outside of the tracking volume. The S mass points for which signal events are generated are chosen to cover the kinematically accessible range of masses for the two production modes. Furthermore, signal events are generated with mass hypotheses close to expected peaking background sources. These are needed to correctly model the signal efficiency close to regions with tighter signal selection requirements, see Sect. 5.5.

Table 4.2.: Lifetime values for simulated signal events.

$c\tau_S$ (cm)			
0.001	0.003	0.005	0.007
0.01	0.025	0.0	0.1
0.25	0.5	1	2.5
5	10	25	50
100	200	400	700
1000	2500	5000	10000

Lifetime hypotheses above 100 cm and up to 400 cm are simulated for muon and pion final-state channels for S masses up to 0.3 GeV/c^2. The longer lifetime hypotheses up to 10000 cm are simulated for the electron S final-state channels for S masses up

to 0.2 GeV/c^2. These samples are simulated to cover the experimental sensitivity to the benchmark models in these regions specifically. At given values of the coupling parameters, the lifetimes of the dark scalar and ALP increase rapidly below the di-muon threshold. This is due to the small total decay rate as only the $\gamma\gamma$ and e^+e^- final states are kinematically accessible.

Table 4.3.: Mass values for simulated signal events.

m_S (GeV/c^2)						
0.025	0.050	0.070	0.100	0.120	0.150	0.200
0.214	0.220	0.250	0.280	0.282	0.285	0.290
0.300	0.350	0.400	0.440	0.445	0.450	0.455
0.457	0.460	0.462	0.465	0.485	0.487	0.489
0.500	0.507	0.509	0.511	0.800	0.990	0.995
1.000	1.005	1.030	1.039	1.040	1.050	1.100
1.250	1.348	1.352	1.500	1.600	1.650	1.653
1.700	1.703	1.750	1.753	1.797	1.800	1.847
1.850	1.947	1.950	2.000	2.250	2.500	2.750
2.990	3.000	3.010	3.140	3.150	3.155	3.160
3.200	3.250	3.500	3.647	3.653	3.748	3.750
4.000	4.250	4.380	4.500	4.750	4.780	

Signal kinematics

Figure. 4.1 shows the distributions of the generated momenta and polar angles of the particles in the signal process for a given S mass. The two-body B-meson decay leads to large and similar momenta of the S and the K^+ (K^{*0}) in $B^+ \to K^+S$ ($B^0 \to K^{*0}\,S$). The S daughters feature a broad momentum distribution, identical for both charges due to the symmetric decay. The $K^{*0} \to K^+\pi^-$ decay leads to a distribution of K^+ and π^- momenta with a prominent first peak at low and a smaller second peak at larger momenta. The momentum of the K^+ in $B^0 \to K^{*0}\,S$ is generally smaller compared to the K^+ in $B^+ \to K^+S$ due to the sharing of the momentum of the K^{*0} between K^+ and π^-. The lighter π^- have the smallest momenta of all signal particles.

The polar angle distribution of all signal particles follows the boost of the cms system with a preference for the forward direction due to the asymmetric beam energies. There are no significant differences between the production modes; therefore, only $B^+ \to K^+S$ is shown as an example.

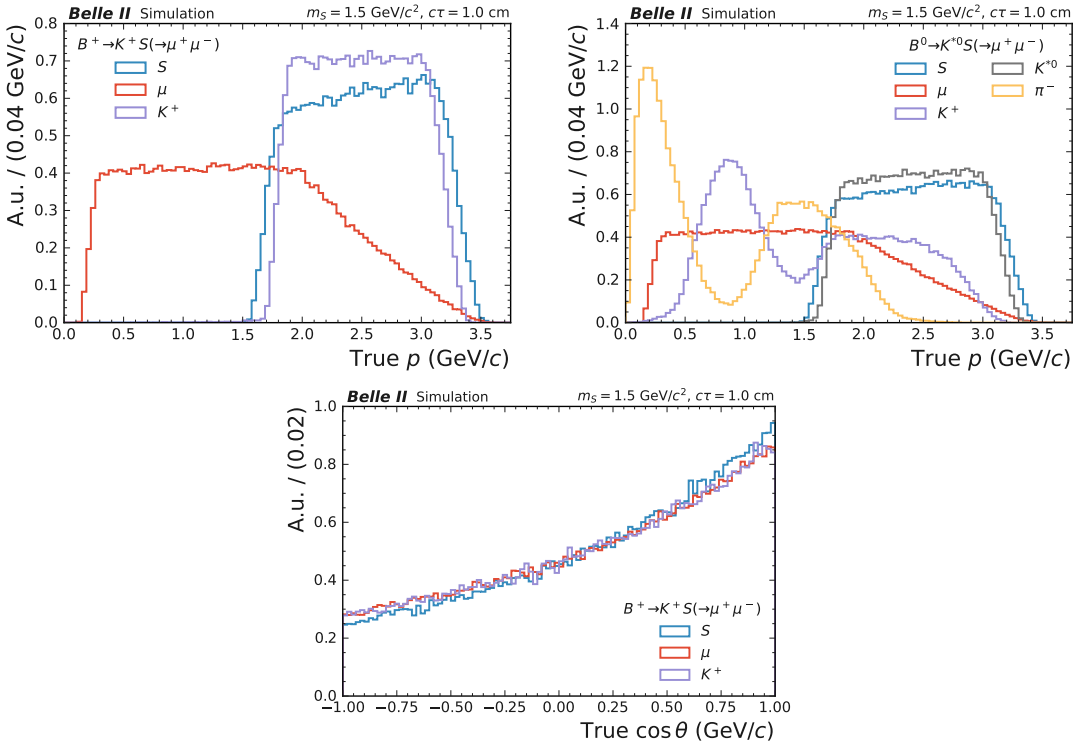

Figure 4.1.: Generator-level distributions for different signal particles of (upper left) momentum in $B^+ \to K^+S$, (upper right) $B^0 \to K^{*0}(\to K^+\pi^-)S$ and (lower) polar angle in $B^+ \to K^+S$.

The variation of the generated momenta for different signal particles and of the opening angle between the S daughters with the S mass are shown in Fig. 4.2. Momenta of the S and the prompt K^+ ($K^+\pi^-$) in $B^+ \to K^+S$ ($B^0 \to K^{*0}\,S$) are shifted to smaller values for larger S masses. The S daughter momenta vary weakly for small to intermediate S masses and tend to larger values for heavy S. The opening angle between the S daughters is very narrow at small S masses due to its Lorentz-boost

and increases with the S mass. For a two-body decay of a particle x the opening angle between the decay products can be approximated as $\alpha \approx 2m_x/p_x$.

The figures show the kinematics for the $S \to \mu^+\mu^-$ final state. The findings apply to the other final states channels, with larger (smaller) S daughter momenta, at a given S mass, for smaller (larger) daughter masses.

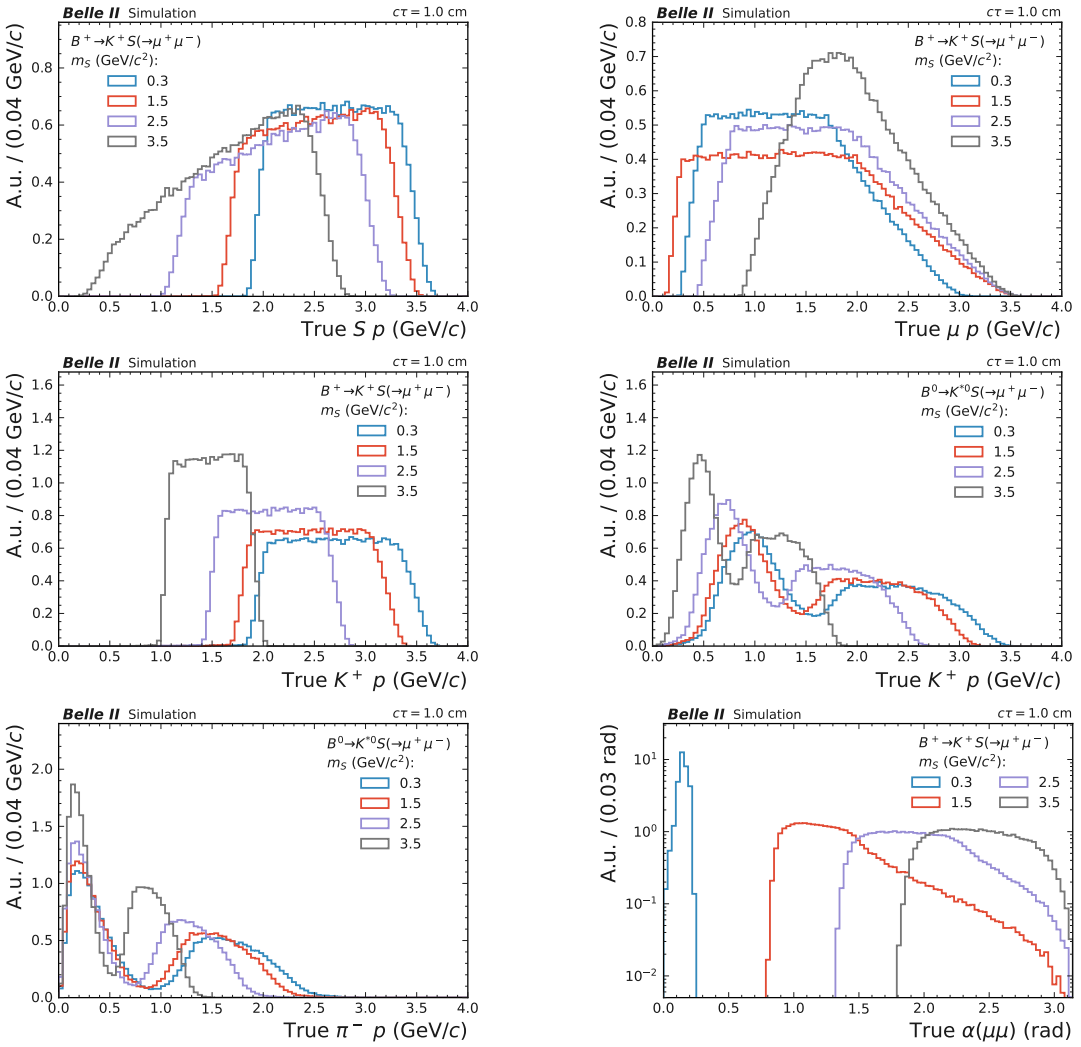

Figure 4.2.: Generator-level distributions at different S masses for (upper left) S momentum in $B^+ \to K^+S$, (upper right) S daughter momentum in $B^+ \to K^+S$, (middle left) prompt kaon momentum in $B^+ \to K^+S$, (middle right) prompt kaon momentum in $B^0 \to K^{*0}(\to K^+\pi^-)S$, (lower left) prompt pion momentum in $B^0 \to K^{*0}(\to K^+\pi^-)S$ and (lower right) opening angle between both S daughters. S daughter momenta for both charges are shown together resulting in two entries per event.

5. Event selection

This chapter describes the selection requirements that are applied to the dataset to separate the long-lived particle signal process from background sources.

5.1. Long-lived particle reconstruction

The daughters of the S are reconstructed as tracks from detector hits in the tracking subdetectors, detailed in Sec. 3.3.1. Long lifetimes of the S lead to decay vertices that are displaced from the IP. This means that the daughter particles travel a smaller distance in the tracking detectors and leave fewer hits to reconstruct the tracks. Figure 5.1 shows the performance of the S daughter track reconstruction in quantities important for determining signal properties. The momenta and polar angles of the tracks are required to calculate the momentum and mass of the S, as well as of the signal B-meson. The opening angle between the two S daughter tracks is calculated from the difference of azimuthal and polar angles of both tracks and used directly in the determination of the mass S. Furthermore, it plays a role in the efficiency to find the S daughter tracks and to find the common vertex, which is detailed later. The normalised difference between the generated and reconstructed values (residuals) is shown for different regions of S vertex displacement from the IP in the transverse plane, dr, to gauge the reconstruction performance. A worsening of the resolution in all three is clearly visible for larger vertex displacements. The effect is most visible for the momentum and polar angle of the S daughters as soon as the S decays outside the inner detector region (above approximately 15 cm) where only CDC hits are available. The opening angle is less affected in the first CDC region, suggesting that tracks that start not too far within the CDC still retain a decent azimuthal angle

reconstruction. The angular quantities show a symmetric worsening of the resolution for larger S vertex displacements with no clear bias developing. However, the momentum residuals at large vertex displacements are asymmetric with longer tails towards larger reconstructed momenta. Furthermore, the peak position of the distribution is biased towards smaller reconstructed momenta at very large vertex displacements.

The worsening of the tracking performance is the result of the reduced number of detector hits available to reconstruct the tracks. The fewer hits a charged particle produces, the shorter the reconstructed track, and the smaller the lever arm to accurately determine its curvature and hence momentum. The momentum asymmetry and bias could be an effect of not properly taking into account differences between the magnetic-field strength close to the IP and at the displaced vertex. The resolution of position-related quantities such as track angles becomes worse with fewer hits as well, especially without hits in the highly resolving vertex detectors. By the design of the CDC, the hits have a poorer resolution in z compared to $x - y$, leading to a poorer determination of the polar angle.

The S candidate is reconstructed from two oppositely charged tracks that correspond to its daughter particles. The decay vertex of the S is determined with a kinematic fit. Figure 5.2 shows the efficiency of reconstructing the candidate S and finding its decay vertex as a function of the S vertex dr for different S masses.

The S reconstruction efficiency corresponds to the efficiency to reconstruct both daughter tracks. It is defined with simulated signal events as:

$$S \text{ reconstruction efficiency} = \frac{\#\text{reconstructed } S}{\#\text{generated } S}. \tag{5.1}$$

The reconstruction efficiency is large for small vertex displacements and decreases rapidly for larger displacements. For displacements greater than around 80 cm the efficiency of finding both tracks is almost zero. The Belle II track reconstruction algorithms as introduced in Sec. 3.3.1 assume that the tracks originate close to the IP and lose efficiency if the track does not pass close to it. The largest values of the reconstruction efficiency close to the IP lie around 75-80%. They do not reach close to 100%, since no requirements are placed on the tracks to lie within the detector

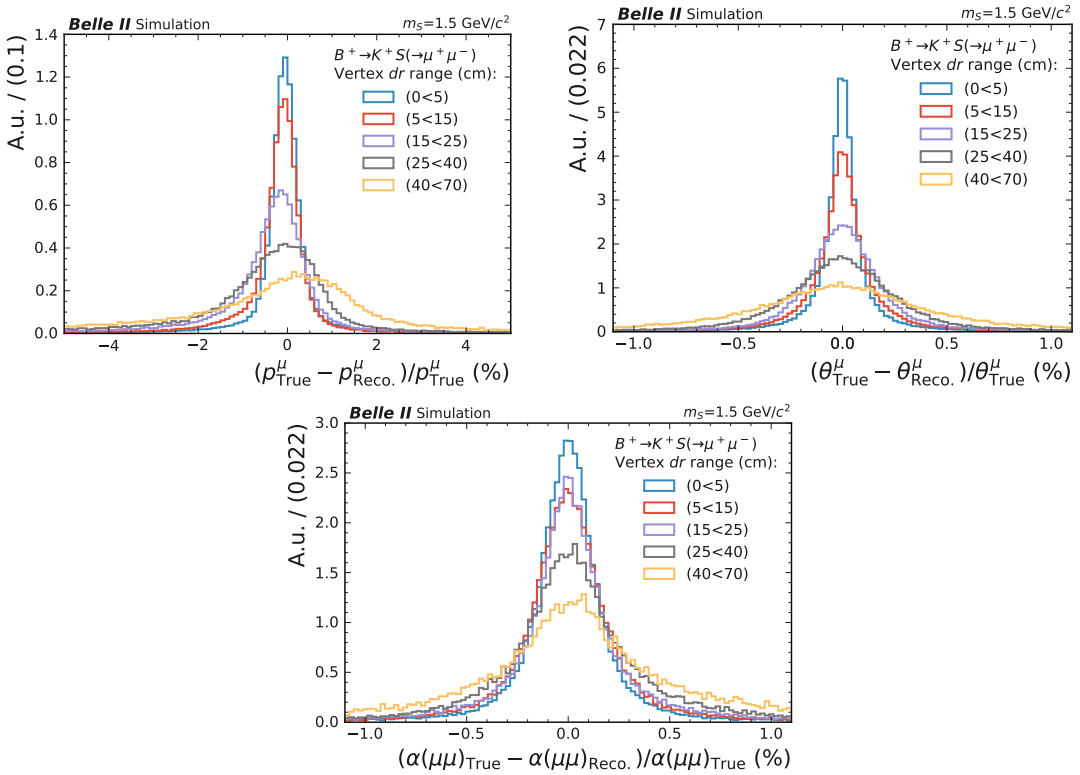

Figure 5.1.: Normalised difference between generated and reconstructed (upper left) S daughter momentum p^μ, (upper right) S daughter polar angle θ^μ and (lower) three dimensional angle between both S daughters $\alpha(\mu\mu)$ for different regions of S vertex displacement. Daughter momenta and polar angles are shown for both charges, resulting in two entries per event.

acceptance. The reconstruction loses efficiency less quickly for very small S masses. The behaviour as a function of the vertex displacement is similar between intermediate to large S masses. This is due to the small opening angle between the daughters at small S masses. At larger vertex displacements, the small opening angle of both tracks leads to track directions close to the original S direction. Therefore, it is more likely that both tracks can be extrapolated to pass close to the IP where the S was produced in the decay of the B-meson. On the other hand, at larger S masses and hence daughter opening angles, at least one of the two tracks often does not pass close to the IP and is not reconstructed.

The S vertexing efficiency corresponds to the efficiency to determine the common geometrical origin of both S daughter tracks and is defined as:

$$S \text{ vertexing efficiency} = \frac{\#\text{reconstructed and vertexed } S}{\#\text{reconstructed } S}, \tag{5.2}$$

from the subset of S decays generated where both daughter tracks were found by the tracking. No requirements are placed on the accuracy of the decay vertex, apart from a successful kinematic fit. The S vertexing efficiency approaches one for small S vertex displacements and slowly decreases linearly to 40-60% at 80 cm vertex displacement, where it is capped by the almost zero reconstruction efficiency. The vertex-fit algorithm assumes, similarly to the track finding, that the tracks originate from close to the IP when estimating a starting point for the kinematic fit. This bias affects the efficiency more strongly at large vertex displacements. Furthermore, one can assume that the reduced resolution of the daughter track properties at large S vertex displacements reduces the efficiency to find the common origin of the tracks. The vertexing efficiency is highest for small S masses and again decreases for larger S mass. Smaller opening angles lead to tracks that are closer together and pass close to the IP more often, thereby reducing the effect of the internal bias of the fit and enabling the algorithm to find a common origin (even if not an accurate one).

Figures 5.1 and 5.2 show the $B^+ \to K^+ S(\to \mu^+\mu^-)$ channel but the findings also apply to the neutral S production mode, and other final-state channels.

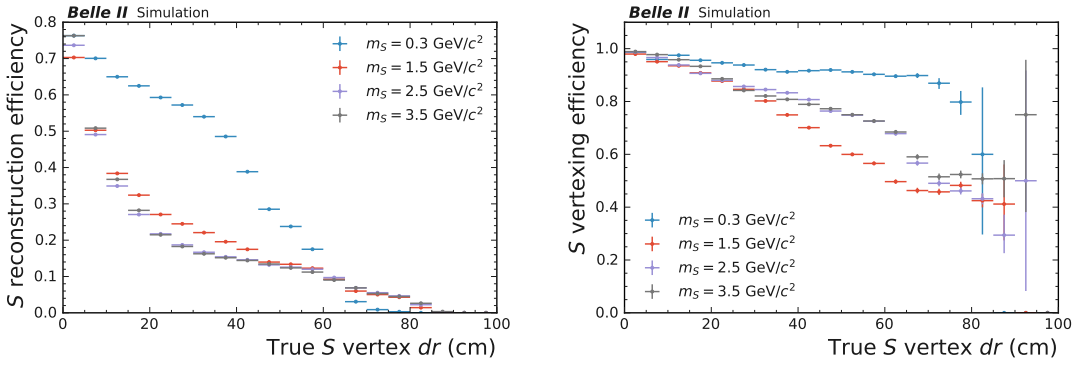

Figure 5.2.: Efficiency at different S masses (left) to find both S daughter tracks and (right) to find the common vertex of both tracks as a function of the S decay vertex dr.

5.2. Overview of observables

The B-meson candidate is formed by combining the S candidate with a kaon candidate. In the $B^+ \to K^+ S$ production, channel a track under the kaon mass hypothesis is used. In the $B^0 \to K^{*0}(\to K^+\pi^-)S$ production channel, two oppositely charged tracks under the kaon and pion mass hypotheses are combined, a common vertex is determined using a kinematic fit, and a mass window around the K^{*0} is imposed. Kaon candidate tracks are termed *prompt* in the following, as they are expected to originate from the S production point.

Kinematic features based on these candidates are identified to separate the signal from background processes. These are introduced in the following, including details on variations of the signal shape in the features with the S mass, lifetime, production channel or final-state channel. If not mentioned explicitly, a feature is (almost) invariant.

B-meson candidate

- Beam constrained mass M_{bc}: Defined as $M_{\mathrm{bc}} = \sqrt{s/4 - |p_B^*|^2}$ with the reconstructed momentum of the B-meson candidate in the cms system p_B^* and the

collision energy in the cms $\sqrt{s}$. Separates reconstructed B-mesons from combinatorial backgrounds such as those coming from light-quark pair events. B-mesons peak around m_B, while combinatorial backgrounds follows an ARGUS shape [156]. Small variations of M_{bc} with S lifetime are due to the reduced momentum resolution of S daughter particles at long lifetimes.

- Energy difference ΔE: Defined as $\Delta E = E_B^* - \sqrt{s}/2$ with the reconstructed energy of the B-meson candidate in the cms system E_B^*. Separates reconstructed B-mesons from combinatorial background and misidentified final states in B-meson events due to the inclusion of the final-state mass hypotheses to calculate the energy. Correctly identified B-meson candidates peak around zero, misidentified B-meson background peaks offset of zero, and combinatorial backgrounds follow a linearly falling distribution. ΔE varies more strongly with the S lifetime, leading to longer tails to large ΔE values due to reduced resolution and bias of the S daughter momentum.

Event shape

- Fox-Wolfram $R2$: The second Fox-Wolfram moment [157], [158] normalised by the zeroth moment $R2 = H_2/H_0$ which are defined as:

$$H_l = \sum_{i,j} p_i p_j P_l(\cos \alpha_{ij}), \tag{5.3}$$

summing over pairs (i, j) of all track and photon momenta $p_{i/j}$ in an event with the angle between pairs α_{ij} and the Legendre polynomials P_l of order l. The Fox-Wolfram $R2$ ratio distinguishes between spherical and collimated distributions of momenta in an event. B-meson events, such as the signal process, feature a more spherical distribution because of the small Lorentz-boost of the B-mesons. Light-quark-pair events feature a larger Lorentz-boost along the beam direction because of the smaller invariant masses of light quarks. The observable tends to small (large) values for B-meson signal (light-quark-pair background) with values ranging between zero and unity. It varies with the

S mass and the production mode. The former is due to different angles between its daughter particles. The latter is due to the additional prompt pion in $B^0 \to K^{*0}(\to K^+\pi^-)S$.

S candidate

- Reconstructed mass $M(x^+x^-)$: The invariant mass of the S candidate calculated from the two daughter tracks. A modified version of the reconstructed S mass is used to determine the signal yield. The signal process peaks in $M(x^+x^-)$ while combinatorial backgrounds feature a locally smooth low-order polynomial shape. Both are detailed in Chap. 6. Backgrounds from SM resonances can also peak in $M(x^+x^-)$ and need to be suppressed. The reconstructed mass varies mostly with S mass and lifetime, the latter due to the daughter momentum resolution and bias leading to a longer tail towards large reconstructed masses for large S lifetime. At given S mass, the resolution of $M(x^+x^-)$ can vary between different channels due to different daughter mass and hence daughter momentum.

- Decay vertex dr: The distance of the reconstructed S decay vertex from the IP in the transverse plane. The vertex dr distribution is exponentially decreasing for the backgrounds and the signal. However, backgrounds fall more strongly as most backgrounds arise from nearly prompt processes. Long-lived SM backgrounds, such as K^0_S, feature longer tails. Signal candidates feature real displacement with an exponential distribution that varies with S lifetime and mass due to its Lorentz-boost.

- Pointing angle 2d $\cos\theta^S_{\text{pointing}}$: The angle between the vector connecting the IP with the reconstructed S decay vertex and the S momentum vector reconstructed from the momenta of both daughter tracks in the transverse plane. The two-body signal decay process features a pointing decay vertex as the sum of the momenta of both daughters has to point in the same direction as the S mother particle, leading to a 2d $\cos\theta^S_{\text{pointing}}$ distribution peaking at unity. Backgrounds can feature non-pointing vertices due to different effects. In most of the cases, the two tracks do not originate from the same mother particle and randomly

cross to yield a common vertex. Furthermore, more than two particles can be produced in the decay of a particle, and the vertex is reconstructed from only two. The latter arises often in the case of hadronic interactions of beam particles, or particles from the e^+e^- collision, with detector material leading to a true displaced but non-pointing decay vertex. The pointing angle varies with S lifetime and mass due to different Lorentz-boost. For small lifetimes and therefore vertex displacements, the separation power of 2d $\cos\theta^S_{\text{pointing}}$ is reduced. At long lifetimes, it is affected by the reduced momentum resolution.

- Significance Of Distance: The significance of the three-dimensional distance between the IP and the reconstructed S vertex. The vertex fit returns an estimate of the uncertainty of the vertex location. This is used to normalise the vertex displacement distance, and thereby determine its significance. Larger displacements feature larger significances as the vertex distance grows faster than its uncertainty.

S daughter tracks

- Displaced track PID: The particle identification scores of both S daughter tracks. The PID response depends on the momentum and polar angle of the tracks. Hence, small variations arise between S masses and between the final states at a given S mass. The latter is due to different final-state masses leading to different track momentum. More details, including lifetime dependence, are discussed in Sec. 5.3.

- Transverse momentum p_{T}: The momentum of the tracks in the transverse plane. The transverse momentum varies with the S mass and the final state at a given S mass. The p_{T} resolution decreases and a bias towards large reconstruction values appears with larger S lifetimes.

- Polar angle θ: The polar angle of the tracks. The polar angle resolution decreases with the S lifetime. For these selections, the polar angle is determined at the inner surface of either the CDC or ECL by following the helix of the track from

the IP to the respective subdetector. This is done to account for the production of the track at a displaced vertex to verify whether the track is in the acceptance of a subdetector.

Kaon candidate

- Prompt track PID: The particle identification scores of the prompt tracks. The PID response varies between both production modes due to a larger prompt kaon momentum when produced directly in the B-meson decay in $B^+ \to K^+ S$ compared to further down the decay chain in $B^0 \to K^{*0}(\to K^+\pi^-)S$. The efficiency of the PID requirement for a prompt track decreases for very large S masses due to the same reason, reduced prompt track momentum.

- Reconstructed K^{*0} mass $M_{K^{*0}}$: Invariant mass of the two prompt tracks $K^+\pi^-$ that form the K^{*0} candidate.

Figures 5.3, 5.4 and 5.5 show typical background distributions together with signal distributions for different S masses in observables that are used in the selection. The requirements on the S daughter transverse momentum and polar angle are introduced to improve agreement between data and simulation in background enriched regions of the phase space, detailed in Sec. 8.1 and are not optimised for signal sensitivity. The PID distributions for the S daughter tracks are shown in Sec. 5.3. The reconstructed S mass distribution is shown only around one S mass hypothesis, as with more mass samples the range would be too large to show the shape of the signal due to its high resolution.

The distributions include candidates with a loose pre-selection of:

- $M_{\text{bc}} > 5.27\,\text{GeV}/c$,
- $|\Delta E| < 0.05\,\text{GeV}$,
- Fox-Wolfram $R2 < 0.45$,
- S vertex $dr > 0.05\,\text{cm}$,

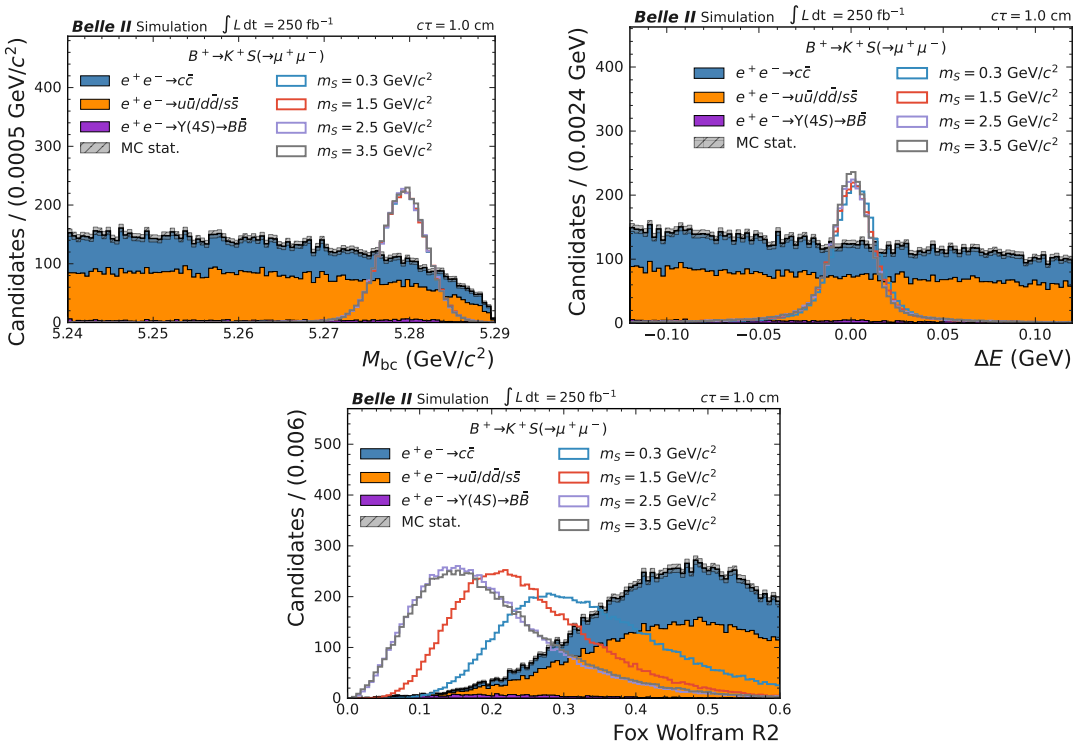

Figure 5.3.: Distributions of background and signal for different S masses in $B^+ \to K^+S(\to \mu^+\mu^-)$ for (upper left) M_{bc}, (upper right) ΔE and (lower) Fox-Wolfram $R2$. The signal is normalised arbitrarily.

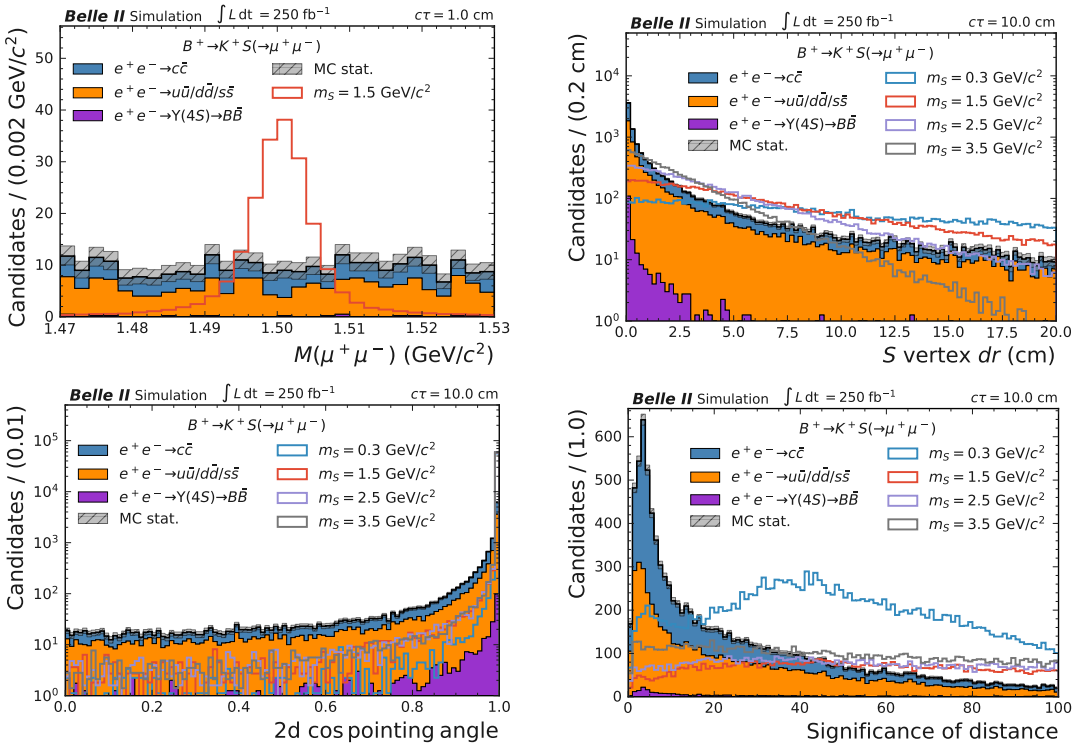

Figure 5.4.: Distributions of background and signal for different S masses in $B^+ \to K^+S(\to \mu^+\mu^-)$ for (upper left) M_S, (upper right) vertex dr, (lower left) 2d $\cos\theta^S_{\text{pointing}}$ and (lower right) Significance Of Distance. The signal is normalised arbitrarily.

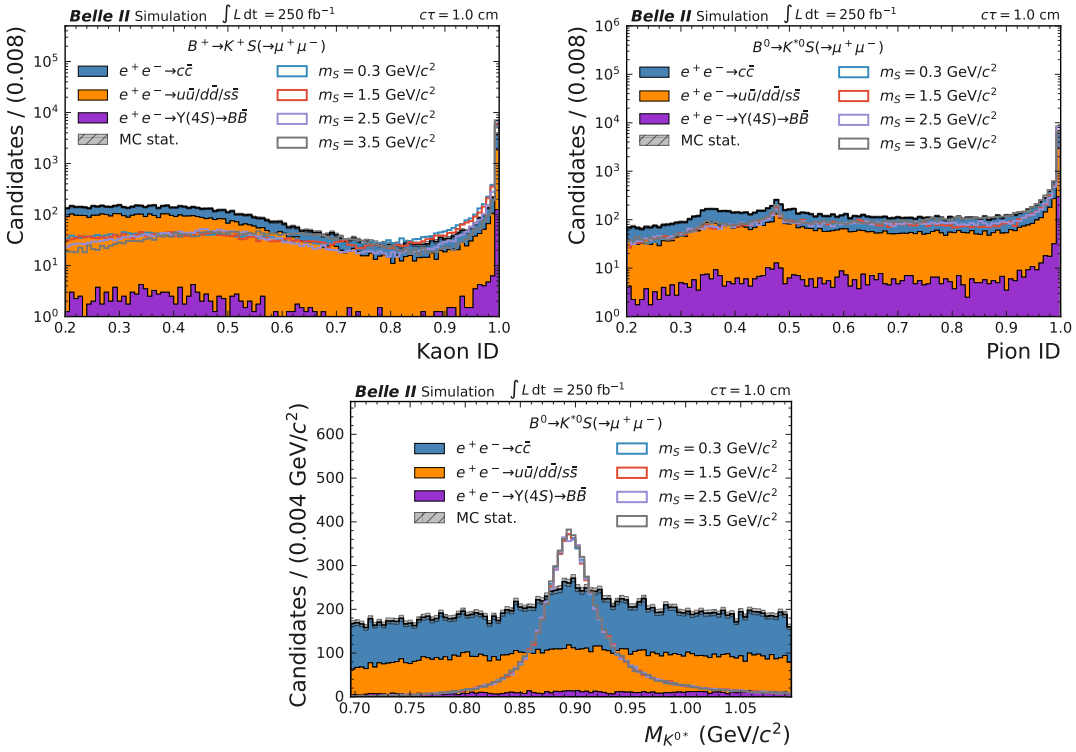

Figure 5.5.: Distributions of background and signal for different S masses for (upper left) the prompt kaon PID score in $B^+ \to K^+S(\to \mu^+\mu^-)$, (upper right) the prompt pion PID score in $B^0 \to K^{*0}(\to K^+\pi^-)S(\to \mu^+\mu^-)$ and (lower) $M_{K^{*0}}$ in $B^0 \to K^{*0}(\to K^+\pi^-)S(\to \mu^+\mu^-)$. The signal is normalised arbitrarily.

- 2d $\cos\theta^{S}_{\text{pointing}} > 0$,
- Prompt kaon/pion global PID > 0.2,
- $(0.696 < M_{K^{*0}} < 1.096)\,\text{GeV}/c^2$.

PID requirements on the S daughter tracks are applied to all distributions except for the PID scores themselves, which are detailed in Sec. 5.3

Further selections include track quality requirements of $|dz| < 2\,\text{cm}$, $|dr| < 0.5\,\text{cm}$, and $p_\text{T} > 150\,\text{MeV}$ applied to select the prompt tracks. The selection ensures that the prompt tracks emerge from close to the IP and reduces fake tracks that arise mostly from beam background particles producing detector hits. No such requirements are placed on the S daughter tracks, as they are expected to originate from a displaced vertex, leading to a higher rate of fake tracks and susceptibility to beam background levels.

K^0_S are naturally long-lived and have the same pointing characteristic and decay to a pair of oppositely charged pions. This means that K^0_S poses an irreducible background to a search for an S in the $S \to \pi^+\pi^-$ channel, and a veto is applied. S candidates with $(489 < M_{\pi^+\pi^-} < 507)\,\text{MeV}/c^2$ are rejected from the search. A significant number of real K^0_S are reconstructed in the other final-state channels even after PID selections due to residual misidentification coupled with the high abundance of K^0_S in B- and D-meson decays and in $e^+e^- \to s\bar{s}$ events. They are rejected by recomputing the reconstructed S mass assuming the pion mass for the daughters and vetoing the same mass region.

No vertex fit is applied on the B-meson candidate, since it is found to not improve the signal resolution and to lead to increased background contamination from B-meson processes at large S masses.

The event-shape variable, Fox-Wolfram $R2$, is calculated from photon candidates with an ECL cluster energy of $E > 50\,\text{MeV}$ and tracks with a minimum transverse momentum $p_\text{T} > 100\,\text{MeV}/c$.

Electron tracks in the $S \to e^+e^-$ channel are corrected for possible energy loss in the detector material due to bremsstrahlung. The four-momentum of photon clusters that are matched to the electron track as potential bremsstrahlung are added to it.

The bremsstrahlung correction takes the helix parameters at the IP and extrapolates to the ECL. Photon cluster candidates within a cone of a certain angular threshold around the extrapolated track are chosen if their energy is below another threshold value. Photons are selected by requiring more than one crystal in the ECL cluster, a timing of the ECL cluster below 200 ns after the collision time and excluding the outermost 5° of the ECL in forward and backward directions. If a photon is matched to more than one electron, the photon is used multiple times and added to all of them. The energy and angular thresholds are determined for three different momentum bins, shown in Tab. 5.1.

Table 5.1.: Parameters used in the bremsstrahlung correction.

Electron momentum (GeV/c)	Photon energy (GeV)	Angular threshold (rad)
< 0.6	< 0.09	< 0.137
$0.6 < p < 1.0$	< 0.9	< 0.074
> 1.0	< 1.2	< 0.063

The correction increases the resolution of the reconstructed S mass, M_{bc}, and ΔE. The effect is most pronounced for small S vertex displacements and is reduced for longer lifetimes because the electrons have to pass less material and the matching using helix parameters at the IP ceasing efficiency.

After these preselections, the combinatorial background from light-quark-pair events dominates. The largest contribution to this background is from $e^+e^- \to c\bar{c}$ events. The backgrounds in the $S \to \pi^+\pi^-$ channels are the largest because pions are the most abundant particle type in e^+e^- collision events at Belle II and the dominance of combinatorial backgrounds within the selection. The contribution from $e^+e^- \to \tau^+\tau^-$ events is reduced to a negligible level.

5.3. Displaced final state identification

The invariant mass of the S candidate depends on the mass hypothesis chosen for the two daughter tracks. As this analysis considers four final states of the S, there are four candidates per combination of oppositely charged tracks. This leads to a different

reconstructed mass M_S of a track combination in each final-state channel. The signal yield is determined using the reconstructed mass of the S candidates. The correct S daughters will be chosen by identification of the final-state particles.

5.3.1. Discrimination via kinematic constraints

The first level of discrimination between the S decay channels is achieved using the energy difference ΔE. Figure 5.6 shows ΔE distributions of the correct and the three wrong S daughter mass hypotheses. The ΔE values are shifted towards values larger than zero if the assumed daughter mass is larger than its true mass, and vice versa. The kaon mass is large compared to the other three mass hypotheses, resulting in a well-separated distribution. Thus, selections in ΔE can be used to sufficiently minimise misidentification between events in $S \to K^+K^-$ and the other three final-state channels. The three lighter hypotheses are too close to each other to disentangle the final states via ΔE. Additional specific particle identification requirements must be used. Different S masses and lifetimes are not found to change the ΔE discrimination between final states.

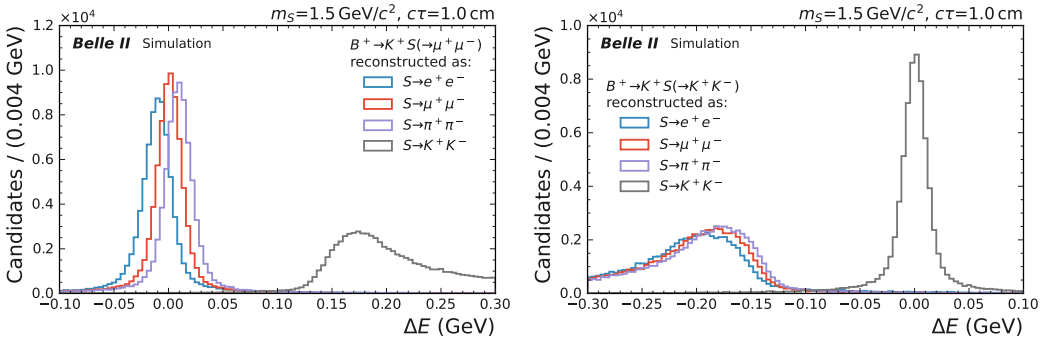

Figure 5.6.: ΔE distributions for generated (left) $S \to \mu^+\mu^-$ and (right) $S \to K^+K^-$ events reconstructed in all four final-state channels.

5.3.2. Particle identification via detector likelihoods

Specific PID information is used to distinguish between the electron, muon and pion final-state channels. Furthermore, the backgrounds in each final-state channel from other particle sources are reduced, thereby increasing the sensitivity of the search.

TOP performance for LLPs

A degradation of the PID performance is observed for daughter tracks of LLPs, especially from the TOP detector. The TOP detector heavily relies on the timing information of the tracks arrival to calculate the likelihood for different particle hypothesis. The TOP PID response assumes the production of the track close to the IP. If the track originates from a LLP decay, the true production time is greater than the assumed time. The time difference depends on the S vertex distance and velocity, i.e. its lifetime and Lorentz-boost. The difference is largest for heavy S with a long lifetime. An example of TOP performance degradation for a high mass and lifetime sample is shown in Fig. 5.7. The binary π/K particle identification is calculated using the TOP likelihoods only:

$$\text{TOP only binary } \pi/K\text{ID} = \frac{L_\pi^{\text{TOP}}}{L_\pi^{\text{TOP}} + L_K^{\text{TOP}}}. \tag{5.4}$$

Compared are true pions from S decays at a small S mass and lifetime to pions from a large S mass and lifetime sample. The TOP detector is found to be able to correctly classify pions from the small S mass and lifetime sample as more pion-like than kaon-like. For the large S mass and lifetime sample, TOP misidentifies a large number of true pions as being closer to the heavier kaon hypothesis because of their later timing. The events correctly identified in the long-lifetime sample are decays of S still close to the IP, as the displacement of the vertex follows an exponential distribution.

Identification of $S \to e^+e^-/\mu^+\mu^-/\pi^+\pi^-$

Three PID definitions are studied to discriminate between the lighter S final states. The default global PID score was introduced in Sec. 3.3.3 using the likelihoods of the

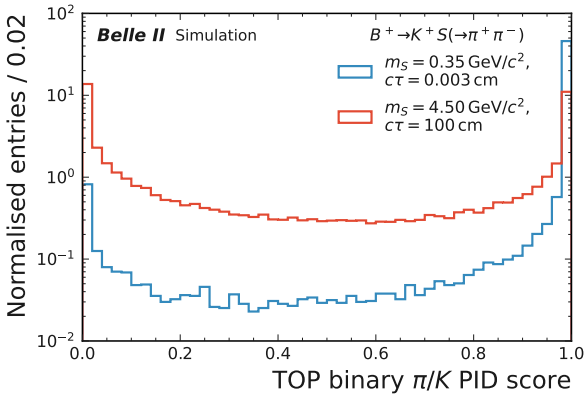

Figure 5.7.: TOP only binary pion/kaon classification for an S sample with small mass and lifetime compared to a large mass and lifetime sample.

CDC, ECL, TOP, ARICH and KLM under the charged-particle hypotheses $e/\mu/\pi/K/p/d$. A modified version of the global PID score excluding the TOP likelihood is studied due to the performance degradation observed for TOP. Furthermore, a trinary PID score with excluded TOP likelihoods is studied to minimise the effect of misidentifying S daughters as heavier particles. Only the likelihoods of the three light hypotheses important for the analysis are included in the trinary PID score:

$$\mathrm{PID}_{\mathrm{Trinary}} = \frac{\mathcal{L}_{\mathrm{Test}}}{\mathcal{L}_e + \mathcal{L}_\mu + \mathcal{L}_\pi}. \tag{5.5}$$

An S candidate passes the PID requirement of a given final state if both S daughter tracks have a PID score under the given final-state hypothesis above a certain threshold. Threshold values are studied from 0.1 to 0.9 in steps of 0.1.

An alternative PID strategy: choosing the final state based on the highest value of the sum of the PID scores of both tracks, was tried and found to yield worse performance.

PID efficiency The PID efficiency ϵ_{PID} for a given final state, $S \to x^+x^-$, is defined as:

$$\epsilon_{\text{PID}} = \frac{\#\text{correctly identified } S \to x^+x^-}{\#\text{true } S \to x^+x^-}, \tag{5.6}$$

where the number of true candidates is given by the number of generated signal candidates in the final state that pass the other pre-selection criteria.

The PID efficiencies for the three lighter final states as a function of S mass and lifetime using the three PID definitions are shown in Fig. 5.8. PID definitions excluding the TOP likelihood are found to retain a large efficiency across different S lifetimes, visible mostly for the pion and electron final states. The muon PID is dominated by the KLM, which is less sensitive to large lifetimes, as the decay vertices have to be in front of the KLM to be able to find the S daughter particles as tracks. The trinary PID excluding the TOP likelihood is found to yield a higher and more uniform efficiency across S masses compared to the global PID excluding TOP. Therefore, the trinary PID without TOP is used to identify the three light S final-state channels.

Figure of merit for the sensitivity A figure of merit quantifying the signal sensitivity of a given selection requirement is needed to find the optimal selection value. The figure of merit due to Punzi [159]:

$$f_{\text{Punzi}} = \frac{\epsilon_{\text{Sig.}}}{\left(\sigma/2 + \sqrt{n_{\text{Bkg.}}}\right)}, \tag{5.7}$$

where $\epsilon_{\text{Sig.}}$ is the signal efficiency and $n_{\text{Bkg.}}$ the number of remaining background candidates, is used here with $\sigma = 5$ in order to optimise for discovery. This figure of merit has the advantage that no absolute signal normalisation is required, only the signal efficiency and the number of background candidates. The number of background candidates depends on the selection requirements and the dataset size. The absolute signal normalisation depends on the cross section of the signal process and is often a free parameter of the theory on which experimental bounds are set. Using f_{Punzi} to optimise the signal selection effectively yields the minimally detectable signal cross section for a given dataset size.

The optimisation at a given S mass is done by counting background candidates in

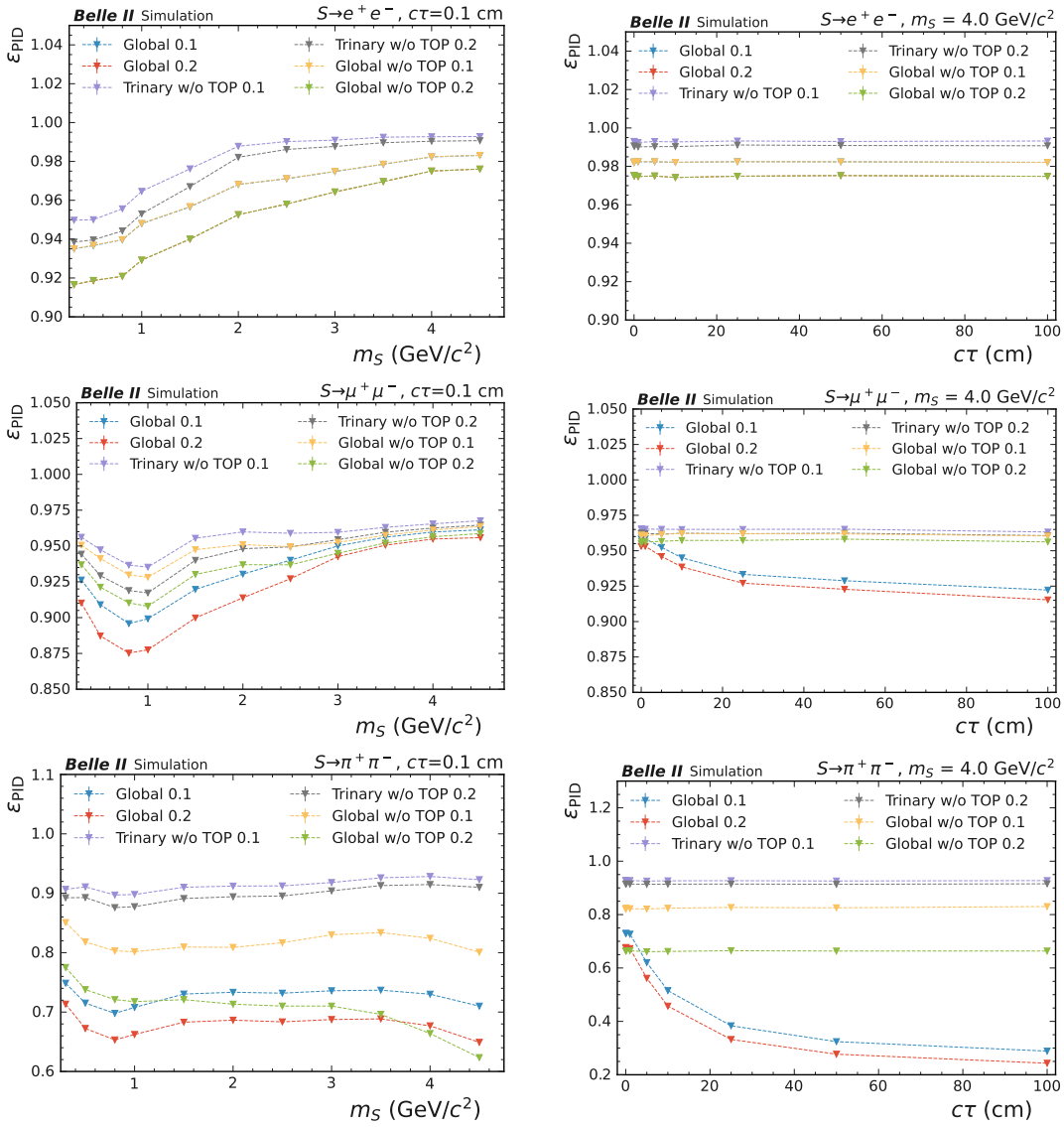

Figure 5.8.: PID efficiencies for different PID definitions as function of (left) S mass and (right) S lifetime for (upper) $S \to e^+e^-$, (middle) $S \to \mu^+\mu^-$ and (lower) $S \to \pi^+\pi^-$. The number in the legend corresponds to the PID score threshold applied to both S daughter tracks.

the reconstructed S mass distribution using a window of $(-3\sigma_{\text{Sig.}}, +3.5\sigma_{\text{Sig.}})$ around the tested S mass, with the signal S mass resolution $\sigma_{\text{Sig.}}$. The asymmetric window accounts for longer tails towards larger reconstructed masses observed for larger S lifetimes. The statistical uncertainties on the signal efficiency and the number of background candidates are symmetrised and propagated to f_{Punzi}.

The figure of merit is shown in Fig. 5.9 for different thresholds in the trinary PID score without TOP as a function of S mass for the three light final states. Additionally,

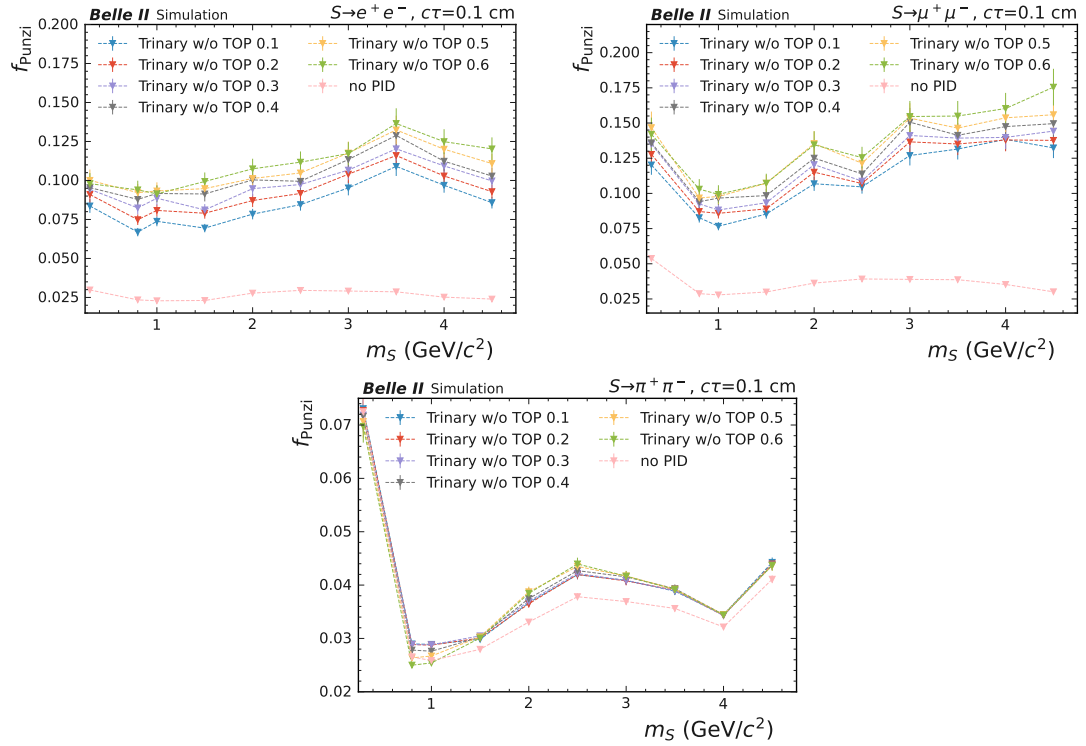

Figure 5.9.: Figure of merit f_{Punzi} as a function of S mass for different minimum PID requirements for (upper left) $S \to e^+e^-$, (upper right) $S \to \mu^+\mu^-$ and (lower) $S \to \pi^+\pi^-$.

the figure of merit value without application of a PID requirement is shown. Including a PID requirement is found to increase f_{Punzi} for all studied S masses and final states, especially for the electron and muon final states. This highlights the background

rejection properties of PID in addition to discrimination between the S final states. The sensitivity improvements are smallest in the pion final state because most tracks in the combinatorial background originate from true pions that pass the pion PID selection. The background in the muon and pion final-state channels is highest around a reconstructed S mass of $M_S \approx 1\,\text{GeV}/c^2$ leading to a dip in f_{Punzi} around that mass.

The value of f_{Punzi} is increasing for tighter selections in the PID score. However, the difference between the selection values is not large, especially given the size of the statistical uncertainties resulting from the low remaining background numbers with tight selections. Above a medium selection value, the gain in f_{Punzi} is small and further selections in the kinematic features introduced earlier will be optimised. Furthermore, it is desirable to retain a small but finite number of background candidates to perform the statistical interpretation using likelihood fits. The same conclusion is drawn when studying the figure of merit as a function of S lifetime. For all three final states studied, a medium PID selection of trinary PID score without TOP > 0.4 applied on both tracks is chosen.

PID purity The signal PID purity is defined as:

$$\epsilon_{\text{PID}} = \frac{\#\text{correctly identified } S \to x^+x^-}{\#\text{identified } S \to x^+x^-}. \tag{5.8}$$

This PID purity is calculated only for signal events, as the primary aim is to disentangle the different S decay channels. The background is determined directly using data. This means that, as long as the remaining backgrounds have a low absolute normalisation, it is of low concern whether the true particle type of a candidate corresponds to the identified one.

The purity of the signal for the selection is shown in Fig. 5.10 as a function of S mass and lifetime for the three light final-state channels. The high purity of the final states identified yields a negligible expected cross-feed between signal channels. This is especially the case since this analysis is a rare new-physics search: at most only a few signal candidates would be found, given existing bounds.

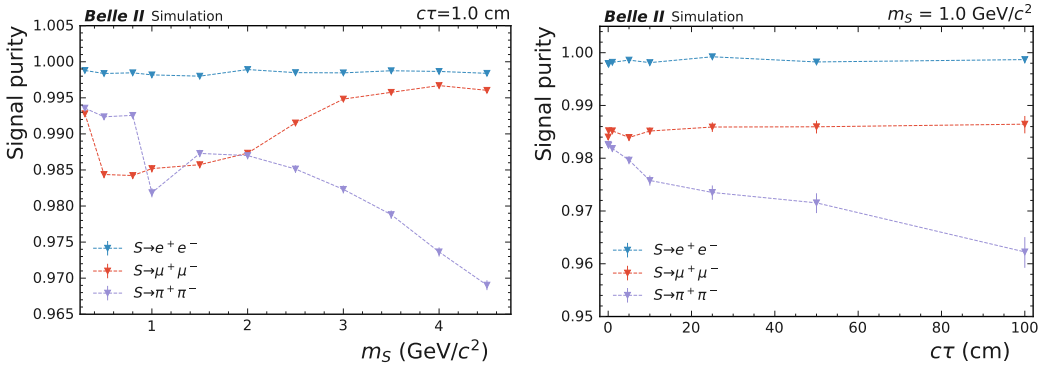

Figure 5.10.: Signal purity with the PID requirements for the three light final-state channels as a function of (left) S mass and (right) S lifetime.

Identification of $S \to K^+K^-$

Further PID selections are applied to $S \to K^+K^-$ final-state channels, as there are more backgrounds from non-kaon tracks. The lighter signal final-state channels could feed into the kaon channel by misidentification from the TOP detector as heavier particles. The ΔE selection prevents this, and PID scores with the TOP likelihood included can be used. However, a deteriorating effect is observed at large lifetimes due to kaons being misidentified as even heavier protons or deuterons. The binary kaon/pion PID including only the kaon and pion likelihoods is not affected by this, as the late timing will not lead to misidentification of signal kaons as lighter pions. True pions from backgrounds most often do not originate from a true displaced vertex and are not misidentified as kaons.

The performance of the global and binary PID in the $S \to K^+K^-$ channel is shown in Fig. 5.11 using the f_{Punzi} figure of merit. The global kaon PID is found to yield higher performance compared to the binary kaon/pion PID for most of the studied phase-space since more background particle types can be rejected. At large S masses and lifetimes the TOP begins to misidentify signal kaons, resulting in a loss of signal efficiency and, in turn, also in f_{Punzi}. Toward larger lifetimes, the loss in signal efficiency leads to a region where no PID, with correspondingly much larger backgrounds, yields a comparable f_{Punzi} value. The binary PID is still able to reject

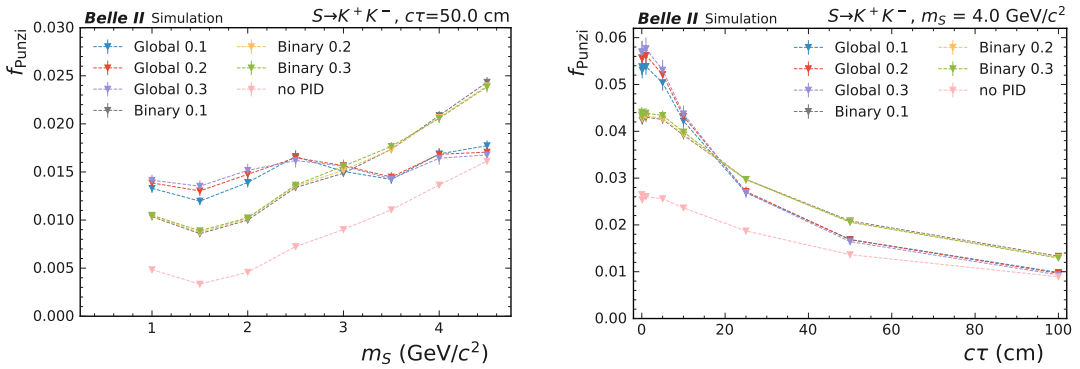

Figure 5.11.: Figure of merit f_{Punzi} in the $S \to K^+K^-$ channel for binary and global PID definitions as a function of (left) S mass and (right) S lifetime.

pions, the largest background source, while retaining the signal efficiency, resulting in a better performance compared to the global PID.

The turnover point where the binary PID starts to perform better is reached for $m_S > 3\,\text{GeV}/c^2$ and $c\tau_S > 15\,\text{cm}$. The global (binary) PID is applied to all candidates when a signal hypothesis with parameters below (above) this point is tested. A selection of global/binary$_{K,\pi} > 0.2$ is used on both S daughter tracks in the kaon final-state channel as larger values are not found to further increase the sensitivity.

PID distribution The distributions of PID scores for the S daughter tracks are shown for background and signal in Fig. 5.12. Shown is the PID score of the positively charged S daughter track, the differences to the negative charge are minimal.

5.4. Selection optimisation

The signal selection is further optimised by studying f_{Punzi} for the kinematic quantities introduced in Sec. 5.2. A range of selection values is scanned in each of the observables. At each scan step, the remaining signal efficiency and background candidates in the Punzi window are determined to compute the value of f_{Punzi}. This is done separately for different signal hypotheses such as masses, lifetimes, or channels. The selection

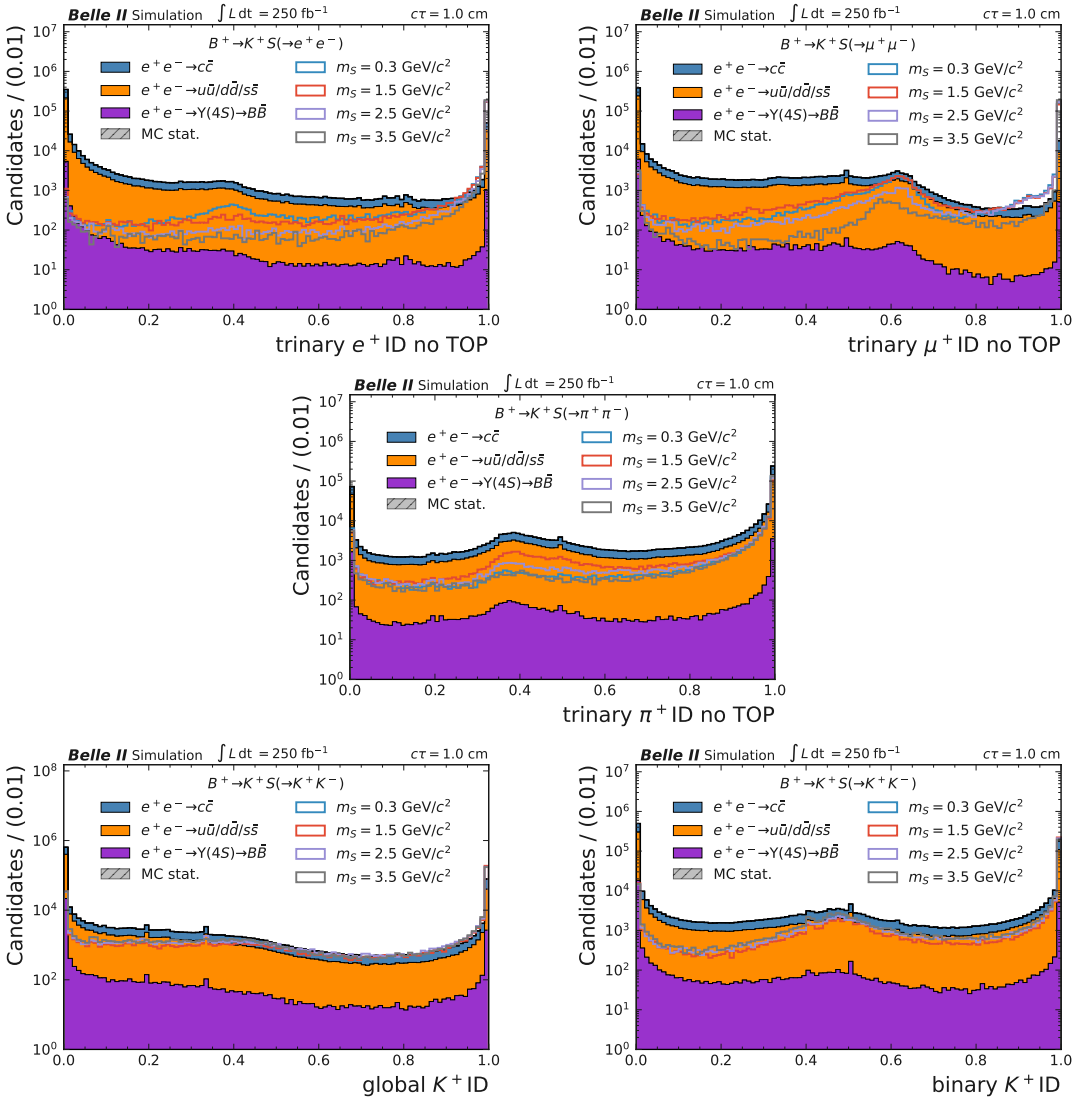

Figure 5.12.: Distributions of background and signal for different S masses in $B^+ \to K^+S$ in the trinary PID scores in (upper left) $S \to e^+e^-$, (upper right) $S \to \mu^+\mu^-$, (middle) $S \to \pi^+\pi^-$, and in the $S \to K^+K^-$ channel in the (lower right) global PID score and (lower right) binary kaon/pion PID score. The signal is normalised arbitrarily.

in one observable is optimised by finding the selection value that maximises f_{Punzi}. The selection is then used as an additional pre-selection to the optimisation of the selection in the following observables. This optimisation procedure usually yields a preference for tight selections in the observables that are used early in the chain. The procedure is repeated a few times with variations in the order of observables and comparing the resulting optimal selection requirements to minimise the effect. As a result, slightly looser selection values than preferred by the one-dimensional f_{Punzi} optimisation scheme are chosen.

More strict requirements on ΔE, Fox-Wolfram $R2$, and 2d $\cos\theta^{S}_{\text{pointing}}$ are preferred in the pion final-state channels compared to the other channels. This is due to larger backgrounds, which leads to an increase in sensitivity with tighter selections despite the reduced signal efficiency. All selections that are used in $B^+ \to K^+S$ are also used in $B^0 \to K^{*0}(\to K^+\pi^-)S$ as the signal distributions in the observables used for the selection do not differ considerably between the production modes and in order not to complicate the selection. The resulting optimised selection values are summarised in Sec. 5.6.

Optimisation example

An example of one optimisation step using M_{bc} is given below. The M_{bc} distributions of signal and background are shown in Fig. 5.13 with a f_{Punzi} scan to find the optimal selection value. The combinatorial background from light-quark-pair events constitutes the main background source after rejection of B-meson decays close to the IP and of K^0_S. These candidates follow an ARGUS-like distribution [156] that drops toward the kinematic endpoint around $5.9\,\text{GeV}/c^2$. On the other hand, true B-meson candidates, such as the overlaid signal, feature a clear peak around the mass of the B-meson. The scan of possible selection values starts at the pre-selection of $5.24\,\text{GeV}/c$ and finds an increasing f_{Punzi} up to an optimal point around $M_{\text{bc}} > 5.27\,\text{GeV}/c^2$ around the start of the signal distribution. The loss of signal efficiency for even tighter cuts is too large to be balanced by a further decrease in background levels, resulting in the sudden drop in f_{Punzi}.

Several of these one-dimensional f_{Punzi} optimisation scans are performed for dif-

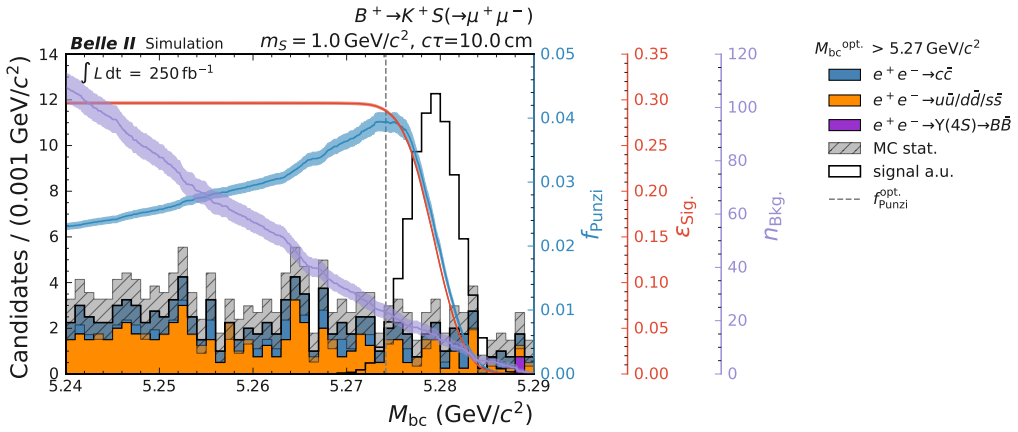

Figure 5.13.: M_{bc} distribution with f_{Punzi} scan for $S \to \mu^+\mu^-$ with $m_S = 1.0\,\mathrm{GeV}/c^2$ and $c\tau_S = 10\,\mathrm{cm}$. The signal is normalised arbitrarily.

ferent S masses and lifetimes, as well as analysis channels. The optimal and chosen selection values on M_{bc} are shown in Fig. 5.14 across S masses and lifetimes for the $B^+ \to K^+S(\to \mu^+\mu^-)$ channel. Furthermore, the value of f_{Punzi} and of the quantities that are used for its calculation, the number of background candidates and signal efficiency in the optimisation windows with the optimal and chosen selection values are shown. The chosen single selection value of $M_{\mathrm{bc}} > 5.27\,\mathrm{GeV}/c^2$ for all S masses and lifetimes is found to produce a comparable sensitivity, within statistical uncertainty, to an S mass and lifetime-dependent selection.

5.5. Peaking backgrounds

Peaking background sources are found in the reconstructed mass distributions after heavily suppressing combinatorial backgrounds by application of the PID requirements and optimised selections in kinematic quantities previously introduced.

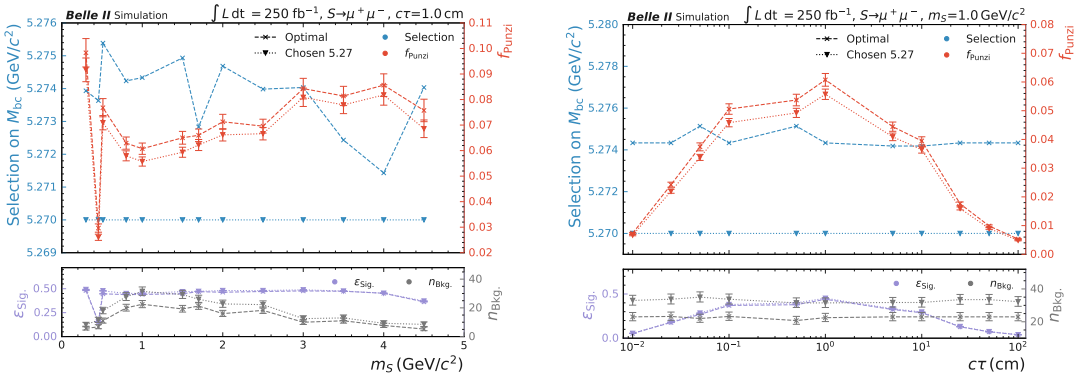

Figure 5.14.: Comparison of f_{Punzi} for optimal and chosen selection values in M_{bc} for different S (left) masses and (right) lifetimes.

5.5.1. Resonances

Many of these backgrounds can be attributed to two-body decays of SM resonances in true B-meson decays that pass the selections in the B-meson kinematic features. These candidates pass the minimum vertex displacement selection mainly due to being part of the far end of the lifetime tail of prompt decays, as well as misreconstruction of S vertex positions. Due to the large branching fractions of these decays compared to the searched for signal process, a couple of candidates pass the selection and cause peaks.

A veto in the reconstructed S mass around these peaking backgrounds leads to complete insensitivity to some mass regions and is not necessary. The vertex displacement distribution of these falls rapidly beyond the minimum displacement selection. The backgrounds are rejected by tightening the selection of minimum vertex displacement to $dr > 0.2\,\text{cm}$ in S mass regions around the resonances. The tighter dr selection is applied to all candidates in the signal extraction window when a signal mass hypothesis is tested that lies within the determined peaking background windows. This is done to avoid kinks in the reconstructed mass distribution within a given signal extraction window. Some of the identified peaking two-body SM decays identified are also not to two particles of the same type, e.g. $D^0 \to K^-\pi^+$. These decays enter

the signal selection by misidentification of final-state particles. The mass hypothesis of the misidentified track is wrong, shifting the windows in the reconstructed S mass where the peaking background is found away from the true mass of the resonance. The resonance suppression windows are used in both production modes as the same backgrounds are found. A summary of these backgrounds and the S mass windows in which the suppression is used is given in Sec. 5.6.

Resonance background example

The reconstructed mass distribution in the $B^+ \to K^+S(\to \mu^+\mu^-)$ channel after PID and optimised selections is shown in Fig. 5.15 with and without additional rejection of the peaking background. Clearly visible is the peak around the J/ψ resonance. Further shown are the reconstructed S mass and dr distributions where event generator information is used to restrict to true J/ψ being identified as S candidates. These originate from the SM $B^+ \to J/\psi\,(\to \mu^+\mu^-)K^+$ process, which has a large branching fraction of 10^{-3} [1]. The J/ψ produced in a B-meson decay in association with a kaon and decaying to $\mu^+\mu^-$ mimics the $B^+ \to K^+S(\to \mu^+\mu^-)$ signal process and enters the selection. In the signal extraction using the reconstructed S mass distribution these candidates would be counted as signal and hence need to be suppressed. The vertex dr distribution of these candidates shows that they can be suppressed to negligible levels using a tighter selection of $dr > 0.2\,\text{cm}$.

5.5.2. Photon conversions

The reconstructed S mass distribution in the electron final-state channels feature clear peaks toward very low masses from photon conversions $\gamma \to e^+e^-$ in detector material. Most of the photon conversion candidates are from the dominant light-quark-pair background processes which enter the kinematic selection as combinatorics. The photon conversion vertex features true displacement and passes the minimum vertex displacement and vertex quality requirements. Photon conversions should peak at zero S candidate mass as photons are massless. However, due to the difficult reconstruction of two tracks with very small opening angles, there is a long tail up to

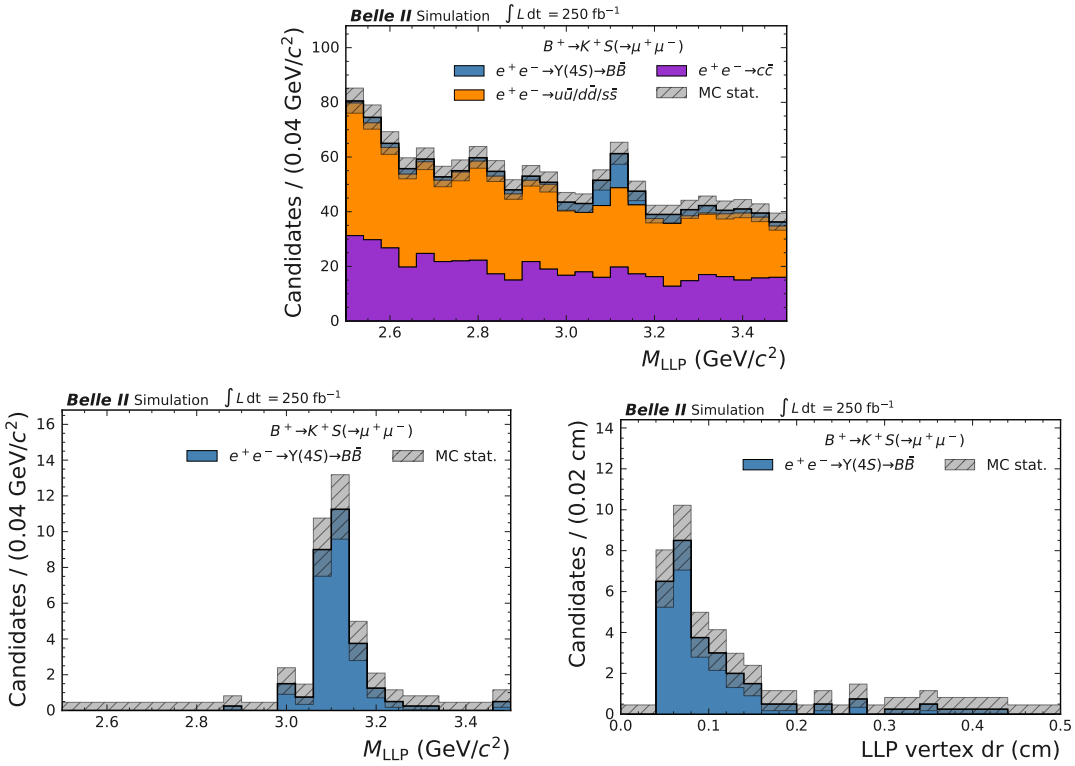

Figure 5.15.: (Upper) Reconstructed mass distribution in $B^+ \to K^+S(\to \mu^+\mu^-)$ after PID and optimised selection requirements, and true J/ψ candidates distributions in $B^+ \to K^+S(\to \mu^+\mu^-)$ of (lower left) the reconstructed mass and (lower right) the vertex dr.

S masses around $50\,\mathrm{MeV}/c^2$. This background is reduced by a veto in the S vertex dr distribution around the PXD and SVD layers when signal masses below $50\,\mathrm{MeV}/c^2$ are tested.

The signal selections are loosened to the pre-selection values with the PID requirements in place to study the dr distribution of photon conversion in simulation enriched with photon conversions. Example distributions of the vertex dr around PXD layer one and SVD layer one are shown in Fig. 5.16 for background S candidate masses below $50\,\mathrm{MeV}/c^2$ in the $B^+ \to K^+S(\to e^+e^-)$ channel. Clear peaks are found around most layers. No significant photon conversion peak is observed in the dr region of the second, not fully instrumented PXD layer, even when restricted to the azimuthal angle range in which modules are installed. The vertex reconstruction is found to bias the photon conversion vertex towards active detector material locations and to the IP. This leads to no photon conversions found around the beam pipe. The vertex dr regions around the second PXD layer and the beam pipe are therefore not vetoed. Photon conversions have not yet been studied extensively in Belle II data at the time of this study. As a precaution towards potential larger photon conversion backgrounds in data compared to simulation the regions around SVD layer three and four are vetoed, albeit no large contributions are found in simulation. The veto windows around the active layers are chosen as: $(1.3, 1.55)\,\mathrm{cm}$, $(3.75, 4.25)\,\mathrm{cm}$, $(7.5, 8.4)\,\mathrm{cm}$, $(9.8, 10.8)\,\mathrm{cm}$, and $(13, 14)\,\mathrm{cm}$.

The effectiveness of the veto is shown in Fig. 5.17 with a comparison of the S mass distribution in $B^+ \to K^+S(\to e^+e^-)$ for small S candidate masses before and after adding the photon conversion rejection.

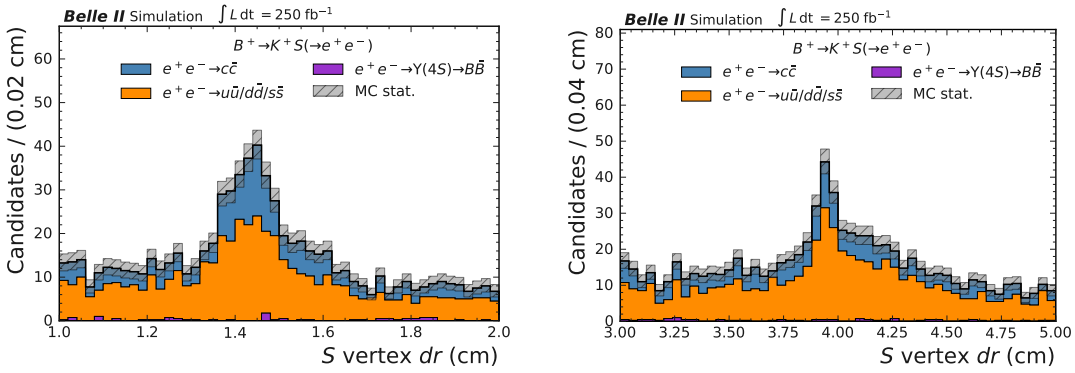

Figure 5.16.: S vertex dr distribution in $B^+ \to K^+S(\to e^+e^-)$ for reconstructed masses below $50\,\mathrm{MeV}/c^2$ around (left) PXD layer 1, (right) SVD layer 1.

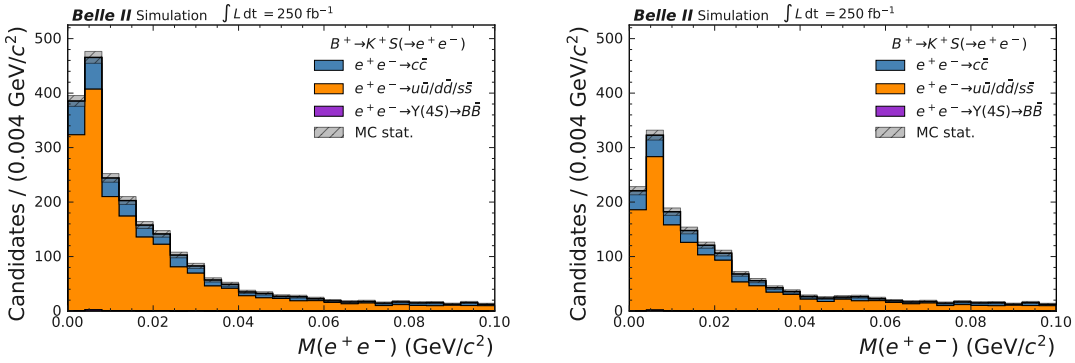

Figure 5.17.: Reconstructed S mass distribution in $B^+ \to K^+S(\to e^+e^-)$ (left) without and (right) with veto around active PXD and SVD layers.

5.6. Selection summary

A summary of the selection requirements used in the analysis is shown in Tab. 5.2. The peaking backgrounds that were identified and are rejected by a tighter S vertex dr selection are summarised in Tab. 5.3.

If multiple candidates in an event pass these requirements, the candidate with the minimum absolute value of ΔE is chosen. In case multiple candidates arise due to different final-state hypotheses passing the PID selection, this selects the mass hypothesis closest to the expectation. The fraction of events with more than one candidate is less than 0.5%.

The signal efficiencies after all selections averaged across the searched range of S masses for the different analysis channels and for different S lifetimes are shown in Fig. 5.18. The values are corrected for differences between the data and simulation, described in more detail in Sec. 7.2. The signal efficiency is reduced around the K^0_S mass region due to the rejected window in $M_{\pi^+\pi^-}$ and in regions where the peaking backgrounds are suppressed, mostly for short lifetimes. For small (large) lifetimes, the signal efficiency is smaller (larger) at large S masses due to the Lorentz-boost and the minimum vertex displacement requirement. Furthermore, a variation of the signal efficiency with the S mass arises due to fixed selection requirements on kinematic features that have an S mass dependence such as Fox-Wolfram $R2$, and the PID, whose efficiency drops mostly for low-momentum prompt tracks at very large S masses. The difference between production channels is dominated by the additional selections in the PID of the prompt pion and in $M_{K^{*0}}$ for $B^0 \to K^{*0} S$. The differences between the final-state channels are mainly due to the PID efficiencies and the tighter selections in $S \to \pi^+\pi^-$. The lifetime dependence is due to the minimum vertex displacement selection, which reduces the signal efficiency at small lifetimes and the decrease in track finding and vertexing efficiency to larger lifetimes.

The distributions of the remaining backgrounds in the reconstructed S mass distribution are shown for the $B^+ \to K^+ S$ channels in Fig. 5.19 and for the $B^0 \to K^{*0} S$ channels in Fig. 5.20

Table 5.2.: Selection requirements for the analysis. The same selection requirements apply for the $B^+ \to K^+ S$ and $B^0 \to K^{*0} S$ production modes.

Type	Observable	e^+e^-	$\mu^+\mu^-$	$\pi^+\pi^-$	K^+K^-
Prompt tracks	d_0 (cm)	< 0.5			
	$\lvert z_0 \rvert$ (cm)	< 2			
	p_T (GeV/c)	> 0.15			
	$\mathrm{PID}_\mathrm{global}$	> 0.2			
B-meson	M_bc (GeV/c^2)	> 5.27			
	$\lvert \Delta E \rvert$ (GeV)	< 0.05	< 0.05	< 0.035	< 0.05
Cont. sup.	R2	< 0.45	< 0.45	< 0.35	< 0.45
S	Vertex dr (cm)	> 0.05 and > 0.2 around peaking backgrounds, see Tab. 5.3			
	2d $\cos\theta^S_\mathrm{pointing}$	> 0.95	> 0.95	> 0.99	> 0.95
	Sign. of distance	> 3			
	$M_{\pi^+\pi^-}$ (MeV/c^2)	$\notin (489, 507)$			
S tracks	p_T (GeV/c)	> 0.25, see Sec. 8.1			
	θ (°)	> 32 at ECL surface and < 150 at CDC surface, see Sec. 8.1			
	PID	$\mathrm{PID}_\mathrm{Trinary} > 0.4$			global / binary$_{K,\pi} > 0.2$, see Sec. 5.3.2
K^{*0}	$M_{K^+\pi^-}$ (MeV/c^2)	$\in (796, 996)$			

Table 5.3.: Selection requirements to reject resonance and γ conversion background sources. The resonance background windows correspond to S masses with a tighter vertex dr selection and are given in GeV/c^2. The γ conversion windows correspond to S vertex dr which are vetoed for S masses below $50\,\mathrm{MeV}/c^2$ and are given in cm. The same windows are used for the $B^+ \to K^+ S$ and $B^0 \to K^{*0}\, S$ production modes.

Source	e^+e^-	$\mu^+\mu^-$	$\pi^+\pi^-$	K^+K^-
D^0	$(1.00, 1.35)$	$(1.70, 1.80)$	$(1.65, 1.85)$	$(1.75, 1.95)$
J/ψ	$(3.00, 3.15)$		—	—
$\psi(2S)$	$(3.65, 3.75)$		—	—
η_c	—	—	$(2.85, 3.15)$	$(2.80, 3.20)$
χ_{c1}, $\eta_c(2S)$	—	—	$(3.40, 3.80)$	
ϕ	—	—	—	$(1.00, 1.04)$
γ conv.	$(1.30, 1.55)$, $(3.75, 4.25)$, $(7.50, 8.40)$, $(9.80, 10.80)$, $(13.00, 14.00)$	—	—	—

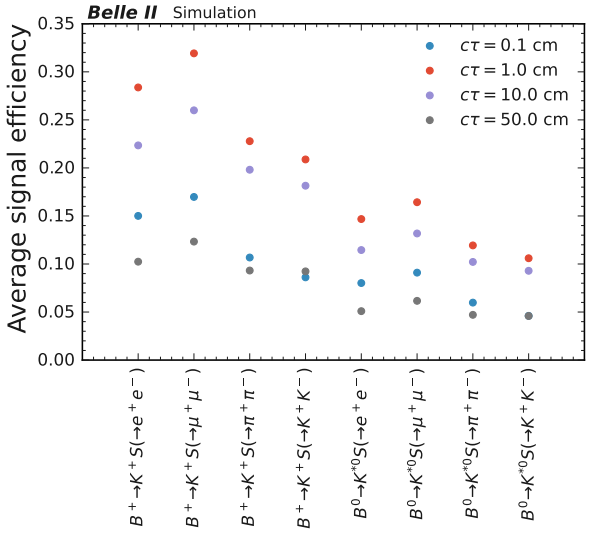

Figure 5.18.: Signal efficiency after all selections averaged across the S mass range for the different analysis channels and for example S lifetime values.

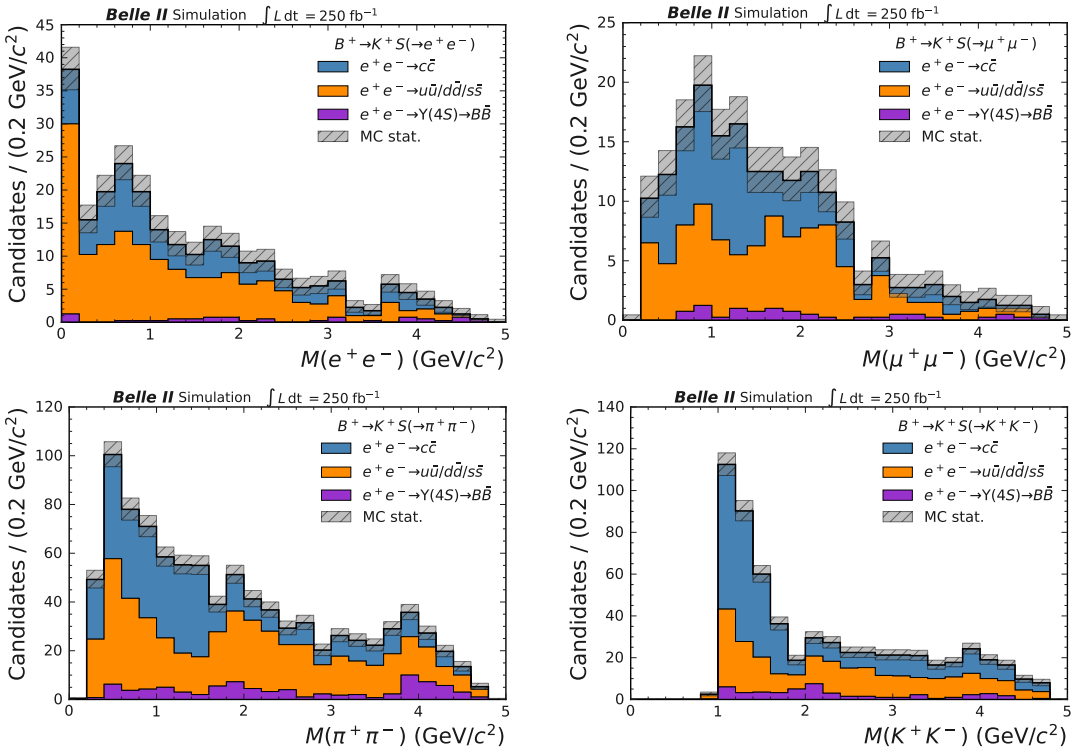

Figure 5.19.: Reconstructed S mass distributions after all selections for remaining backgrounds in $B^+ \to K^+S$ with (upper left) $S \to e^+e^-$, (upper right) $S \to \mu^+\mu^-$, (lower left) $S \to \pi^+\pi^-$ and (lower right) $S \to K^+K^-$.

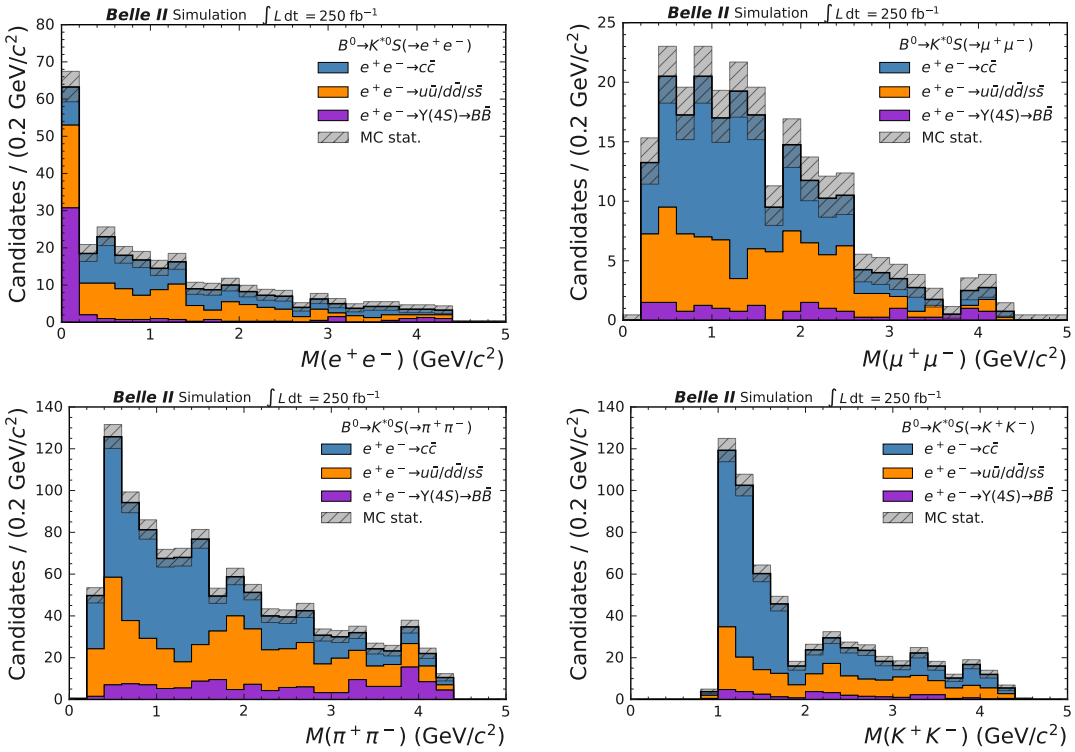

Figure 5.20.: Reconstructed S mass distributions after all selections for remaining backgrounds in $B^0 \to K^{*0} S$ with (upper left) $S \to e^+e^-$, (upper right) $S \to \mu^+\mu^-$, (lower left) $S \to \pi^+\pi^-$ and (lower right) $S \to K^+K^-$.

6. Signal and background modelling

This chapter introduces the models that are used to describe the shape of the signal and background distributions in the reconstructed S mass.

The reconstructed mass of the S candidate is used to search for a signal-like excess of candidates on top of the SM background. Templates describing the shape of the signal and the background in the reconstructed S mass are defined using simulation. Analytical probability density functions (pdfs) are chosen as templates. The pdfs are fitted to the distribution in data to extract the number of signal and background candidates for a given S mass. The implementation of the pdfs is provided by the ZFIT framework [160]. The framework uses the maximum likelihood method [161] to estimate the pdf parameter values that best describe a given dataset. All fits of pdfs to distributions in observables throughout this work are performed directly to the dataset without prior grouping of the data into separate bins.

6.1. Signal model

The signal S mass distribution is modelled by a Double Sided Crystal Ball (DSCB) pdf, defined as:

$$f(x;\mu,\sigma,\alpha_l,\alpha_r,n_l,n_r) = \begin{cases} A_l(B_l - \frac{x-\mu}{\sigma})^{-n_l} & \text{for } \frac{x-\mu}{\sigma} < -\alpha_l, \\ \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) & \text{for } -\alpha_l \leq \frac{x-\mu}{\sigma} \leq \alpha_r, \\ A_r(B_r - \frac{x-\mu}{\sigma})^{-n_r} & \text{for } \frac{x-\mu}{\sigma} > \alpha_r, \end{cases} \tag{6.1}$$

where:

$$A_{l/r} = \left(\frac{n_{l/r}}{|\alpha_{l/r}|}\right)^{n_{l/r}} \exp\left(-\frac{|\alpha_{l/r}|^2}{2}\right),$$
$$B_{l/r} = \frac{n_{l/r}}{|\alpha_{l/r}|} - |\alpha_{l/r}|. \tag{6.2}$$

The core of the DSCB is a Gaussian distribution with mean μ and width σ. The σ parameter is interpreted as the signal resolution. The function transitions to the left and right sides of the Gaussian peak into exponential tails. The points of transition between both are controlled by the $\alpha_{l/r}$ parameters given in units of the Gaussian width σ. From these points on, the tails decrease exponentially with an order parameter $n_{l/r}$. This pdf is a generalisation of the standard Crystal Ball [162], [163] with non-Gaussian tails on both sides instead of one. The exponential order parameter of a side and its corresponding transition parameter both describe the tail and are hence strongly correlated. This leads to an ambiguity of the values determined and to large uncertainties in $n_{l/r}$ from the fit. The order parameters are fixed to a constant value $n_{l/r} = 3$ to increase the stability of the estimation of the values of the pdf parameters. Therefore, the variations in the tails are described by a change in the values of $\alpha_{l/r}$. This results in a more flexible and robust pdf compared to a setup with fixed transition parameters. The DSCB is found to well describe the reconstructed S mass distribution for all combinations of signal model parameters tested, as shown in Fig. 6.3 and Fig. 6.4 in Sec. 6.1.2.

The values of the $\sigma, \alpha_l, \alpha_r$ parameters describe the shape of an excess of signal candidates and are determined for each signal hypothesis tested using simulated signal events. The number of signal-like candidates in a dataset is described by the parameter $N_{\text{Sig.}}$ that acts as a normalisation of the signal pdf. In the signal extraction fit, only the signal yield parameter value is estimated, whereas the shape parameters are kept at their determined values.

6.1.1. Mass definition

The observable which is used for the signal extraction is the reduced mass $M'(x^+x^-)$ defined as:

$$M'(x^+x^-) = \sqrt{M^2(x^+x^-) - 4m_x^2}, \tag{6.3}$$

with the reconstructed mass of the S candidate $M(x^+x^-)$ and the invariant mass of the S daughters m_x. The reduced mass corresponds to twice the S daughter momentum in the S rest-frame.

This definition is chosen to simplify the modelling of the signal width σ close to kinematic thresholds where the S mass approaches twice the invariant mass of the final-state particles. In this region, the largest part of $M(x^+x^-)$ is described by the invariant mass of the final-state particles. The invariant masses are known with precision far exceeding the resolution of the track momenta and opening angle, which determine the width of $M(x^+x^-)$. Therefore, the width σ in the $M(x^+x^-)$ distribution decreases exponentially close to a threshold. On the other hand, the width in the reduced mass behaves linearly close to a threshold and approaches a stable point.

Consistency checks are performed and show that both mass definitions yield almost identical values for $\alpha_{l/r}$. Therefore, the shape of the distribution tails is not changed by the choice of reconstructed mass definition. The widths using both definitions approach the same values in the limit $M_S \gg m_{\text{finalstate}}$. Figure 6.1 shows the estimated values of the σ and α_l parameters close to the K^+K^- kinematic threshold, separately for M_S and $M'(x^+x^-)$ using the $B^+ \to K^+S(\to K^+K^-)$ channel.

The overall shape of the remaining background spectra is unchanged when moving from $M(x^+x^-)$ to $M'(x^+x^-)$. The distribution in $M'(x^+x^-)$ is more washed out on the lower mass side, as it starts from zero instead of twice the final-state particle mass. The differences are greater for heavy S final states such as $S \to K^+K^-$. Both definitions are shown for the $B^+ \to K^+S(\to K^+K^-)$ channel in Fig. 6.2.

6.1.2. Distributions in the reduced mass

Reconstructed $M'(x^+x^-)$ distributions and fits of the DSCB pdf for generated signal events are shown for the four different S final states in Fig. 6.3. All distributions are

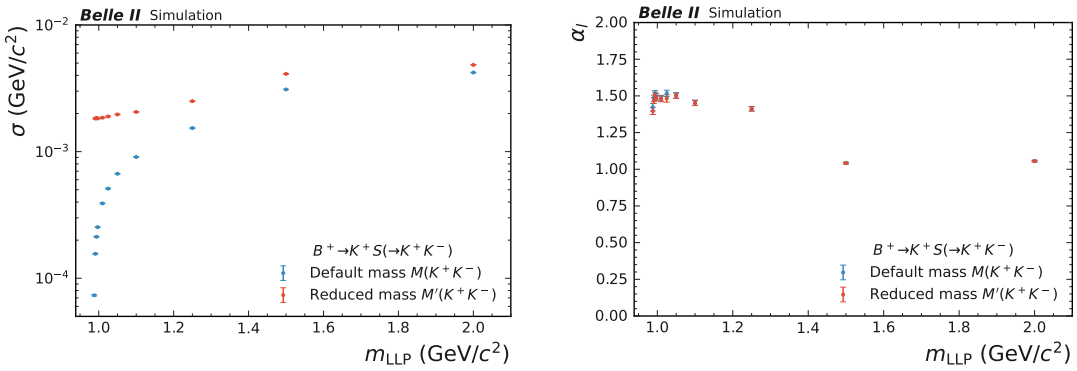

Figure 6.1.: Values of the DSCB parameters estimated by a fit of the DSCB pdf to simulated signal events. (Left) σ and (right) α_l in the $B^+ \to K^+S(\to K^+K^-)$ channel for S masses close to the kinematic threshold, extracted from the reconstructed M_S and $M'(x^+x^-)$ distributions.

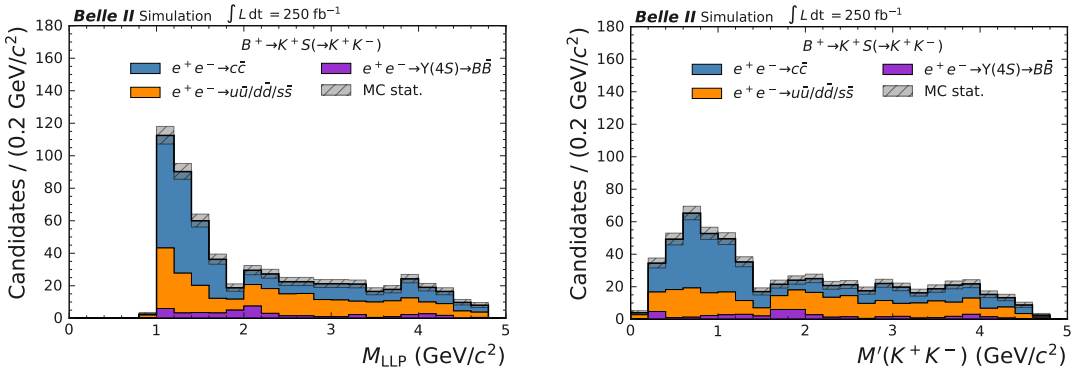

Figure 6.2.: Reconstructed mass distribution of backgrounds in the $B^+ \to K^+S(\to K^+K^-)$ channel with the (left) M_S and (right) $M'(x^+x^-)$ definitions.

well described by the chosen pdf. The distributions for the heavier final states are very similar to each other. The $S \to e^+e^-$ distribution features a longer tail toward small reconstructed S mass values due to non-ideal bremsstrahlung recovery. This longer tail is modelled in the DSCB through a smaller α_l value.

The distributions of $M'(x^+x^-)$ with a fit of the DSCB pdf for different lifetimes are shown in Fig. 6.4. They are quite symmetric for small lifetimes, where most of the S decays occur close to the interaction point. The tail towards large reconstructed masses becomes more pronounced for longer S lifetimes. This is modelled by a gradually decreasing value of α_r for an increasing lifetime of the S. The width parameter σ is almost unaffected by differences in lifetime.

The differences between the reconstructed S mass distributions in $B^+ \to K^+S$ and $B^0 \to K^{*0}\,S$ are found to be negligible. Therefore, the same DSCB parameter values are used in the signal extraction fits for both production modes.

The reconstructed mass distribution for S masses around $20\,\mathrm{MeV}/c^2$ becomes difficult to model, due to experimental resolution effects. The $S \to e^+e^-$ channels are probed down to $25\,\mathrm{MeV}/c^2$ where the signal peak in $M'(x^+x^-)$ is still pronounced and can be adequately modelled by the DSCB pdf.

6.1.3. Interpolation of parameter values

The DSCB parameter values for the signal extraction fits are determined with fits of the pdf to distributions of simulated signal events at various model parameter and channel combinations. The signal extraction is performed in steps given by half of the determined signal resolution σ. Signal samples are simulated for all tested lifetimes and analysis channels with a coarser spacing in the S mass, detailed in Sec. 4.3.2. Therefore, the determined DSCB parameters have to be interpolated between S masses to determine the values for the entire tested phase space. The interpolation is performed with a fourth order Chebyshev polynomial, recursively defined as:

$$\begin{aligned} T_{n+1}(x) &= 2\,x\,T_n(x) - T_{n-1}(x), \\ T_1 &= x. \end{aligned} \tag{6.4}$$

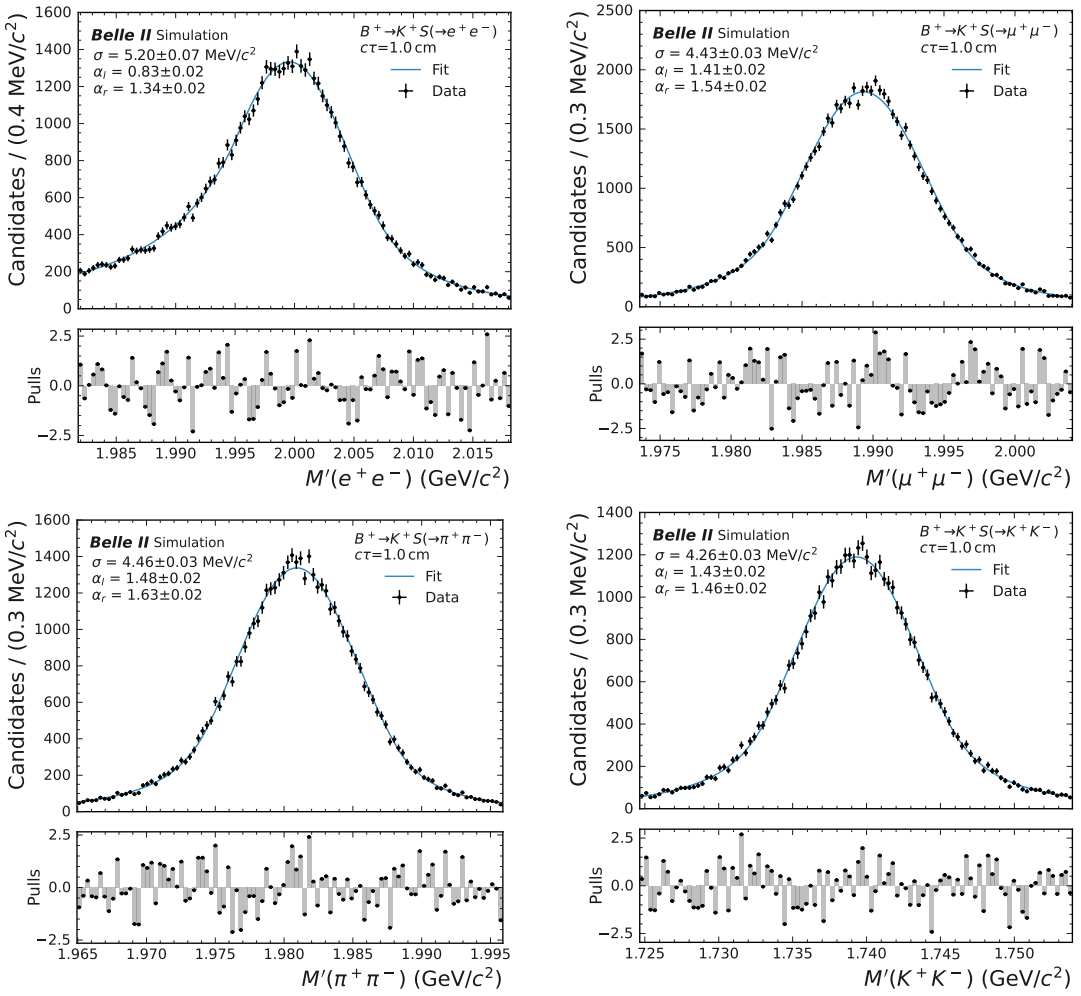

Figure 6.3.: Reconstructed $M'(x^+x^-)$ distributions and fit of the DSCB pdf for simulated signal samples with $m_S = 2\,\mathrm{GeV}/c^2$ and $c\tau_S = 1\,\mathrm{cm}$ in the $B^+ \to K^+S$ production mode for the (upper left) $S \to e^+e^-$, (upper right) $S \to \mu^+\mu^-$, (lower left) $S \to \pi^+\pi^-$ and (lower right) $S \to K^+K^-$ final states.

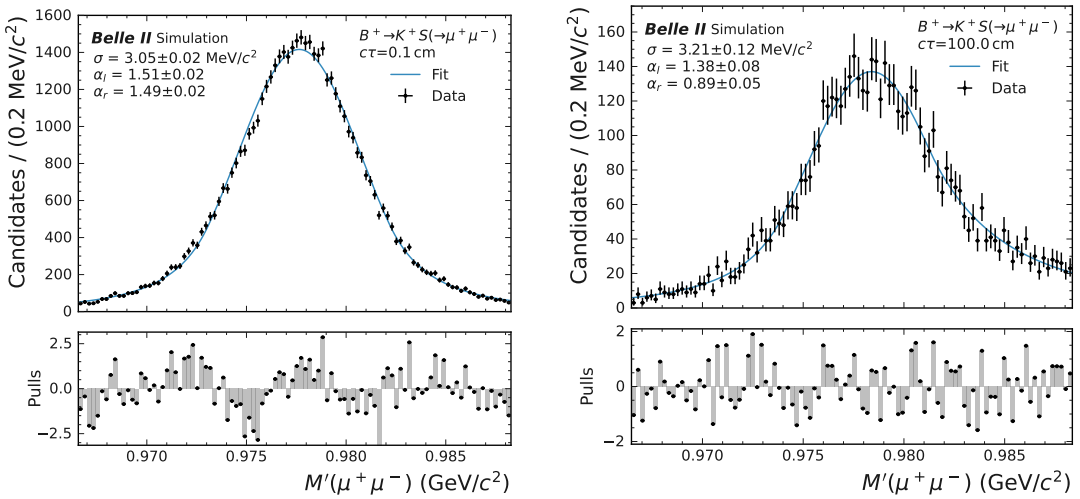

Figure 6.4.: Reconstructed $M'(x^+x^-)$ distributions and DSCB fit for simulated signal samples with $m_S = 1\,\text{GeV}/c^2$ in $B^+ \to K^+S(\to \mu^+\mu^-)$ for (left) $c\tau_S = 0.1\,\text{cm}$ and (right) $c\tau_S = 100\,\text{cm}$.

with coefficients c_i scaling the different polynomial orders.

The function is fit to the determined parameter values and found to describe their variation well for most of the phase space. The procedure is shown in Fig. 6.5 for the $S \to \mu^+\mu^-$ channels with a lifetime hypothesis of $c\tau_S = 1\,\text{cm}$. The $\alpha_{l/r}$ parameters exhibit a strongly decreasing behaviour for S masses close to the kinematic thresholds in the different final states. This is not described by the fourth-order interpolation. Instead, a quadratic interpolation is used between the parameter values of the smallest three to four simulated S mass samples in each final state close to the thresholds.

6.2. Background model

The reconstructed $M'(x^+x^-)$ spectra of the remaining generic backgrounds exhibit a locally smooth behaviour. The background is modelled using Chebyshev polynomial pdfs. The overall normalisation is given by the background yield estimator $N_{\text{Bkg.}}$. The coefficients as well as $N_{\text{Bkg.}}$ are estimated directly in the signal extraction fit.

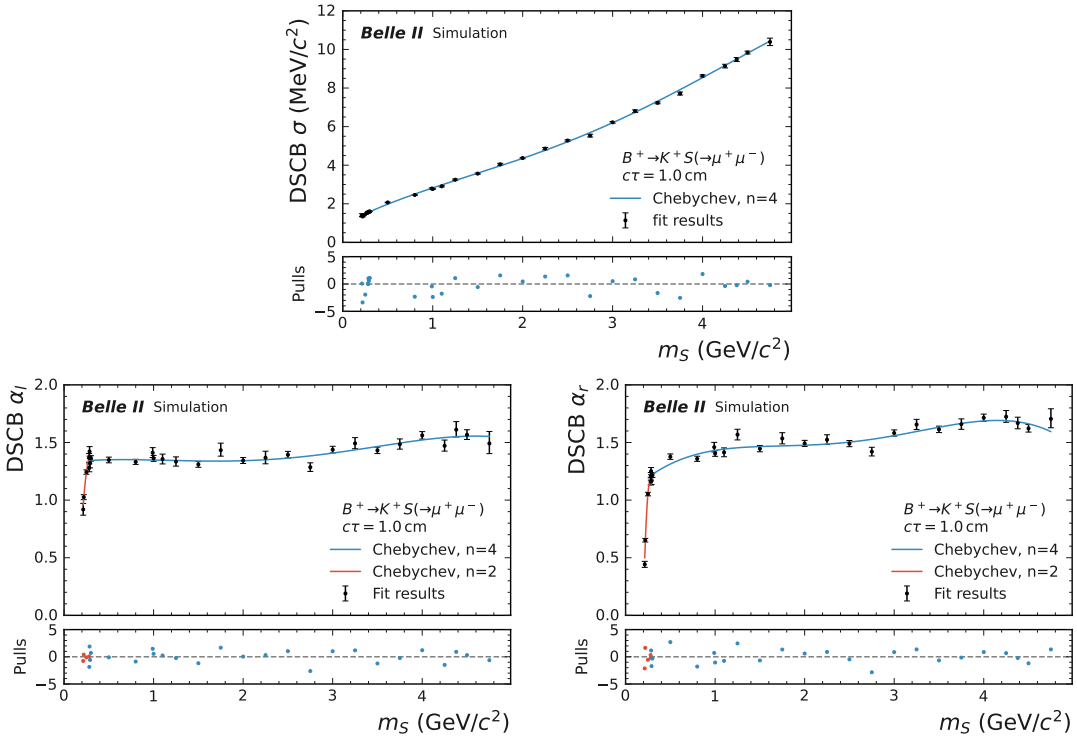

Figure 6.5.: Interpolation of the signal DSCB parameters (upper) σ, (lower left) α_l and (lower right) α_r between different S masses for $B^+ \to K^+ S(\to \mu^+\mu^-)$ with $c\tau_S = 1\,\text{cm}$.

This minimises the dependence on the background simulation.

6.2.1. Determination of polynomial order

The polynomial order used for the signal extraction is determined in background simulation. An order that is too small leads to poor modelling of the distribution. Too large of an order can result in numerical instabilities in the minimisation procedure. A further problem can arise due to over-fitting where a potential signal excess gets modelled as part of the background.

The polynomial order is chosen up to and not including the coefficient that is not statistically significant to describe the shape in simulation. Coefficient scans are shown in Fig. 6.6 for the $S \to \mu^+\mu^-$ and $S \to \pi^+\pi^-$ final-state channels.

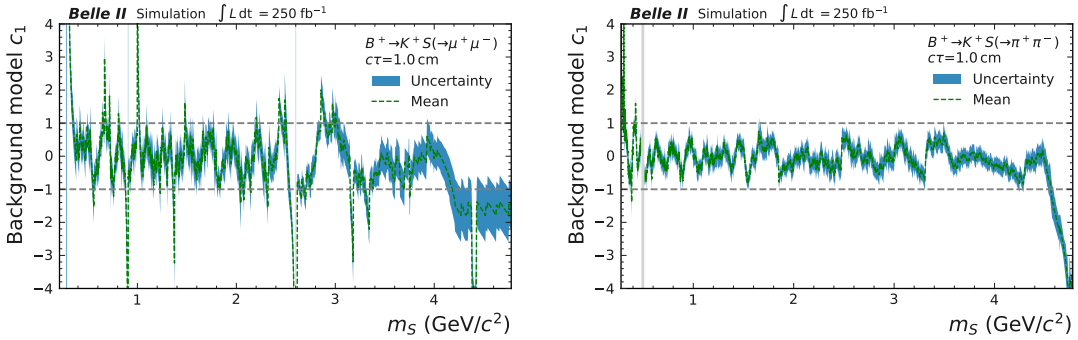

Figure 6.6.: The linear Chebyshev coefficient c_1 of the background model extracted from fits to background only simulation in the (left) $B^+ \to K^+S(\to \mu^+\mu^-)$ and (right) $B^+ \to K^+S(\to \pi^+\pi^-)$ channels at S masses spaced apart by half the signal resolution σ.

The extracted coefficients favour a small polynomial order for all final states and production modes. The low number of background candidates is not able to form structures which would require a large order. In a large part of the phase space c_1 is compatible with zero within its uncertainty, suggesting that a constant pdf is sufficient to describe the distribution. However, some regions prefer a finite linear coefficient. The first order Chebyshev polynomial is chosen as a background pdf for the signal extraction.

Due to the technical implementation of the Chebyshev polynomial within ZFIT, c_1 values outside of the range $[-1, 1]$ unavoidably cause a linear Chebyshev pdf to cross the x-axis at some value within the fit window. This can happen if a statistical fluctuation leads to a significantly lower number of background candidates at one end of the fit window compared to the other. This is more likely to occur at the endpoints of the S mass range because the fit window is capped to one side due to kinematics. The crossing of the background pdf yields a negative background prediction in a part of the fit window. However, the total background yield in the full fit window is always positive. Negative background predictions are nonphysical and lead to an overestimation of the signal yield when testing for the signal in these regions. In this analysis, these cases were first observed in fits to experimental data, whereupon the background model was re-tested and adjusted inf simulation. The linear coefficient c_1 of the background model is therefore constrained in the range $[-1, 1]$ yielding non-negative background predictions everywhere.

6.2.2. Photon conversions

The very low S mass region in the $S \to e^+e^-$ channel is populated by photon conversions. These candidates lead to an exponentially rising background below approximately $20\,\mathrm{MeV}/c^2$. Since the signal extraction windows are much larger than the signal width, they can extend into this region. An exponential pdf is added to the linear Chebyshev model to form the background model when S mass hypotheses below $40\,\mathrm{MeV}/c^2$ are tested. The exponential pdf is defined as:

$$f(x;\lambda) = \exp(\lambda x). \tag{6.5}$$

The values of the exponential coefficient λ are also determined directly in the signal extraction fit.

6.2.3. Examples background distributions

Background distributions with a fit of the background model are shown in Fig. 6.7 for the $S \to \mu^+\mu^-$ and $S \to e^+e^-$ final-state channels. For the latter, the exponential

pdf can be seen for the low S mass region. The size of the statistical uncertainties of the parameters estimated by the fit must be evaluated using a simulated dataset with a size comparable to the experimental dataset. Thus, a sub-sample of $250\,\mathrm{fb}^{-1}$ of the available background simulation is used instead of scaling the total sample down.

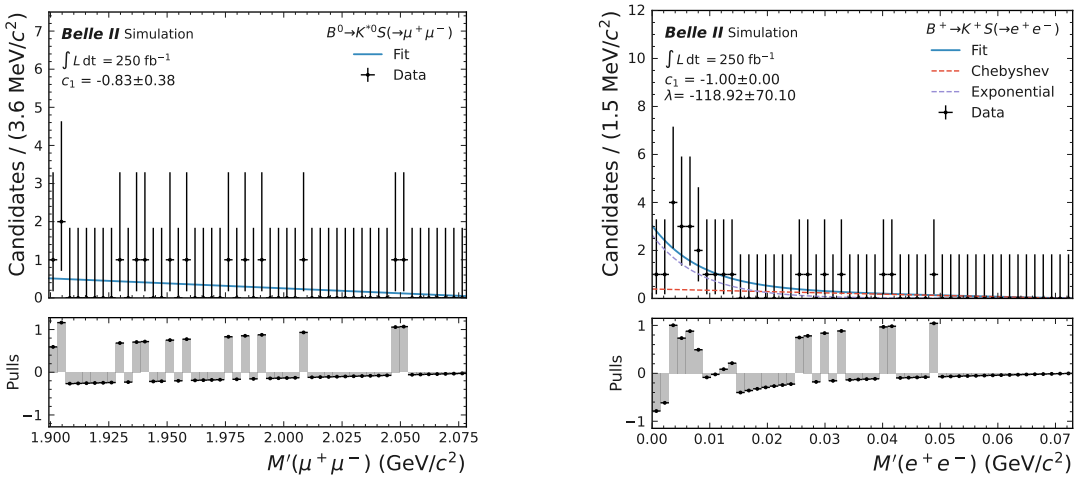

Figure 6.7.: Fits of the linear Chebyshev background model to background distributions in (left) $B^+ \to K^+S(\to \mu^+\mu^-)$ and (right) $B^+ \to K^+S(\to e^+e^-)$ in fit windows corresponding to a searched-for S at $m_S = 2\,\mathrm{GeV}/c^2$ and $m_S = 0.025\,\mathrm{GeV}/c^2$, respectively.

Summary of model parameters

A summary of the pdf parameters for the signal and background templates and their constraints is shown in Tab. 6.1.

Table 6.1.: Summary of signal and background pdf parameters used in the fits to the reconstructed $M'(x^+x^-)$ distribution. Details of the parameters are given in the text.

	Process	Constant	Range	Determination
$N_{\text{Sig.}}$	Signal	×	≥ 0	Data
μ	Signal	✓	Scan point	–
σ	Signal	✓	> 0	Simulation
$\alpha_{l/r}$	Signal	✓	> 0	Simulation
$n_{l/r}$	Signal	✓	3	Simulation
$N_{\text{Bkg.}}$	Background	×	≥ 0	Data
c_1	Background	×	$[-1, 1]$	Data
λ	Background	×	< 0	Data

7. Correction factors

This chapter describes correction factors that are applied to the quantities derived from simulation to produce values that are closer to the experimental data.

7.1. K_S^0 control sample

The main corrections to signal quantities are introduced because of an overestimation of the LLP performance in simulations compared to the data. A control sample of candidates that resemble the signal process is chosen to compare the performance in data and simulation and derive correction factors based on the differences. The closest SM particle to the S that is searched for is the K_S^0. It is naturally long-lived and decays to a pair of oppositely charged pions. Furthermore, K_S^0 are a source of a peaking non-reducible background to the signal process. The reconstructed S mass region around the K_S^0 is hence excluded from the search and can be safely studied in data before looking at candidates in the signal region.

Correction factors for properties of tracks that are determined in simulation are usually derived for tracks that originate from close to the IP. This is sufficient for most analyses dealing with short-lived or prompt objects such as B-mesons or τ-leptons. Consequently, reconstruction algorithms are optimised for such scenarios. The correct modelling of performance parameters for LLPs has to be studied as a function of the displacement of the decay vertex from the IP. The K_S^0 candidates are reconstructed in both data and simulation, and the performance parameters are compared. Discrepancies between data and simulation are then propagated as additional correction factors to the S on top of the nominal corrections.

The K_S^0 candidates are reconstructed in the $D^{*+} \to D^0(\to K_S^0\pi^+\pi^-)\pi^+$ decay

chain. The D-mesons are abundantly produced at Belle II via the $e^+e^- \to c\bar{c}$ process as well as in tree-level decays of B-mesons.

LLP corrections based on the performance of K_S^0 are determined using the same dataset used for the LLP search, introduced in Sec. 4.2. Simulated events are taken from the $e^+e^- \to \Upsilon(4S) \to B\overline{B}$ and $e^+e^- \to q\bar{q}$, where $q \in (u, d, s, c)$, samples that are used to study the backgrounds to the signal process, described in Sec. 4.3.1.

Reconstruction

The K_S^0 candidate is reconstructed from two oppositely charged tracks under the pion hypothesis. The decay vertex is determined with a kinematic fit. All signal selections that are applied to the S candidate in the main search described in Sec. 5.6 are applied to the K_S^0 candidate as well: the requirements on the minimum vertex displacement in the transverse plane, on the significance of the three-dimensional vertex distance and on how pointing the reconstruct vertex is. Furthermore, both K_S^0 daughter tracks have to pass the selection in the trinary PID score without TOP for the pion hypothesis, detailed in Sec. 5.3.2

All other pion tracks are selected using the track cleanup selections to minimise the impact of beam backgrounds, $d_0 < 0.5\,\text{cm}$ and $|z_0| < 2\,\text{cm}$. The two pions from the D^0 decay that accompany the K_S^0 are required to pass the minimum transverse momentum selection of $p_\text{T} > 150\,\text{MeV}/c$. No such selection is placed on the last pion track from the decay of the D^{*+} as it tends to small momentum due to the small mass difference between the D^{*+} and D^0.

The D^{*+} decay process is cleanly selected using requirements on the mass of the D^0 candidate, $M_D > 1.85\,\text{GeV}/c^2$, and the mass difference between D^0 and D^{*+} candidates, $(144 < \Delta_M < 147)\,\text{MeV}/c^2$. This kinematic selection allows to enrich the sample in K_S^0 candidates without further selections that could bias the result, such as on the PID scores of the non-K_S^0 pion tracks.

Normalisation

The main idea of the study is to derive LLP specific correction factors as a function of the displacement of the vertex from the IP in addition to the standard corrections, which only cover differences between data and simulation close to the IP. Thus, the LLP corrections from the K_S^0 sample are applied on top of the near-IP corrections. To this end, the number of K_S^0 candidates found in simulation is normalised to match the number in experimental data within a region slightly displaced, but still close to the IP. The region of $(0.5, 1.4)\,\text{cm}$ is chosen in the three-dimensional K_S^0 vertex distance from the IP to reject most prompt backgrounds while still being covered by the near-IP corrections. Normalising the simulation in this region decouples the effects purely due to displacement of the K_S^0 decay vertex from differences between data and simulation near the IP.

Sideband subtraction

The well reconstructed K_S^0 candidates form a peak in the reconstructed K_S^0 mass distribution. The remaining background from combinatorics as well as from (severely) misreconstructed K_S^0 candidates are found to follow a nearly constant shape. The distribution in simulation is shown in Fig. 7.1. A subtraction method is used to reduce

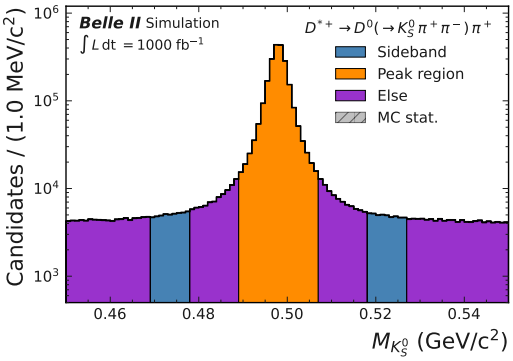

Figure 7.1.: Reconstructed K_S^0 mass distribution in simulation with peak and sideband regions highlighted.

the impact of these backgrounds on the comparison between data and simulation. The K_S^0 candidates that lie within the reconstructed K_S^0 mass peak region are counted for each bin of a distribution in a specific kinematic feature. Then, the number of K_S^0 candidates found in the mass sideband regions is subtracted. The mass peak and sideband regions have the same width and are highlighted in Fig. 7.1.

Agreement between data and simulation

The distributions in data and simulation are shown in Fig. 7.2 for the two D-meson observables, and in Fig. 7.3 for the observables used to select the K_S^0 candidate and its daughters. In these figures, normalisation in the small vertex distance region and sideband subtraction are used. The agreement between data and simulation is generally good, with the simulation well reproducing the shape of the distributions in data. However, there is a lower number of K_S^0 candidates reconstructed in data compared to simulation. This discrepancy is concentrated in the peak region of the reconstructed K_S^0 mass distribution. Furthermore, the peak of the K_S^0 mass distribution in data is found to be slightly wider compared to the estimation in simulation. Both effects are assumed to originate from an overestimation of LLP performance in simulation as a function of the vertex distance. The corresponding correction factors are derived in the following.

7.1.1. LLP correction factors

Correction factors that depend on the displacement of the vertex are derived for the reconstruction efficiency and for the signal pdf parameters, introduced in Sec. 6.1, using the K_S^0 control sample.

The K_S^0 candidates reconstructed in data and simulation are divided into regions of three-dimensional K_S^0 vertex distance from the IP. Instead of the simple subtraction method introduced above, a fit of template pdfs to the reconstructed K_S^0 mass distribution is used to determine the K_S^0 yields. A separate fit is performed to the distributions in data and simulation for each region of vertex distance. The K_S^0 signal peak is modelled with a DSCB pdf template as used for the S signal in this search. A

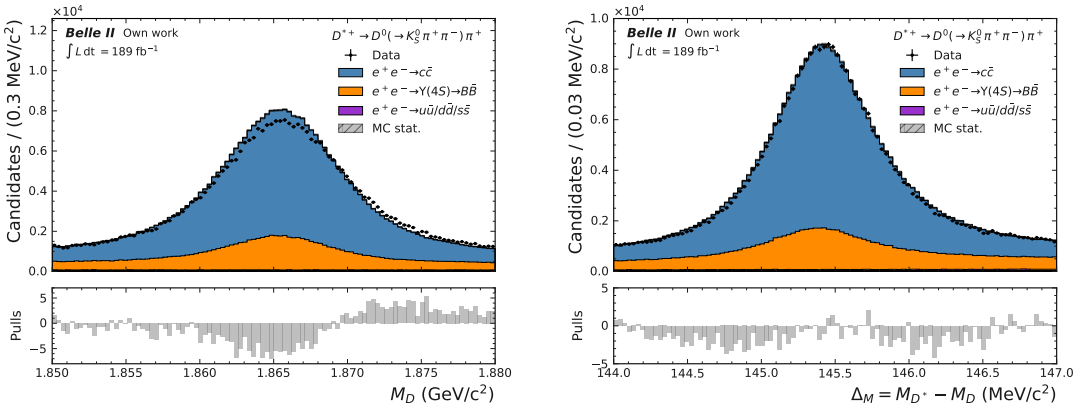

Figure 7.2.: Reconstructed distributions of the (left) D^0 mass M_D and (right) difference between D^{*+} and D^0 masses Δ_M in data and simulation.

constant pdf is used to describe the backgrounds. The K_S^0 yield and the values of the DSCB parameters with associated uncertainties are estimated by the fits to data and simulation. The ratios between data and simulation for each parameter of interest are normalised to unity within the region of small K_S^0 vertex distance from the IP. Fits of the DSCB pdf to the reconstructed K_S^0 mass distributions in data and simulation are shown in Fig. 7.4. These correspond to K_S^0 vertex distances from the IP between $(5.2, 7)$ cm.

No dependence on the type of charged particle is assumed and the correction factors are applied to all S decay channels. Systematic uncertainties corresponding to the full size of the correction factors are applied to the S parameters that are corrected, detailed in Chap. 9. This is found to cover variations of the correction factors between different regions of K_S^0 momentum and polar angle.

LLP efficiency

The ratio of the K_S^0 yields estimated by the fits in data and simulation is an estimator of the relative difference in the efficiency of reconstructing K_S^0. The resulting values for the different regions of K_S^0 vertex distance from the IP are shown in Fig. 7.5. The K_S^0 reconstruction efficiency in data is found to be smaller than estimated in

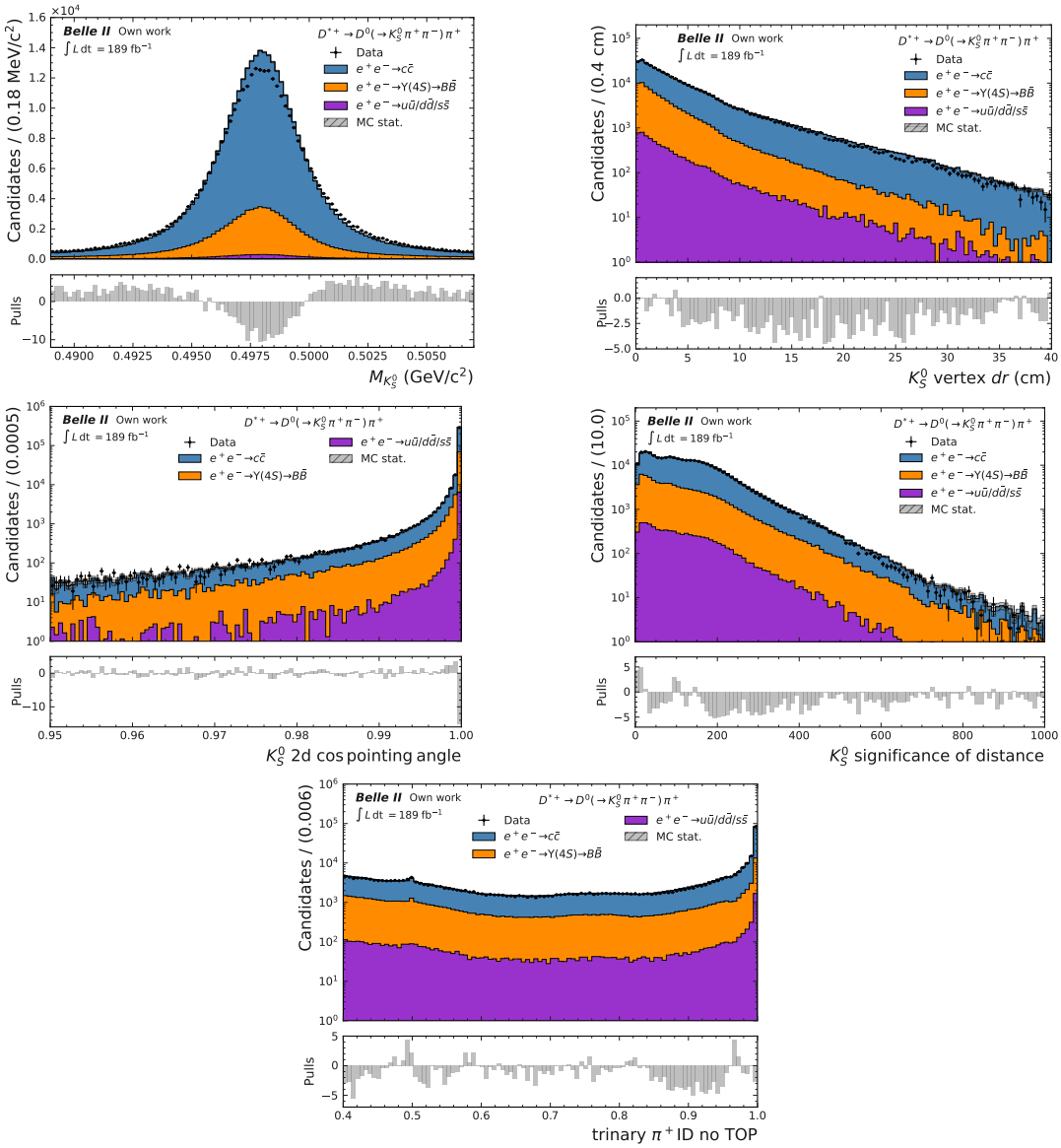

Figure 7.3.: Reconstructed distributions of the (upper left) K_S^0 mass $M_{K_S^0}$ and (upper right) K_S^0 vertex dr, (middle left) K_S^0 pointing angle, (middle right) K_S^0 Significance Of Distance and (lower) trinary pion ID score without TOP for the K_S^0 daughter pion in data and simulation.

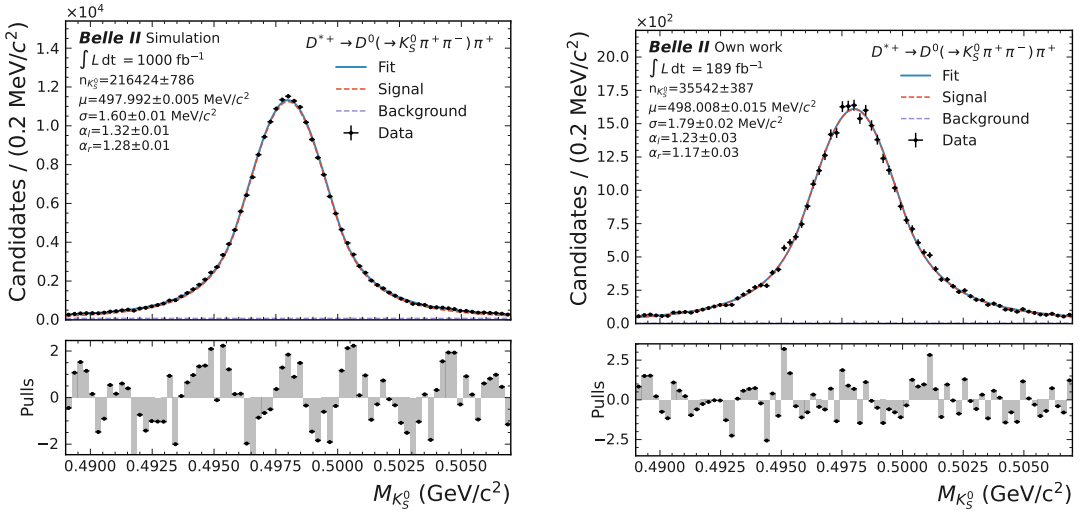

Figure 7.4.: Fits of the DSCB pdf to the reconstructed K_S^0 mass distributions in (left) simulation and (right) data for K_S^0 vertex distances from the IP between $(5.2, 7)$ cm.

simulation, and the difference increases with the vertex distance. This trend has a linear shape, and a straight line is fit to the values. The slope is estimated by the fit to -0.44%, which is used to correct the S efficiency determined in simulation. The size of the correction applied at a specific S mass and lifetime depends on the mean three dimensional S vertex distance from the IP:

$$\epsilon_{\text{correction}}(\text{distance}) = -0.44\%/\,\text{cm} < \text{distance} > . \tag{7.1}$$

The efficiency of an S signal with a mean vertex distance from the IP of 10 cm would hence be corrected by -4.4%.

LLP template model parameters

The reconstructed K_S^0 mass peak in data is found to be wider than estimated in simulation. This is due to a too optimistic assumption on the single-track momentum resolution in simulation. A correction factor is derived by comparing the estimated

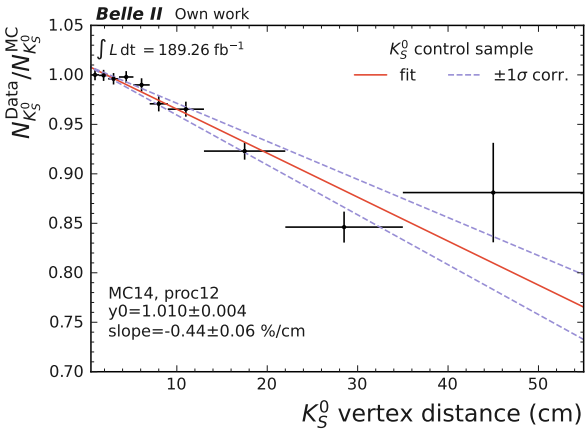

Figure 7.5.: Ratio of fitted K^0_S yields in data and simulation binned in the K^0_S vertex distance from the interaction region normalised to the first bin with linear fit to determine the efficiency correction factor.

values of the DSCB parameters σ, α_l and α_r from fits to distributions in data and simulation as a function of K^0_S vertex displacement from the IP. The ratios between data and simulation for the tail parameters α_l and α_r are shown in Fig. 7.6. Smaller values of the tail parameters correspond to longer tails signalling worse resolution. The ratio between data and simulation for α_l as a function of K^0_S vertex distance shows smaller values in data compared to simulation, following a linear trend. A correction factor similar to the efficiency is derived from the slope estimated by a straight line fit:

$$\alpha_{l\,\text{correction}}(\text{distance}) = -0.71\%/\,\text{cm} < \text{distance} > . \tag{7.2}$$

The right tail parameter α_r is found to be well modelled in simulation with a constant trend of the ratio as a function of K^0_S vertex distance from the IP. Therefore, no correction factor is applied to α_r.

The ratios between data and simulation for the DSCB width parameter σ estimated by the fits are shown in Fig. 7.7. The σ values in data are higher than those in simulation, observed by ratios greater than unity. This translates to a positive correction

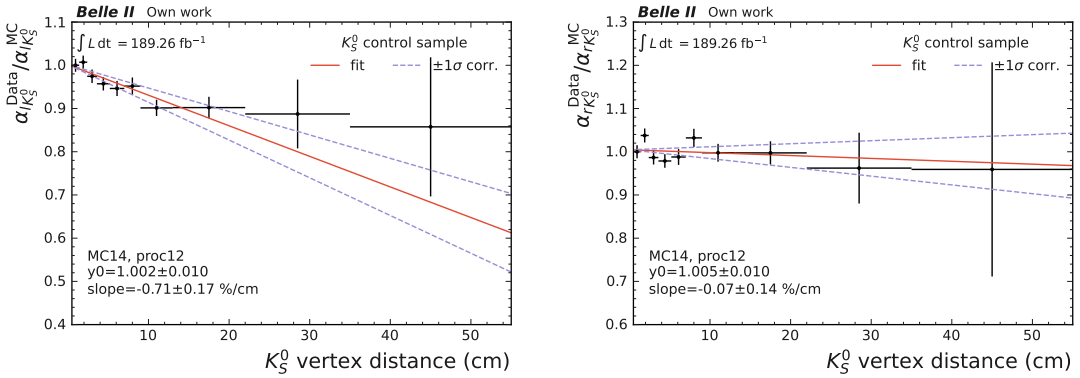

Figure 7.6.: Ratio of DSCB tail parameters (left) α_l and (right) α_r estimated by fits to the reconstructed K_S^0 mass distributions in data and simulation binned in the K_S^0 vertex distance from the IP normalised to the first bin. Shown with a fit of a linear Chebyshev pdf to α_l and a constant pdf to α_r.

factor for σ. A linear trend of the ratio is observed as a function of the K_S^0 vertex distance from the IP up to a value of 9 cm. The ratio values reach a plateau region and remain approximately constant for vertex distances higher than that. A possible explanation lies in the modelling of the distribution using the chosen pdf. The estimation of the Gaussian core in simulation does not deteriorate further. Instead, the increasing shape discrepancy at larger vertex distances is absorbed by the tail parameter α_l. Here, the difference between data and simulation is found to increase further. The slope of a straight line estimated by a fit to the region up to 9 cm of K_S^0 vertex distance, and a constant factor for the larger distances, are used to derive a correction factor for σ. A weighted mean for a given sample is determined by splitting the signal sample between S vertex distance below and above 9 cm. Each simulated signal event with small distances is weighted by its vertex distance multiplied by the slope of the straight line. For large distances, the simulated event is weighted by a

constant.

$$\sigma_{\text{correction}}(\text{distance}) = \frac{\sum_{i \in \text{signal sample}} w(i)}{N_{\text{signal sample}}}, \tag{7.3}$$

$$\text{where } w(i) = \begin{cases} 2.2\% \cdot \text{distance}(i) & \text{for distance}(i) < 9\,\text{cm}, \\ 13\% & \text{for distance}(i) > 9\,\text{cm}. \end{cases} \tag{7.4}$$

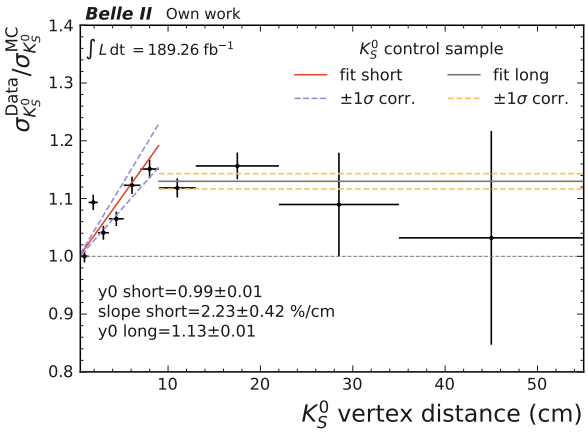

Figure 7.7.: Ratio of DSCB width parameter σ estimated by fits to the reconstructed K^0_S mass distributions in data and simulation binned in the K^0_S vertex distance from the IP normalised to the first bin. A fit of a linear (constant) Chebyshev pdf to short (long) vertex distances is shown.

K^0_S daughter particle identification

Standard correction factors that cover the efficiency differences of the PID selections between data and simulation only take into account tracks that originate from close to the IP. Therefore, the modelling of the PID selection efficiency must also be validated as a function of the K^0_S vertex distance. The K^0_S sample allows to study the efficiency of the selection on the trinary PID score without TOP for the pion final state. The K^0_S yield is determined in data and simulation separately with and without this PID

requirement on both K_S^0 daughter tracks in each of the K_S^0 vertex distance bins:

$$\frac{\epsilon_{\text{PID}}^{\text{Data}}}{\epsilon_{\text{PID}}^{\text{Sim.}}}(\text{distance}) = \frac{N_{K_S^0}^{\text{Data, PID sel.}}(\text{distance})/N_{K_S^0}^{\text{Data, total}}(\text{distance})}{N_{K_S^0}^{\text{Sim., PID sel.}}(\text{distance})/N_{K_S^0}^{\text{Sim., total}}(\text{distance})}. \tag{7.5}$$

The ratio in the different K_S^0 vertex distance bins is shown in Fig. 7.8 where the first bin is again normalised to one.

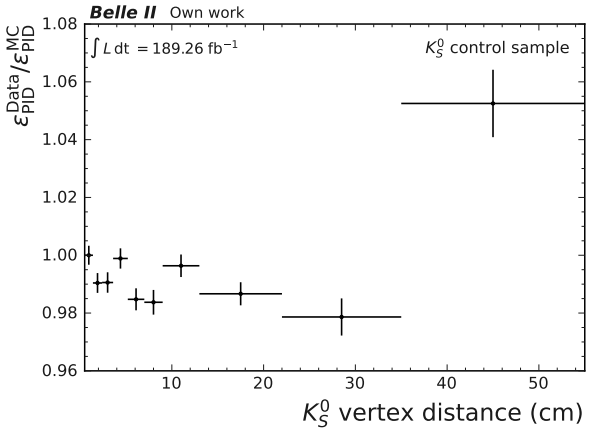

Figure 7.8.: Ratio between the efficiencies of the PID requirements on the K_S^0 daughter tracks in data and simulation binned in the K_S^0 vertex distance from the IP normalised to the first bin.

The PID efficiency is found to be stable across the K_S^0 vertex distance bins, and no further efficiency correction is applied. Unfortunately, there is no suitable SM long-lived particle decaying to electrons, muons, or kaons that could be used to study the respective PID score requirements as a function of vertex displacement.

7.2. Further correction factors

Further correction factors in this analysis are the aforementioned near-IP correction factors applied to the momentum of tracks and to the PID requirement efficiencies and misidentification rates. There is no correction for the efficiency for near-IP tracks

estimated in simulation. However, an uncertainty is applied to cover differences between near-IP tracking efficiencies in data and simulation, detailed in Sec. 9.1.1. The application of the resulting correction factors to the presented search is performed as part of this work.

7.2.1. Particle identification

The efficiencies and misidentification rates of PID requirements are different between data and simulation. Studies in control channels are performed to determine the correction factors on efficiencies and misidentification rates in bins of the charge, momentum and polar angle of the tracks. Corrections are derived for:

- the global kaon PID applied to the prompt kaon and displaced kaons at small S masses and lifetimes,
- the binary kaon PID applied to displaced kaons at large S masses and lifetimes,
- the global pion PID applied to prompt pions,
- the trinary electron PID excluding TOP applied to displaced electron,
- the trinary muon PID excluding TOP applied to displaced muons,
- the trinary pion PID excluding TOP applied to displaced pions.

The correction factors for the PID requirements on electrons and muons are determined using control samples enriched in events containing $J/\psi \to e^+e^-/\mu^+\mu^-$ decays. The PID control channels for the hadronic final-state particles are $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$ as well as $K^0_S \to \pi^+\pi^-$ and $\Lambda^0 \to p^+\pi^-$. Although long-lived SM particles are used to derive these corrections, no specific study as a function of the displacement is performed. Tracks for these PID control studies are selected with requirements on the d_0 and z_0 of the tracks, further biasing the sample towards small vertex displacements. The size of the PID correction factors for the signal efficiency and their associated uncertainties are shown in Fig. 7.9 as a function of the S mass for the different analysis channels. The determination of the uncertainty associated

with the PID correction factors is detailed in Sec. 9.1.3. The dependence on the S lifetime is small. The size of the correction factors and their uncertainties are larger for the $B^0 \to K^{*0} S$ channels due to the smaller prompt track momenta. Here, PID performance is reduced, and the discrepancy between data and simulation is larger. The difference between the correction factors in the two leptonic S final-state channels is found to be small.

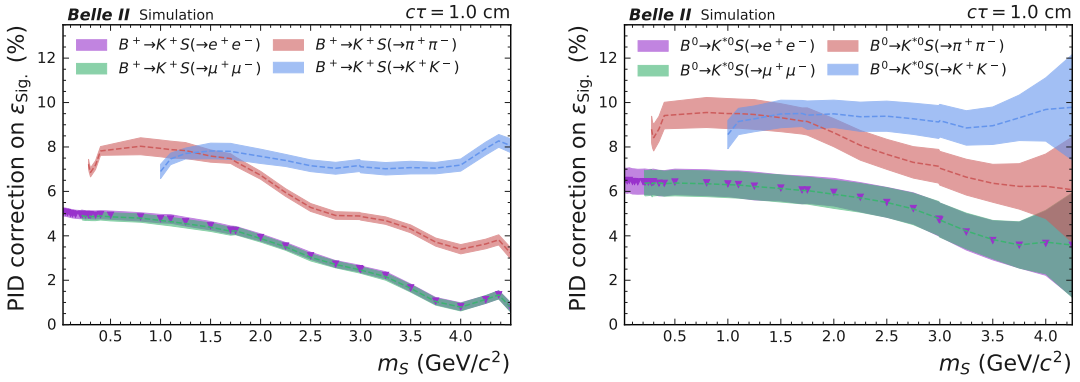

Figure 7.9.: PID correction factors on the signal efficiency and associated uncertainties as a function of S mass for the different S final-state channels in (left) $B^+ \to K^+ S$ and (right) $B^0 \to K^{*0}(\to K^+\pi^-)S$. The correction factors are normalised to the uncorrected signal efficiency values.

The PID efficiency correction factors are used to correct the signal efficiency determined from the generated signal samples and to correct the prediction of backgrounds from simulation in the sideband validation studies. All tracks that make up the signal B-meson candidate are selected with a PID requirement. Hence, separate PID efficiency correction factors are applied to each. A total correction factor per event is determined by multiplying the separate correction factors.

The misidentification rate correction factors are only used for the correction of simulated background events in the sideband validation studies described in Chap. 8 and in the final figures in the signal region shown in Chap. 11. Corrections to the misidentification rate are only considered for $\pi \to K$ and $K \to \pi$. Additional corrections for the leptonic final-state channels are expected to further increase agreement be-

tween data and simulation for background processes. However, they have no effect on the resulting upper limits on the signal process, as the background is predicted directly using data in the reconstructed S mass sidebands when extracting the signal. Additional cross-feed between signal final-state channels due to potentially larger misidentification rates in data compared to simulation is expected to have negligible impact on the search due to the very high signal purity estimated in simulation.

7.2.2. Track momentum scale

The momentum of tracks in experimental data needs to be adjusted to account for the finite knowledge on the inhomogeneity of the magnetic field in the detector. In simulation there is perfect knowledge of the magnetic field and no correction is required. The correction factor in experimental data is determined by calibrating the reconstructed mass of the $D-$meson to its average known value. The D-meson candidates are identified by selecting $D^{*+} \to D^0(\to K^+\pi^-)\pi^+$ events with kinematic requirements similar to those used in this work for the K^0_S control sample. The extracted correction is a scale factor to the track momentum which is binned in the polar angle and charge of the tracks, separately for two data taking sub-periods. A validation is performed using D-mesons from $D^0 \to K^+\pi^-\pi^+\pi^-$, $D^+ \to K^+\pi^+\pi^-$, $J/\psi \to \mu^+\mu^-$ and $K^0_S \to \pi^+\pi^-$. No dependence on the type of charged particles is assumed. The correction factors are of the order $(0.05 - 0.3)\%$, with the largest corrections in the forward region.

8. Validation

This chapter presents the validation of the modelling of background processes in simulation. A SM physics process similar to the signal is validated as well. The predictions from simulation are compared to experimental data in regions enriched with these processes.

The precision of the signal and background simulation within the signal region, defined by the selection requirements summarised in Sec. 5.6, is studied in dedicated sideband regions. The sideband regions are constructed to be kinematically close to the signal region, but enriched in the processes that are validated. Furthermore, they should contain a negligible amount of signal compared to the background. The signal selection allows to define different sidebands targeted to study the modelling of all major background sources, and of a SM process similar to the signal. The prediction of the simulated backgrounds in the sidebands are corrected for differences between data and simulation in particle identification efficiencies and fake rates using the corrections introduced in Sec. 7.2.1.

The background contribution is determined directly in fits to data using large fit windows in the reconstructed S mass distribution. Simulation is used to optimise the signal selection requirements and to choose the order of the polynomial function that describes the background in the reconstructed S mass distribution. The signal extraction method works as intended, as long as the background can be described locally by the polynomial function and does not feature peaking signal-like structures. Therefore, no corrections to the analysis in the signal region are derived from small differences between data and simulation in the sidebands.

8.1. Combinatorial backgrounds

Combinatorial background, mostly from light-quark-pair production, is the dominant remaining background in the signal region. The regions around the B-meson kinematic selections in M_{bc} and ΔE are well suited to study this background. The shape of the combinatorial background in these variables is well studied and detailed in Sec. 8.3.1. A rectangular sideband region is chosen to study the modelling of the combinatorial background from light-quark-pair processes:

$$\begin{aligned} &M_{\mathrm{bc}} \in (5.20, 5.27)\,\mathrm{GeV}/c^2, \\ &\Delta E \in (-0.15, 0.15)\,\mathrm{GeV} \text{ and} \\ &\quad \begin{cases} \notin (-0.05, 0.05)\,\mathrm{GeV}/c & \text{for } S \to e^+e^- / \mu^+\mu^- / K^+K^- \\ \notin (-0.035, 0.035)\,\mathrm{GeV}/c & \text{for } S \to \pi^+\pi^- \end{cases} \end{aligned} \tag{8.1}$$

The simulation predicts more candidates reconstructed in the combinatorial background sideband in the $S \to \mu^+\mu^-$ final-state channel as observed in data, see left part of Fig. 8.1. This discrepancy is traced back to S daughter tracks with polar angles in the far forward region and with low transverse momenta.

Tracks with small momenta can curl within the CDC and do not reach the outer detectors, leading to worse PID performance. The reason for the disagreement in the forward region is not fully understood. However, the Belle II collaboration is aware of this discrepancy in PID performance and has started investigating. Additional selections on the S daughter track polar angle and transverse momentum p_{T}:

$$\begin{aligned} p_{\mathrm{T}} &> 250\,\mathrm{MeV}/c, \\ \theta_{\mathrm{CDC}} > 32^\circ &\text{ and } \theta_{\mathrm{ECL}} < 150^\circ, \end{aligned} \tag{8.2}$$

are placed to improve the agreement between data and simulation in all analysis channels. The selection in the polar angle is chosen based on subdetector acceptance. It corresponds to the coverage of the barrel ECL in the forward direction and of the CDC in the backward direction. As described in Sec. 5.2, the track helices are extrapolated to the subdetector surfaces, where the polar angle is determined to account for the

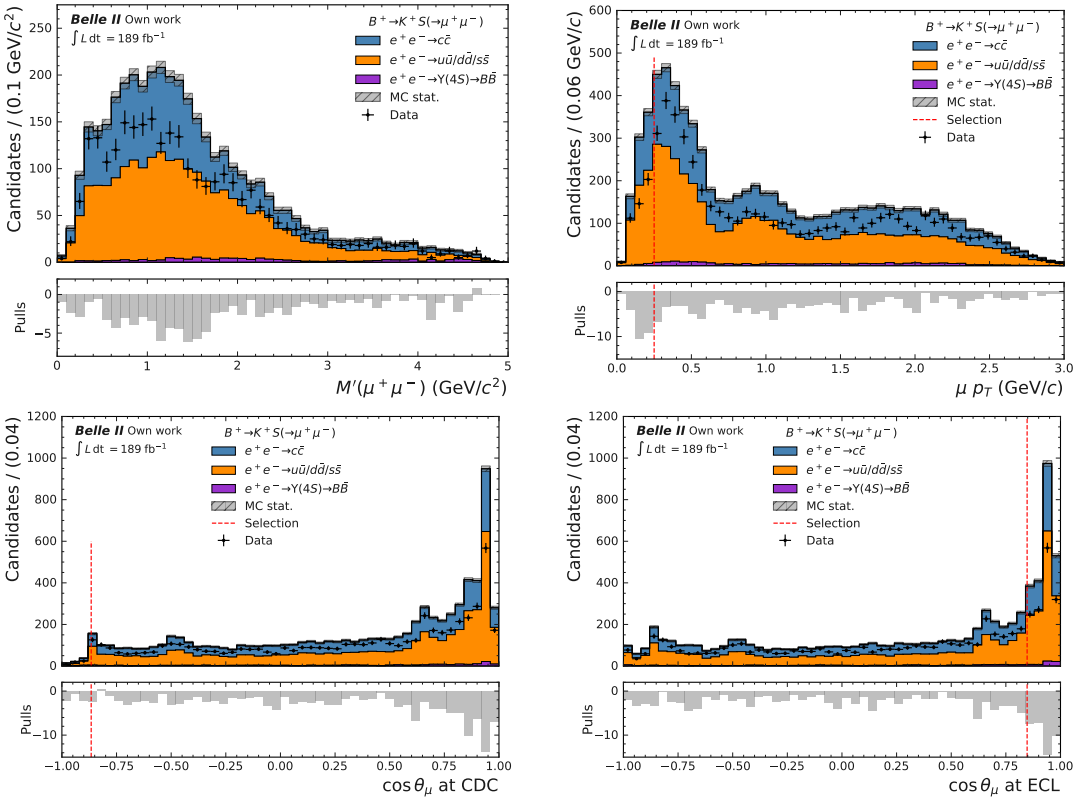

Figure 8.1.: Distributions from data and simulation for $B^+ \to K^+S(\to \mu^+\mu^-)$ in (upper left) $M'(x^+x^-)$ and S daughter (upper right) transverse momentum and (lower left) polar angle at the CDC surface and (lower right) polar angle at the ECL surface without additional selections on the lower variables. Track variables are shown for both charges together.

production at a displaced vertex. The agreement between data and simulation is good in the backward direction. However, in this region, a selection is also added to ensure that the S daughter tracks are within the acceptance of the CDC. The impact of the additional selection on the signal efficiency can be estimated from Fig. 8.2, showing the p_{T} distributions at different S masses and the polar angle distributions at both the IP and the detector surfaces.

The difference between the polar angle determined at the IP and at the detector surfaces is small for signal and background. The distribution in the latter peaks slightly higher towards large (forward) values.

The additional requirements suppress a large fraction of the remaining backgrounds, concentrated at small momenta and large polar angles of the displaced tracks. The impact of these requirements on the sensitivity to the benchmark models is evaluated. An increased constraining power is found for small S masses, while a small amount of sensitivity is lost for larger S masses.

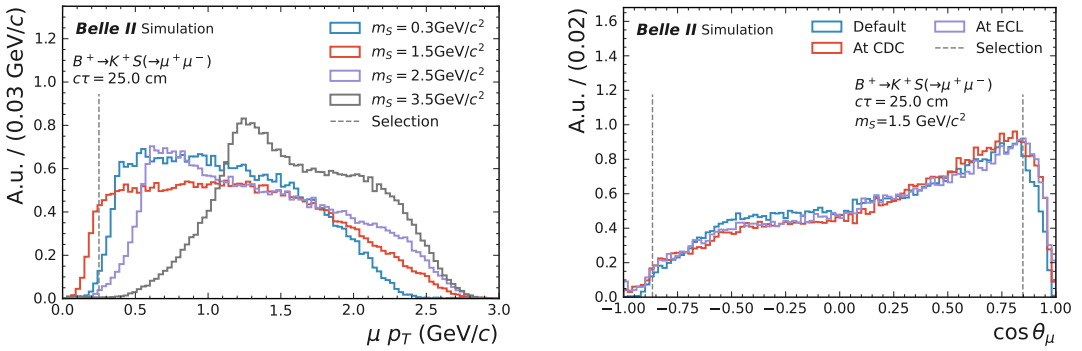

Figure 8.2.: Simulated signal distributions of the S daughter tracks in $B^+ \to K^+ S(\to \mu^+\mu^-)$ in the (left) transverse momentum p_{T} for different S masses and (right) $\cos\theta$ for the default polar angle and at the CDC and ECL surfaces. The vertical lines correspond to the additional selections that are placed to improve agreement between data and simulation. Both track charges are shown together.

The comparison between data and simulation for the combinatorial background sideband in $M'(x^+x^-)$, vertex dr and the two observables used to define the side-

band is shown in Fig. 8.3 for $B^+ \to K^+ S(\to \mu^+\mu^-)$. These distributions include the selections on the S daughter p_{T} and polar angle, substantially improving the agreement between the data and the simulation. Generally, the agreement in the combinatorial background sideband is good for all analysis channels and across selection variables. Small differences between data and prediction include an excess of data at large PID score values for the prompt kaons. This effect is not seen for the prompt pion in $B^0 \to K^{*0} S$, suggesting that it is due to residual differences between data and simulation in the kaon PID score. The shape of the reconstructed $M'(x^+x^-)$ distributions is well reproduced by the simulation. An overall excess of candidates in $B^0 \to K^{*0}(\to K^+\pi^-)S(\to e^+e^-)$ is observed in data compared to the expectation by simulation. It takes the form of a normalisation difference and is not localised in a specific region of one of the selection variables or the reconstructed S mass distribution. As such, it does not affect the background determination directly in the fit to data.

8.2. Backgrounds from resonances

Background contributions that peak in the reconstructed S mass distribution include the long-lived K^0_S and prompt two-body resonances such as J/ψ.

8.2.1. Peaking resonances from prompt decays

Residual peaking background sources are detailed in Sec. 5.5. These backgrounds originate from almost prompt decays of two-body SM resonances. They are rejected by tightening the S vertex dr requirement in windows in the reconstructed S mass around the resonances. Smaller vertex displacements in these mass regions can be used as a sideband to study peaking backgrounds close to the signal selection in simulation and data. The sideband is defined by applying the full signal region selection, and inverting the dr requirement:

$$dr < 0.2\,\mathrm{cm}, \tag{8.3}$$

within the peaking background S mass windows.

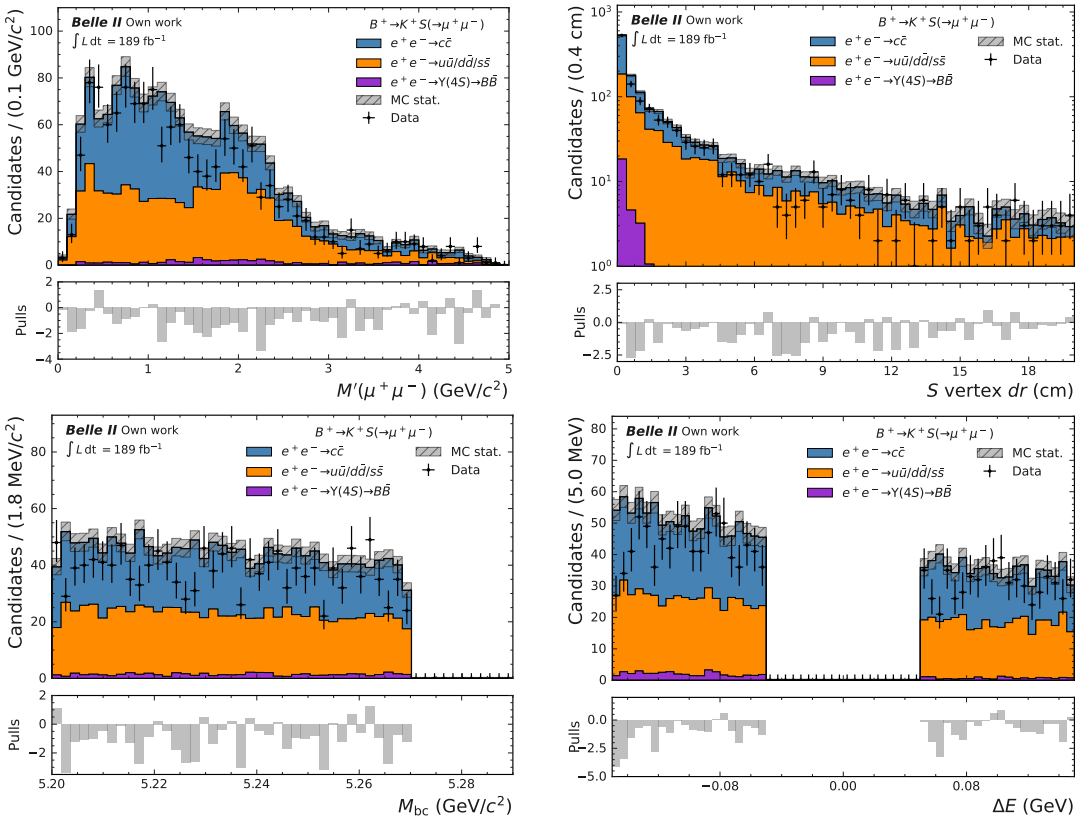

Figure 8.3.: Distributions from data and simulation for $B^+ \to K^+S(\to \mu^+\mu^-)$ in (upper left) $M'(x^+x^-)$, (upper right) S vertex dr, (lower left) M_{bc} and (lower right) ΔE using the selections of the combinatorial background sideband.

The comparison between data and simulation for the J/ψ background in $B^+ \to K^+S(\to \mu^+\mu^-)$ and the D^0 background in $B^+ \to K^+S(\to \pi^+\pi^-)$ is shown in Fig. 8.4 for M_S and the S vertex dr. The distributions of peaking backgrounds in the selection variables, as well as for the reconstructed S mass, are found to be well predicted by simulation in all analysis channels. Backgrounds from B-meson decays, such as true D^0 in $B^+ \to K^+S(\to \pi^+\pi^-)$ in data show the expected peaking behaviour in M_{bc} and ΔE and populate smaller values of Fox-Wolfram $R2$. The widths of the reconstructed S mass distributions and the roll-off in vertex dr are well modelled by simulation for the peaking backgrounds. Hence, the rejection of peaking backgrounds with a tighter selection in dr works as expected from simulation. The size of the mass windows and the tight dr selection value itself do not need to be adjusted.

8.2.2. Peaking K_S^0 background

Long-lived K_S^0 are an irreducible background due to their displaced vertex and two-body decay, similar to $S \to \pi^+\pi^-$. They are produced in B-meson decays which can pass the kinematic selections and in light-quark-pair processes selected due to combinatorics. The K_S^0 mass peak region is vetoed in $M_{\pi^+\pi^-}$ in all channels and is used as a sideband to study the K_S^0 background that passes all other selections:

$$M_{\pi^+\pi^-} \in (489, 507)\,\mathrm{GeV}/c^2. \tag{8.4}$$

The comparison between data and simulation in the sideband enriched with K_S^0 is shown in Fig. 8.5 for the reconstructed S mass and the vertex dr using the $B^+ \to K^+S(\to \pi^+\pi^-)$ channel. The agreement between data and simulation is very good, the distributions of the S (K_S^0) mass and of all selection variables are well reproduced. Part of the overall deficit in K_S^0 reconstructed in data with respect to simulation can be connected to the difference in reconstruction efficiency as a function of the decay vertex distance studied in Sec. 7.1.1. It amounts to a correction of 4.5% given the average vertex distance of K_S^0 selected in the sideband region.

The tail of the reconstructed S mass distribution is reduced, as expected by simulation, at the edges of the window depicted in $M_{\pi^+\pi^-}$. This means that the size

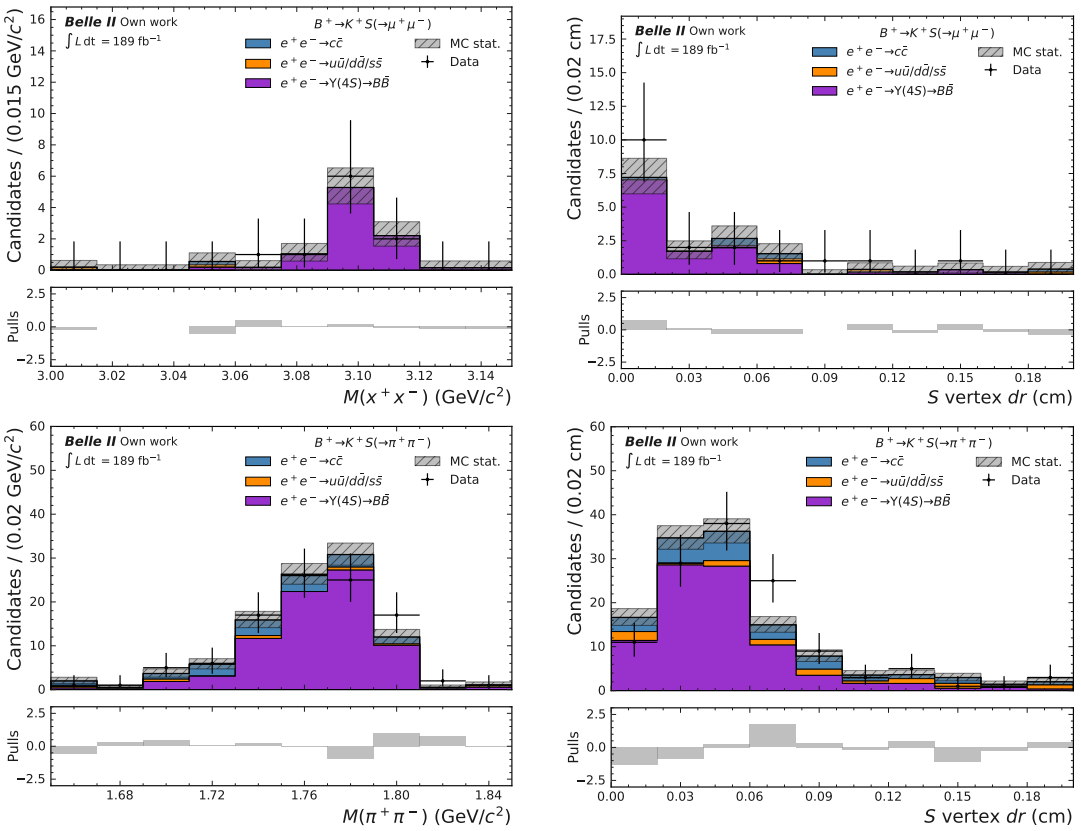

Figure 8.4.: Distributions from data and simulation for $B^+ \to K^+S(\to \mu^+\mu^-)$ in (upper) and (lower) $B^+ \to K^+S(\to \pi^+\pi^-)$ in (left) $M'(x^+x^-)$ and (right) S vertex dr using the selections of the peaking background sideband.

of the K^0_S veto region, which is equivalent to the depicted window, is adequate. The same conclusions hold for the $B^0 \to K^{*0}\, S$ production channel. The statistics in the S decay channels other than those to pions are small because true K^0_S are only selected when both pions are simultaneously misidentified as another particle.

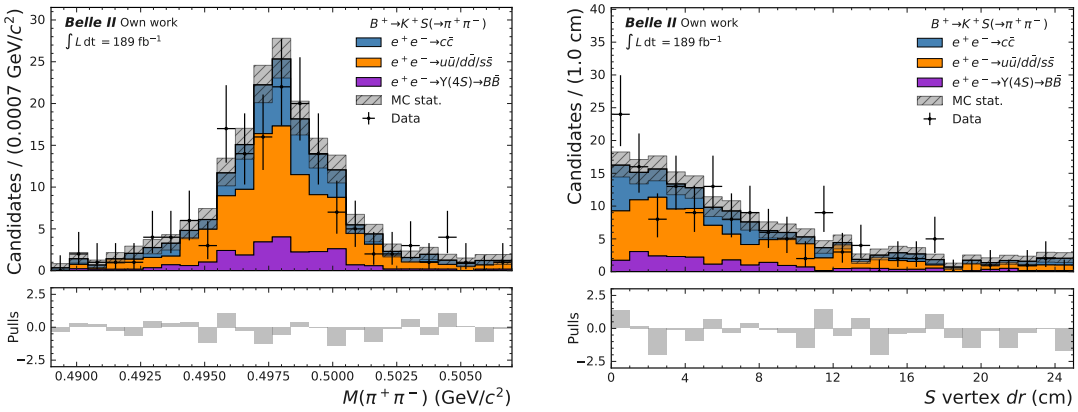

Figure 8.5.: Distributions from data and simulation for $B^+ \to K^+ S(\to \pi^+\pi^-)$ in (left) $M'(x^+x^-)$, (right) S vertex dr using the selections of the K^0_S background sideband.

8.3. $B^+ \to K^0_S \pi^+$ process

The SM process $B^+ \to K^0_S \pi^+$ is kinematically close to the signal process, both including a B-meson decay to a long-lived particle and a charged pseudoscalar meson. This process can be used to study background processes directly within the selection requirements specific to the B-meson and S candidates. The global PID score requirement on the prompt track is changed to select pions instead of kaons:

$$\mathrm{PID}_\pi > 0.2. \tag{8.5}$$

A signal predicted by the dark scalar or ALP models could enter this sideband if the S was produced in $b \to d$ instead of $b \to s$. However, the probability of this is small due to the large CKM suppression of $|V_{td}|^2/|V_{ts}|^2$. The process with a decay

of a B-meson to K^{*0} and K_S^0 does not exist in the SM. Therefore, the corresponding neutral S production channel is not used for validation studies.

The comparison between data and simulation for the $B^+ \to K_S^0\pi^+$ validation sideband is shown in Fig. 8.6 for the reconstructed K_S^0 mass, the K_S^0 vertex dr, M_{bc}, and ΔE. The agreement between data and simulation for the $B^+ \to K_S^0\pi^+$ sideband is found to be very good. The M_{bc} and ΔE distributions show a peaking behaviour, also expected from the S signal process, and are well described by simulation. This indicates that the corresponding signal selections work as intended for a potential S signal. The efficiency correction based on the difference between data and simulation as a function of the vertex distance amounts to 4.5%. This can cover part of the small observed recess of candidates K_S^0 found in data.

8.3.1. Branching fraction validation

The kinematic similarities between the $B^+ \to K_S^0\pi^+$ process and the S signal allow for further validations using this sideband region. The estimation of the signal efficiency in simulation, the modelling of the signal process, and the determination of signal yields by fits can be studied by measuring the branching fraction of the $B^+ \to K_S^0\pi^+$ process in data and simulation.

The extraction of the signal yield in the reconstructed S (K_S^0) mass distribution cannot be directly performed due to the peaking contribution from K_S^0 produced in light-quark-pair events. These cannot be separated from the $B^+ \to K_S^0\pi^+$ contribution using $M(\pi^+\pi^-)$ alone, as can be seen in Fig. 8.6. In the S search, all background sources that peak in the mass distribution have been vetoed (in case of K_S^0) or reduced to negligible levels (by tightening the dr selection). Therefore, extraction of the signal yield in the reconstructed mass distribution from a peak over a smooth linear background is possible.

The $B^+ \to K_S^0\pi^+$ yield is determined by a two-dimensional fit to the M_{bc} and ΔE distributions, in which light-quark-pair and $B\overline{B}$ events can be disentangled by their shapes. A large sample of the $B^+ \to K_S^0\pi^+$ process is simulated with the same settings as used for the S signal, discussed in Sec. 4.3.2. Templates for the $B^+ \to K_S^0\pi^+$ shape are constructed by fits of a Crystal Ball pdf to the M_{bc}, and of a DSCB pdf to

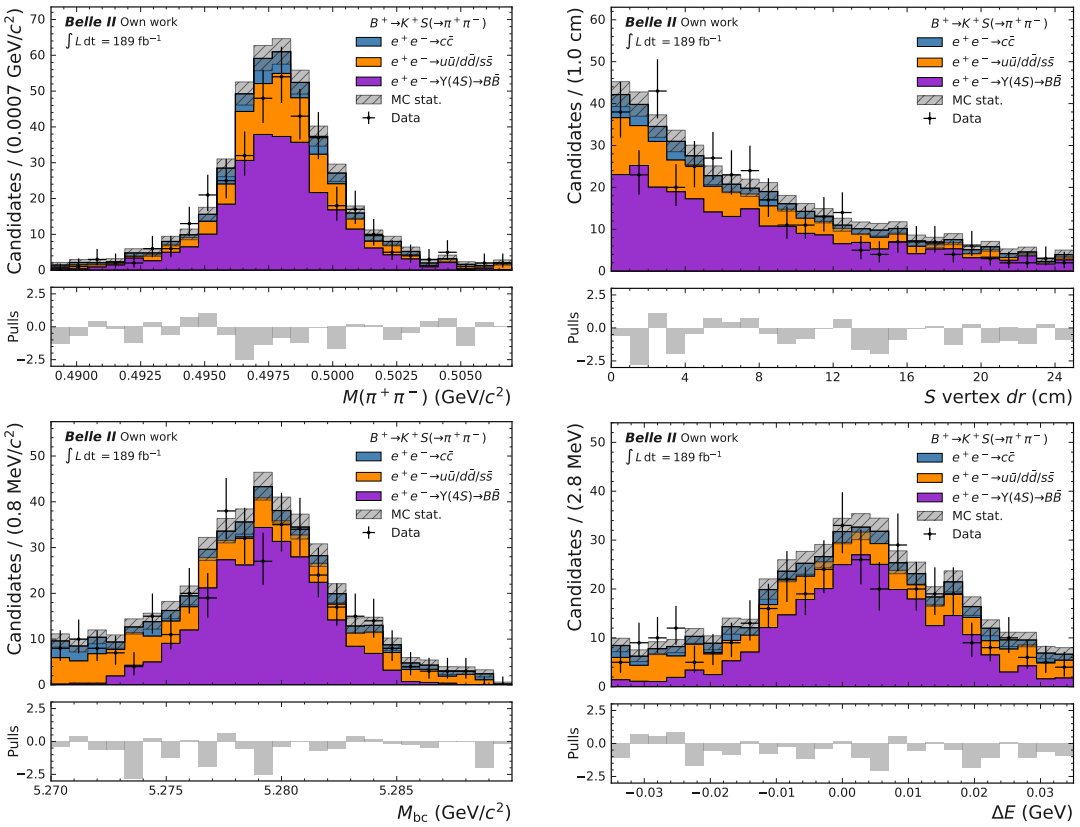

Figure 8.6.: Distributions from data and simulation for $B^+ \to K^+S(\to \pi^+\pi^-)$ in (upper left) $M'(x^+x^-)$, (upper right) S vertex dr, (lower left) M_{bc} and (lower right) ΔE using the selections of the $B^+ \to K^0_S\pi^+$ sideband.

the ΔE distributions in the dedicated simulation, after application of the sideband selection requirements, see Fig. 8.7. The ΔE distribution is found to not be accurately described by a single Crystal Ball pdf due to the longer tails present on both sides of the peak.

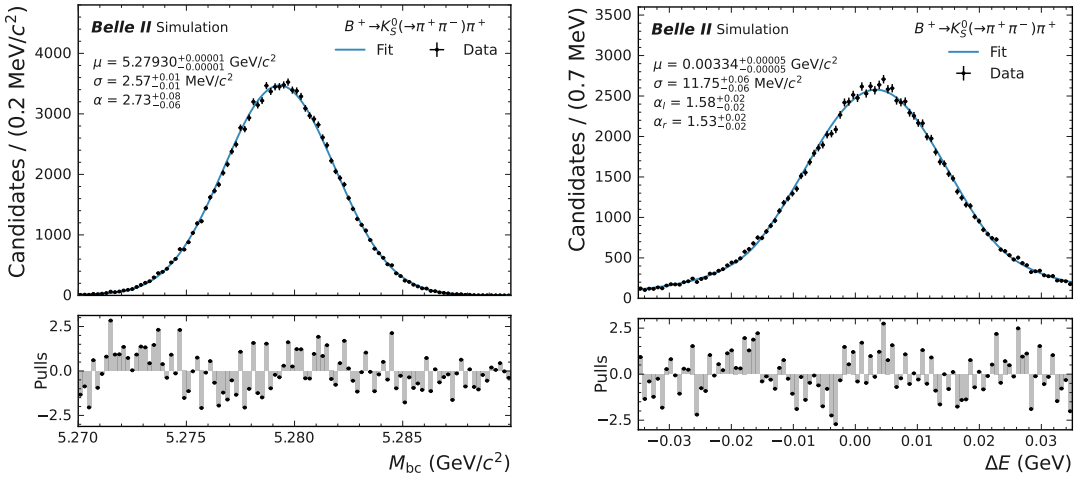

Figure 8.7.: $B^+ \to K^0_S\pi^+$ (left) $M_{\rm bc}$ and (right) ΔE distributions in simulation after sideband selections with a fit of template pdfs.

The $M_{\rm bc}$ distribution of the light-quark-pair background is modelled by an Argus pdf [156]:

$$\text{Argus}(m, m_0, c) = m \cdot \sqrt{1 - \left(\frac{m}{m_0}\right)^2} \cdot \exp\left[c \cdot \left(1 - \left(\frac{m}{m_0}\right)^2\right)\right], \tag{8.6}$$

where m_0 is the maximally allowed mass or cut-off value and c describes how peaky the distribution is. The background shape in ΔE is modelled by a linear Chebyshev polynomial pdf.

The shape parameters of the $B^+ \to K^0_S\pi^+$ pdfs in the fit to extract the branching fraction are set to the values estimated using the dedicated simulation. The $B^+ \to K^0_S\pi^+$ normalisation as an estimator for the yield is estimated by the branching fraction fit. All background pdf parameters are estimated by the branching fraction fit

to determine the background shape and normalisation directly in the dataset. A single $B^+ \to K^0_S\pi^+$ yield parameter controls the normalisation of the pdfs in M_{bc} and ΔE simultaneously. The same is done for the background yield. The two-dimensional fit is carried out by minimising the sum of the negative log-likelihoods in both observables.

The $B^+ \to K^0_S\pi^+$ branching fraction $\mathcal{B}$ is calculated from the $B^+ \to K^0_S\pi^+$ yield $N_{\mathrm{Sig.}}$ estimated by the fit using:

$$N_{\mathrm{Sig.}} = 2N_{BB}\epsilon_{B^+\to K^0_S\pi^+}\mathcal{B}, \tag{8.7}$$

where N_{BB} describes the number of $B\overline{B}$ pairs in the dataset and thereby depends on the integrated luminosity. The factor two takes into account that both B-mesons in an event can decay via the searched-for process. The $B^+ \to K^0_S\pi^+$ efficiency $\epsilon_{B^+\to K^0_S\pi^+}$ is determined using the dedicated simulation, within the sideband selection:

$$\begin{aligned} \epsilon_{B^+\to K^0_S\pi^+} &= 11.99 \pm 0.03\%, \\ \epsilon^{\mathrm{corrected}}_{B^+\to K^0_S\pi^+} &= 10.70 \pm 0.03\%, \end{aligned} \tag{8.8}$$

where the uncertainties are statistical, and the second value has been corrected for differences between data and simulation. Corrections are taken into account for the PID efficiency, as well as for the K^0_S efficiency as a function of the vertex distance.

The $B^+ \to K^0_S\pi^+$ branching fraction is measured separately in sideband simulation, see Fig. 8.8 and in data, see Fig. 8.9. The uncorrected efficiency is used for the estimation in simulation. The efficiency corrected for differences between data and simulation is used in the fit to data.

The resulting values are:

$$\begin{aligned} \mathcal{B}(B^+ \to K^0_S\pi^+)_{\mathrm{Simulation}} &= (11.2 \pm 0.3) \cdot 10^{-6}, \\ \mathcal{B}(B^+ \to K^0_S\pi^+)_{\mathrm{Data}} &= (10.8 \pm 0.8) \cdot 10^{-6}, \end{aligned} \tag{8.9}$$

where the uncertainties are statistical only. The values are consistent with each other and agree well with the PDG average of $(11.9 \pm 0.4) \cdot 10^{-6}$ [1].

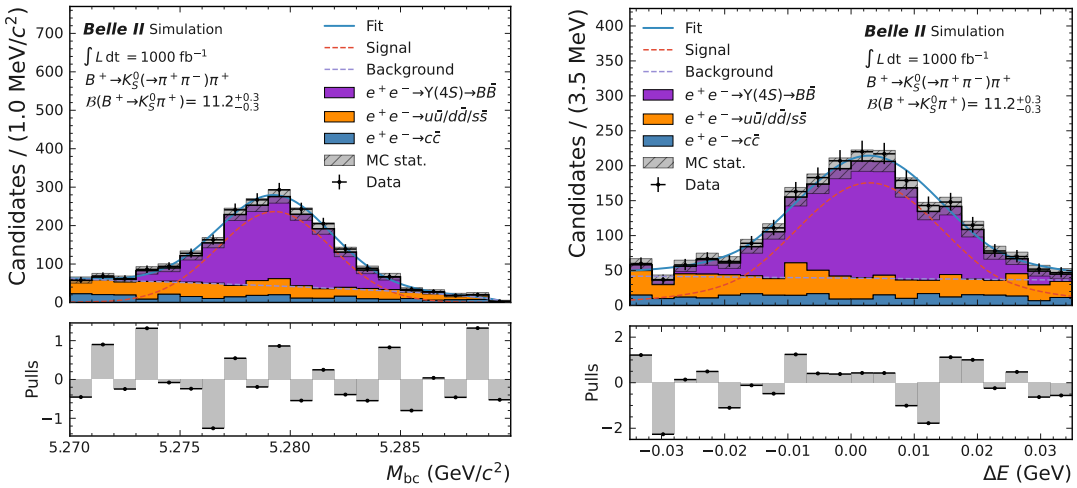

Figure 8.8.: Distributions of (left) M_{bc} and (right) ΔE in simulation using the $B^+ \to K^0_S\pi^+$ sideband selections with a simultaneous fit in both variables using template pdfs.

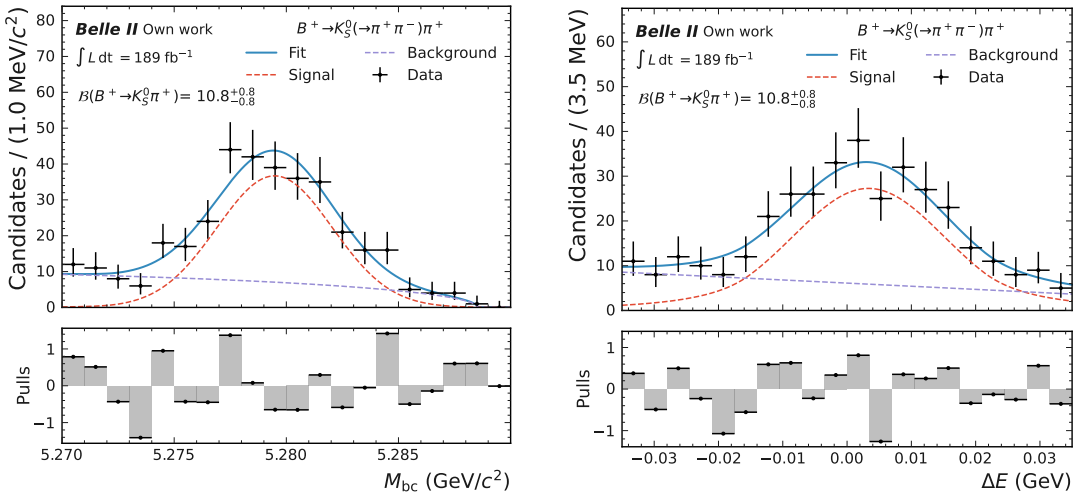

Figure 8.9.: Distributions of (left) M_{bc} and (right) ΔE in data using the $B^+ \to K^0_S\pi^+$ sideband selections with a simultaneous fit in both variables using template pdfs.

8.3.2. Lifetime validation

The fact that the S signal features a non-zero lifetime is used mainly for the S specific selections, such as the minimum vertex displacement, and to determine the signal efficiency and template parameters as a function of lifetime. The objective of the analysis is to search for an excess of signal candidates in the reconstructed mass distribution. Hence, no attempt is made to precisely measure the lifetime. The lifetime and other signal parameters have to be determined in a separate specific measurement if a significant signal-like excess is found. Therefore, a dedicated measurement of the K^0_S lifetime in $B^+ \to K^0_S\pi^+$ is beyond the scope of this validation. However, it is worth studying the agreement between the vertex displacement distributions in data and simulation in more detail.

An exponential pdf is used to fit the three-dimensional S (K^0_S) vertex distance distribution in sideband simulation, sideband data, and the dedicated $B^+ \to K^0_S\pi^+$ simulation to determine the decay parameter λ, see Fig. 8.10. Only candidates with three-dimensional vertex distance above 5 cm are used in the fit, since the minimum displacement selection in the two-dimensional vertex distance (dr) distorts the former distribution at small values to slightly non-exponential behaviour.

The resulting values for the vertex distance decay constant $\lambda^{\text{Distance}}$ are:

$$\begin{aligned} \lambda^{\text{Distance}}_{\text{Sideband simulation}} &= -0.090 \pm 0.003\,\text{cm}^{-1}, \\ \lambda^{\text{Distance}}_{\text{Sideband data}} &= -0.107 \pm 0.009\,\text{cm}^{-1}, \\ \lambda^{\text{Distance}}_{B^+\to K^0_S\pi^+\text{ simulation}} &= -0.0946 \pm 0.0005\,\text{cm}^{-1}, \end{aligned} \tag{8.10}$$

where the uncertainty is statistical only, agree with each other, suggesting that the vertex displacement distribution is well described by the simulation. The main comparison is between the determined values from sideband simulation and data. It is worth noting that the distributions in sideband data and simulation contain not only the $B^+ \to K^0_S\pi^+$ process but also K^0_S from light-quark-pair events, which pass the selections by combinatorics. However, the B-meson process dominates the sample, and the K^0_S kinematics are similar between both processes. This results in agreement within statistics with the decay constant determined using the dedicated $B^+ \to K^0_S\pi^+$

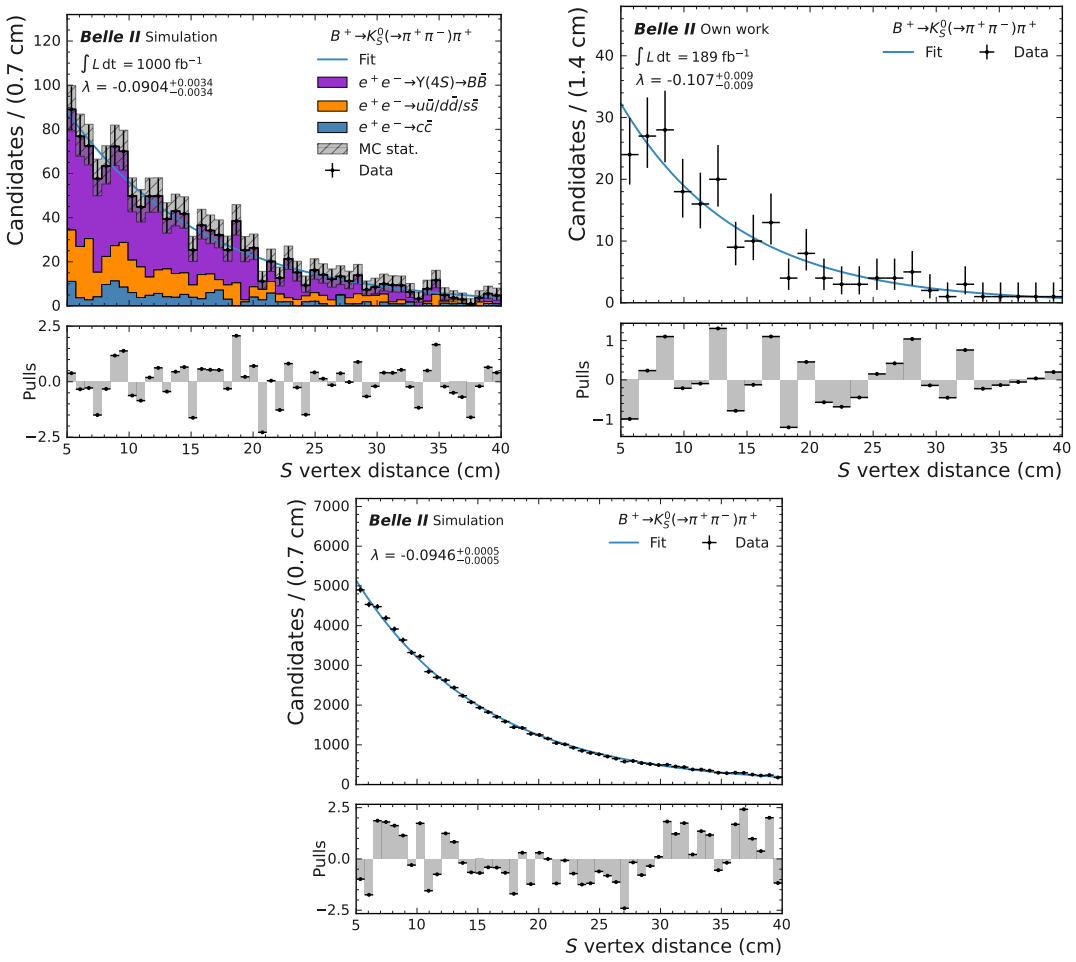

Figure 8.10.: Three dimensional vertex distance distributions of the K^0_S (S) candidates in (upper left) simulation, (upper right) data and (lower) $B^+ \to K^0_S\pi^+$ simulation using the $B^+ \to K^0_S\pi^+$ sideband selections with fits of an exponential pdf.

simulation.

Closely related to the vertex distance is the flight time of the S (K^0_S) candidate calculated from its vertex distance and momentum. The exponential decay parameters for the flight time are measured analogously to the distance by a fit of an exponential pdf to the distributions in sideband simulation, sideband data, and $B^+ \to K^0_S\pi^+$ simulation, shown in Fig. 8.11. Candidates are included in the fit if their flight time is greater than 150 ps to avoid shaping effects due to the selection in vertex dr.

The resulting values for the flight time decay constant $\lambda^{\text{Flight time}}$ are:

$$\begin{aligned} \lambda^{\text{Flight time}}_{\text{Sideband simulation}} &= -5.3 \pm 0.2\,\text{ns}^{-1}, \\ \lambda^{\text{Flight time}}_{\text{Sideband data}} &= -6.2 \pm 0.6\,\text{ns}^{-1}, \\ \lambda^{\text{Flight time}}_{B^+\to K^0_S\pi^+ \text{ simulation}} &= -5.16 \pm 0.03\,\text{ns}^{-1}, \end{aligned} \tag{8.11}$$

where the uncertainty is statistical only. The values are found to be consistent with each other. Again, the sideband data and simulation also contain K^0_S not from the $B^+ \to K^0_S\pi^+$ process.

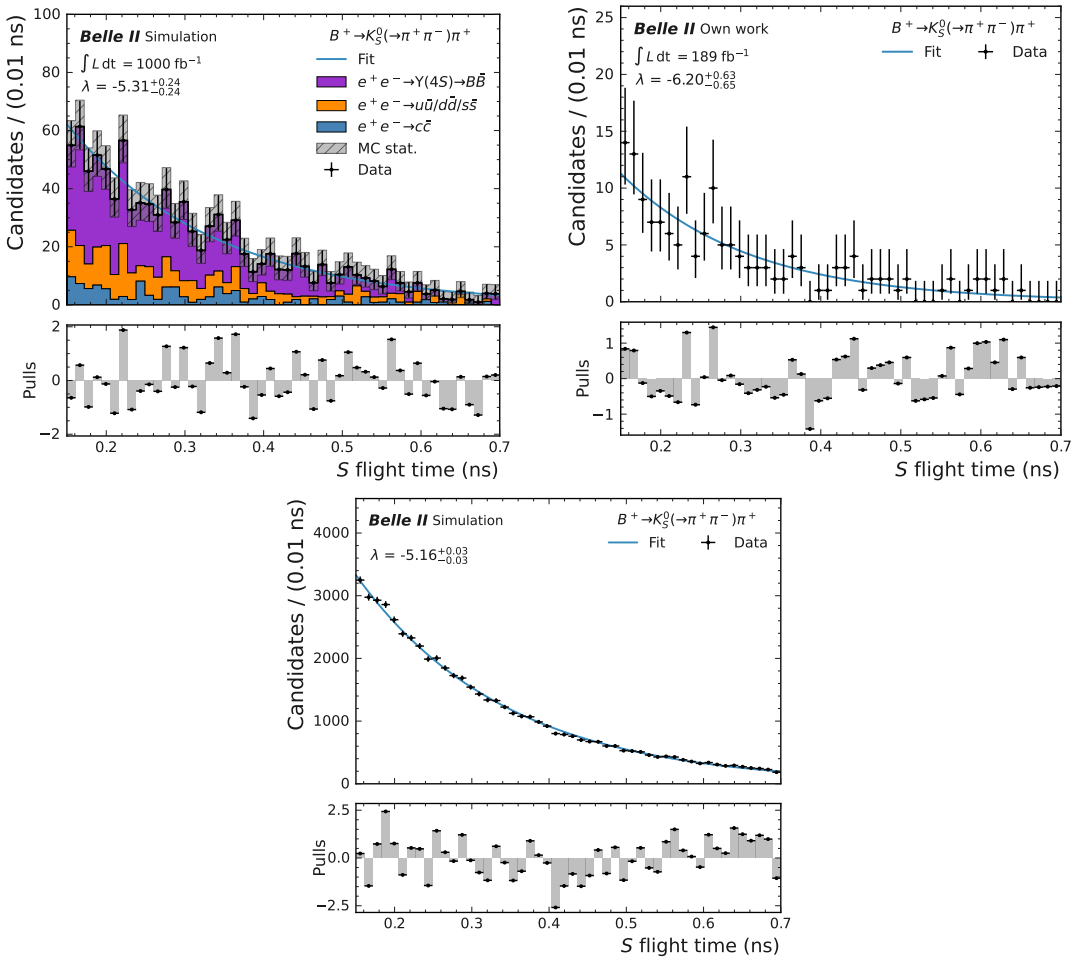

Figure 8.11.: Flight time distributions of the K^0_S (S) candidates in (upper left) simulation, (upper right) data and (lower) $B^+ \to K^0_S\pi^+$ simulation using the $B^+ \to K^0_S\pi^+$ sideband selections with fits of an exponential pdf.

9. Systematic uncertainties

This chapter describes the sources of systematic uncertainties that are considered in this search.

The parameter of interest for this search is the product of branching fractions for the S production in $B \to KS$ and its decay via $S \to x^+x^-$. The signal efficiency and signal pdf parameter values are determined using simulation. Both are subject to systematic uncertainties that have an effect on the resulting signal branching fraction. The largest sources of systematic uncertainties arise due to the finite knowledge of the size of correction factors derived from differences between data and simulation. No uncertainties are considered for the background as the contribution is determined directly in fits to data using large fit windows in the reconstructed mass distribution.

9.1. Signal efficiency

The signal efficiency and the number of B-meson pairs in the dataset are used as multiplicative factors to determine the signal branching fraction from the estimated signal yield, see Sec. 10.2 for details on the fit model. Therefore, both affect the signal branching fraction in the same way and are detailed in the following. The resulting sizes of systematic uncertainties are summarised at the end of the section.

9.1.1. Near-IP tracking efficiency

The difference in track-finding efficiency between data and simulation for particles that originate from close to the IP is studied using a dedicated control sample enriched with $e^+e^- \to \tau^+\tau^-$. Events are selected with one tau decay via $\tau^- \to e^-\bar{\nu}_e\nu_\tau/\mu^-\bar{\nu}_\mu\nu_\tau$

and the other via $\tau^+ \to \pi^+\pi^-\pi^+\bar{\nu}_\tau$. The efficiency is determined in data and simulation separately. The number of candidates where all three pion tracks are found is compared to the number where only two pion tracks are found. The difference between data and simulation amounts to 0.3%, and is used as an uncertainty associated with the efficiency to find a near-IP track. An uncertainty is considered for the signal efficiency, depending on the number of reconstructed signal tracks. There are three signal tracks in the $B^+ \to K^+S$ production mode and four in the $B^0 \to K^{*0}\,S$ mode, resulting in an uncertainty of:

$$\Delta\epsilon_{\text{Sig.}}(\text{track finding}) = \begin{cases} 0.9\% & \text{for } B^+ \to K^+S, \\ 1.2\% & \text{for } B^0 \to K^{*0}\,S. \end{cases} \tag{9.1}$$

The uncertainty is independent of the S mass, lifetime, and decay channel.

9.1.2. S efficiency correction

The tracking efficiency for charged particles produced in the decay of a LLP decreases with the distance of the decay vertex from the IP. A correction factor is derived from the K^0_S control channel, described in Sec. 7.1.1, and accounts for differences between data and simulation as a function of the vertex distance. The S efficiency correction is calculated by normalising to the region of small vertex distances. Therefore, the uncertainty of the correction is used in addition to the standard tracking efficiency uncertainty for near-IP tracks. The value of the correction factor is found to vary between regions of K^0_S momentum and polar angle. An uncertainty of the same size as the nominal correction factor is found to cover these differences. The size of the correction factor is adopted as uncertainty of the S signal due to possible differences between the S and K^0_S kinematics.

9.1.3. Particle identification

The PID correction factors have associated statistical and systematic uncertainties derived in the respective control channels. For each signal track, 200 variations of the correction factor are drawn from multivariate Gaussian distributions, separately

for the systematic and statistical components. The systematic components across the momentum and polar angle regions are considered to be fully correlated. The statistical uncertainty is treated as uncorrelated. Per variation, a total PID correction factor is determined by multiplying the varied correction factors associated with each signal track. The uncertainty of the correction factor is calculated by comparing the standard deviation of the sample of varied total PID correction factors to the nominal total PID correction factor:

$$\Delta\epsilon_{\text{Sig.}}(\text{PID}) = \frac{\sigma(\text{correction variations})}{\text{nominal correction}}. \tag{9.2}$$

9.1.4. Number of B-meson pairs

The overall number of B-meson pairs and its uncertainty in the dataset used for this analysis is:

$$N_{BB} = (198.0 \pm 3.0) \times 10^6. \tag{9.3}$$

The number of B-meson pairs is determined by selecting hadronic events consisting of $e^+e^- \to q\bar{q}$, where $q \in (u, d, s, c)$, and $e^+e^- \to \Upsilon(4S) \to B\overline{B}$ processes. The same selection is applied to data collected at the $\Upsilon(4S)$ resonance and at slightly lower energies, where only $e^+e^- \to q\bar{q}$ contributes. The number of B-meson pairs is extracted by subtracting the number of events in both datasets and taking into account the difference in sample size. The value corresponds to a relative uncertainty of 1.5% in the calculation of the signal branching fraction from the estimated signal yield.

A second uncertainty originates from the ratio between the number of charged and neutral B-meson pairs in the decay of the $\Upsilon(4S)$. The value used in this search is taken from a recent measured by Belle [164]:

$$f^{\pm}/f^{00} = 1.065 \pm 0.052. \tag{9.4}$$

It is used to calculate the number of charged and neutral B-meson pairs for the determination of signal branching fractions in the $B^+ \to K^+S$ and $B^0 \to K^{*0}S$ production modes. The uncertainty of $f^{\pm}/f^{00}$ propagates to an uncertainty associated

with the number of charged and neutral B-meson pairs of 2.5%.

The total uncertainty of the number of charged or neutral B-meson pairs used in this analysis is therefore:

$$\Delta\epsilon_{\text{Sig.}}(N_{BB}^{\pm/00}) = \sqrt{\Delta^2\epsilon_{\text{Sig.}}(N_{BB}) + \Delta^2\epsilon_{\text{Sig.}}(f^{\pm}/f^{00})} = 2.9\%. \tag{9.5}$$

9.1.5. Track momentum correction

The momentum correction factors for tracks from close to the IP in data have associated uncertainties. These take the form of upward and downward variations of the nominal corrections. All tracks used for the reconstructed signal B-meson and for the event shape Fox-Wolfram $R2$ are corrected. Signal candidates can migrate into or out of the region defined by the selection requirements when the signal track momenta are modified. A variation of the track momentum correction within its uncertainty can have this effect. This can lead to a difference in signal efficiency between data and simulation.

Simulated signal events are used to estimate this source of uncertainty. There is no such correction for tracks in simulation because of the perfect knowledge of the magnetic field. Hence, the size of the uncertainty is determined by comparing the signal efficiency with the variations of the track momentum correction factor to the nominal case without correction:

$$\Delta\epsilon_{\text{Sig.}}(\text{scale}) = \frac{|\epsilon_{\text{Sig.}}(\text{scale up}) - \epsilon_{\text{Sig.}}(\text{scale down})|}{\epsilon_{\text{Sig.}}(\text{nominal})}. \tag{9.6}$$

9.1.6. Track momentum resolution

The resolution is studied in a control sample enriched with $e^+e^- \to \mu^+\mu^-$ events. The reconstructed momenta of the positively and negatively charged muons are compared with each other in data and simulation. The difference between data and simulation is attributed to an overestimation of the track momentum resolution in the latter. A smearing of the track momentum in simulation is performed to cover this difference. Variations in the size of the smearing can lead to signal candidates migrating in and out of the selection requirements, analogously to the track momentum scale.

The uncertainty is estimated by computing the relative difference between the signal efficiency in simulation with different smearing scenarios:

$$\Delta\epsilon_{\text{Sig.}}(\text{smear}) = \frac{|\epsilon_{\text{Sig.}}(\text{large smear}) - \epsilon_{\text{Sig.}}(\text{small smear})|}{\epsilon_{\text{Sig.}}(\text{medium smear})}. \tag{9.7}$$

No smearing of track momenta is applied for the nominal values derived from simulation.

9.1.7. Sample size

The number of signal events simulated for each combination of model parameters is finite. This gives rise to a statistical uncertainty of the signal efficiency. The efficiency is calculated by comparing the number of simulated signal candidates to the number of reconstructed and selected signal candidates. The uncertainty is determined using a Clopper-Pearson confidence interval, symmetrised by taking the average. The large sample sizes ensure good coverage of the symmetrised interval.

9.1.8. Interpolation

The signal efficiency for each possible S mass value is determined by linear interpolation between the efficiency values computed at S masses where signal events are simulated. The interpolation procedure can yield efficiency values that do not perfectly match the actual efficiency due to the finite spacing of simulated signal samples. This translates to a further source of uncertainty associated with the signal efficiency. Its size is estimated at each simulated signal sample m_i by excluding the sample from the interpolation and comparing the directly computed efficiency at the sample to the efficiency interpolated between the other samples:

$$\Delta\epsilon_{\text{Sig.}}(\text{interpolation})(m_i) = \frac{|\epsilon_{\text{Sig.}}(m_i) - \epsilon_{\text{Sig.}}(\text{Interpolated}_{m_{i-1}}^{m_{i+1}}(m_i))|}{\epsilon_{\text{Sig.}}(m_i)}. \tag{9.8}$$

9.1.9. Fit

Detailed studies, presented in Sec. 10.6.1, find a small bias on the signal yield estimator in the fit. An additional uncertainty of 3%, independent of kinematics and analysis channel, is considered for the signal efficiency to cover this effect.

9.1.10. Signal efficiency uncertainty results

The resulting sizes of the different sources of uncertainty of the signal efficiency together with the total uncertainty are shown for different S mass hypotheses in:

- Figure 9.1 for $S \to \mu^+\mu^-$ for two different S lifetimes and both production modes, and in
- Figure 9.2 for $B^+ \to K^+S$ with the other three S final-state channels.

The total uncertainty is computed by adding the uncertainties from different sources in quadrature. It is linearly interpolated to determine the uncertainty for the S mass hypotheses between the simulated signal samples. The uncertainties associated with the signal efficiency due to variations in track momentum correction and smearing are combined.

The uncertainties associated with the number of $B\overline{B}$ pairs and the fit bias are constant and are the dominant sources of uncertainty for a large part of the parameter space. The K^0_S efficiency correction uncertainty is typically large and is dominant for small S masses and long lifetimes. The mass dependence is due to the Lorentz boost of the S which depends on its mass and leads to a decreasing average vertex distance with increasing mass. The correction and the corresponding uncertainty of the same size depend on the average vertex distance. The largest sub-leading uncertainties are attributed to the near-IP track finding efficiency and the signal efficiency interpolation procedure. Smaller contributions arise due to the finite sample sizes. The uncertainty due to the PID requirements is small for most of the parameter space except for the production in $B^0 \to K^{*0}(\to K^+\pi^-)S$ with very large S mass hypotheses. Here, the phase space for the K^{*0} is small, leading to small momenta for the two prompt

tracks, see Sec. 4.3.2. The uncertainty associated with the track momenta is negligible compared to the other sources.

Small variations of the uncertainties associated with the LLP efficiency correction are visible. These arise due to selections that have an effect on the mean value of the S vertex distance. The $M_{\pi^+\pi^-}$ veto in the K^0_S region, as well as the tighter vertex dr selection around SM resonances, have this effect. The uncertainty of interpolation increases in these regions as well, because of the rapid change in signal efficiency induced by the additional selections.

The typical total uncertainty of the S signal efficiency is around a few percent. However, it can be as large as 15% for large mean vertex distances. The size of the uncertainties for different S final-state channels is similar.

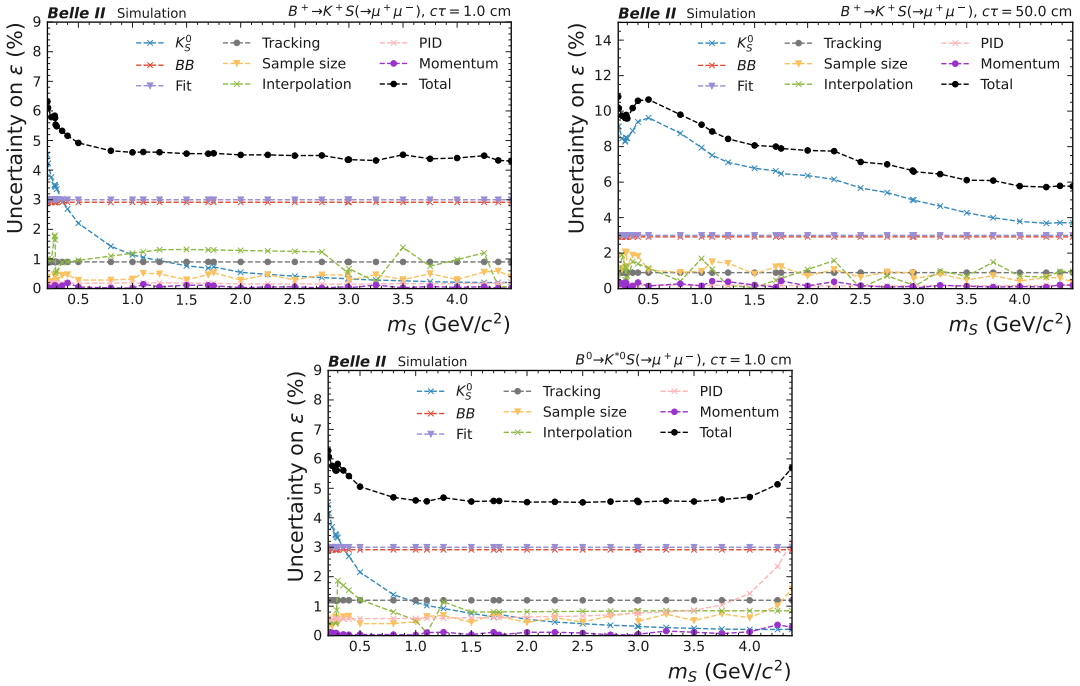

Figure 9.1.: Uncertainty of the signal efficiency and its components as a function of S mass. In $B^+ \to K^+ S(\to \mu^+\mu^-)$ with (upper left) a lifetime of $c\tau_S = 1\,\text{cm}$, (upper right) a lifetime of $c\tau_S = 50\,\text{cm}$, and (lower) in $B^0 \to K^{*0}(\to K^+\pi^-)S(\to \mu^+\mu^-)$ with a lifetime of $c\tau_S = 50\,\text{cm}$.

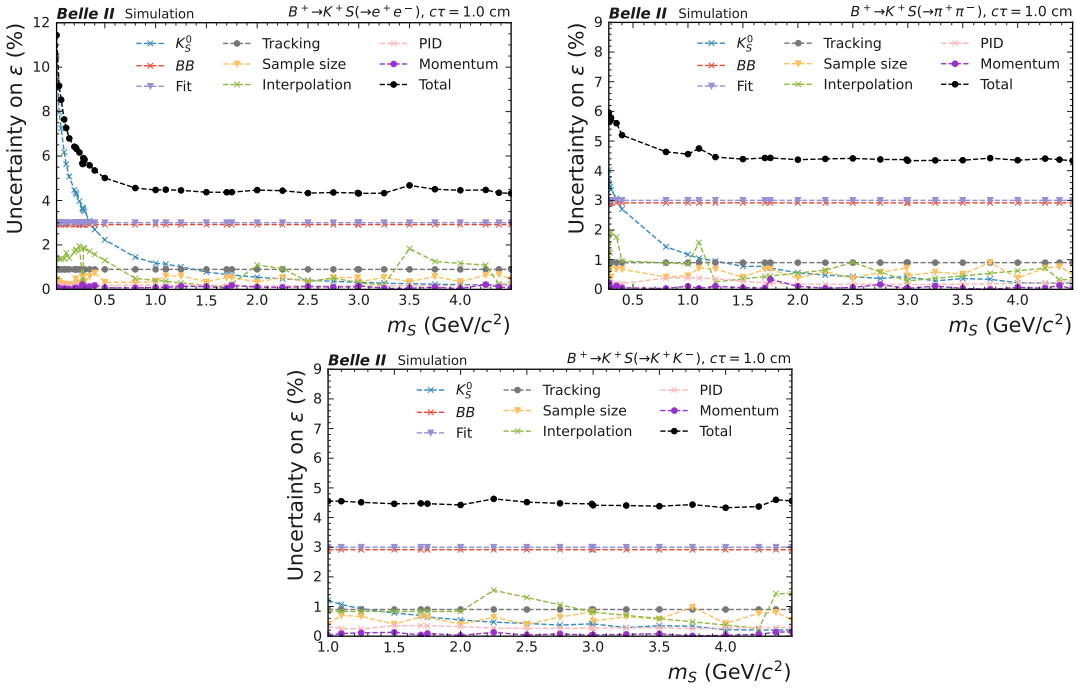

Figure 9.2.: Uncertainty of the signal efficiency and its components as a function of S mass. (Upper left) in $B^+ \to K^+ S(\to e^+ e^-)$, (upper right) in $B^+ \to K^+ S(\to \mu^+ \mu^-)$ and (lower) in $B^+ \to K^+ S(\to K^+ K^-)$.

9.2. Signal model

The second type of systematic uncertainties considered for this search affects the template parameters of the signal pdf. An uncertainty associated with the shape parameters translates to different possible signal widths and tails, which can affect the signal yield and hence the signal branching fraction. They arise, for example, due to differences in the shape of the signal peak in the reconstructed S mass distribution determined in simulation and data.

9.2.1. Track momentum correction

The uncertainty of the pdf parameters of the signal associated with the track momentum correction is determined by comparing the pdf parameters with different sizes of the correction. The sizes of the two different correction scenarios correspond to the uncertainty of the correction. The signal pdf is fitted to distributions of simulated signal events, separately with the corrections on the track momenta and without correction. The size of the uncertainty is estimated as the difference between the pdf parameters using the two sizes of corrections relative to the nominal value:

$$\begin{aligned}
\Delta\sigma(\text{scale}) &= \frac{|\sigma(\text{scale up}) - \sigma(\text{scale down})|}{\sigma(\text{nominal})}, \\
\Delta\alpha_l(\text{scale}) &= \frac{|\alpha_l(\text{scale up}) - \alpha_l(\text{scale down})|}{\alpha_l(\text{nominal})}, \\
\Delta\alpha_r(\text{scale}) &= \frac{|\alpha_r(\text{scale up}) - \alpha_r(\text{scale down})|}{\alpha_r(\text{nominal})}.
\end{aligned} \tag{9.9}$$

The effect of the track momentum correction on the S peak position is not taken into account. The peak position is a constant parameter in the determination of the signal yield, and the expected magnitude of its variation due the uncertainty associated with the track momentum correction is small. A precise measurement must follow in a dedicated study when a significant signal is found.

9.2.2. Track momentum resolution

The primary motivation of smearing the track momentum is to match the width of the reconstructed mass distributions in simulation to data. The uncertainty of the signal pdf parameter values due to the momentum resolution is computed analogously

using the two variations of the momentum smearing relative to the central smearing:

$$\begin{aligned}\Delta\sigma(\text{smear}) &= \frac{|\sigma(\text{smear up}) - \sigma(\text{smear down})|}{\sigma(\text{central smear})},\\ \Delta\alpha_l(\text{smear}) &= \frac{|\alpha_l(\text{smear up}) - \alpha_l(\text{smear down})|}{\alpha_l(\text{central smear})},\\ \Delta\alpha_r(\text{smear}) &= \frac{|\alpha_r(\text{smear up}) - \alpha_r(\text{smear down})|}{\alpha_r(\text{central smear})}.\end{aligned} \tag{9.10}$$

9.2.3. S shape correction

The track momentum correction and resolution uncertainties cover the differences between data and simulation for tracks that originate from close to the IP. Correction factors for signal template parameters that take into account additional differences for displaced tracks are estimated from the K^0_S control sample in Sec. 7.1.1. Uncertainties corresponding to the full sizes of the correction factors are considered on the signal pdf parameters. This is done following the same argumentation as for the S efficiency correction, namely variations between K^0_S phase-space regions.

9.2.4. Sample size

The finite number of signal events simulated for each S mass, lifetime, and channel combination propagates to a statistical uncertainty of the pdf parameter values. The statistical uncertainty of the parameter values estimated by the fits to signal simulation is used to cover this effect.

9.2.5. Signal model uncertainty results

The resulting sizes of the different sources of uncertainty associated with the signal pdf parameters together with the total uncertainties for different S mass hypotheses in $B^+ \to K^+ S(\to \mu^+\mu^-)$ are shown in Fig. 9.3 for a short S lifetime and in Fig. 9.4 for a long S lifetime. Total uncertainties are linearly interpolated between S samples simulated at different masses.

The variations between S masses are smoothed for uncertainties associated with the finite sample size, the track momentum correction, and the track momentum smearing. This is done by fitting a straight line to the values determined at different S masses. Such a procedure is adopted to simplify the application of uncertainties to the signal extraction. It is applicable due to the fact that the signal pdf uncertainties have a small impact on the resulting upper limit on the signal branching fraction.

The uncertainties due to the K_S^0 based correction factors are the dominant source of uncertainty of the pdf parameters, especially for the Gaussian width σ. The size of the correction and hence the uncertainty depend on the average vertex distance. All other sources are found to depend weakly on the lifetime of the S. The differences in the size of the shape uncertainties between final states are found to be small. The same uncertainties are considered for both production modes as the differences between the two is found to be small, and the same pdf parameter values are used.

The interpolation between S masses for the shape parameters is performed with fits of higher-order polynomials, see Sec. 6.1.3. The resulting uncertainty associated with the interpolation procedure is negligible compared to the other sources and is not shown.

The typical total uncertainties of the pdf parameters are around a few percent and as high as 15% for σ and α_l at large average vertex distances. The uncertainty of α_r is smaller since there is no additional correction based on the vertex distance.

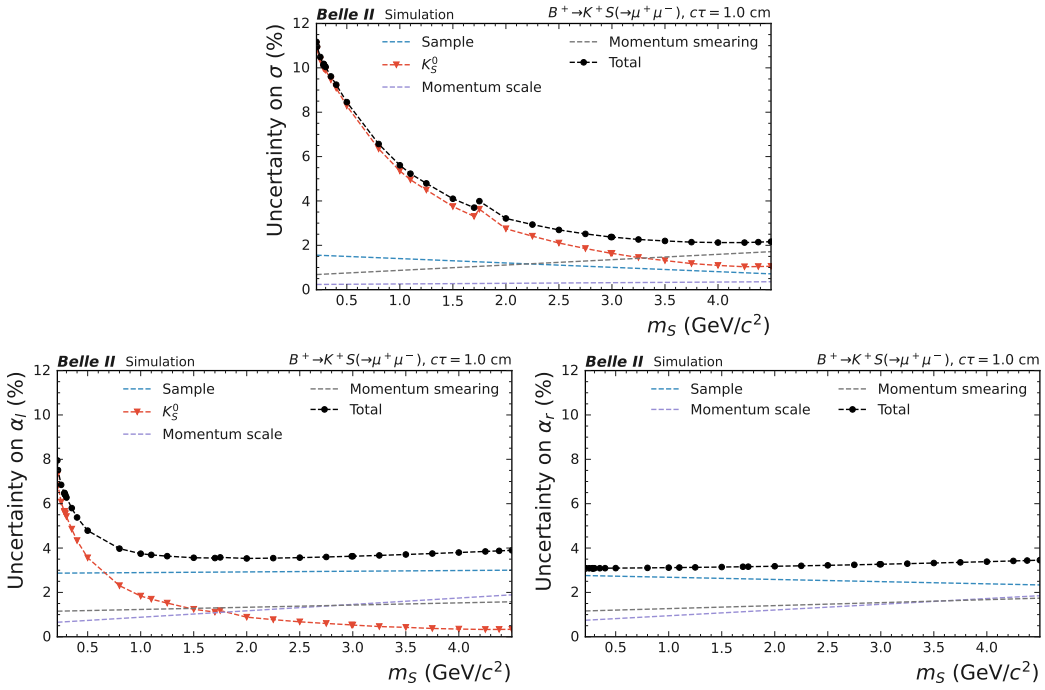

Figure 9.3.: The signal shape uncertainty and its different components as a function of the S mass. Shown in the $B^+ \to K^+S(\to \mu^+\mu^-)$ channel with a lifetime of $c\tau_S = 1\,\text{cm}$ for (upper left) σ, (upper right) α_l and (lower) α_r.

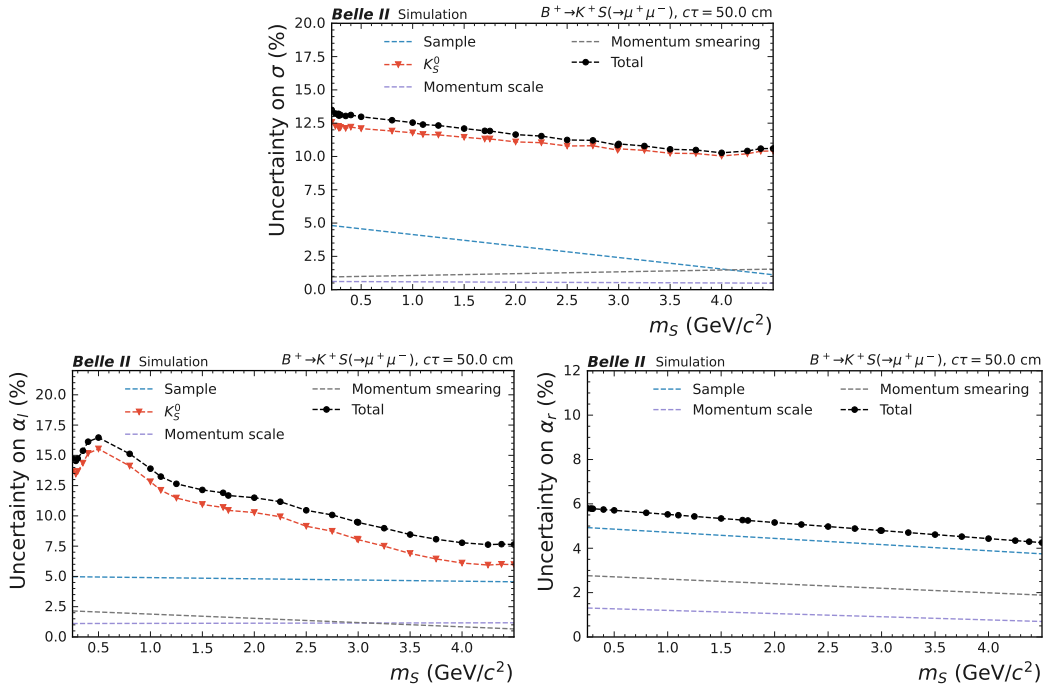

Figure 9.4.: The signal shape uncertainty and its different components as a function of the S mass. Shown in the $B^+ \to K^+ S(\to \mu^+\mu^-)$ channel with a lifetime of $c\tau_S = 50\,\text{cm}$ for (upper left) σ, (upper right) α_l and (lower) α_r.

10. Statistical treatment

This chapter describes the statistical methods that are used to determine the number of signal candidates in data and to derive upper limits on the signal branching fraction.

10.1. Mass scan

The signal is searched for as a peaking excess of candidates on the linear SM background in the reconstructed M' distribution. Templates used to describe the signal and background components are introduced in Chap. 6. The M' distribution is scanned in steps of half the signal width σ to ensure that no possible signal is missed. The spacing thereby increases towards larger mass hypotheses where the signal resolution deteriorates.

The smallest scan point for the $S \to e^+e^-$ channels is at $m_S = 25\,\mathrm{MeV}/c^2$. The scans for the heavier final states start one σ above the corresponding kinematic thresholds to minimise residual modelling difficulties close to the threshold.

The largest testable S mass hypotheses are given by the difference between the B-meson and the kaon masses. This results in $m_S \leq 4.78\,\mathrm{GeV}/c^2$ for $B^+ \to K^+S$ and $m_S \leq 4.38\,\mathrm{GeV}/c^2$ for $B^0 \to K^{*0}\,S$.

The scan point values are determined for a single low lifetime hypothesis and used for all lifetimes even though there is a small difference in σ between lifetimes. This facilitates the translation of upper limits on the signal branching fraction at different lifetimes into limits on the coupling parameter in the benchmark models, detailed in Sec. 11.4.

The total numbers of scan points for the different channels are summarised in Tab. 10.1.

Table 10.1.: Number of mass scan points in the different analysis channels.

	$B^+ \to K^+ S$	$B^0 \to K^{*0} S$
$S \to e^+e^-$	1905	1829
$S \to \mu^+\mu^-$	2067	1981
$S \to \pi^+\pi^-$	2013	1926
$S \to K^+K^-$	1509	1430

The M' windows in which signal extraction fits are performed are constructed around the tested S mass hypothesis. The minimum size of each window is set to $\pm 20\sigma(m_S)$ with the signal width at the tested mass m_S. The window size is iteratively increased in steps of 10% in case the fit window contains less than ten candidates until at least ten candidates are found to increase the fit stability.

The size of the windows is chosen much larger as the signal width to include the mass sidebands around the tested S mass value. The values of the background normalisation $N_{\text{Bkg.}}$, and of the shape parameters, the linear coefficient c_1 and the exponential coefficient λ, are determined directly in data. The large lever arms of the fit window facilitate this determination. This ensures smooth modelling around the tested S mass and minimises the dependence on the background simulation. A too small window could result in a potential excess of signal candidates being absorbed into the background description, leading to a reduced separation power. The windows are truncated if, by construction, they extend into an unphysical region such as negative or larger than the kinematically allowed S masses. All other windows are symmetric around the tested S mass.

The scan of the M' distribution is performed separately for different S lifetime hypotheses and analysis channels. Further scans are done using a combined extraction in all relevant and kinematically accessible analysis channels, again separately for different lifetimes, detailed in Sec. 10.2.1.

The signal width at a scan point m_S can vary between final states, e.g. close to the di-kaon threshold where the S width is smallest in the kaon final state. This must be taken into account for the spacing between scan points in the combined extraction. The distance between scan points at m_S is determined by half of the signal width in

the final state with the smallest width:

$$d(m_S, m_S + 1) = \frac{\min(\sigma_{\mu^+\mu^-}(m_S), \sigma_{\pi^+\pi^-}(m_S), \sigma_{K^+K^-}(m_S))}{2}. \tag{10.1}$$

10.2. Fit model

The signal extraction fit in the production mode i and the final state j at a given mass m_S and lifetime $c\tau_S$ is performed by fitting the combined signal+background model to the $M'(x^+x^-)$ distribution in the window around m_S constructed as explained above.

The Gaussian mean of the signal DSCB pdf is set fixed to the test mass:

$$\mu = m_S. \tag{10.2}$$

The remaining pdf parameters of the signal, σ, $\alpha_{l/r}$, are set to their values determined for the specific combination of lifetime and final state, and interpolated between signal samples simulated at different masses to m_S. They are corrected for the observed differences between LLP performance in data and simulation. The same values are used for both production modes.

Three more fit parameters are defined; the signal efficiency $\epsilon_{\text{Sig.}}$, the number of B-meson pairs N_{BB}, and the signal branching fraction $\mathcal{B}$. The branching fraction parameter describes both parts of the underlying process, the production rate of the S by the decay of a B-meson, and the decay rate of the S in the final state considered x:

$$\mathcal{B} = \mathcal{B}(B \to K\,S) \times \mathcal{B}(S \to x^+x^-). \tag{10.3}$$

The signal efficiency parameter is set to the value determined for the given lifetime and channel by interpolation between the simulated signal samples to m_S:

$$\epsilon_{\text{Sig.}} = \epsilon_{\text{Sig.}}^{i,j,c\tau_S}(m_S). \tag{10.4}$$

The signal efficiency is corrected for differences between data and simulation in the LLP reconstruction performance and in the PID selection efficiency. The number of

B-meson pairs is set to the mean value determined for the dataset used in this search, $N_{BB} = 198.0 \times 10^6$.

The DSCB pdf is extended by a signal yield parameter $N_{\text{Sig.}}$ which depends on these parameters

$$N_{\text{Sig.}} = 2 N_{BB} f^{\text{charged/neutral}} \epsilon_{\text{Sig.}} \mathcal{B}. \tag{10.5}$$

The factor two takes into account that both B-mesons in an event can decay via the signal process. The number of B-meson pairs of specific flavour as needed for the two production modes are determined using the fractions f^{charged} for $B^+ \to K^+ S$ and f^{neutral} for $B^0 \to K^{*0} S$:

$$\begin{aligned} f^{\text{charged}} &= \frac{f^{\pm}/f^{00}}{1 + f^{\pm}/f^{00}}, \\ f^{\text{neutral}} &= \frac{1}{1 + f^{\pm}/f^{00}}, \end{aligned} \tag{10.6}$$

with $f^{\pm}/f^{00}$ as introduced in Eq. 9.4. Furthermore, a factor 2/3 is included for the $B^0 \to K^{*0} S$ production mode to account for the fraction of K^{*0} decaying into a charged kaon and pion [1].

As the signal efficiency is determined using simulated signal candidates that assume the kinematics of a scalar particle, the $\mathcal{B}$ parameter relates to a scalar or pseudoscalar LLP.

The $\mathcal{B}$ parameter and thus also $N_{\text{Sig.}}$ are constrained to values $N_{\text{Sig.}} \geq 0$. This increases the fit stability, especially in windows with a small number of candidates.

The linear Chebyshev pdf (plus exponential for the low-mass $S \to e^+ e^-$ region) describing the background contribution is extended by a yield parameter $N_{\text{Bkg.}}$. The background yield and the linear coefficient are restricted to $N_{\text{Bkg.}} \geq 0$ and $c_1 \in [-1, 1]$. The extended background model is added to the signal model to form the signal+background model.

Fits of the signal+background model to M' distributions in background-only simulation are shown in Fig. 10.1. The first example shows a fit window in which no signal component is found. The distribution is described solely by the background pdf. In the second example, a small but statistically insignificant signal component is found,

which is described by the signal pdf component of the combined fit model.

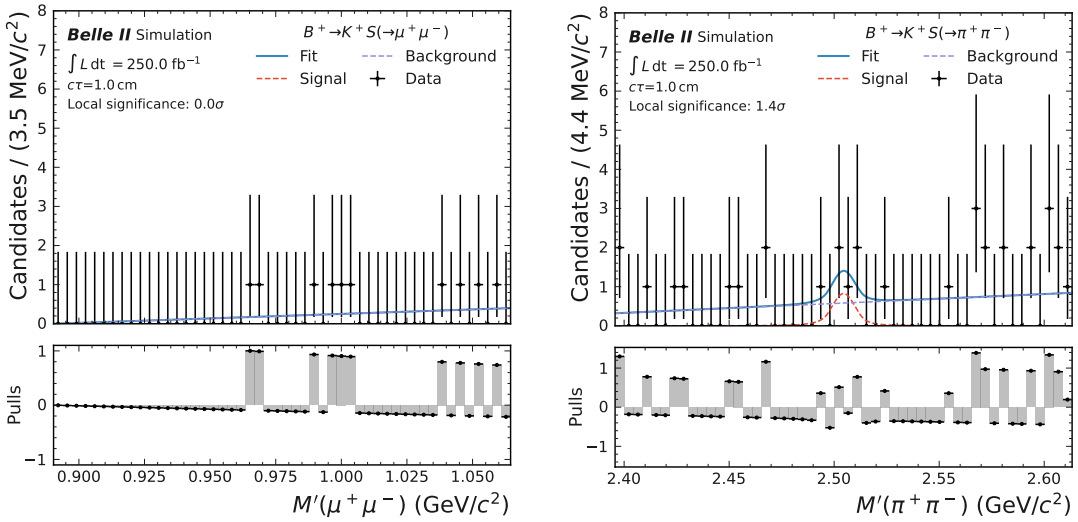

Figure 10.1.: Reconstructed M' distributions for (left) $B^+ \to K^+S(\to \mu^+\mu^-)$ and (right) $B^+ \to K^+S(\to \pi^+\pi^-)$ in background simulation with a fit of the combined signal+background model.

10.2.1. Combined fit

The dark scalar and ALP models predict the production and decay rate of LLPs in each analysis channel. A model-dependent combined fit is performed by simultaneously extracting the signal yield in multiple channels. The splitting of the total signal yield into individual channels in the combined fit is constrained to the predicted values from theory. This constrains the BSM models more strongly compared to a single-channel fit. A combined fit model is constructed from multiple analysis channels.

The signal and background pdfs for the combined fit are the same as for a single-channel extraction. Separate and independent background yield and shape parameters are added for each channel. The signal shape parameters in each channel are set to their individual values. The single free signal model parameter in the combined fit is the overall branching fraction $\mathcal{B}_{\text{Comb.}}$, describing the sum over all kinematically

accessible production channels i and decay channels j at a given S mass:

$$\mathcal{B}_{\text{Comb.}} = \sum_{i,j} \mathcal{B}_{i,j} \tag{10.7}$$

The signal models in each channel are extended by dependent signal yield parameters:

$$N^{i,j}_{\text{Sig.}}(m_S, c\tau_S) = 2N_{BB} f^{\text{charged/neutral}} A^{i,j}(m_S) \epsilon^{i,j}_{\text{Sig.}}(m_S, c\tau_S) \mathcal{B}_{\text{Comb.}}, \tag{10.8}$$

where $A^{i,j}(m_S)$ is the model prediction of the production and decay rate of S in channel i, j at a given test mass:

$$\begin{aligned} A^{\pm,j}(m_S) &= \mathcal{B}(B^+ \to K^+ S) \times \mathcal{B}(S \to j^+ j^-), \\ A^{0,j}(m_S) &= \mathcal{B}(B^0 \to K^{*0}\, S) \times \mathcal{B}(S \to j^+ j^-). \end{aligned} \tag{10.9}$$

The values for both models are shown in Sec. 2.2. Within the K^0_S veto region, the fractions $A^{i,\pi}(m_S)$ for the dark scalar interpretation are set to zero since the pion channels are excluded. The lifetime influences the overall production and decay rate of dark scalars and ALPs but not the splitting into the individual channels. The efficiency $\epsilon^{i,j}_{\text{Sig.}}(m_S, c\tau_S)$ is determined separately for each channel and lifetime by interpolating between simulated signal samples at different S masses. The channels are correlated with each other by the signal yields that depend on a common free parameter $\mathcal{B}_{\text{Comb.}}$.

The combined signal+background model of each channel is applied to the candidates found in the corresponding fit window of the channel to form a negative log-likelihood. The combined fit is performed by minimising the sum of negative log-likelihoods of all channels active at m_S.

The channels included in the different mass regions of the combined fit are summarised for the dark scalar model in Tab. 10.2 and for the ALP model in Tab. 10.3. The $S \to e^+e^-$ rate is negligible for S masses above the di-muon threshold in both BSM models. Therefore, no significant information is obtained by adding the $S \to e^+e^-$ channels to the combined fit in this region. There are no predictions for the dark scalar decay rates in hadrons for $m_S > 2\,\text{GeV}/c^2$, resulting in the drop of the corresponding channels from the combined extraction. The model-independent

results obtained from single-channel fits extend to S masses beyond $2\,\mathrm{GeV}/c^2$. The only prediction used in the model-independent results is the spin-0 kinematics of the S that affects the signal efficiency within the selection. The hadronic S decay channels are not used in the combined fit for the ALP, as the ALP model does not predict decays to two mesons.

Table 10.2.: Analysis channels included in the dark scalar model-dependent combined fit per S mass region.

m_S	K^+	K^{*0}	e^+e^-	$\mu^+\mu^-$	$\pi^+\pi^-$	K^+K^-
$(0.025\,\mathrm{GeV}/c^2, 2m_\mu)$	✓	✓	✓	×	×	×
$(2m_\mu, 2m_\pi)$	✓	✓	×	✓	×	×
$(2m_\pi, 2m_K)$	✓	✓	×	✓	✓	×
$(2m_K, 2.00\,\mathrm{GeV}/c^2)$	✓	✓	×	✓	✓	✓
$(2.00, 4.38)\,\mathrm{GeV}/c^2$	✓	✓	×	✓	×	×
$(4.38, 4.78)\,\mathrm{GeV}/c^2$	✓	×	×	✓	×	×

Table 10.3.: Analysis channels included in the ALP model-dependent combined fit per a mass region.

m_a	K^+	K^{*0}	e^+e^-	$\mu^+\mu^-$
$(0.025\,\mathrm{GeV}/c^2, 2m_\mu)$	✓	✓	✓	×
$(2m_\mu, 4.78\,\mathrm{GeV}/c^2)$	✓	✓	×	✓
$(4.38, 4.78)\,\mathrm{GeV}/c^2$	✓	×	×	✓

10.2.2. Treatment of systematic uncertainties

The maximum likelihood method allows the treatment of systematic uncertainties by implementing nuisance parameters (NPs) [165] in the fit model. Each systematic uncertainty is represented by a NP. Both the mean value of a NP and the size of its uncertainty are determined in control studies. The effect of the uncertainty is implemented by constraining the NP with a Gaussian prior. The width of the constraint term is set to the size of the systematic uncertainty. The NP is initially set to its mean value and can be varied by the fit. The constraint term increasingly penalises

the likelihood if the NP value deviates (being pulled) from the mean. This allows the fit to adjust a NP to better describe the underlying dataset.

The sources of systematic uncertainty relevant for this search manifest themselves as uncertainties on:

- the number of B-meson pairs $\rightarrow N_{BB}$,
- the signal efficiency $\rightarrow \epsilon_{\text{Sig.}}$,
- the signal shape parameters $\rightarrow \sigma$, α_l, α_r.

These quantities are present in the fit model as constant terms and are elevated to NPs. This is done by constructing the corresponding Gaussian constraint terms, including them into the likelihood, and letting the fit vary the parameters.

As an example, if there is a signal-like excess of candidates in data that is broader than the expected signal width σ, the fit can increase the NP within its constraint to better describe the excess. This is in accordance with the finite knowledge on the signal width determined from simulation.

Correlation between systematic uncertainties

The combined fit in multiple channels requires that certain sources of systematic uncertainty be treated as correlated. As the most prominent example, N_{BB} should acquire the same numerical value in all channels, since the individual samples are selected from the same dataset. If there was an independent N_{BB} NP for every channel, they could be pulled to different values or even in different directions from the mean. This is solved by replacing the individual N_{BB} parameters in Eq. 10.8 by a common NP that simultaneously affects all channels.

The uncertainty on

1. N_{BB} is fully correlated between all channels,
2. the ratio between the number of charged to neutral B-meson pairs $f^{\pm}/f^{00}$ is taken into account by adding it as a separate NP in the fit. This reflects the full anticorrelation between both production modes. An increase in the number

of charged B-meson pairs due to a positive pull of the $f^{\pm}/f^{00}$ NP results in a smaller number of neutral B-meson pairs.

3. $\epsilon_{\text{Sig.}}$ derived from the K^0_S efficiency correction are fully correlated between all channels,

4. the signal pdf parameters due to track momentum resolution variations are fully correlated between all channels,

5. the signal pdf parameters derived from the K^0_S control sample are fully correlated between all channels.

The uncertainties due to the limited number of simulated events and interpolation between simulated samples are uncorrelated. All remaining sources of uncertainty are treated as uncorrelated to reduce complexity of the fit model as their impact is small compared to the above mentioned cases.

10.3. Signal significance

The significance of a possible signal excess is determined using the numerical values of the likelihoods $\mathcal{L}$. The background-only and combined signal+background models are fitted separately to the distribution of candidates in the extraction window. The ratio of the maximised log-likelihood values quantifies by how much the signal+background hypothesis describes the data better than the background-only hypothesis. Since twice the likelihood ratio is expected to be distributed like χ^2 and the logarithm ratio can be written as difference, the significance in units of standard deviations can be calculated as [1]:

$$S = \sqrt{2\left(\log\hat{\mathcal{L}}_{\text{Sig.+Bkg.}} - \log\hat{\mathcal{L}}_{\text{Bkg.}}\right)}, \tag{10.10}$$

where the $\hat{}$ denotes maximised quantities.

10.4. Look-elsewhere effect

The look-elsewhere effect emerges when an experiment such as the signal extraction at a test mass is repeated for many different points. The probability that there is a significant outcome in one of the extractions due to a statistical fluctuation increases with the number of scan points. The p_0-value is related to the significance of an excess and describes the probability that, given the background-only hypothesis, an outcome as extreme as the one obtained occurs.

For a single extraction, it is labelled local p_0-value. It describes the outcome as being independent of all other extraction points. The look-elsewhere effect is taken into account by weighting the local p_0-value by a trials factor down to a larger (less significant) global p_0-value. The method to compute the trials factor used in this search was proposed by the authors of Ref. [166].

The mass scans at different lifetime hypotheses are treated as separate searches for the look-elsewhere effect. A potential signal with a specific lifetime in data would be found by extracting with lifetime hypotheses slightly above and below the real lifetime. The trials factor is determined for the mass scan with the lifetime hypothesis where the extracted signal significance is the largest. The 'global' in the weighted p_0-value therefore refers to the search at multiple S mass scan points for a given lifetime hypothesis.

The signal significance is repeatedly extracted from background pseudo-experiment datasets to represent statistical fluctuations. The values of the background pdf parameters $N_{\text{Bkg.}}(m_S)$ and $c_1(m_S)$ are obtained from background-only fits to background simulation in fit windows around m_S. The background pdf is constructed using the coefficient value $c_1(m_S)$. Candidates are drawn from the background model to form a background pseudo-experiment dataset. The number of candidates drawn is randomly chosen from a Poisson distribution around a mean value set to $N_{\text{Bkg.}}(m_S)$. The combined signal+background model is fitted to the pseudo-experiment dataset to estimate the signal yield $N_{\text{Sig.}}$. As is the case for the regular signal extraction fit, the background parameters are also directly estimated by the fit. This procedure is performed $n_{\text{Scans}} = 2000$ times per scan point to simulate the same number of mass

scans.

The trials factor is determined by counting the number of crossings of N_{Up} above a certain threshold significance S_{Test} in each scan and taking the average across scans:

$$< N_{\mathrm{Up}}(S_{\mathrm{Test}}) >= \frac{\sum_{\mathrm{Scans}} N_{\mathrm{Up}}(S_{\mathrm{Test}})}{n_{\mathrm{Scans}}}. \tag{10.11}$$

The correlation between neighbouring scan points due to overlapping extraction windows and close spacing is corrected for by counting a crossing above threshold only if the significance at the previous mass point was below S_{Test}.

A pseudo-experiment mass scan and the distribution of the number of crossings above a threshold significance of 2.5σ across all scans are shown in Fig. 10.2. A Gaussian pdf is fitted to the distribution to extract the mean and width. This results in around ten crossings above the threshold that are expected in a typical mass scan.

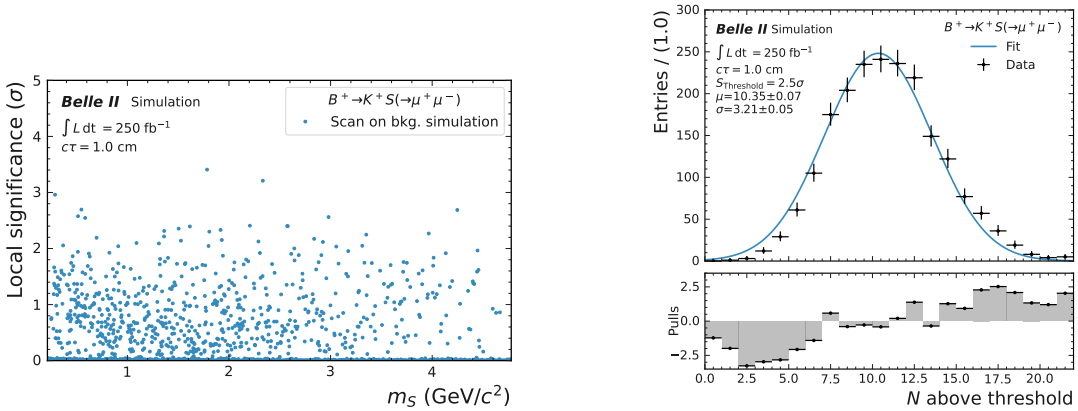

Figure 10.2.: (Left) distribution of the local signal significances in a pseudo-experiment scan in $B^+ \to K^+S(\to \mu^+\mu^-)$. (Right) number of crossings above $S_{\mathrm{Test}} = 2.5\sigma$ for 2000 pseudo-experiment mass scans in $B^+ \to K^+S(\to \mu^+\mu^-)$ with a fit of a Gaussian pdf.

The global p_0-value is then determined by:

$$p_0\text{-value}^{\text{Global}}(S_{\text{Test}}) = p_0\text{-value}^{\text{Local}} + < N_{\text{Up}}(S_{\text{Test}}) > \exp\left(\frac{S^2_{\text{Test}} - S^2_{\text{Local}}}{2}\right). \tag{10.12}$$

A statistical uncertainty due to the finite number of pseudo-experiment scans is propagated by computing the standard deviation of the number of crossings above threshold and recomputing the global p_0-value with $< N_{\text{Up}} > \pm\sigma(N_{\text{Up}})$.

The dependence of the global significance on the precise threshold value is studied by varying S_{Test} within reasonable values and is found to be small. By community standards, a global signal significance of 5σ or greater is required for the discovery of a process. The global significance of a local signal excess with a 5σ significance using different threshold significances is shown in Fig. 10.3. The statistical uncertainty increases for larger S_{Test} due to the smaller number of crossings above the threshold and therefore a less precise mean value. The upper uncertainty on the global significance for the highest threshold value would exceed the threshold value S_{Test}, where the prescription is no longer valid.

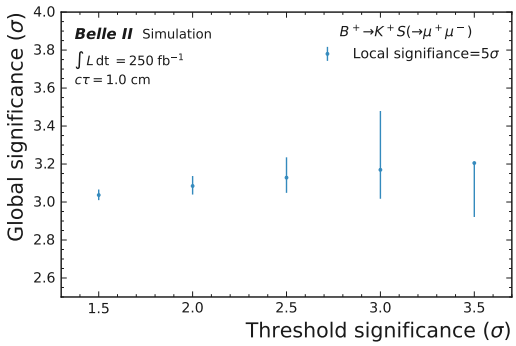

Figure 10.3.: The global significance for a signal excess with a local significance of 5σ computed with different threshold values.

A threshold value of $S_{\text{Test}} = 2.5$ is used for the look-elsewhere effect treatment in the search in experimental data by directly counting the number of points above the

threshold in data. Hypothetical large local signal significance translate to:

$$\begin{aligned} 5\sigma\,(\text{Local}) &\rightarrow 3.13^{+0.11}_{-0.08}\sigma\,(\text{Global}), \\ 4\sigma\,(\text{Local}) &\rightarrow 1.41^{+0.19}_{-0.15}\sigma\,(\text{Global}). \end{aligned} \tag{10.13}$$

A local significance of 5σ corresponds to a global signal significance above 3σ, which would be evidence for the signal process. A smaller local significance of 4σ corresponds to a global significance of 1.4σ, compatible with a statistical fluctuation.

10.5. Upper limits

Upper limits are set on the $\mathcal{B}$ parameter using the modified frequentist CL_S method [167]. The asymptotic approximation [168] is used to reduce extensive CPU time for the limit setting. The confidence level of the limits is set to the standard value of the community of $\alpha = 0.05$. Limit setting is performed with the HEPSTATS [169] framework, which is part of the SCIKIT-HEP project [170]. HEPSTATS interfaces with the fit models and results of ZFIT.

10.6. Fit validation

The fitting procedure used to extract the number of signal candidates is validated using pseudo-experiments with datasets sampled from the signal and background models.

10.6.1. Bias and coverage

The fidelity of the signal yield estimator is studied by determining its bias and the coverage of its uncertainty. A potential bias in the extraction method would manifest itself as a statistically significant deviation in the number of signal candidates found by the fit to the true number. The uncertainty on the signal yield determined by the fit should also correctly cover statistical fluctuations in the dataset. An overcoverage is identified as a significant deviation of the extracted signal pull width to values below unity; an undercoverage corresponds to a value above unity.

Injected signal

The fitting procedure is studied with pseudo-experiment datasets that include signal candidates injected with different signal strengths. The fit is supposed to recover the number of injected signal candidates on a statistical average. The background pseudo-experiment datasets are drawn according to the same method as for the look-elsewhere effect. The signal pdf is constructed at a given m_S with fixed parameter values from the interpolation procedure, similar to the signal extraction fit. Signal strengths of approximately two, five, and eight standard deviations are injected into the background pseudo-experiment dataset to test the procedure for different significance levels. The number of signal candidates to be injected to reach the approximate signal strength depends on the background level at given m_S and is calculated using:

$$\begin{aligned} S &= \frac{N_{\text{Sig.}}(m_S)}{\sqrt{N_{\text{Sig.}}(m_S) + N^{\pm 2\sigma}_{\text{Bkg.}}(m_S)}}, \\ N_{\text{Sig.}}(m_S) &= \frac{1}{2}\left(S^2 + \sqrt{S^4 + 4N^{\pm 2\sigma}_{\text{Bkg.}}(m_S)S^2}\right), \end{aligned} \tag{10.14}$$

where $N^{\pm 2\sigma}_{\text{Bkg.}}(m_S)$ refers to the yield in background-only simulation in a small window of two signal widths around the injected signal mass. This is used because the significance of an excess depends on the background yield actually below the signal peak and not on the background yield in the full fit window that is employed for the data-driven background estimation. The signal component of the pseudo-experiment dataset is constructed by drawing candidates from the signal pdf. The number of candidates for each pseudo-experiment is randomly chosen from an independent Poisson distribution with mean $N_{\text{Sig.}}(m_S)$. Both the signal and background components are combined, and the signal+background model is fitted to it. This is repeated $n = 2000$ times to represent statistical fluctuations in both background and signal. Signal+background fits to pseudo-experiment data with injected signal are shown in Fig. 10.4 for the one-dimensional extraction in separate analysis channels.

For the validation of the combined fit, background candidates are drawn individually in each analysis channel using the same method. The combined fit is set up to find

a signal that follows the predicted splitting into the analysis channels. The mean number of signal candidates to be injected into the channel with the highest predicted signal fraction is determined using Eq. 10.14. The predictions from the BSM model are then used to calculate the number of signal candidates to be injected into the remaining channels. A combined fit using the dark scalar model with injected signal is shown in Fig. 10.5. The figure shows a fit to a dataset with a dark scalar signal injected at a mass of $1.5\,\mathrm{GeV}/c^2$. Here, the predicted decay rate of the dark scalar into $S \to K^+K^-$ is dominant. Therefore, the largest number of signal candidates is injected into the $S \to K^+K^-$ channels and subsequently found by the fit. A slightly larger number of signal candidates is injected and found in $B^+ \to K^+S$ compared to $B^0 \to K^{*0}S$. The difference between the dark scalar production rates in both channels is small at this dark scalar mass. However, the K^{*0} is reconstructed only in its charged decay channel, leading to a larger signal fraction in $B^+ \to K^+S$.

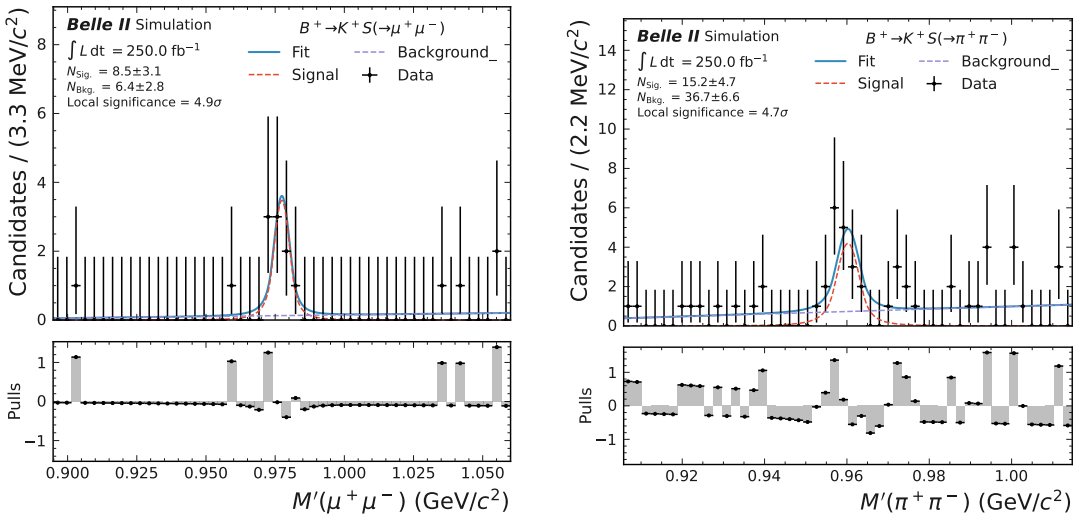

Figure 10.4.: Distributions of pseudo-experiment datasets with injected signal candidates for (left) $B^+ \to K^+S(\to \mu^+\mu^-)$ and (right) $B^+ \to K^+S(\to \pi^+\pi^-)$ with a fit of the combined signal+background model.

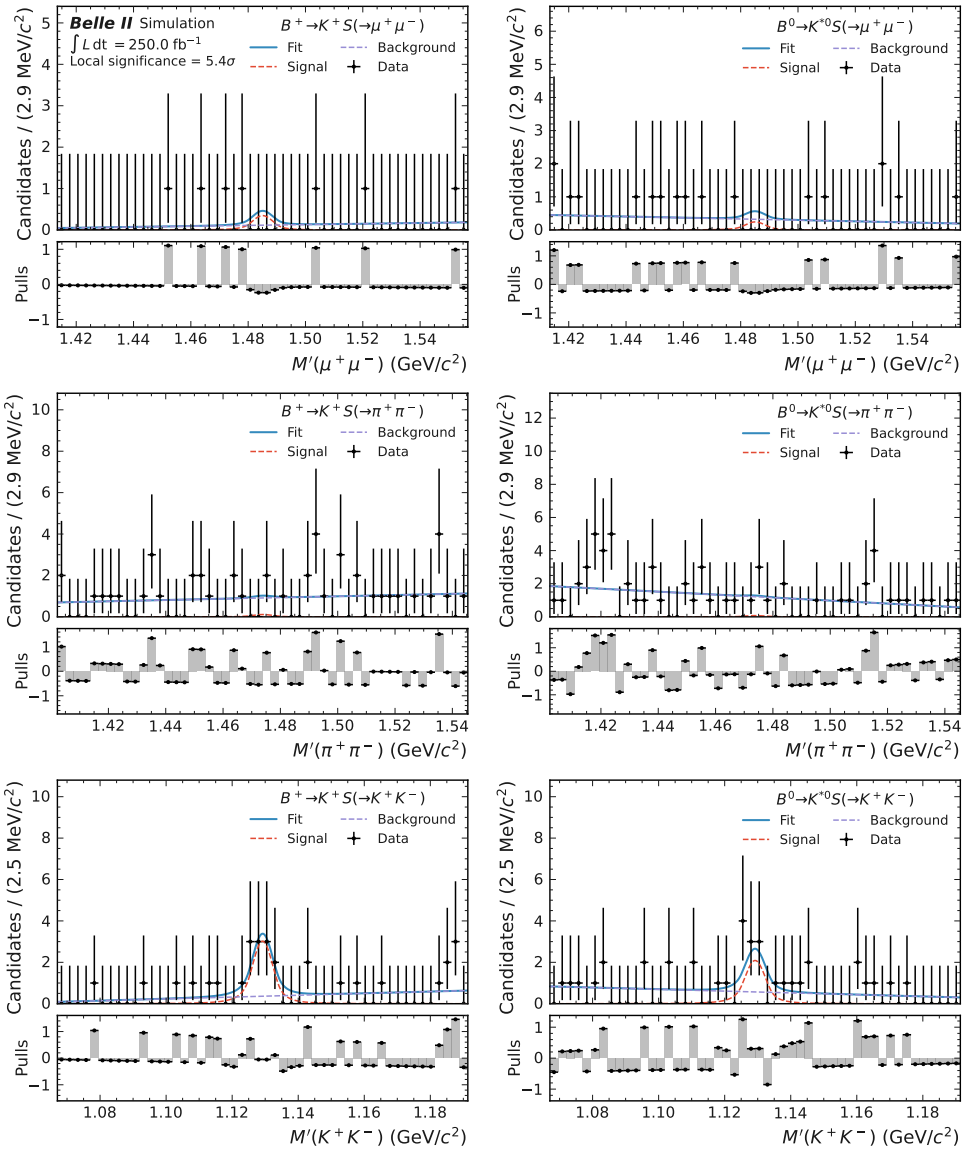

Figure 10.5.: Distributions of pseudo-experiment datasets with injected signal candidates for the combined fit using the dark scalar hypothesis with a fit of the combined signal+background model.

Fit bias

The bias test is performed by studying the difference between the extracted (fit) and the true (injected) number of signal candidates. A relative difference is studied to not depend on the absolute number of candidates:

$$\frac{\Delta(N_{\text{Sig.}})}{N_{\text{Sig.}}}(m_S) = \frac{N_{\text{Sig.}}^{\text{fit}}(m_S) - < N_{\text{Sig.}}^{\text{injected}}(m_S) >}{< N_{\text{Sig.}}^{\text{injected}}(m_S) >}, \tag{10.15}$$

where the angle brackets denote the mean of the Poisson distribution from which the number of injected signal candidates is drawn. A Gaussian pdf is fitted to the distribution of $(\Delta(N_{\text{Sig.}})/N_{\text{Sig.}})(m_S)$ and the mean is extracted to check for consistency with zero. Bias distributions for single-channel and combined fits are shown in Fig. 10.6.

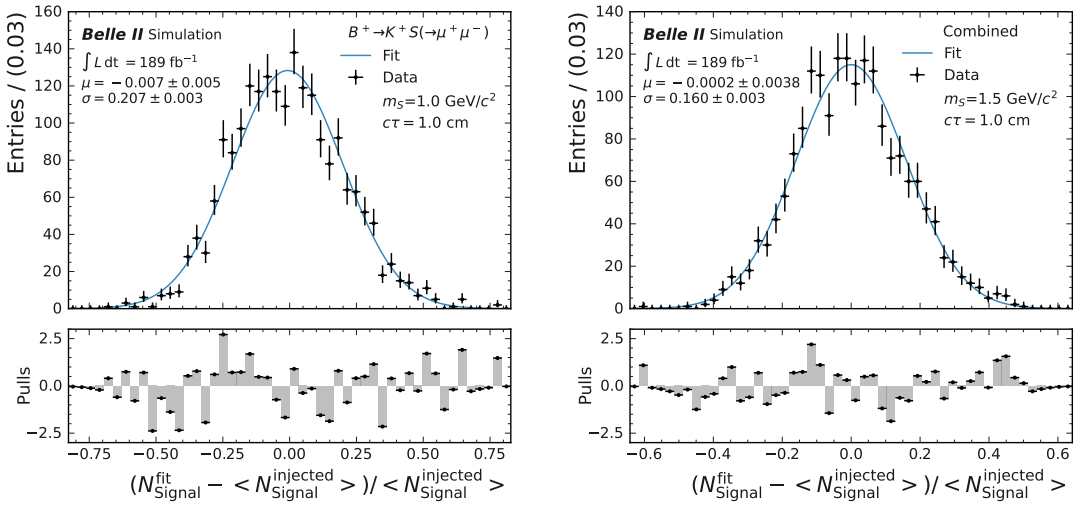

Figure 10.6.: Distributions of the bias on the signal yield estimator with fit of a Gaussian pdf for (left) $B^+ \to K^+S(\to \mu^+\mu^-)$ with injected signal at $m_S = 1\,\text{GeV}/c^2$ and (right) for the combined fit using the dark scalar hypothesis with signal candidates injected at $m_S = 1.5\,\text{GeV}/c^2$, both corresponding to an approximate injected signal strength of 5σ.

The bias is studied for different injected signal masses, lifetimes, signal strengths, and analysis channels. The results for an approximate injected signal strength of 5σ

are shown in Fig. 10.7 for $B^+ \to K^+S(\to \mu^+\mu^-)$ and the combined fit using the dark scalar hypothesis. No significant bias of the signal yield estimator is found for both single-channel and combined fits for most of the tested parameter space. Small negative biases are found close to the kinematic endpoints, with a larger bias in the combined fit.

The kinematic endpoints of the reconstructed S mass distributions are challenging for the fit method. It relies on the large fit window size to both sides of the tested signal hypothesis to determine the background directly in data. Close to the endpoints, there is only one side available, leading to a reduced power to correctly constrain the background model. A similar effect is observed with negative background predictions in a part of the fit window when the linear coefficient is allowed outside the $[-1, 1]$ interval, which happens with a higher probability in these regions. This is due to the negligible amount of background expected at the endpoints. Therefore, the coefficient is constrained between $[-1, 1]$, producing a non-negative background yield in all parts of the fit window. There is a close-to-zero true background contamination at kinematic endpoints. Two choices exist for a general smooth linear background model that is constrained by background candidates on the other side of the fit window. The first choice is an overestimation of the signal yield at the endpoint, which happens for a negative background prediction below the mass point tested. The second one is an underestimation with the chosen non-negative background model below the test point, resulting in a negative bias. The latter is chosen as a more conservative approach, and the resulting bias is covered by an additional systematic uncertainty.

The effect is larger for the combined fit, as each channel individually features a small negative bias that propagates to the summed combined signal yield. This is visible with the negative bias being larger at small S masses where $S \to \mu^+\mu^-$ and $S \to \pi^+\pi^-$ contribute, compared to the high mass end of the spectrum where only $S \to \mu^+\mu^-$ is considered. Furthermore, a small negative bias is also visible around the kinematic threshold of $S \to K^+K^-$, where the same effect occurs, and the branching fraction in this channel is large. The bias in the combined fit using the ALP model is expected to be smaller or equal to the dark scalar model due the number of active channels.

A small bias of up to two percent is found across the spectrum away from the endpoints for small injected signal strengths. This could be due to the non-negative constraint on the signal yield parameter. The size of the bias is reduced towards larger injected signal strengths. Since the local significance of 2σ is far from the level of discovery of 5σ, the bias for the low injected signal strength is considered tolerable.

The same observations are found for the other channels. A flat additional systematic uncertainty on the signal efficiency of 3% (4%) is considered for all fits in the single-channel (combined) signal extractions to cover for the bias on the signal yield estimator.

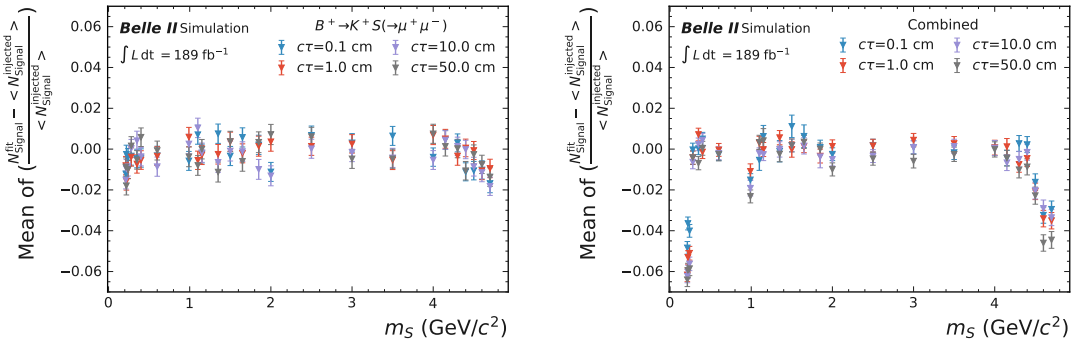

Figure 10.7.: The determined bias on the signal yield estimator across injected masses and lifetimes in (left) $B^+ \to K^+S(\to \mu^+\mu^-)$ and (right) the combined fit using the dark scalar hypothesis with an approximate injected signal strengths of 5σ.

Fit uncertainty coverage

The coverage is studied using the distribution of pulls. The pulls are computed by weighting the difference of the determined number to the true number of signal candidates by the uncertainty estimated by the fit Δ:

$$\text{Pull}_{\text{Sig.+Bkg}} = \frac{N^{\text{fit}}_{\text{Sig.}}(m_S) - < N^{\text{injected}}_{\text{Sig.}}(m_S) >}{\Delta(N^{\text{fit}}_{\text{Sig.}}(m_S))}. \tag{10.16}$$

A Gaussian pdf is fitted to the distribution to determine the width of the pull distribution. Pull distributions for $B^+ \to K^+S(\to \mu^+\mu^-)$ and the combined fit using the dark scalar hypothesis for an approximate injected signal strength of 5σ are shown in Fig. 10.8.

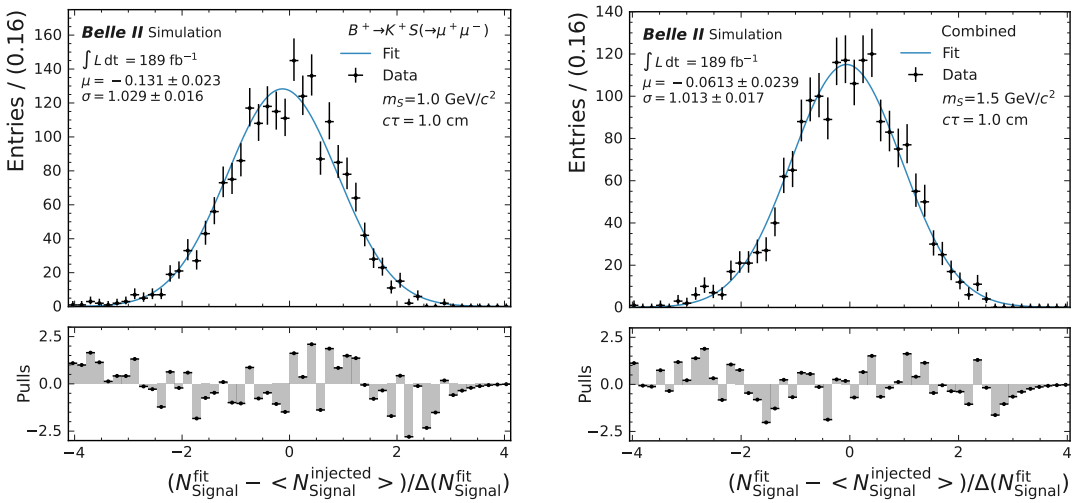

Figure 10.8.: Distributions of the coverage of the uncertainty on the signal yield estimator with fit of a Gaussian pdf for (left) $B^+ \to K^+S(\to \mu^+\mu^-)$ with injected signal at $m_S = 1\,\mathrm{GeV}/c^2$ and (right) for the combined fit using the dark scalar hypothesis with signal candidates injected at $m_S = 1.5\,\mathrm{GeV}/c^2$, both corresponding to an approximate injected signal strength of 5σ.

The corresponding widths of the pull distributions across signal masses and lifetimes are shown in Fig. 10.9. The widths are mostly compatible with unity within the statistical uncertainty estimated by the Gaussian fit. This confirms good coverage of the statistical fluctuations induced by repeatedly sampling a Poisson number of signal candidates. There is a trend to a small undercoverage of the uncertainty on the signal yield estimator for a large part of the mass spectrum indicated by the central values of the widths above unity. The combined fit shows a larger undercoverage for the kinematic endpoints, presumably related to the same effects observed that lead to the negative bias in these regions. Coverage improves for larger injected signal strengths

similar to the bias. The systematic uncertainty that was introduced before due to the fit procedure is deemed sufficient to cover these effects. It should be noted that the uncertainty of the signal yield estimator is approximated by the HESSE method using the second-derivative matrix at the minimum. This method can yield non-optimal values for likelihoods that do not follow a perfect quadratic behaviour in the parameter of interest. It is not used at any point in the derivation of upper limits or signal significances.

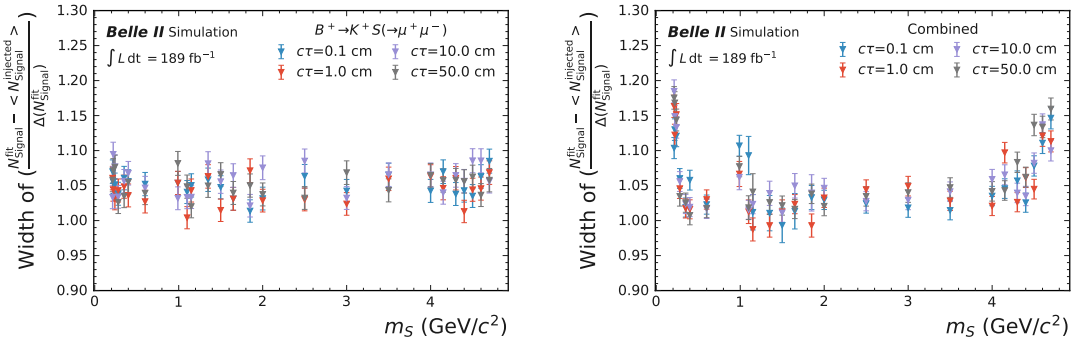

Figure 10.9.: The determined coverage of the uncertainty on the signal yield estimator across injected masses and lifetimes in (left) $B^+ \to K^+S(\to \mu^+\mu^-)$ and (right) the combined fit using the dark scalar hypothesis with an approximate injected signal strength of 5σ.

10.6.2. Scan point spacing

The spacing between scan points is set to half the signal width σ. A signal is injected with a specific m_S hypothesis and extracted at all scan points in a range corresponding to $m_S \pm 4\sigma$. This procedure is repeated 2000 times to determine the mean and standard deviation of the local signal significances extracted at the extraction points. The results for $B^+ \to K^+S(\to \mu^+\mu^-)$ and $B^+ \to K^+S(\to \pi^+\pi^-)$ with signal injected at $m_S = 1\,\text{GeV}/c^2$ are shown in Fig. 10.10. Extracting exactly at the injected mass yields the largest local significance. The mean local significance is already noticeably reduced when extracting one full signal width away from the injection point. The local significance decreases further with the distance to the injected mass and is on

average compatible with zero at a distance of around $3.5-4\sigma$. A spacing of $\sigma/2$ leads to a maximum distance to a potential true signal of $\sigma/4$, where the mean extracted local significance is close to its maximum.

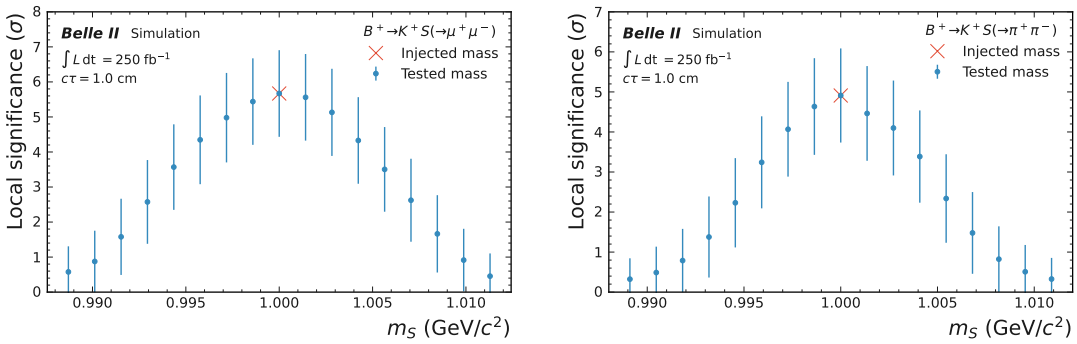

Figure 10.10.: Mean extracted significance and standard deviation across 2000 pseudo-experiment signal extractions at and around the injected signal in (left) $B^+ \to K^+S(\to \mu^+\mu^-)$ and (right) $B^+ \to K^+S(\to \pi^+\pi^-)$.

10.6.3. Fit window size

The nominal fit window size is set to $\pm 20\sigma$ around the tested mass point and is enlarged in the case of low statistics. The impact of variations in the size of the fit windows on the upper limits on the signal branching fraction is shown in Fig. 10.11. The expected CL_S upper limits are determined with the nominal window size, a smaller size of $\pm 10\sigma$ and a larger size of $\pm 30\sigma$. The difference between the nominal and varied fit window sizes are normalised to the nominal upper limit. The sizes of the one- and two-standard deviation bands of the expected nominal CL_S upper limit are normalised as well. The limits are found to be robust against variations of the fit window size. Small differences exist, but are negligible compared to the size of expected statistical fluctuations.

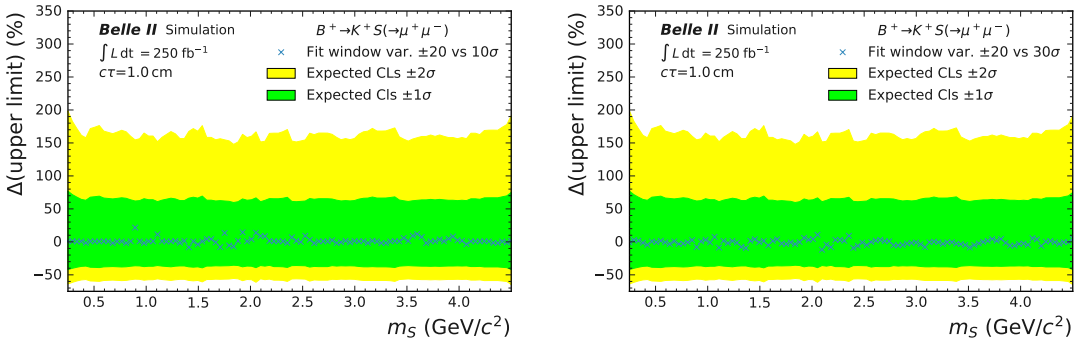

Figure 10.11.: Normalised difference between the expected CL_S upper limits on the signal branching fraction in $B^+ \to K^+S(\to \mu^+\mu^-)$ with the nominal fit window size and (left) a smaller and (right) larger window size. The one and two standard deviation bands of the expected upper limit are shown for the nominal fit window size.

10.6.4. Systematic uncertainties

The systematic uncertainties that are considered in this search affect the signal efficiency and signal shape. The impact of the systematic uncertainties on the results is shown in Fig. 10.12.

The profiles of the negative log-likelihoods with and without systematic uncertainties in $B^+ \to K^+S(\to \mu^+\mu^-)$ are shown as a function of the signal branching fraction estimator at a test mass and lifetime. The values are shifted with respect to the best fit so that the minimum of the profile is at zero. Systematic uncertainties are switched off by setting the corresponding NPs to their nominal values and constant in the fit. The effect of systematic uncertainties is a broadening of the likelihood profile, corresponding to a reduced power to constrain the value of the branching fraction.

The effect on the upper limit on the signal branching fraction is shown as a normalised difference between the expected CL_S upper limit with and without systematic uncertainties as a function of the S mass and for different lifetime hypotheses. The upper limit including systematic uncertainties is larger and hence weaker than the one without uncertainties. The trend of the difference follows the sizes of systematic uncertainties detailed in Chap. 9. The effect is strongest for small masses and large

lifetimes where the LLP correction factors and associated uncertainties are large due to the large average S vertex distance. The systematic uncertainties weaken the upper limit by typically less than 0.5% and up to around 2%. This is small compared to the size of expected statistical fluctuations described by the bands around the expected CL_S upper limit shown before in Fig. 10.11.

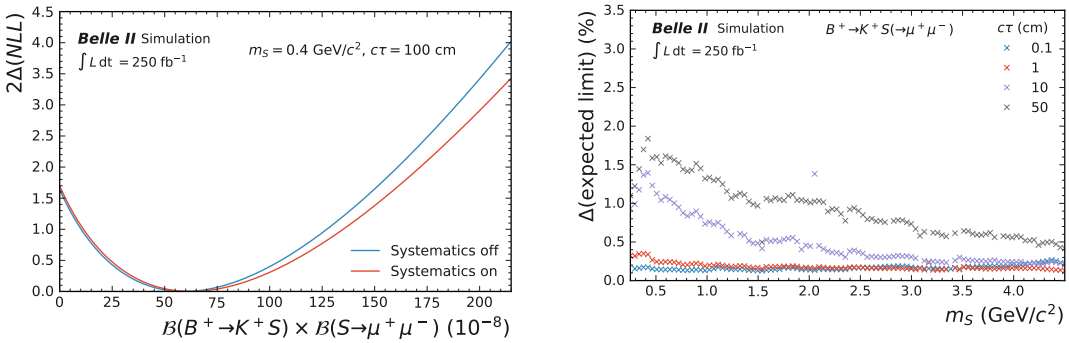

Figure 10.12.: (Left) profile of the negative log-likelihood as a function of the signal branching fraction with and without active NP. (Right) normalised difference between the expected upper limits with and without active NP as a function of tested S mass for different lifetime hypotheses.

10.6.5. Counting experiment

The CL_S upper limits are compared with an alternative method to derive limits using a simple Poisson counting experiment, both using a confidence level of 95%. The counting experiment assumes a zero background, which corresponds to the conservative assumption that each candidate is due to the signal process. The upper limits of the counting experiment are determined with the number of candidates found in a window of $\pm 3\sigma$ around the test mass, and using Table 40.4 from the statistics part of Ref. [1]. The much smaller window compared to the fit is not used to constrain the background, but to count signal candidates within the expected signal peak.

The expected CL_S upper limits together with the one- and two-standard deviation bands are shown with the alternative limits derived from the counting experiment for

$B^+ \to K^+S(\to \mu^+\mu^-)$ and $B^+ \to K^+S(\to \pi^+\pi^-)$ in Fig. 10.13. The upper limits of the counting experiment are weaker than the CL_S upper limits as expected by the conservative construction and compatible with the latter within their bands. The difference between the upper limits of the CL_S and counting experiment is greater in $B^+ \to K^+S(\to \pi^+\pi^-)$. This is due to the larger background contamination in this channel, leading to the zero background assumption to be more conservative.

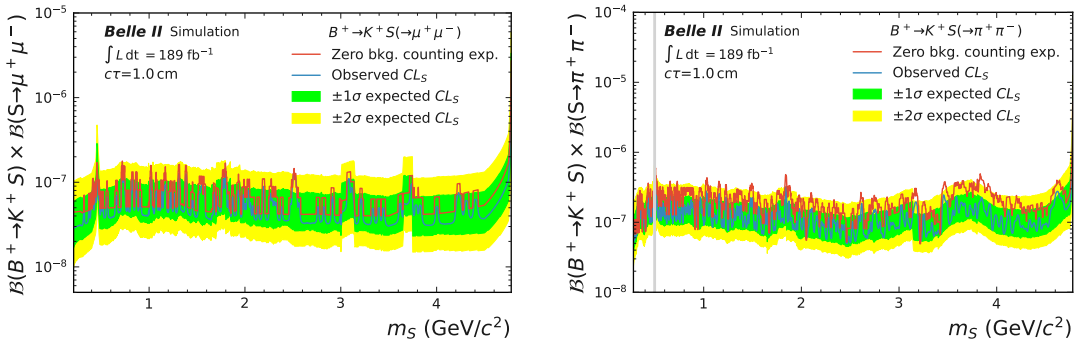

Figure 10.13.: Expected CL_S limits with the one and two standard deviation bands and limits from the counting experiment detailed in the text for (left) $B^+ \to K^+S(\to \mu^+\mu^-)$ and (right) $B^+ \to K^+S(\to \pi^+\pi^-)$.

11. Results

This chapter presents the results of the search for the S signal and describes the translation of the results into bounds on the benchmark models that predict the signal process.

11.1. Reconstructed mass distribution

The S signal is expected to yield a peaking excess of candidates above a smooth background in the reconstructed $M'(x^+x^-)$ distribution. The $M'(x^+x^-)$ distributions in the signal regions of the experimental data and simulation are shown for the $B^+ \to K^+S$ channels in Fig. 11.1 and for the $B^0 \to K^{*0}(\to K^+\pi^-)S$ channels in Fig. 11.2. The signal regions are defined by the selection requirements summarised in Sec. 5.6. The agreement between data and simulation is good across all channels, as expected from the validation of the background modelling in the sideband regions detailed in Chap. 8. Small differences exist in the form of as a general trend of fewer candidates observed in data than expected by simulation. This difference is more pronounced in the $B^+ \to K^+S$ production mode and manifests itself as a normalisation discrepancy. The shape of the $M'(x^+x^-)$ distribution in the signal region is very well modelled across all channels. No obvious excess that could hint at the signal process is visible in the $M'(x^+x^-)$ distributions. It should be noted that the bin size of the histograms is much larger than the width of the predicted signal distributions. The largest remaining source of background is combinatorial and originates from $e^+e^- \to q\overline{q}$ events, where $q \in (u, d, s, c)$.

A normalisation difference for the light-quark-pair processes was also found in the sideband studies, see Sec.8.1, and hints towards a mismodelling of the underlying

processes in simulation. Part of the reason for the mismodelling can be found in the PYTHIA step within the event generation. The Belle II simulation currently uses default values for the PYTHIA internal parameters. The collaboration will carry out dedicated studies to tune these parameters for Belle II conditions, which could help reduce the discrepancy in a future search. The dataset collected at an energy slightly below the $\Upsilon(4S)$ resonance can be used. It allows the study of light-quark-pair production with kinematics close to the $\Upsilon(4S)$ dataset without the $B\overline{B}$ contribution.

However, the difference in normalisation poses no problem for the signal extraction performed in this work, as the background contribution is determined directly from the distribution in data.

11.2. Signal extraction

The signal yield and hence the branching fraction for an S signal at m_S are estimated by fitting the signal+background pdf templates to the M' distribution in windows around the mass hypothesis. The calculation of the branching fraction is detailed in Sec. 10.2. The S mass is scanned separately for each analysis channel and for different lifetime hypotheses. Further combined fits using the dark scalar and ALP hypothesis are performed. The combined fit procedure is introduced in Sec. 10.2.1.

The resulting local signal significances are shown for single channel fits as a function of the scanned S mass for a lifetime hypothesis of $c\tau_S = 1\,\text{cm}$ in Fig. 11.3 or the $B^+ \to K^+S$ channels and in Fig. 11.4 for the $B^0 \to K^{*0}(\to K^+\pi^-)S$ channels. The significance scans are consistent with expectations of statistical fluctuations in the background. These are shown in the discussion of the look-elsewhere effect in Sec. 10.4. The local significances are very small for most of the parameter space tested, with small peaks building up around local fluctuations. This is expected from the study of the local significance in the vicinity of a signal-like excess detailed in Sec. 10.6.2.

The lifetime has a small effect on the signal pdf parameters and, thereby, on the fit window sizes. The difference in signal efficiency due to the S lifetime has no impact on the estimated number of signal candidates and local significances. However, it affects

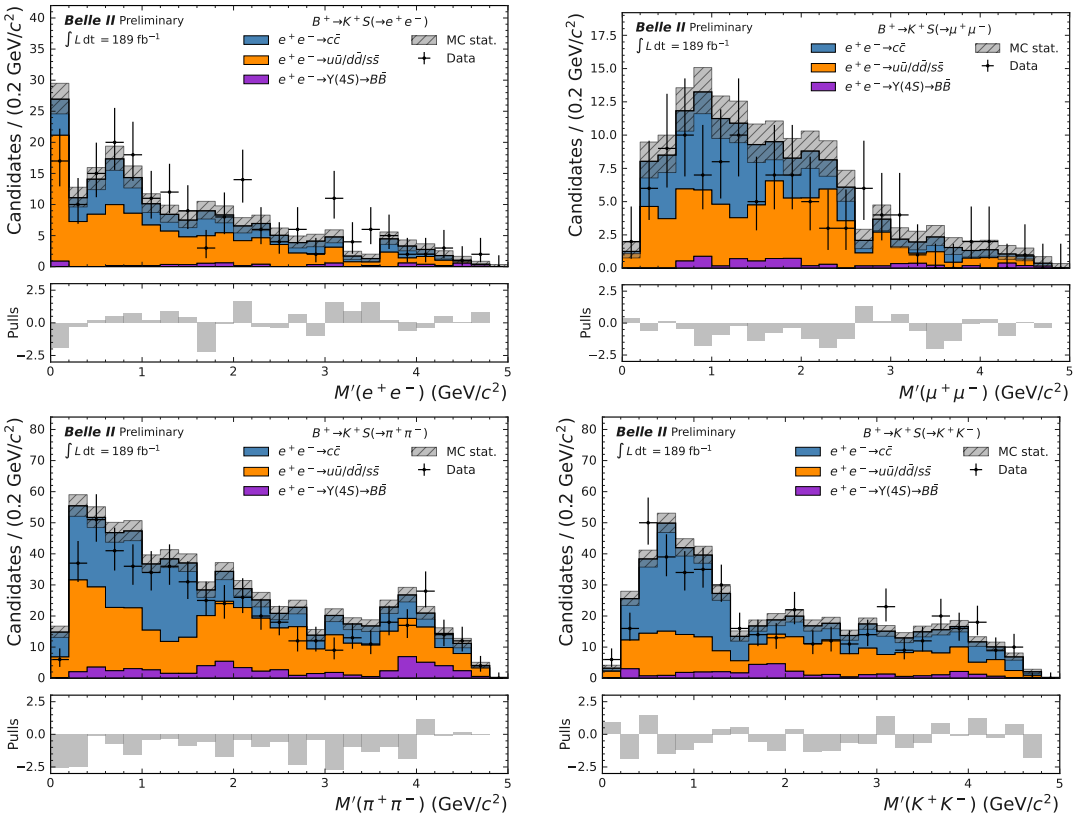

Figure 11.1.: Distributions of $M'(x^+x^-)$ in the signal region in data and simulation for $B^+ \to K^+S$ for (upper left) $S \to e^+e^-$, (upper right) $S \to \mu^+\mu^-$, (lower left) $S \to \pi^+\pi^-$ and (lower right) $S \to K^+K^-$.

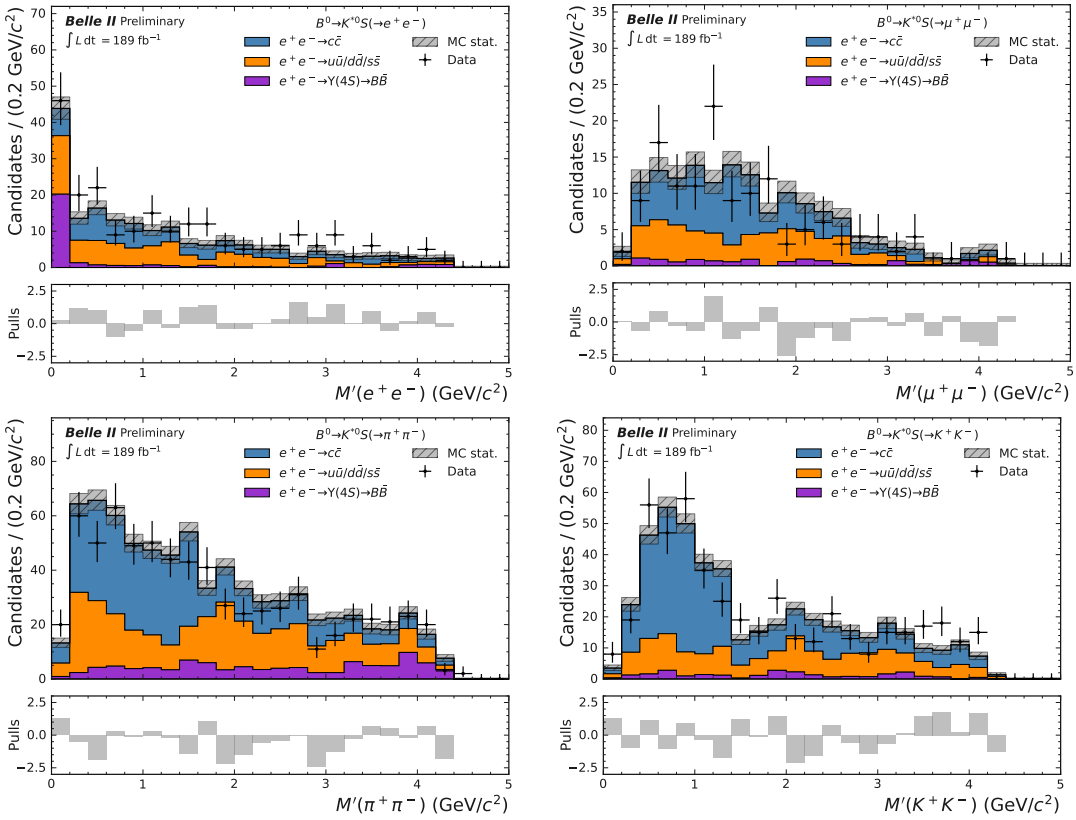

Figure 11.2.: Distributions of $M'(x^+x^-)$ in the signal region in data and simulation for $B^0 \to K^{*0}(\to K^+\pi^-)S$ for (upper left) $S \to e^+e^-$, (upper right) $S \to \mu^+\mu^-$, (lower left) $S \to \pi^+\pi^-$ and (lower right) $S \to K^+K^-$.

the determined branching fraction values. The differences in local signal significance between lifetime hypothesis do not exceed 0.7σ.

The highest local signal significance for all single channel fits is 3.6σ. It is found near $m_S = 1.061\,\text{GeV}/c^2$ for $B^+ \to K^+S(\to \pi^+\pi^-)$ with a lifetime hypothesis of $c\tau_S = 0.05\,\text{cm}$. For the model-dependent combined fits, the highest local signal significance is 3.3σ. It is found near $m_S = 2.619\,\text{GeV}/c^2$ for a lifetime hypothesis of $c\tau_S = 100\,\text{cm}$. The choice of the benchmark model for the combined fit has a negligible effect on the significance of the signal at this mass point. This is the case because only the $S \to \mu^+\mu^-$ channels contribute to the combined fits in both models. Furthermore, similar values for the rate difference between both production modes are predicted. Taking into account the look-elsewhere effect, described in Sec. 10.4, this corresponds to a global significance of 1.0σ for the single-channel fit and 0.3σ for the combined fit. No evidence for the signal process is found, as both largest global significances agree well with statistical fluctuations.

The fits with the highest local significances are shown for model-independent extractions in the $B^+ \to K^+S$ channels in Fig. 11.5, in the $B^0 \to K^{*0}(\to K^+\pi^-)S$ channels in Fig. 11.6 and for the combined fit using the dark scalar hypothesis in Fig. 11.7.

11.3. Model-independent bounds

Model-independent upper limits are set using the CL_S method at the 95% confidence level on the product of signal branching fractions $\mathcal{B}(B \to KS) \times \mathcal{B}(S \to x^+x^-)$ as a function of the mass S separately for all analysis channels and considered lifetime hypotheses.

The upper limits as a function of S mass in $B^+ \to K^+S(\to \mu^+\mu^-)$ are shown for two different lifetime hypotheses, $c\tau_S = 0.25\,\text{cm}$ and $c\tau_S = 25\,\text{cm}$ in Fig. 11.8.

Depending on the remaining background levels and signal efficiencies, the observed limits on the signal branching fraction vary between $\mathcal{O}(10^{-7} - 10^{-6})$ for these lifetime values. The observed limits are all well within the bands around the expected limits signalling the expected size of fluctuations in the sensitivity. No limit is observed to

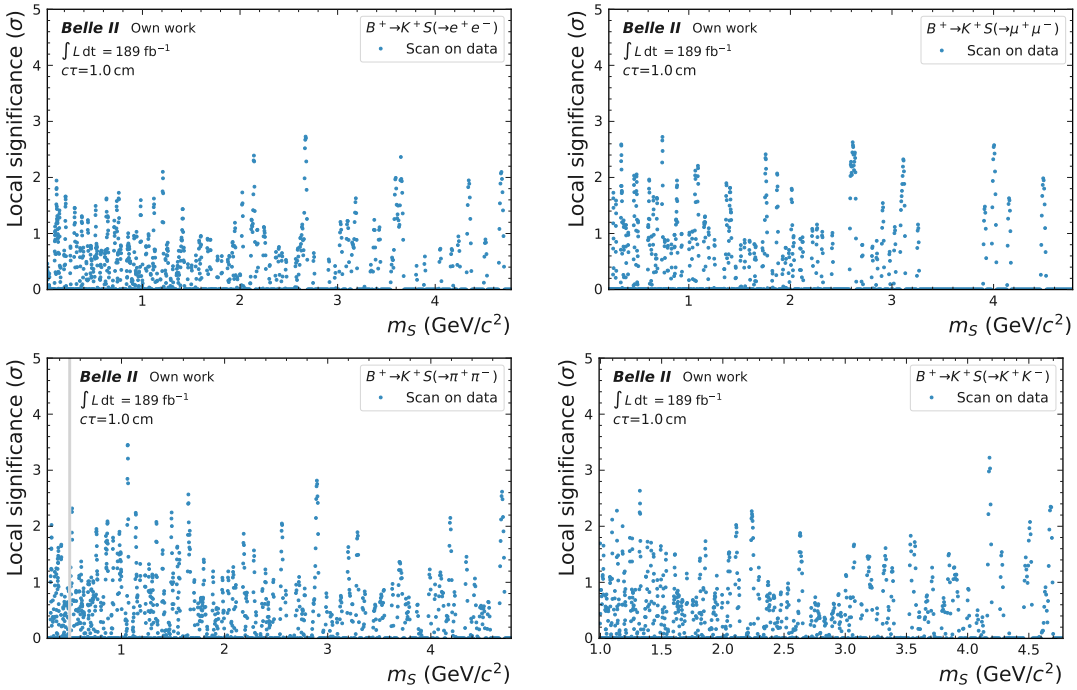

Figure 11.3.: Local signal significances from fits to data in $B^+ \to K^+S$ for (upper left) $S \to e^+e^-$, (upper right) $S \to \mu^+\mu^-$, (lower left) $S \to \pi^+\pi^-$ and (lower right) $S \to K^+K^-$ for a lifetime of $c\tau_S = 1\,\text{cm}$. The region corresponding to the fully-vetoed K^0_S for $S \to \pi^+\pi^-$ is marked in gray.

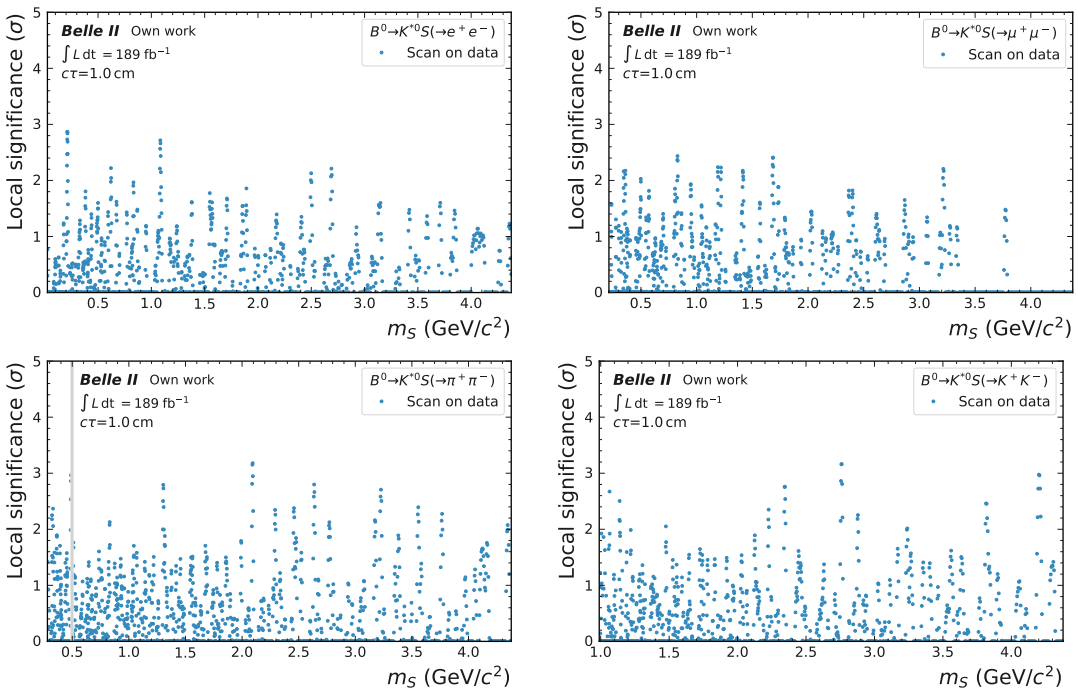

Figure 11.4.: Local signal significances from fits to data in $B^0 \to K^{*0}(\to K^+\pi^-)S$ for (upper left) $S \to e^+e^-$, (upper right) $S \to \mu^+\mu^-$, (lower left) $S \to \pi^+\pi^-$ and (lower right) $S \to K^+K^-$ for a lifetime of $c\tau_S = 1\,\text{cm}$. The region corresponding to the fully-vetoed K^0_S for $S \to \pi^+\pi^-$ is marked in gray.

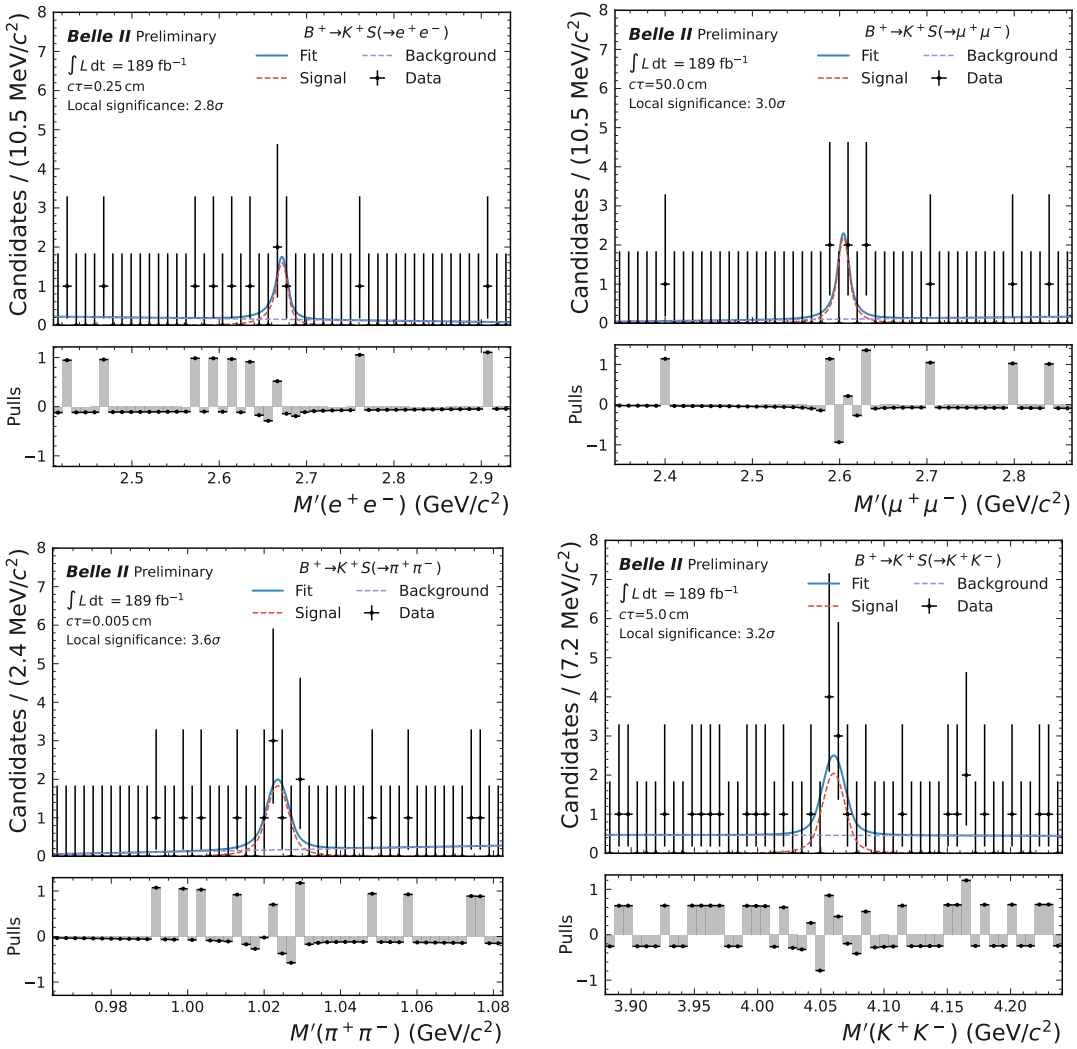

Figure 11.5.: Fits to data with largest local signal significances in $B^+ \to K^+S$ for (upper left) $S \to e^+e^-$, (upper right) $S \to \mu^+\mu^-$, (lower left) $S \to \pi^+\pi^-$ and (lower right) $S \to K^+K^-$

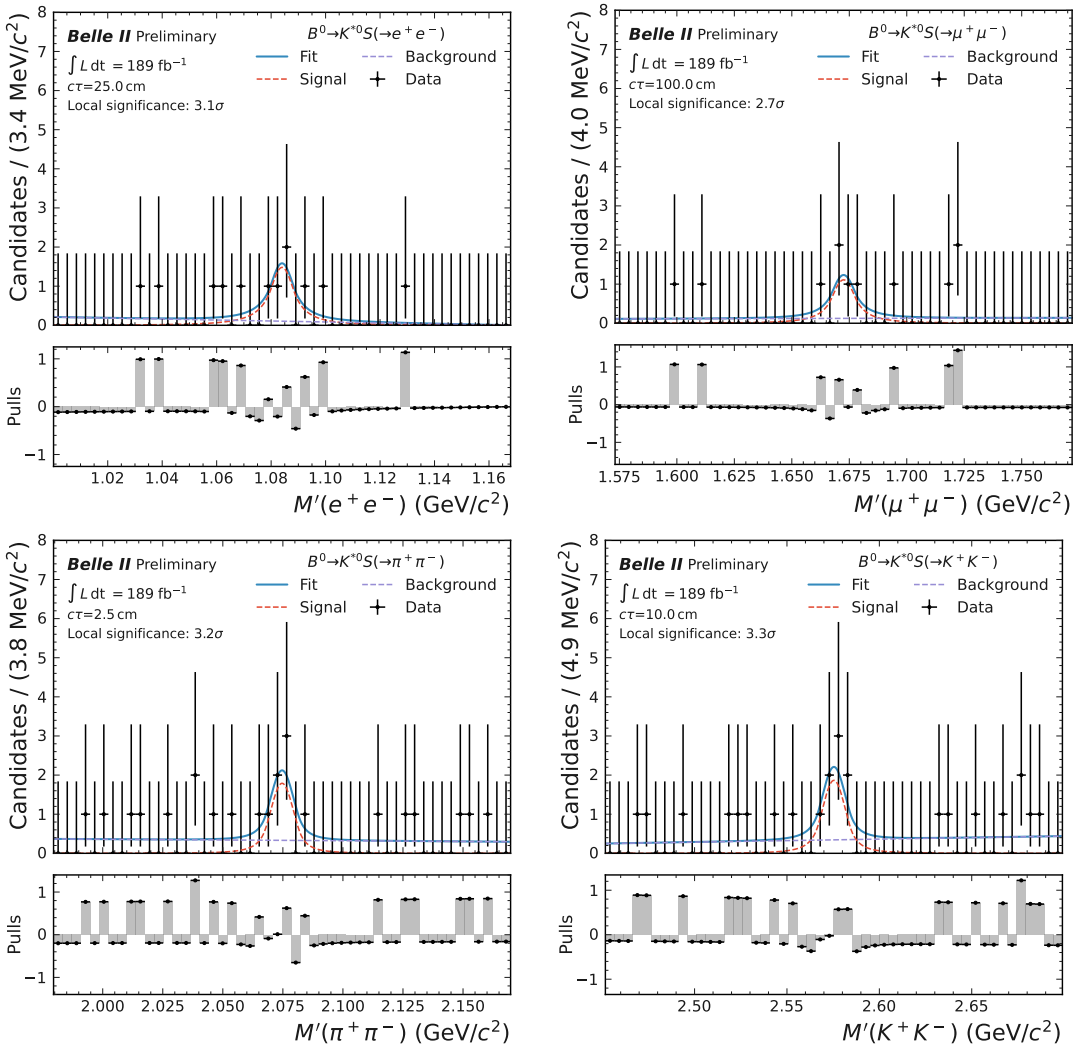

Figure 11.6.: Fits to data with largest local signal significances in $B^0 \to K^{*0}(\to K^+\pi^-)S$ for (upper left) $S \to e^+e^-$, (upper right) $S \to \mu^+\mu^-$, (lower left) $S \to \pi^+\pi^-$ and (lower right) $S \to K^+K^-$.

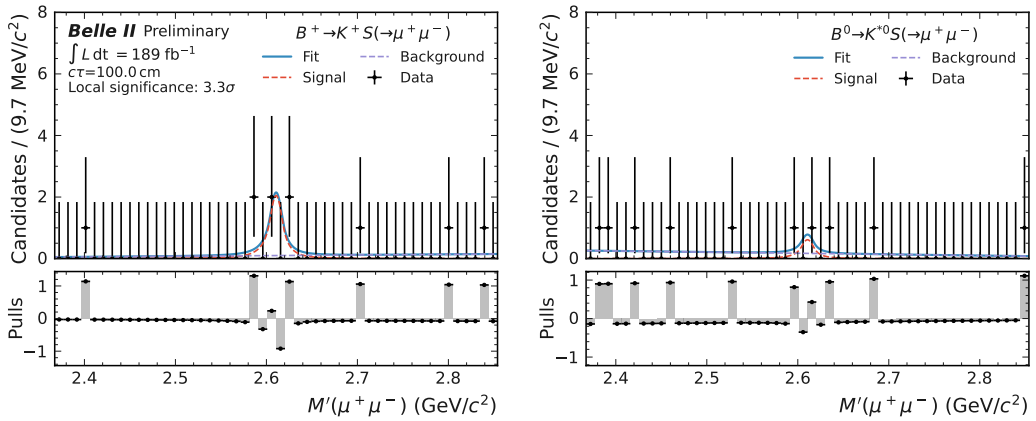

Figure 11.7.: Fits to data with largest local signal significance in the combined fit.

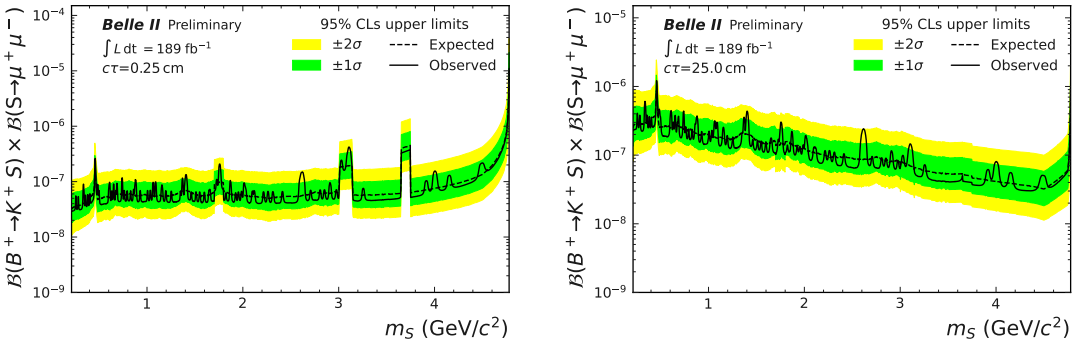

Figure 11.8.: CL_S upper limits on the product of signal branching fractions for $B^+ \to K^+S(\to \mu^+\mu^-)$ for (left) a lifetime hypothesis of $c\tau_S = 0.25\,\text{cm}$ and (right) $c\tau_S = 25\,\text{cm}$.

be significantly weaker compared to the expected limit. This is in accordance with the fact that no evidence of the signal process is found in the scans of the signal significances.

The general trend of the upper limits as a function of S mass is due to the signal efficiency. At small lifetimes the sensitivity is largest, corresponding to stronger upper limits for small S masses due to the larger Lorentz-boost needed to pass the minimum displacement selection. For longer lifetimes the effect is reversed; small Lorentz boom at large masses results in more S decays within the tracking acceptance. The upper limit scans for small lifetime hypotheses also show a drop in sensitivity around the regions where the vertex dr requirement is tightened to reduce the tails of two-body resonance backgrounds. In the $S \to \mu^+\mu^-$ final state this is the case for S masses around $(1.7-, 1.8)\,\mathrm{GeV}/c^2$ for the D^0 background and $(3.00, 3.15)\,\mathrm{GeV}/c^2$ for J/ψ and $(3.65, 3.75)\,\mathrm{GeV}/c^2$ for $\psi(2S)$ backgrounds. The effect on the signal efficiency is less pronounced for longer lifetime hypotheses due to the longer average flight distance of the S with a smaller impact of the tighter vertex dr selection. At the D^0, the effect is suppressed compared to the J/ψ and $\psi(2S)$ regions by the smaller S mass and therefore larger Lorentz-boost. The K^0_S background rejection is implemented as a veto in $M_{\pi^+\pi^-}$ and manifests itself as a region of reduced signal efficiency and therefore weaker upper limits. In $S \to \pi^+\pi^-$ the veto is a fully excluded S mass region around the K^0_S mass. The weakening of the upper limits at the largest S masses is explained by the smaller phase space available for the prompt kaon. The prompt track reconstruction and identification efficiency decrease with momentum.

The model-independent upper limits constitute the main result of the search presented in this thesis. An overview of the upper limits observed in all analysis channels for three different lifetime hypotheses is shown in Fig. 11.9. The collection of all observed and expected model-independent upper limits derived in this search is shown in App. A.

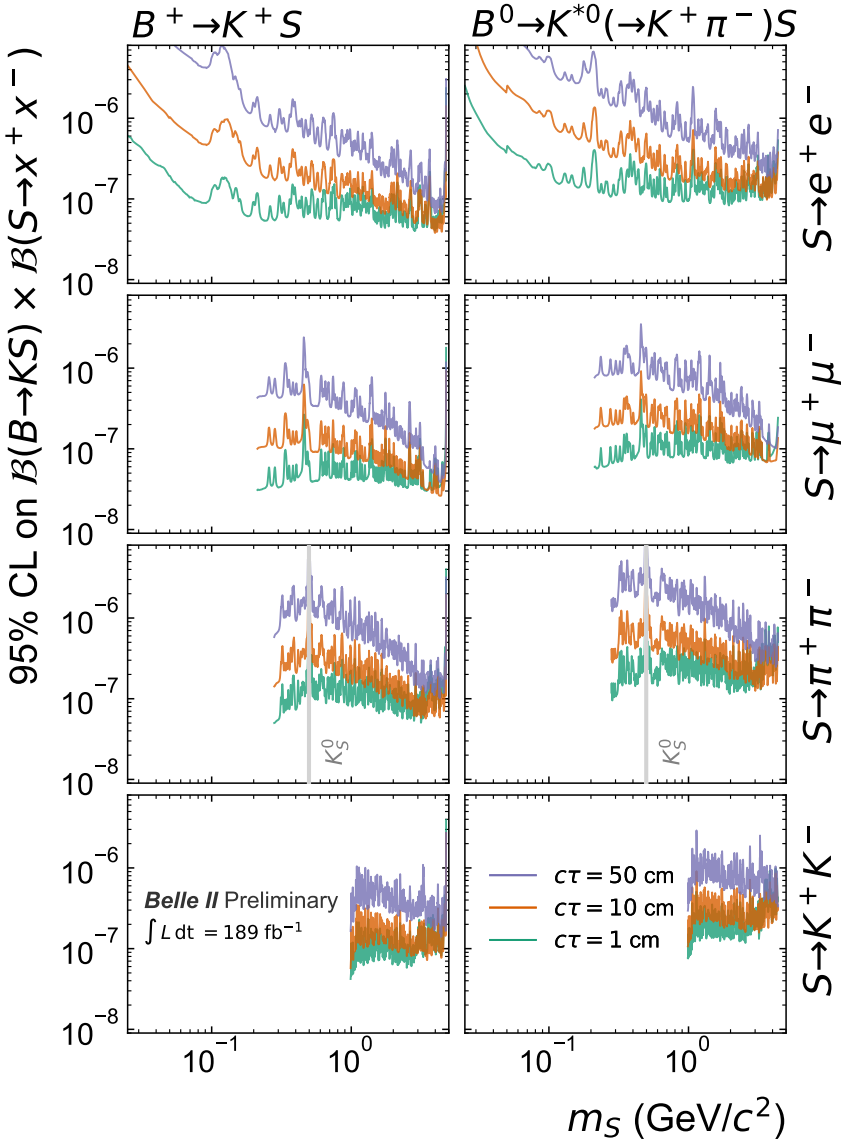

Figure 11.9.: Model-independent CL_S observed upper limits on the product of signal branching fractions for three different lifetime hypotheses in all analysis channels.

11.4. Model-dependent bounds

Model-dependent upper limits are set on the combined product of signal branching fractions $\sum(\mathcal{B}(B \to KS) \times \mathcal{B}(S \to x^+x^-))$ with the sum over all active analysis channels as a function of S mass separately for the lifetime hypotheses considered and the two benchmark models. Predictions of relative rates in the different analysis channels are used in the combined fits to constrain the signal yields in the individual channels.

11.4.1. Translation procedure

The upper limits on the combined branching fraction are translated into bounds on the parameters of the benchmark model. The parameters are the mass of the dark scalar m_S or ALP m_A and their coupling parameters, the mixing angle θ and g_Y, respectively. The bounds on the branching fraction can be translated into bounds on the coupling parameters using results for different lifetime hypotheses. The couplings in the benchmark models control both the lifetimes and the production rates of the mediators, as described in Sec. 2.2. Small coupling values are related to long lifetimes and low production rates. Therefore, at a given mass m_S each tested lifetime hypothesis corresponds to a specific predicted value of the coupling parameter and of the production rate and therefore the branching fraction.

The translation of the bounds is performed individually for each value of m_S. The predicted values of the branching fractions at m_S are computed as a function of the lifetime and thus also of the coupling parameter. The observed upper limits on the branching fraction at m_S are linearly interpolated between all lifetime hypotheses tested. Each lifetime or coupling value with a predicted branching fraction larger than the interpolated observed limit is excluded. A graphical illustration of the method for a specific mediator mass value is shown in Fig. 11.10 using the dark scalar model. The translation for the ALP model works analogously with g_Y instead of θ

The resulting bounds on the dark scalar mixing angle as a function of dark scalar mass are shown in Fig. 11.11 for a translation using the model-independent upper limits from single channel fits, and from the model-dependent upper limits from the combined fits in all active analysis channels using the predicted rates. The expected

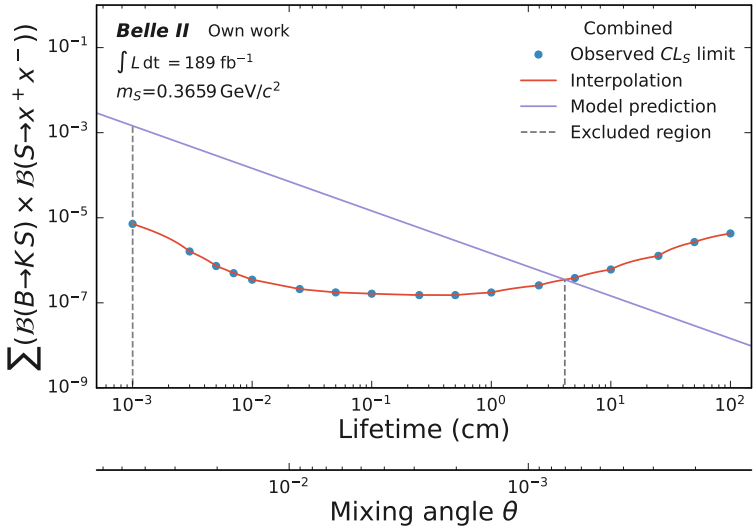

Figure 11.10.: Translation of the bounds on the signal branching fraction into a bound on benchmark model parameters at a given S mass point using an interpolation of the experimental limits at different lifetime hypotheses and the predicted branching fraction.

CL_S upper limits derived from the experimental data are used to facilitate the comparison of sensitivity between channels without statistical fluctuations.

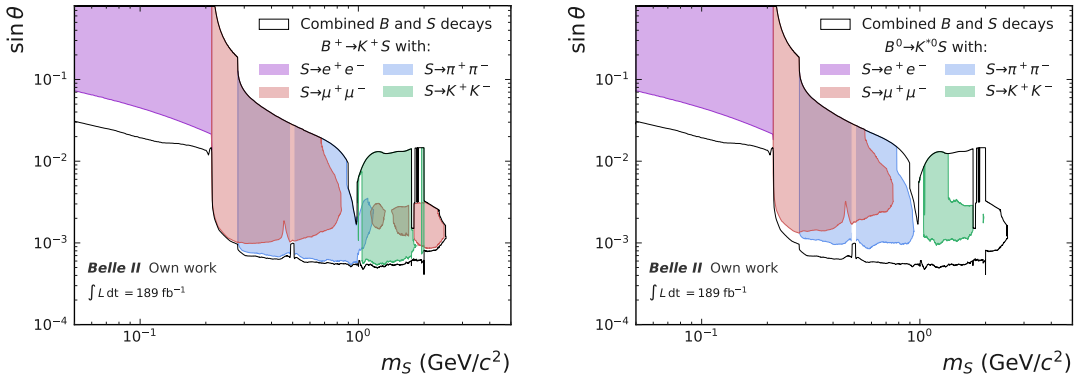

Figure 11.11.: Bounds on the dark scalar model from the single channel fits in (left) $B^+ \to K^+S$ and (right) $B^0 \to K^{*0}(\to K^+\pi^-)S$ compared to the bounds from the combined fit using both production modes and all decay modes in the $\sin\theta$ versus m_S plane.

Below $2m_\mu$ the bounds come only from the $S \to e^+e^-$ channels. Above $2m_\mu$ the respective channels with the largest final-state particle mass dominate the bounds due to the Yukawa-like couplings of the dark scalar. For masses above $m_S > 2\,\mathrm{GeV}/c^2$ only the $S \to \mu^+\mu^-$ final-state channels contribute as explained in Sec. 2.2.

The $B^+ \to K^+S$ channels contribute stronger bounds on the dark scalar compared to $B^0 \to K^{*0}(\to K^+\pi^-)S$ due to the higher signal efficiency and the higher production rate above around $m_S \geq 1.5\,\mathrm{GeV}/c^2$ as shown in Fig. 2.1. The signal efficiency in $B^0 \to K^{*0}(\to K^+\pi^-)S$ is lower due to the additional prompt track, which has a finite reconstruction and identification efficiency, the additional selection in $M_{K^{*0}}$ and the smaller momentum of the prompt tracks. Of the single-channel results, only $B^+ \to K^+S(\to \mu^+\mu^-)$ shows sensitivity to the dark scalar benchmark model above $2\,\mathrm{GeV}/c^2$.

The comparison between single-channel and combined bounds shows the additional constraining power of the combination of information. The largest gain in sensitivity to the benchmark model comes from the combination of different S decay channels.

Adding a second production mode further increases the sensitivity, which is most visible for $S \to e^+e^-$ below the di-muon threshold. The inclusion of the two exclusive production modes is important for future interpretations of the model-independent limits, as detailed in Sec. 2.2.

In the ALP model, only the electron and muon channels contribute below and above $2m_\mu$, respectively.

11.4.2. Results

The resulting bounds from this search on the benchmark models are shown in Fig. 11.12 for the dark scalar and in Fig. 11.13 for the ALP mediator. The bounds are derived by translating the observed upper limits using combined model-dependent fits. The existing bounds of other experiments, introduced in Sec. 2.2.3, are shown in addition.

This search is able to constrain the mixing angle in the dark scalar model down to $\sin\theta > 0.00037$ around $m_S = 2\,\mathrm{GeV}/c^2$. In the ALP model, the coupling values are constrained down to $g_Y > 0.0003$ for masses around $(2 < m_a < 2.5)\,\mathrm{GeV}/c^2$. The ALP model can be constrained up to higher mediator masses because the ALP branching fraction into muons decreases lees steeply compared to the dark scalar. This can be seen in Fig. 2.1 and Fig. 2.2. For mediator masses above $2\,\mathrm{GeV}/c^2$ both models are only constrained by this search using the muon final-state channels.

The Belle II bounds show reduced sensitivity around S masses where the mediator mixes with a SM resonance. The dark scalar mixes with the f_{980} leading to an increased rate for $S \to \pi\pi$ via the f_{980} around $m_{f_{980}}$. The ALP mixes with the η, resulting in enhanced $a \to \gamma\gamma$ and $a \to \pi\pi\pi$ rates around m_η. The enlarged total decay rates lead to shorter lifetimes. For larger couplings values this results in a reduced signal efficiency due to the minimum displacement selection. Furthermore, the enhancement of the ALP decay rates into channels that are not probed by the search results in weaker constraints. The effect is constrained to a narrower region for the ALP due to the small inherit width of the η.

Another visible effect in the Belle II bounds is the K_S^0 veto. The $S \to \pi^+\pi^-$ final-state channel contributes largely to the Belle II sensitivity in this mass region. The veto results in contributions only from $S \to \mu^+\mu^-$ and therefore a reduced efficiency

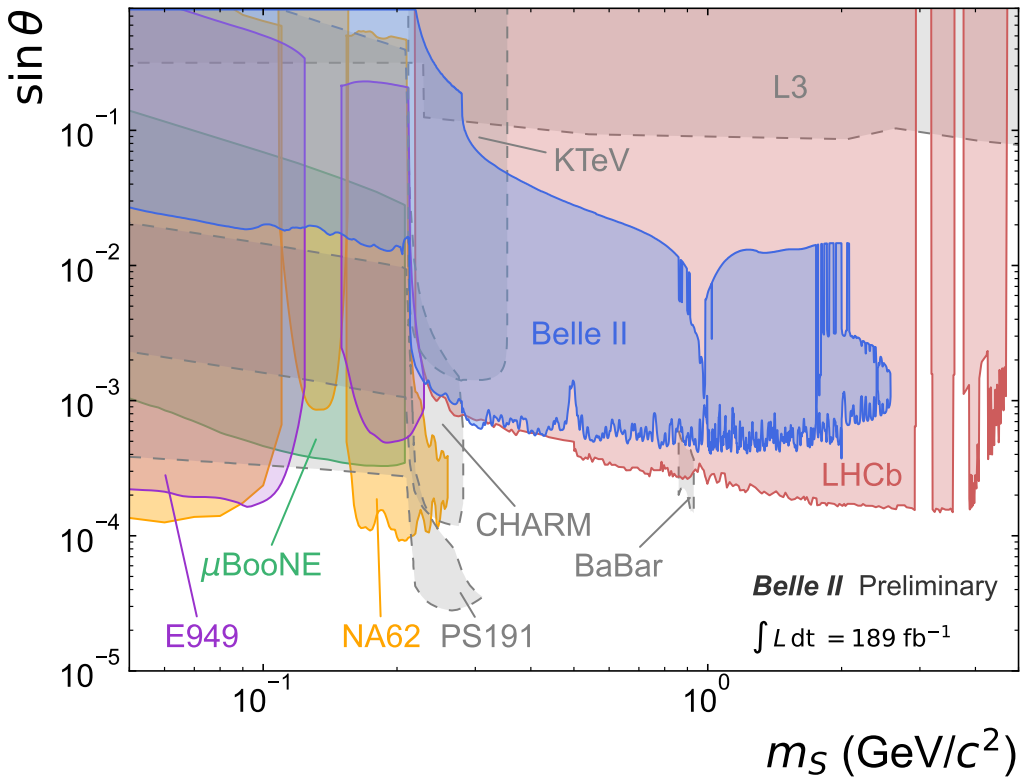

Figure 11.12.: Bounds on the dark scalar model from the combined fit in the $\sin\theta$ versus m_S plane including results from other experiments.

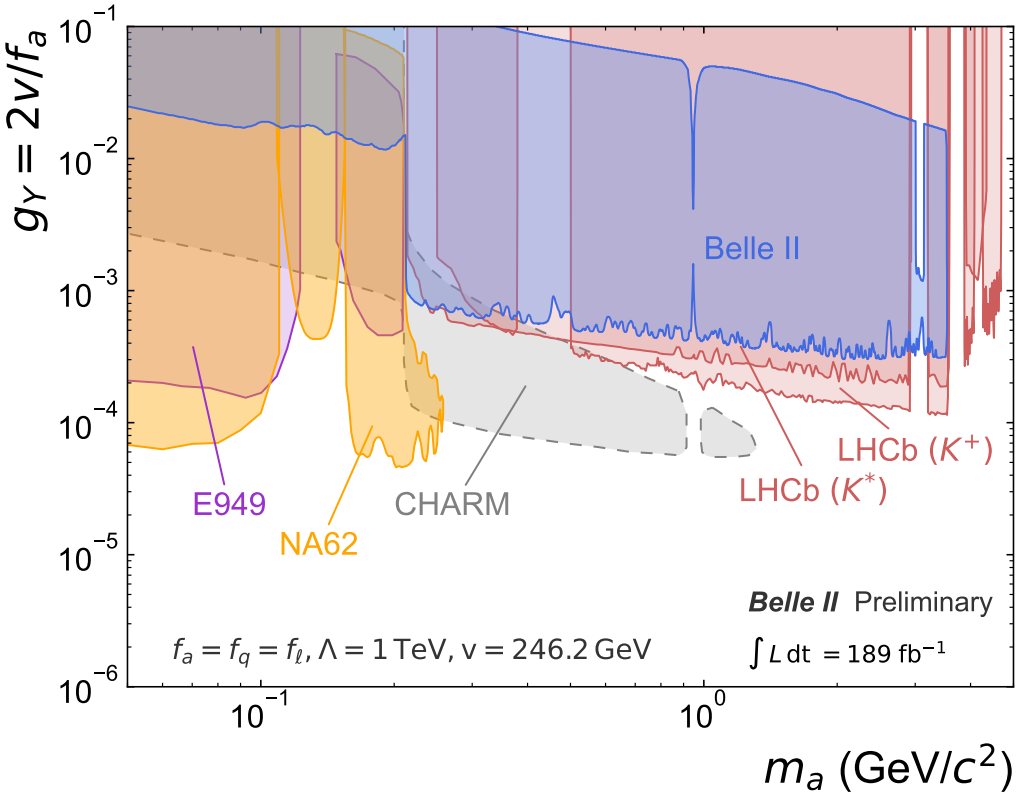

Figure 11.13.: Bounds on the ALP model from the combined fit in the g_Y versus m_a plane including results from other experiments.

and loss of sensitivity.

The LHCb collaboration searches for $S \to \mu^+\mu^-$ and the resulting bounds in the benchmark models are leading for a large part of the parameter space. This is mainly due to the very large production rate of B-mesons at LHCb. The results of this search surpass these bounds at small S masses slightly above 2μ. The LHCb limits also feature cut-out regions where a veto was applied to reject two-body resonance backgrounds, similar to the ones found in this work. These regions are not fully vetoed in this search, instead a tighter minimum vertex displacement selection is applied. This recovers some of the sensitivity and results in the first bounds in the ALP model around the J/ψ mass, set by the work.

The LHCb search includes prompt S decays, resulting in no loss in efficiency for smaller lifetimes in the mixing regions. In these regions, the large boost of B-mesons, produced in proton-proton collisions at the LHC, lead to higher efficiencies for smaller lifetimes due to less decays of S outside of acceptance. These two effects recover the sensitivity of LHCb in the mixing regions, despite the smaller muon branching fraction.

The difference between the large boost at LHCb and the small boost of $\beta\gamma \approx 0.28$ at Belle II results in a better expected scaling of Belle II constraints for long lifetimes with the size of the dataset. The LHCb collaboration also plans to include hadronic decays of the S in future iterations of the search [171].

The large dependence of experimental bounds on the lifetime predicted by the benchmark models can also be seen by comparing the CHARM bounds interpreted in the dark scalar and ALP models. The geometry of the CHARM detector with the reconstruction of a possible signal decay vertex in a decay volume, displaced from the production in a fixed-target collision, leads to a sensitivity to a specific range of lifetime values. The different shape of the bounds in the two benchmark models is a result of the different dependence of the lifetime on the mediator mass and coupling parameter.

BaBar performed an inclusive search for LLPs that were not specifically targeted at the signal process. This leads to a larger signal efficiency translating to a competitive limit in the dark scalar model in a narrow region of phase space.

Below the threshold for $S \to \mu^+\mu^-$ searches from beam-dump experiments dominate

the bounds in the benchmark models and are not expected to be superseded by future Belle II searches. However, the limits derived in this search constitute the strongest bounds in the benchmark models from direct searches at e^+e^- colliders.

12. Summary and outlook

A search for a long-lived spin-0 particle, S, in e^+e^- collision events at the Belle II experiment was presented. The production of S through $B^+ \to K^+S$ and $B^0 \to K^{*0}\,S$ and decays via $S \to e^+e^-/\,\mu^+\mu^-/\,\pi^+\pi^-/\,K^+K^-$ are considered. A dataset corresponding to an integrated luminosity of $189\,\text{fb}^{-1}$ of e^+e^- collisions at the $\Upsilon(4S)$ resonance energy is analysed. No evidence for the signal process is found. Model-independent upper limits at the 95% confidence level are derived on the product of branching fractions $\mathcal{B}(B \to KS) \times \mathcal{B}(S \to x^+x^-)$ with $x \in (e, \mu, \pi, K)$ as a function of S mass for different S lifetime hypotheses between $(0.001,\ 400)\,\text{cm}$. Additional upper limits for up to 10000 cm are placed for S masses below the di-muon threshold. The upper limits extend down to the order of $\mathcal{O}(10^{-7})$. These are the first upper limits for hadronic final states from $B \to KS$, and the most constraining upper limits from a direct search for $S \to e^+e^-$ at e^+e^- colliders. The results are interpreted in two benchmark models, a dark scalar mixing with the SM Higgs boson and a pseudoscalar ALP with fermion couplings. The model-dependent bounds are competitive with existing experimental constraints.

Summary of the methods

The energy and momentum of the reconstructed B-meson are required to be close to their expected values which greatly reduces misidentified $B\overline{B}$ and combinatorial backgrounds. The latter are further reduced by selecting events with a spherical momentum distribution. The prompt background is reduced by requirements on the displacement and quality of the S vertex. A window around the K^0_S mass is vetoed. PID requirements are placed on all signal tracks. The TOP likelihood is removed from the PID score of displaced electrons, muons, and pions to prevent misidentification

due to later time-of-arrival. Two-body decays of SM resonances are suppressed by a tighter selection on the S vertex displacement. Photon conversion contributions are minimised by a veto of active material locations for the S decay vertex at small S masses.

The signal distribution in the reconstructed S mass is modelled using a DSCB pdf. A linear polynomial pdf describes the remaining backgrounds. At small S masses, an exponential pdf is added to account for a background due to converted photons.

The signal efficiency as well as the signal pdf parameters are corrected for differences between data and simulation. A correction is derived as a function of the displacement of the S vertex using a control sample enriched in K_S^0. Further corrections are applied to the near-IP tracking efficiency, track momentum scale, and PID efficiencies.

The modelling of background in simulation is validated in separate sideband regions close to the expected signal but enriched in combinatorial backgrounds, K_S^0, and in the two-body SM resonances. The $B^+ \to K_S^0\pi^+$ process, similar to the signal, is confirmed to be well modelled using a dedicated sideband region.

Systematic uncertainties that affect the signal efficiency and the signal pdf parameters are considered. For longer lifetimes and smaller masses, the largest source of uncertainty is the size of the LLP correction factors. Further large uncertainties associated with the signal efficiency are due to the number of $B\overline{B}$ pairs in the dataset and a bias in the fit method. The signal pdf parameters receive uncertainties due to the LLP correction and the track momentum estimation.

The signal branching fraction is determined with maximum likelihood fits to the reconstructed S mass distribution. Large fit windows around the tested S mass are used to estimate the background yield and shape directly in data. Combined fits that include all relevant analysis channels are performed using the signal rates predicted by the two benchmark models. Systemic uncertainties take the form of NPs in the likelihood. The look-elsewhere effect is taken into account by computation of a trials factor. The fit method is validated using simulated pseudo-experiments. Upper limits on the signal branching fraction are derived using the CL_S method. The model-independent upper limits in each channel constitute the main result of this work. The combined fits are then used to determine model-dependent bounds on the benchmark

models.

Conclusions

What can be concluded in view of the questions that were raised at the beginning of this thesis? No evidence is found for the signal, and hence for a possible connection between the SM and a dark sector. However, it can be concluded that Belle II is sensitive to traces of light LLPs. The first such a search is detailed here. Future iterations of the search will benefit greatly from this groundwork and from various possible improvements.

The model-independent results presented in this thesis are derived separately as a function of S mass for various lifetime hypotheses and the eight exclusive channels. In addition to their current use to constrain the two exemplar BSM models, they can be interpreted in any model for new physics that predicts the signal process.

Outlook

Over the lifetime of the experiment, a dataset corresponding to up to $50\,\mathrm{ab}^{-1}$ is expected to be collected, an increase by a factor of 250. At a fixed S lifetime this will increase the sensitivity of the search by an approximate factor of $\sqrt{250} \approx 16$ in the case where systematic uncertainties are not dominant. Furthermore, larger lifetimes of the dark scalar and ALP can be probed with more data, which is limited by the coupling between lifetime and production rate.

The signal efficiency will increase with improvements in the reconstruction of displaced tracks and vertexing. A CDC algorithm with high displaced track finding efficiency is available but is not used due to large fake rates, which can be reduced by a dedicated study. Additionally, a replacement of the current track finding with a graph-based neural network is studied by the collaboration and shows great potential.

The selection can be improved with multivariate analysis techniques that take into account correlations between observables. An artificial neural network that includes the figure of merit in its loss function [172] is an option. Different selections or neural network trainings can be used in regions of S vertex displacement. Furthermore, the

two-body resonance and photon conversion suppressions can be relaxed with progress in their validation.

The TOP performance for displaced tracks could be recovered by using the displaced vertex in its PID determination. However, the expected time of arrival for a displaced track depends on the vertex distance and the Lorentz-boost of the mother particle, the latter determined by using its mass. A displaced TOP PID likelihood would hence depend on the prior assignment of mass hypotheses to the daughter tracks.

A three-dimensional fit using M_{bc} and ΔE in addition to $M'(x^+x^-)$ can be studied. Variations in signal shapes with model parameters are small. This could improve the background modelling and the bias of the signal yield estimator at the endpoints of the S mass spectrum. The downside is the increased complexity of the fit model, which is challenging with low background rates.

The systematic uncertainties do not limit the presented search; however, a reduction can have a significant impact for larger dataset sizes. Separate LLP correction factors can be determined in regions of K^0_S momentum and polar angle, with early studies showing reduced uncertainties. The uncertainty of the fraction of charged and neutral $B\overline{B}$ pairs in the decay of the $\Upsilon(4S)$ will be reduced by a dedicated measurement at Belle II [173].

The scope of the search can be extended as well. Other kaon states in association with the S can be probed, the natural next step being $B^0 \to K^0 S$. The expected gain in signal statistics is not large as K^0_S are reconstructed with an efficiency of around 50%. However, the backgrounds are likely much smaller due to the two displaced vertices. The K^0_L modelling and performance are not yet well understood. Additionally, final states that include neutral particles such as $B^+ \to K^{*+}(\to K^+\pi^0)S$ can be considered. In these cases, a worse detector resolution is expected compared to fully charged final states.

The sensitivity to the benchmark models can be improved with the $S \to \tau^+\tau^-$ channel, which is dominant for $m_S > 2m_\tau$. The channel is experimentally challenging due to neutrinos in the decay of the tau lepton. This leads to nonpointing displaced vertices, which can be used as selection requirement itself, and worse energy and momentum resolution.

Furthermore, dark scalar decays to four mesons [90] and ALP decays to three mesons [96] can be considered, the former suffering from a small predicted rate.

Almost straightforward experimental extensions are flavour-violating decays of the S and the production through $B \to \pi S$ by adjusting the corresponding PID requirements. The former are typically not part of minimal BSM models, and the latter is less motivated due to CKM suppression.

An extension to vector (gauge boson) mediators is possible by simulating signal events with adjusted kinematics and following the same steps as described in this work.

A search for $B \to K(S \to \gamma\gamma)$ with zero as well as finite S lifetimes is being performed at Belle II. The dark scalar model can also be extended by a dark photon and two non-degenerate dark matter states [74]. Two searches at Belle II aim at this scenario, one with a method similar to the search presented in this thesis and the other with a neural network method to find a more general non-SM anomaly.

A summary and outlook of dark Higgs-like scalar searches including other experiments can be found in Ref. [174].

Related searches at Belle II will progress as well. The $b \to s\nu\overline{\nu}$ process is expected to be measured to a precision of up to 8% relative to the SM prediction [175]. A search for this process at Belle II is underway with a focus on two-body decay kinematics. The search will be sensitive to S signals with invisible decays to dark matter or decays outside the detector.

Bibliography

[1] R. L. Workman *et al.*, Particle Data Group collaboration, "Review of Particle Physics," *PTEP*, vol. 2022, p. 083C01, 2022. DOI: 10.1093/ptep/ptac097 (cit. on pp. 11, 17, 18, 76, 125, 148, 153, 168).

[2] K. Garrett and G. Duda, "Dark Matter: A Primer," *Adv. Astron.*, vol. 2011, p. 968 283, 2011. DOI: 10.1155/2011/968283. arXiv: 1006.2483 [hep-ph] (cit. on p. 11).

[3] N. Aghanim *et al.*, Planck collaboration, "Planck 2018 results. VI. Cosmological parameters," *Astron. Astrophys.*, vol. 641, A6, 2020, [Erratum: Astron.Astrophys. 652, C4 (2021)]. DOI: 10.1051/0004-6361/201833910. arXiv: 1807.06209 [astro-ph.CO] (cit. on p. 11).

[4] G. Bertone, D. Hooper, and J. Silk, "Particle dark matter: Evidence, candidates and constraints," *Phys. Rept.*, vol. 405, pp. 279–390, 2005. DOI: 10.1016/j.physrep.2004.08.031. arXiv: hep-ph/0404175 (cit. on p. 11).

[5] G. Steigman, B. Dasgupta, and J. F. Beacom, "Precise Relic WIMP Abundance and its Impact on Searches for Dark Matter Annihilation," *Phys. Rev. D*, vol. 86, p. 023 506, 2012. DOI: 10.1103/PhysRevD.86.023506. arXiv: 1204.3622 [hep-ph] (cit. on p. 15).

[6] L. J. Hall, K. Jedamzik, J. March-Russell, and S. M. West, "Freeze-In Production of FIMP Dark Matter," *JHEP*, vol. 03, p. 080, 2010. DOI: 10.1007/JHEP03(2010)080. arXiv: 0911.1120 [hep-ph] (cit. on p. 15).

[7] J. Hisano, S. Matsumoto, and M. M. Nojiri, "Explosive dark matter annihilation," *Phys. Rev. Lett.*, vol. 92, p. 031 303, 2004. DOI: 10.1103/PhysRevLett.92.031303. arXiv: hep-ph/0307216 (cit. on p. 15).

[8] B. Shuve and I. Yavin, "Dark matter progenitor: Light vector boson decay into sterile neutrinos," *Phys. Rev. D*, vol. 89, no. 11, p. 113 004, 2014. DOI: 10.1103/PhysRevD.89.113004. arXiv: 1403.2727 [hep-ph] (cit. on p. 16).

[9] G. Steigman and M. S. Turner, "Cosmological Constraints on the Properties of Weakly Interacting Massive Particles," *Nucl. Phys. B*, vol. 253, pp. 375–386, 1985. DOI: 10.1016/0550-3213(85)90537-1 (cit. on p. 16).

[10] N. Arkani-Hamed, D. P. Finkbeiner, T. R. Slatyer, and N. Weiner, "A theory of dark matter," *Phys. Rev. D*, vol. 79, p. 015 014, 1 Jan. 2009. DOI: 10.1103/PhysRevD.79.015014 (cit. on p. 16).

[11] L. Roszkowski, E. M. Sessolo, and S. Trojanowski, "WIMP dark matter candidates and searches − current status and future prospects," *Rept. Prog. Phys.*, vol. 81, no. 6, p. 066 201, 2018. DOI: 10.1088/1361-6633/aab913. arXiv: 1707.06277 [hep-ph] (cit. on p. 16).

[12] P. Fayet, "Constraints on Light Dark Matter and U bosons, from ψ, Υ, K^+, π^0, *eta* and *eta′* decays," *Phys. Rev. D*, vol. 74, p. 054 034, 2006. DOI: 10.1103/PhysRevD.74.054034. arXiv: hep-ph/0607318 (cit. on p. 16).

[13] P. Fayet, "U-boson production in e^+e^- annihilations, ψ and Υ decays, and Light Dark Matter," *Phys. Rev. D*, vol. 75, p. 115 017, 2007. DOI: 10.1103/PhysRevD.75.115017. arXiv: hep-ph/0702176 (cit. on p. 16).

[14] R. Essig, P. Schuster, and N. Toro, "Probing Dark Forces and Light Hidden Sectors at Low-Energy e^+e^- Colliders," *Phys. Rev. D*, vol. 80, p. 015 003, 2009. DOI: 10.1103/PhysRevD.80.015003. arXiv: 0903.3941 [hep-ph] (cit. on p. 16).

[15] J. Jaeckel and A. Ringwald, "The Low-Energy Frontier of Particle Physics," *Ann. Rev. Nucl. Part. Sci.*, vol. 60, pp. 405–437, 2010. DOI: 10.1146/annurev.nucl.012809.104433. arXiv: 1002.0329 [hep-ph] (cit. on p. 16).

[16] B. Batell, J. Pradler, and M. Spannowsky, "Dark Matter from Minimal Flavor Violation," *JHEP*, vol. 08, p. 038, 2011. DOI: 10.1007/JHEP08(2011)038. arXiv: 1105.1781 [hep-ph] (cit. on p. 16).

[17] R. Essig *et al.*, "Working Group Report: New Light Weakly Coupled Particles," in *Snowmass 2013: Snowmass on the Mississippi*, Oct. 2013. arXiv: 1311.0029 [hep-ph] (cit. on p. 16).

[18] G. Arcadi, M. Dutra, P. Ghosh, M. Lindner, Y. Mambrini, M. Pierre, S. Profumo, and F. S. Queiroz, "The waning of the WIMP? A review of models, searches, and constraints," *Eur. Phys. J. C*, vol. 78, no. 3, p. 203, 2018. DOI: 10.1140/epjc/s10052-018-5662-y. arXiv: 1703.07364 [hep-ph] (cit. on p. 16).

[19] M. Pospelov, A. Ritz, and M. B. Voloshin, "Secluded WIMP Dark Matter," *Phys. Lett. B*, vol. 662, pp. 53–61, 2008. DOI: 10.1016/j.physletb.2008.02.052. arXiv: 0711.4866 [hep-ph] (cit. on p. 16).

[20] B. Batell, M. Pospelov, and A. Ritz, "Probing a secluded U(1) at B factories," *Phys. Rev. D*, vol. 79, p. 115 008, 11 Jun. 2009. DOI: 10.1103/PhysRevD.79.115008 (cit. on p. 16).

[21] W. Altmannshofer, S. Gori, M. Pospelov, and I. Yavin, "Neutrino Trident Production: A Powerful Probe of New Physics with Neutrino Beams," *Phys. Rev. Lett.*, vol. 113, p. 091 801, 2014. DOI: 10.1103/PhysRevLett.113.091801. arXiv: 1406.2332 [hep-ph] (cit. on p. 16).

[22] F. F. Deppisch, P. S. Bhupal Dev, and A. Pilaftsis, "Neutrinos and Collider Physics," *New J. Phys.*, vol. 17, no. 7, p. 075 019, 2015. DOI: 10.1088/1367-2630/17/7/075019. arXiv: 1502.06541 [hep-ph] (cit. on p. 16).

[23] G. Zhou, J. Y. Günther, Z. S. Wang, J. de Vries, and H. K. Dreiner, "Long-lived sterile neutrinos at Belle II in effective field theory," *JHEP*, vol. 04, p. 057, 2022. DOI: 10.1007/JHEP04(2022)057. arXiv: 2111.04403 [hep-ph] (cit. on p. 16).

[24] A. M. Abdullahi *et al.*, "The present and future status of heavy neutral leptons," *J. Phys. G*, vol. 50, no. 2, p. 020 501, 2023. DOI: 10.1088/1361-6471/ac98f9. arXiv: 2203.08039 [hep-ph] (cit. on p. 16).

[25] F. Abudinén *et al.*, Belle II collaboration, "Search for Axion-Like Particles produced in e^+e^- collisions at Belle II," *Phys. Rev. Lett.*, vol. 125, no. 16, p. 161 806, 2020. DOI: 10.1103/PhysRevLett.125.161806. arXiv: 2007.13071 [hep-ex] (cit. on p. 16).

[26] I. Adachi *et al.*, Belle II collaboration, "Search for an invisible Z' in a final state with two muons and missing energy at Belle II," Dec. 2022. arXiv: 2212.03066 [hep-ex] (cit. on pp. 16, 36).

[27] F. Abudinén *et al.*, Belle II collaboration, "Search for a Dark Photon and an Invisible Dark Higgs Boson in $\mu^+\mu^-$- and Missing Energy Final States with the Belle II Experiment," *Phys. Rev. Lett.*, vol. 130, no. 7, p. 071 804, 2023. DOI: 10.1103/PhysRevLett.130.071804. arXiv: 2207.00509 [hep-ex] (cit. on p. 16).

[28] I. Adachi *et al.*, Belle II collaboration, "Search for Lepton-Flavor-Violating τ Decays to a Lepton and an Invisible Boson at Belle II," *Phys. Rev. Lett.*, vol. 130, no. 18, p. 181 803, 2023. DOI: 10.1103/PhysRevLett.130.181803. arXiv: 2212.03634 [hep-ex] (cit. on pp. 16, 36).

[29] D. Curtin *et al.*, "Long-Lived Particles at the Energy Frontier: The MATHUSLA Physics Case," *Rept. Prog. Phys.*, vol. 82, no. 11, p. 116 201, 2019. DOI: 10.1088/1361-6633/ab28d6. arXiv: 1806.07396 [hep-ph] (cit. on pp. 17, 18).

[30] S. Knapen and S. Lowette, "A guide to hunting long-lived particles at the LHC," Dec. 2022. arXiv: 2212.03883 [hep-ph] (cit. on pp. 17, 18).

[31] F. Abudinén *et al.*, Belle II collaboration, "Precise measurement of the D^0 and D^+ lifetimes at Belle II," *Phys. Rev. Lett.*, vol. 127, no. 21, p. 211 801, 2021. DOI: 10.1103/PhysRevLett.127.211801. arXiv: 2108.03216 [hep-ex] (cit. on pp. 17, 31).

[32] F. J. Abudinen *et al.*, Belle II collaboration, "Measurement of the Ω_c^0 lifetime at Belle II," *Phys. Rev. D*, vol. 107, no. 3, p. L031103, 2023. DOI: 10.1103/PhysRevD.107.L031103. arXiv: 2208.08573 [hep-ex] (cit. on p. 17).

[33] F. Abudinén *et al.*, Belle II collaboration, "Measurement of the Λ_c^+ Lifetime," *Phys. Rev. Lett.*, vol. 130, no. 7, p. 071 802, 2023. DOI: 10.1103/PhysRevLett.130.071802. arXiv: 2206.15227 [hep-ex] (cit. on p. 17).

[34] F. Abudinén *et al.*, Belle II collaboration, "Measurement of the B^0 lifetime and flavor-oscillation frequency using hadronic decays reconstructed in 2019-2021 Belle II data," Feb. 2023. arXiv: 2302.12791 [hep-ex] (cit. on p. 17).

[35] F. Abudinén *et al.*, Belle II collaboration, "Measurement of the branching fraction and CP asymmetry of $B^0 \to \pi^0\pi^0$ decays using 2019-2021 Belle II data," Mar. 2023. arXiv: 2303.08354 [hep-ex] (cit. on p. 17).

[36] F. Abudinén *et al.*, Belle II collaboration, "Measurements of branching fractions and direct CP asymmetries in $B^0 \to K^+\pi^-$, $B^+ \to K^0_S\pi^+$ and $B^0 \to \pi^+\pi^-$ using 2019 and 2020 data," Jun. 2021. arXiv: 2106.03766 [hep-ex] (cit. on p. 17).

[37] A. M. Sirunyan *et al.*, CMS collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV," *JINST*, vol. 13, no. 05, P05011, 2018. DOI: 10.1088/1748-0221/13/05/P05011. arXiv: 1712.07158 [physics.ins-det] (cit. on p. 17).

[38] G. Aad *et al.*, ATLAS collaboration, "ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV," *Eur. Phys. J. C*, vol. 79, no. 11, p. 970, 2019. DOI: 10.1140/epjc/s10052-019-7450-8. arXiv: 1907.05120 [hep-ex] (cit. on p. 17).

[39] L. Lee, C. Ohm, A. Soffer, and T.-T. Yu, "Collider Searches for Long-Lived Particles Beyond the Standard Model," *Prog. Part. Nucl. Phys.*, vol. 106, pp. 210–255, 2019, [Erratum: Prog.Part.Nucl.Phys. 122, 103912 (2022)]. DOI: 10.1016/j.ppnp.2019.02.006. arXiv: 1810.12602 [hep-ph] (cit. on p. 18).

[40] M. Duerr, T. Ferber, C. Hearty, F. Kahlhoefer, K. Schmidt-Hoberg, and P. Tunney, "Invisible and displaced dark matter signatures at Belle II," *JHEP*, vol. 02, p. 039, 2020. DOI: 10.1007/JHEP02(2020)039. arXiv: 1911.03176 [hep-ph] (cit. on pp. 18, 20, 21).

[41] G. Aad *et al.*, ATLAS collaboration, "Search for pairs of muons with small displacements in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," May 2023. arXiv: 2305.02005 [hep-ex] (cit. on p. 18).

[42] G. Aad *et al.*, ATLAS collaboration, "Search for long-lived, massive particles in events with displaced vertices and multiple jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," Jan. 2023. arXiv: 2301.13866 [hep-ex] (cit. on p. 18).

[43] G. Aad *et al.*, ATLAS collaboration, "Search in diphoton and dielectron final states for displaced production of Higgs or Z bosons with the ATLAS detector in $\sqrt{s} = 13$ TeV pp collisions," Apr. 2023. arXiv: 2304.12885 [hep-ex] (cit. on p. 18).

[44] A. Tumasyan *et al.*, CMS collaboration, "Search for long-lived particles using out-of-time trackless jets in proton-proton collisions at $\sqrt{s} = 13$ TeV," Dec. 2022. arXiv: 2212.06695 [hep-ex] (cit. on p. 18).

[45] A. Tumasyan *et al.*, CMS collaboration, "Search for inelastic dark matter in events with two displaced muons and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV," CERN, Geneva, Tech. Rep., 2023 (cit. on p. 18).

[46] A. Tumasyan *et al.*, CMS collaboration, "Search for long-lived heavy neutral leptons with lepton flavour conserving or violating decays to a jet and an electron, muon, or tau lepton," CERN, Geneva, Tech. Rep., 2023 (cit. on p. 18).

[47] D. Liventsev *et al.*, Belle collaboration, "Search for a heavy neutrino in tau decays at Belle," Dec. 2022. arXiv: 2212.10095 [hep-ex] (cit. on p. 18).

[48] J. P. Lees *et al.*, BaBar collaboration, "Search for Long-Lived Particles in e^+e^- Collisions," *Phys. Rev. Lett.*, vol. 114, no. 17, p. 171 801, 2015. DOI: 10.1103/PhysRevLett.114.171801. arXiv: 1502.02580 [hep-ex] (cit. on pp. 18, 27).

[49] J. P. Lees *et al.*, BaBar collaboration, "Search for an Axionlike Particle in B Meson Decays," *Phys. Rev. Lett.*, vol. 128, no. 13, p. 131 802, 2022. DOI:

10.1103/PhysRevLett.128.131802. arXiv: 2111.01800 [hep-ex] (cit. on p. 18).

[50] A. Ariga *et al.*, FASER collaboration, "Technical Proposal for FASER: ForwArd Search ExpeRiment at the LHC," Dec. 2018. arXiv: 1812.09139 [physics.ins-det] (cit. on p. 18).

[51] G. Aielli *et al.*, "Expression of interest for the CODEX-b detector," *Eur. Phys. J. C*, vol. 80, no. 12, p. 1177, 2020. DOI: 10.1140/epjc/s10052-020-08711-3. arXiv: 1911.00481 [hep-ex] (cit. on p. 18).

[52] S. Alekhin *et al.*, "A facility to Search for Hidden Particles at the CERN SPS: the SHiP physics case," *Rept. Prog. Phys.*, vol. 79, no. 12, p. 124 201, 2016. DOI: 10.1088/0034-4885/79/12/124201. arXiv: 1504.04855 [hep-ph] (cit. on p. 18).

[53] M. Bauer, O. Brandt, L. Lee, and C. Ohm, "ANUBIS: Proposal to search for long-lived neutral particles in CERN service shafts," Sep. 2019. arXiv: 1909.13022 [physics.ins-det] (cit. on p. 18).

[54] S. Dreyer *et al.*, "Physics reach of a long-lived particle detector at Belle II," May 2021. arXiv: 2105.12962 [hep-ph] (cit. on p. 18).

[55] A. Heister *et al.*, ALEPH collaboration, "Search for gauge mediated SUSY breaking topologies in e^+e^- collisions at center-of-mass energies up to 209-GeV," *Eur. Phys. J. C*, vol. 25, pp. 339–351, 2002. DOI: 10.1007/s10052-002-1005-z. arXiv: hep-ex/0203024 (cit. on p. 18).

[56] P. Abreu *et al.*, DELPHI collaboration, "A Search for heavy stable and longlived squarks and sleptons in e^+e^- collisions at energies from 130-GeV to 183-GeV," *Phys. Lett. B*, vol. 444, pp. 491–502, 1998. DOI: 10.1016/S0370-2693(98)01443-9. arXiv: hep-ex/9811007 (cit. on p. 18).

[57] V. M. Abazov *et al.*, D0 collaboration, "Search for neutral, long-lived particles decaying into two muons in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$-TeV," *Phys. Rev. Lett.*, vol. 97, p. 161 802, 16 Oct. 2006. DOI: 10.1103/PhysRevLett.97.161802. arXiv: hep-ex/0607028 (cit. on p. 18).

[58] T. Aaltonen *et al.*, CDF collaboration, "Search for Heavy, Long-Lived Neutralinos that Decay to Photons at CDF II Using Photon Timing," *Phys. Rev. D*, vol. 78, p. 032 015, 2008. DOI: 10.1103/PhysRevD.78.032015. arXiv: 0804.1043 [hep-ex] (cit. on p. 18).

[59] F. Bergsma *et al.*, CHARM collaboration, "Search for Axion Like Particle Production in 400-GeV Proton - Copper Interactions," *Phys. Lett. B*, vol. 157, pp. 458–462, 1985. DOI: 10.1016/0370-2693(85)90400-9 (cit. on pp. 18, 27).

[60] S. L. Glashow, J. Iliopoulos, and L. Maiani, "Weak Interactions with Lepton-Hadron Symmetry," *Phys. Rev. D*, vol. 2, pp. 1285–1292, 1970. DOI: 10.1103/PhysRevD.2.1285 (cit. on p. 18).

[61] R. Aaij *et al.*, LHCb collaboration, "Differential branching fractions and isospin asymmetries of $B \to K^{(*)}\mu^+\mu^-$ decays," *JHEP*, vol. 06, p. 133, 2014. DOI: 10.1007/JHEP06(2014)133. arXiv: 1403.8044 [hep-ex] (cit. on p. 18).

[62] R. Aaij *et al.*, LHCb collaboration, "Measurement of CP-Averaged Observables in the $B^0 \to K^{*0}\mu^+\mu^-$ Decay," *Phys. Rev. Lett.*, vol. 125, no. 1, p. 011 802, 2020. DOI: 10.1103/PhysRevLett.125.011802. arXiv: 2003.04831 [hep-ex] (cit. on p. 18).

[63] M. Bordone, G. Isidori, and A. Pattori, "On the Standard Model predictions for R_K and R_{K^*}," *Eur. Phys. J. C*, vol. 76, no. 8, p. 440, 2016. DOI: 10.1140/epjc/s10052-016-4274-7. arXiv: 1605.07633 [hep-ph] (cit. on p. 18).

[64] R. Aaij *et al.*, LHCb collaboration, "Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays," *JHEP*, vol. 08, p. 055, 2017. DOI: 10.1007/JHEP08(2017)055. arXiv: 1705.05802 [hep-ex] (cit. on p. 18).

[65] R. Aaij *et al.*, LHCb collaboration, "Test of lepton universality with $\Lambda_b^0 \to pK^-\ell^+\ell^-$ decays," *JHEP*, vol. 05, p. 040, 2020. DOI: 10.1007/JHEP05(2020)040. arXiv: 1912.08139 [hep-ex] (cit. on p. 18).

[66] R. Aaij *et al.*, LHCb collaboration, "Test of lepton universality in beauty-quark decays," *Nature Phys.*, vol. 18, no. 3, pp. 277–282, 2022. DOI: 10.1038/s41567-021-01478-8. arXiv: 2103.11769 [hep-ex] (cit. on p. 18).

[67] R. Aaij *et al.*, LHCb collaboration, "Tests of lepton universality using $B^0 \to K^0_S \ell^+ \ell^-$ and $B^+ \to K^{*+} \ell^+ \ell^-$ decays," *Phys. Rev. Lett.*, vol. 128, no. 19, p. 191 802, 2022. DOI: 10.1103/PhysRevLett.128.191802. arXiv: 2110.09501 [hep-ex] (cit. on p. 18).

[68] "Test of lepton universality in $b \to s\ell^+\ell^-$ decays," Dec. 2022. arXiv: 2212.09152 [hep-ex] (cit. on p. 18).

[69] F. Abudinén *et al.*, Belle II collaboration, "Search for $B^+ \to K^+ \nu\bar{\nu}$ Decays Using an Inclusive Tagging Method at Belle II," *Phys. Rev. Lett.*, vol. 127, no. 18, p. 181 802, 2021. DOI: 10.1103/PhysRevLett.127.181802. arXiv: 2104.12624 [hep-ex] (cit. on p. 19).

[70] T. Felkl, S. L. Li, and M. A. Schmidt, "A tale of invisibility: constraints on new physics in $b \to s\nu\bar{\nu}$," *JHEP*, vol. 12, p. 118, 2021. DOI: 10.1007/JHEP12(2021)118. arXiv: 2111.04327 [hep-ph] (cit. on p. 19).

[71] "A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery," *Nature*, vol. 607, no. 7917, pp. 52–59, 2022, [Erratum: Nature 612, E24 (2022)]. DOI: 10.1038/s41586-022-04893-w. arXiv: 2207.00092 [hep-ex] (cit. on p. 19).

[72] A. Tumasyan *et al.*, CMS collaboration, "A portrait of the Higgs boson by the CMS experiment ten years after the discovery," *Nature*, vol. 607, no. 7917, pp. 60–68, 2022. DOI: 10.1038/s41586-022-04892-x. arXiv: 2207.00043 [hep-ex] (cit. on p. 19).

[73] E. Izaguirre, G. Krnjaic, and B. Shuve, "Discovering Inelastic Thermal-Relic Dark Matter at Colliders," *Phys. Rev. D*, vol. 93, no. 6, p. 063 523, 2016. DOI: 10.1103/PhysRevD.93.063523. arXiv: 1508.03050 [hep-ph] (cit. on p. 20).

[74] M. Duerr, T. Ferber, C. Garcia-Cely, C. Hearty, and K. Schmidt-Hoberg, "Long-lived Dark Higgs and Inelastic Dark Matter at Belle II," *JHEP*, vol. 04, p. 146, 2021. DOI: 10.1007/JHEP04(2021)146. arXiv: 2012.08595 [hep-ph] (cit. on pp. 20, 195).

[75] A. Berlin and F. Kling, "Inelastic Dark Matter at the LHC Lifetime Frontier: ATLAS, CMS, LHCb, CODEX-b, FASER, and MATHUSLA," *Phys. Rev. D*, vol. 99, no. 1, p. 015 021, 2019. DOI: 10.1103/PhysRevD.99.015021. arXiv: 1810.01879 [hep-ph] (cit. on p. 20).

[76] D. Curtin *et al.*, "Exotic decays of the 125 GeV Higgs boson," *Phys. Rev. D*, vol. 90, no. 7, p. 075 004, 2014. DOI: 10.1103/PhysRevD.90.075004. arXiv: 1312.4992 [hep-ph] (cit. on p. 20).

[77] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, M. Sher, and J. P. Silva, "Theory and phenomenology of two-Higgs-doublet models," *Phys. Rept.*, vol. 516, pp. 1–102, 2012. DOI: 10.1016/j.physrep.2012.02.002. arXiv: 1106.0034 [hep-ph] (cit. on p. 20).

[78] M. Bauer, U. Haisch, and F. Kahlhoefer, "Simplified dark matter models with two Higgs doublets: I. Pseudoscalar mediators," *JHEP*, vol. 05, p. 138, 2017. DOI: 10.1007/JHEP05(2017)138. arXiv: 1701.07427 [hep-ph] (cit. on p. 20).

[79] S. Ipek, D. McKeen, and A. E. Nelson, "A Renormalizable Model for the Galactic Center Gamma Ray Excess from Dark Matter Annihilation," *Phys. Rev. D*, vol. 90, no. 5, p. 055 021, 2014. DOI: 10.1103/PhysRevD.90.055021. arXiv: 1404.3716 [hep-ph] (cit. on p. 20).

[80] D. Goncalves, P. A. N. Machado, and J. M. No, "Simplified Models for Dark Matter Face their Consistent Completions," *Phys. Rev. D*, vol. 95, no. 5, p. 055 027, 2017. DOI: 10.1103/PhysRevD.95.055027. arXiv: 1611.04593 [hep-ph] (cit. on p. 20).

[81] S. P. Martin, "A Supersymmetry primer," *Adv. Ser. Direct. High Energy Phys.*, vol. 18, G. L. Kane, Ed., pp. 1–98, 1998. DOI: 10.1142/9789812839657_0001. arXiv: hep-ph/9709356 (cit. on p. 20).

[82] A. Djouadi, "The Anatomy of electro-weak symmetry breaking. II. The Higgs bosons in the minimal supersymmetric model," *Phys. Rept.*, vol. 459, pp. 1–241, 2008. DOI: 10.1016/j.physrep.2007.10.005. arXiv: hep-ph/0503173 (cit. on p. 20).

[83] U. Ellwanger, C. Hugonie, and A. M. Teixeira, "The Next-to-Minimal Supersymmetric Standard Model," *Phys. Rept.*, vol. 496, pp. 1–77, 2010. DOI: 10.1016/j.physrep.2010.07.001. arXiv: 0910.1785 [hep-ph] (cit. on p. 20).

[84] B. Döbrich, F. Ertas, F. Kahlhoefer, and T. Spadaro, "Model-independent bounds on light pseudoscalars from rare B-meson decays," *Phys. Lett. B*, vol. 790, pp. 537–544, 2019. DOI: 10.1016/j.physletb.2019.01.064. arXiv: 1810.11336 [hep-ph] (cit. on pp. 20, 24, 25).

[85] T. Ferber, A. Filimonova, R. Schäfer, and S. Westhoff, "Displaced or invisible? ALPs from B decays at Belle II," *JHEP*, vol. 131, 2023. DOI: 10.1007/JHEP04(2023)131. arXiv: arXiv:2201.06580 [hep-ph] (cit. on p. 20).

[86] A. Kachanovich, U. Nierste, and I. Nisandzic, "Higgs portal to dark matter and $B \to K^{(*)}$ decays," *The European Physical Journal C*, vol. 80, Jul. 2020. DOI: 10.1140/epjc/s10052-020-8240-z (cit. on pp. 20, 21).

[87] A. Filimonova, R. Schäfer, and S. Westhoff, "Probing dark sectors with long-lived particles at Belle II," *Phys. Rev. D*, vol. 101, p. 095 006, 9 May 2020. DOI: 10.1103/PhysRevD.101.095006 (cit. on pp. 21, 22, 27, 28).

[88] G. Krnjaic, "Probing Light Thermal Dark-Matter With a Higgs Portal Mediator," *Phys. Rev. D*, vol. 94, no. 7, p. 073 009, 2016. DOI: 10.1103/PhysRevD.94.073009. arXiv: 1512.04119 [hep-ph] (cit. on p. 21).

[89] K. Bondarenko, A. Boyarsky, T. Bringmann, M. Hufnagel, K. Schmidt-Hoberg, and A. Sokolenko, "Direct detection and complementary constraints for sub-GeV dark matter," *JHEP*, vol. 03, p. 118, 2020. DOI: 10.1007/JHEP03(2020)118. arXiv: 1909.08632 [hep-ph] (cit. on p. 21).

[90] M. W. Winkler, "Decay and detection of a light scalar boson mixing with the Higgs boson," *Phys. Rev. D*, vol. 99, p. 015 018, 1 Jan. 2019. DOI: 10.1103/PhysRevD.99.015018 (cit. on pp. 22, 26, 27, 195).

[91] J. A. Bailey *et al.*, "$B \to Kl^+l^-$ Decay Form Factors from Three-Flavor Lattice QCD," *Phys. Rev. D*, vol. 93, no. 2, p. 025 026, 2016. DOI: 10.1103/PhysRevD.93.025026. arXiv: 1509.06235 [hep-lat] (cit. on p. 22).

[92] N. Gubernari, A. Kokulu, and D. van Dyk, "$B \to P$ and $B \to V$ Form Factors from B-Meson Light-Cone Sum Rules beyond Leading Twist," *JHEP*, vol. 01, p. 150, 2019. DOI: 10.1007/JHEP01(2019)150. arXiv: arXiv:1811.00983 [hep-ph] (cit. on p. 22).

[93] D. de Florian *et al.*, LHC Higgs Cross Section Working Group collaboration, "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector," vol. 2/2017, Oct. 2016. DOI: 10.23731/CYRM-2017-002. arXiv: 1610.07922 [hep-ph] (cit. on p. 22).

[94] J. Beacham, C. Burrage, D. Curtin, *et al.*, "Physics beyond colliders at CERN: beyond the Standard Model working group report," *Journal of Physics G: Nuclear and Particle Physics*, vol. 47, no. 1, p. 010 501, Dec. 2019. DOI: 10.1088/1361-6471/ab4cd2 (cit. on pp. 24, 25, 27).

[95] G. Hiller, "B physics signals of the lightest CP odd Higgs in the NMSSM at large $\tan\beta$," *Phys. Rev. D*, vol. 70, p. 034 018, 2004. DOI: 10.1103/PhysRevD.70.034018. arXiv: hep-ph/0404220 (cit. on p. 24).

[96] M. J. Dolan, F. Kahlhoefer, C. McCabe, and K. Schmidt-Hoberg, "A taste of dark matter: Flavour constraints on pseudoscalar mediators," *JHEP*, vol. 03, p. 171, 2015, [Erratum: JHEP 07, 103 (2015)]. DOI: 10.1007/JHEP03(2015)171. arXiv: 1412.5174 [hep-ph] (cit. on pp. 24, 25, 195).

[97] B. Batell, M. Pospelov, and A. Ritz, "Multilepton signatures of a hidden sector in rare B decays," *Phys. Rev. D*, vol. 83, p. 054 005, 5 Mar. 2011. DOI: 10.1103/PhysRevD.83.054005 (cit. on p. 24).

[98] M. Kobayashi and T. Maskawa, "CP Violation in the Renormalizable Theory of Weak Interaction," *Prog. Theor. Phys.*, vol. 49, pp. 652–657, 1973. DOI: 10.1143/PTP.49.652 (cit. on p. 24).

[99] M. Freytsis, Z. Ligeti, and J. Thaler, "Constraining the Axion Portal with $B \to Kl^+l^-$," *Phys. Rev. D*, vol. 81, p. 034 001, 2010. DOI: 10.1103/PhysRevD.81.034001. arXiv: 0911.5355 [hep-ph] (cit. on p. 25).

[100] F. Domingo, "Decays of a NMSSM CP-odd Higgs in the low-mass region," *JHEP*, vol. 03, p. 052, 2017. DOI: 10.1007/JHEP03(2017)052. arXiv: 1612.06538 [hep-ph] (cit. on p. 25).

[101] R. Aaij, B. Adeva, M. Adinolfi, *et al.*, LHCb collaboration, "Search for long-lived scalar particles in $B^+ \to K^+\chi(\mu^+\mu^-)$ decays," *Phys. Rev. D*, vol. 95, p. 071 101, 7 Apr. 2017. DOI: 10.1103/PhysRevD.95.071101 (cit. on p. 26).

[102] R. Aaij, B. Adeva, M. Adinolfi, *et al.*, LHCb collaboration, "Search for Hidden-Sector Bosons in $B^0 \to K^{*0}\mu^+\mu^-$ Decays," *Phys. Rev. Lett.*, vol. 115, p. 161 802, 16 Oct. 2015. DOI: 10.1103/PhysRevLett.115.161802 (cit. on p. 26).

[103] M. Acciarri *et al.*, L3 collaboration, "Search for neutral Higgs boson production through the process $e^+e^- \to Z^*H^0$," *Phys. Lett. B*, vol. 385, pp. 454–470, 1996. DOI: 10.1016/0370-2693(96)00987-2 (cit. on p. 27).

[104] A. Alavi-Harati *et al.*, KTeV collaboration, "Search for the Decay $K_L \to \pi^0\mu^+\mu^-$," *Phys. Rev. Lett.*, vol. 84, pp. 5279–5282, 2000. DOI: 10.1103/PhysRevLett.84.5279. arXiv: arXiv:hep-ex/0001006 (cit. on p. 27).

[105] A. V. Artamonov *et al.*, BNL-E949 collaboration, "Study of the decay $K^+ \to \pi^+\nu\bar{\nu}$ in the momentum region $140 < P_\pi < 199$ MeV/c," *Phys. Rev. D*, vol. 79, p. 092 004, 2009. DOI: 10.1103/PhysRevD.79.092004. arXiv: arXiv:0903.0030 [hep-ex] (cit. on p. 27).

[106] D. Gorbunov, I. Krasnov, and S. Suvorov, "Constraints on light scalars from PS191 results," *Phys. Lett. B*, vol. 820, p. 136 524, 2021. DOI: 10.1016/j.physletb.2021.136524. arXiv: arXiv:2105.11102 [hep-ph] (cit. on p. 27).

[107] E. Cortina Gil *et al.*, NA62 collaboration, "Measurement of the very rare $K^+ \to \pi^+\nu\overline{\nu}$ decay," *JHEP*, vol. 06, p. 093, 2021. DOI: 10.1007/JHEP06(2021)093. arXiv: arXiv:2103.15389 [hep-ex] (cit. on p. 27).

[108] E. Cortina Gil *et al.*, NA62 collaboration, "Search for π^0 decays to invisible particles," *JHEP*, vol. 02, p. 201, 2021. DOI: 10.1007/JHEP02(2021)201. arXiv: arXiv:2010.07644 [hep-ex] (cit. on p. 27).

[109] P. Abratenko *et al.*, MicroBooNE collaboration, "Search for a Higgs Portal Scalar Decaying to Electron-Positron Pairs in the MicroBooNE Detector," *Phys. Rev. Lett.*, vol. 127, no. 15, p. 151 803, 2021. DOI: 10.1103/PhysRevLett.127.151803. arXiv: arXiv:2106.00568 [hep-ex] (cit. on p. 27).

[110] A. Tumasyan *et al.*, CMS collaboration, "Search for long-lived particles decaying into muon pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV collected with a dedicated high-rate data stream," *JHEP*, vol. 04, p. 062, 2022. DOI: 10.1007/JHEP04(2022)062. arXiv: 2112.13769 [hep-ex] (cit. on p. 28).

[111] A. Fradette and M. Pospelov, "BBN for the LHC: constraints on lifetimes of the Higgs portal scalars," *Phys. Rev. D*, vol. 96, no. 7, p. 075 033, 2017. DOI: 10.1103/PhysRevD.96.075033. arXiv: 1706.01920 [hep-ph] (cit. on p. 28).

[112] J. R. Ellis and K. A. Olive, "Constraints on Light Particles From Supernova Sn1987a," *Phys. Lett. B*, vol. 193, p. 525, 1987. DOI: 10.1016/0370-2693(87)91710-2 (cit. on p. 28).

[113] M. S. Turner, "Axions from SN 1987a," *Phys. Rev. Lett.*, vol. 60, p. 1797, 1988. DOI: 10.1103/PhysRevLett.60.1797 (cit. on p. 28).

[114] K. Akai, K. Furukawa, and H. Koiso, SuperKEKB collaboration, "SuperKEKB Collider," *Nucl. Instrum. Meth.*, vol. A907, pp. 188–199, 2018. DOI: 10.1016/j.nima.2018.08.017. arXiv: 1809.01958 [physics.acc-ph] (cit. on p. 29).

[115] S. Kurokawa and E. Kikutani, "Overview of the KEKB accelerators," *NIM-A*, vol. 499, no. 1, pp. 1–7, 2003. DOI: https://doi.org/10.1016/S0168-9002(02)01771-0 (cit. on p. 30).

[116] T. Abe, K. Akai, N. Akasaka, *et al.*, "Achievements of KEKB," *Progress of Theoretical and Experimental Physics*, vol. 2013, no. 3, Mar. 2013, 03A001. DOI: 10.1093/ptep/pts102. eprint: https://academic.oup.com/ptep/article-pdf/2013/3/03A001/4440618/pts102.pdf (cit. on p. 30).

[117] M. Bona *et al.*, SuperB collaboration, "SuperB: A High-Luminosity Asymmetric e^+e^- Super Flavor Factory. Conceptual Design Report," May 2007. arXiv: 0709.0451 [hep-ex] (cit. on p. 30).

[118] Y. Funakoshi *et al.*, "The SuperKEKB Has Broken the World Record of the Luminosity," *JACoW*, vol. IPAC2022, pp. 1–5, 2022. DOI: 10.18429/JACoW-IPAC2022-MOPLXGD1 (cit. on p. 30).

[119] W. Altmannshofer *et al.*, Belle II collaboration, "The Belle II Physics Book," *PTEP*, vol. 2019, no. 12, p. 123C01, 2019, [Erratum: PTEP 2020, 029201 (2020)]. DOI: 10.1093/ptep/ptz106. arXiv: 1808.10567 [hep-ex] (cit. on pp. 30, 35, 41).

[120] A. Natochii *et al.*, "Beam background expectations for Belle II at SuperKEKB," in *Snowmass 2021*, Mar. 2022. arXiv: 2203.05731 [hep-ex] (cit. on p. 30).

[121] T. Abe, Belle II collaboration, "Belle II Technical Design Report," 2010. arXiv: 1011.0352 [physics.ins-det] (cit. on pp. 30, 31, 35).

[122] H. Ye, F. Abudinen, K. Ackermann, *et al.*, "Commissioning and performance of the Belle II pixel detector," *NIM-A*, vol. 987, p. 164 875, 2021. DOI: https://doi.org/10.1016/j.nima.2020.164875 (cit. on p. 32).

[123] J. Kemmer and G. Lutz, "New detector concepts," *NIM-A*, vol. 253, no. 3, pp. 365–377, 1987. DOI: https://doi.org/10.1016/0168-9002(87)90518-3 (cit. on p. 32).

[124] B. Wang, F. Abudinen, K. Ackermann, *et al.*, "Operational experience of the Belle II pixel detector," *NIM-A*, vol. 1032, p. 166 631, 2022. DOI: https://doi.org/10.1016/j.nima.2022.166631 (cit. on p. 32).

[125] C. Irmler, K. Adamczyk, L. Aggarwal, *et al.*, "The silicon vertex detector of the Belle II experiment," *NIM-A*, vol. 1045, p. 167 578, 2023. DOI: https://doi.org/10.1016/j.nima.2022.167578 (cit. on p. 32).

[126] G. Rizzo, K. Adamczyk, H. Aihara, *et al.*, "The Belle II Silicon Vertex Detector: Performance and Operational Experience in the First Year of Data Taking," in *Proceedings of the 29th International Workshop on Vertex Detectors (VERTEX2020)*. DOI: 10.7566/JPSCP.34.010003. eprint: https://journals.jps.jp/doi/pdf/10.7566/JPSCP.34.010003 (cit. on p. 33).

[127] N. Taniguchi, "Central Drift Chamber for Belle-II," *Journal of Instrumentation*, vol. 12, no. 06, p. C06014, Jun. 2017. DOI: 10.1088/1748-0221/12/06/C06014 (cit. on p. 33).

[128] V. Bertacchi, T. Bilka, N. Braun, *et al.*, "Track finding at Belle II," *Computer Physics Communications*, vol. 259, p. 107 610, 2021. DOI: https://doi.org/10.1016/j.cpc.2020.107610 (cit. on p. 33).

[129] T. Dong, S. Uno, M. Uchida, H. Ozaki, N. Taniguchi, E. Nakano, and K. Trabelsi, "Calibration and alignment of the Belle II central drift chamber," *NIM-A*, vol. 930, pp. 132–141, 2019. DOI: https://doi.org/10.1016/j.nima.2019.03.072 (cit. on p. 33).

[130] J. Fast, "The Belle II imaging Time-of-Propagation (iTOP) detector," *NIM-A*, vol. 876, pp. 145–148, 2017, The 9th international workshop on Ring Imaging Cherenkov Detectors (RICH2016). DOI: https://doi.org/10.1016/j.nima.2017.02.045 (cit. on p. 34).

[131] K. Inami, "MCP-PMT development for Belle-II TOP counter," *Physics Procedia*, vol. 37, pp. 683–690, 2012, Proceedings of the 2nd International Conference on Technology and Instrumentation in Particle Physics (TIPP 2011). DOI: https://doi.org/10.1016/j.phpro.2012.02.417 (cit. on p. 34).

[132] K. Kojima, "The operation and performance of the TOP detector at the Belle II experiment," *PoS*, vol. EPS-HEP2021, p. 803, 2022. DOI: 10.22323/1.398.0803 (cit. on p. 34).

[133] S. Nishida, I. Adachi, N. Hamada, *et al.*, "Aerogel RICH for the Belle II forward PID," *NIM-A*, vol. 766, pp. 28–31, 2014, RICH2013 Proceedings of the Eighth International Workshop on Ring Imaging Cherenkov Detectors Shonan,

Kanagawa, Japan, December 2-6, 2013. DOI: https://doi.org/10.1016/j.nima.2014.06.061 (cit. on p. 34).

[134] S. Korpar, I. Adachi, N. Hamada, *et al.*, "A 144-channel HAPD for the Aerogel RICH at Belle II," *NIM-A*, vol. 766, pp. 145–147, 2014, RICH2013 Proceedings of the Eighth International Workshop on Ring Imaging Cherenkov Detectors Shonan, Kanagawa, Japan, December 2-6, 2013. DOI: https://doi.org/10.1016/j.nima.2014.05.060 (cit. on p. 34).

[135] M. Yonenaga, I. Adachi, L. Burmistrov, *et al.*, "Performance evaluation of the aerogel RICH counter for the Belle II spectrometer using early beam collision data," *Progress of Theoretical and Experimental Physics*, vol. 2020, no. 9, Aug. 2020, 093H01. DOI: 10.1093/ptep/ptaa090. eprint: https://academic.oup.com/ptep/article-pdf/2020/9/093H01/34833887/ptaa090.pdf (cit. on p. 34).

[136] Belle-ECL, V. Aulchenko, A. Bobrov, *et al.*, "Electromagnetic calorimeter for Belle II," *Journal of Physics: Conference Series*, vol. 587, no. 1, p. 012 045, Feb. 2015. DOI: 10.1088/1742-6596/587/1/012045 (cit. on p. 34).

[137] K. Miyabayashi, "Belle II electromagnetic calorimeter and its performance during early SuperKEKB operation," *Journal of Instrumentation*, vol. 15, no. 10, p. C10016, Oct. 2020. DOI: 10.1088/1748-0221/15/10/C10016 (cit. on p. 35).

[138] V. Bertacchi *et al.*, Belle II Tracking Group collaboration, "Track finding at Belle II," *Comput. Phys. Commun.*, vol. 259, p. 107 610, 2021. DOI: 10.1016/j.cpc.2020.107610. arXiv: 2003.12466 [physics.ins-det] (cit. on p. 35).

[139] T. Aushev, D. Besson, K. Chilikin, *et al.*, "A scintillator based endcap K_L and muon detector for the Belle II experiment," *NIM-A*, vol. 789, pp. 134–142, 2015. DOI: https://doi.org/10.1016/j.nima.2015.03.060 (cit. on p. 35).

[140] S. Yamada, R. Itoh, K. Nakamura, M. Nakao, S. Y. Suzuki, T. Konno, T. Higuchi, Z. Liu, and J. Zhao, "Data Acquisition System for the Belle II Experiment," *IEEE Transactions on Nuclear Science*, vol. 62, no. 3, pp. 1175–1180, 2015. DOI: 10.1109/TNS.2015.2424717 (cit. on p. 35).

[141] V. Bertacchi *et al.*, `Belle II Tracking Group` collaboration, "Track finding at Belle II," *Comput. Phys. Commun.*, vol. 259, p. 107 610, 2021. DOI: `10.1016/j.cpc.2020.107610`. arXiv: `2003.12466 [physics.ins-det]` (cit. on p. 36).

[142] Nils Braun, "Combinatorial Kalman Filter and High Level Trigger Reconstruction for the Belle II Experiment," Ph.D. dissertation, Karlsuhe Institute of Technology, Karlsruhe, 2018. DOI: `https://doi.org/10.1007/978-3-030-24997-7` (cit. on p. 36).

[143] T. Alexopoulos, M. Bachtis, E. Gazis, and G. Tsipolitis, "Implementation of the Legendre Transform for track segment reconstruction in drift tube chambers," *NIM-A*, vol. 592, no. 3, pp. 456–462, 2008. DOI: `https://doi.org/10.1016/j.nima.2008.04.038` (cit. on p. 36).

[144] J. Wagner, "Track Finding with the Silicon Strip Detector of the Belle II Experiment," M.S. thesis, Karlsruhe Institute of Technology (KIT), Karlsruhe, 2017 (cit. on p. 36).

[145] T. Bilka *et al.*, "Implementation of GENFIT2 as an experiment independent track-fitting framework," Feb. 2019. arXiv: `1902.04405 [physics.data-an]` (cit. on p. 36).

[146] J. F. Krohn *et al.*, `Belle-II analysis software Group` collaboration, "Global decay chain vertex fitting at Belle II," *Nucl. Instrum. Meth. A*, vol. 976, p. 164 269, 2020. DOI: `10.1016/j.nima.2020.164269`. arXiv: `1901.11198 [hep-ex]` (cit. on p. 36).

[147] M. D. Nuccio, "Search for Axion Like Particles produced in e^+e^- collisions and photon energy resolution studies at Belle II," Ph.D. dissertation, University of Hamburg, Hamburg, 2021 (cit. on p. 37).

[148] S. Agostinelli *et al.*, `GEANT4` collaboration, "GEANT4: A Simulation toolkit," *Nucl.Instrum.Meth.*, vol. A506, pp. 250–303, 2003. DOI: `10.1016/S0168-9002(03)01368-8` (cit. on p. 42).

[149] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, and N. Braun, Belle-II Framework Software Group collaboration, "The Belle II Core Software," *Comput. Softw. Big Sci.*, vol. 3, no. 1, p. 1, 2019. DOI: 10.1007/s41781-018-0017-9. arXiv: 1809.04299 [physics.comp-ph] (cit. on p. 42).

[150] *Belle II Analysis Software Framework (basf2)*, https://doi.org/10.5281/zenodo.5574115. DOI: 10.5281/zenodo.5574115 (cit. on pp. 42–44).

[151] D. J. Lange, "The EvtGen particle decay simulation package," *Nucl. Instrum. Meth.*, vol. A462, pp. 152–155, 2001. DOI: 10.1016/S0168-9002(01)00089-4 (cit. on p. 42).

[152] S. Jadach, B. F. L. Ward, and Z. Was, "The Precision Monte Carlo event generator KK for two fermion final states in e^+e^- collisions," *Comput. Phys. Commun.*, vol. 130, pp. 260–325, 2000. DOI: 10.1016/S0010-4655(00)00048-5. arXiv: hep-ph/9912214 [hep-ph] (cit. on p. 42).

[153] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An Introduction to PYTHIA 8.2," *Comput. Phys. Commun.*, vol. 191, pp. 159–177, 2015. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph] (cit. on p. 42).

[154] S. Jadach, J. H. Kuhn, and Z. Was, "TAUOLA: A Library of Monte Carlo programs to simulate decays of polarized tau leptons," *Comput. Phys. Commun.*, vol. 64, pp. 275–299, 1990. DOI: 10.1016/0010-4655(91)90038-M (cit. on p. 42).

[155] E. Barberio, B. van Eijk, and Z. Was, "PHOTOS: A Universal Monte Carlo for QED radiative corrections in decays," *Comput. Phys. Commun.*, vol. 66, pp. 115–128, 1991. DOI: 10.1016/0010-4655(91)90012-A (cit. on p. 43).

[156] H. Albrecht *et al.*, ARGUS collaboration, "Exclusive Hadronic Decays of B Mesons," *Z. Phys. C*, vol. 48, pp. 543–552, 1990. DOI: 10.1007/BF01614687 (cit. on pp. 54, 73, 124).

[157] G. C. Fox and S. Wolfram, "Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes," *Phys. Rev. Lett.*, vol. 41, pp. 1581–1585, 23 Dec. 1978. DOI: 10.1103/PhysRevLett.41.1581 (cit. on p. 54).

[158] G. C. Fox and S. Wolfram, "Event shapes in e^+e^- annihilation," *Nuclear Physics B*, vol. 149, no. 3, pp. 413–496, 1979, [Erratum: Nucl. Phys. B157 (1979) 543]. DOI: https://doi.org/10.1016/0550-3213(79)90003-8 (cit. on p. 54).

[159] G. Punzi, "Sensitivity of searches for new signals and its optimization," *eConf*, vol. C030908, L. Lyons, R. P. Mount, and R. Reitmeyer, Eds., MODT002, 2003. arXiv: physics/0308063 (cit. on p. 66).

[160] J. Eschle, A. Puig Navarro, R. Silva Coutinho, and N. Serra, "zfit: Scalable pythonic fitting," *SoftwareX*, vol. 11, p. 100 508, 2020. DOI: https://doi.org/10.1016/j.softx.2020.100508 (cit. on p. 87).

[161] G. Cowan, *Statistical data analysis*, Clarendon Press, 1998, ISBN: 9780198501558 (cit. on p. 87).

[162] J. Gaiser, "Charmonium Spectroscopy from Radiative Decays of the J/ψ and ψ'," Ph.D. dissertation, Stanford University, 1982 (cit. on p. 88).

[163] T. Skwarnicki, "A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances," Ph.D. dissertation, Cracow, INP, 1986 (cit. on p. 88).

[164] S. Choudhury, S. Sandilya, K. Trabelsi, *et al.*, Belle collaboration, "Measurement of the B^+/B^0 production ratio in e^+e^- collisions at the $\Upsilon(4S)$ resonance using $B \to J/\psi(\ell\ell)K$ decays at Belle," *Phys. Rev. D*, vol. 107, p. L031102, 3 Feb. 2023. DOI: 10.1103/PhysRevD.107.L031102 (cit. on p. 133).

[165] G. Bohm and G. Zech, *Introduction to statistics and data analysis for physicists*, DESY, 2010, ISBN: 9783935702416 (cit. on p. 151).

[166] E. Gross and O. Vitells, "Trial factors for the look elsewhere effect in high energy physics," *Eur. Phys. J. C*, vol. 70, pp. 525–530, 2010. DOI: 10.1140/epjc/s10052-010-1470-8. arXiv: 1005.1891 [physics.data-an] (cit. on p. 154).

[167] A. L. Read, "Presentation of search results: the CL_S technique," *Journal of Physics G: Nuclear and Particle Physics*, vol. 28, no. 10, pp. 2693–2704, Sep. 2002. DOI: 10.1088/0954-3899/28/10/313 (cit. on p. 157).

[168] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics," *Eur. Phys. J. C*, vol. 71, p. 1554, 2011, [Erratum: Eur.Phys.J.C 73, 2501 (2013)]. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an] (cit. on p. 157).

[169] M. Marinangeli, J. Eschle, E. Rodrigues, B. Pollack, and H. Schreiner, "scikit-hep/hepstats: Python 3.9 support," version v0.5.0, Jun. 2021. DOI: 10.5281/zenodo.4897800 (cit. on p. 157).

[170] E. Rodrigues *et al.*, "The Scikit HEP Project – overview and prospects," *EPJ Web Conf.*, vol. 245, C. Doglioni, D. Kim, G. A. Stewart, L. Silvestris, P. Jackson, and W. Kamleh, Eds., p. 06 028, 2020. DOI: 10.1051/epjconf/202024506028. arXiv: 2007.03577 [physics.comp-ph] (cit. on p. 157).

[171] D. Craik, P. Ilten, D. Johnson, and M. Williams, "LHCb future dark-sector sensitivity projections for Snowmass 2021," in *Snowmass 2021*, Mar. 2022. arXiv: 2203.07048 [hep-ph] (cit. on p. 189).

[172] P. Feichtinger *et al.*, "Punzi-loss: a non-differentiable metric approximation for sensitivity optimisation in the search for new particles," *Eur. Phys. J. C*, vol. 82, no. 2, p. 121, 2022. DOI: 10.1140/epjc/s10052-022-10070-0. arXiv: 2110.00810 [hep-ex] (cit. on p. 193).

[173] F. Bernlochner, M. Jung, M. Khan, G. Landsberg, and Z. Ligeti, "Novel approaches to determine $B^{\pm}$ and B^0 meson production fractions," Jun. 2023. arXiv: 2306.04686 [hep-ph] (cit. on p. 194).

[174] T. Ferber, A. Grohsjean, and F. Kahlhoefer, "Dark Higgs Bosons at Colliders," May 2023. arXiv: 2305.16169 [hep-ph] (cit. on p. 195).

[175] L. Aggarwal *et al.*, Belle II collaboration, "Snowmass White Paper: Belle II physics reach and plans for the next decade and beyond," Jul. 2022. arXiv: 2207.06307 [hep-ex] (cit. on p. 195).

[176] I. Adachi *et al.*, Belle II collaboration, "Search for a long-lived spin-0 mediator in $b \to s$ transitions at the Belle II experiment," Jun. 2023. arXiv: 2306.02830 [hep-ex] (cit. on p. 238).

A. Model-independent upper limits

Expected and observed 95% upper limits for various S masses and lifetimes are shown in:

- Figure A.1 and A.2 for $B^+ \to K^+ S(\to e^+ e^-)$,
- Figure A.3 and A.4 for $B^0 \to K^{*0}(\to K^+\pi^-)S(\to e^+ e^-)$,
- Figure A.5 and A.6 for $B^+ \to K^+ S(\to \mu^+\mu^-)$,
- Figure A.7 and A.8 for $B^0 \to K^{*0}(\to K^+\pi^-)S(\to \mu^+\mu^-)$,
- Figure A.9 and A.10 for $B^+ \to K^+ S(\to \pi^+\pi^-)$,
- Figure A.11 and A.12 for $B^0 \to K^{*0}(\to K^+\pi^-)S(\to \pi^+\pi^-)$,
- Figure A.13 and A.14 for $B^+ \to K^+ S(\to K^+K^-)$,
- Figure A.15 and A.16 for $B^0 \to K^{*0}(\to K^+\pi^-)S(\to K^+K^-)$.

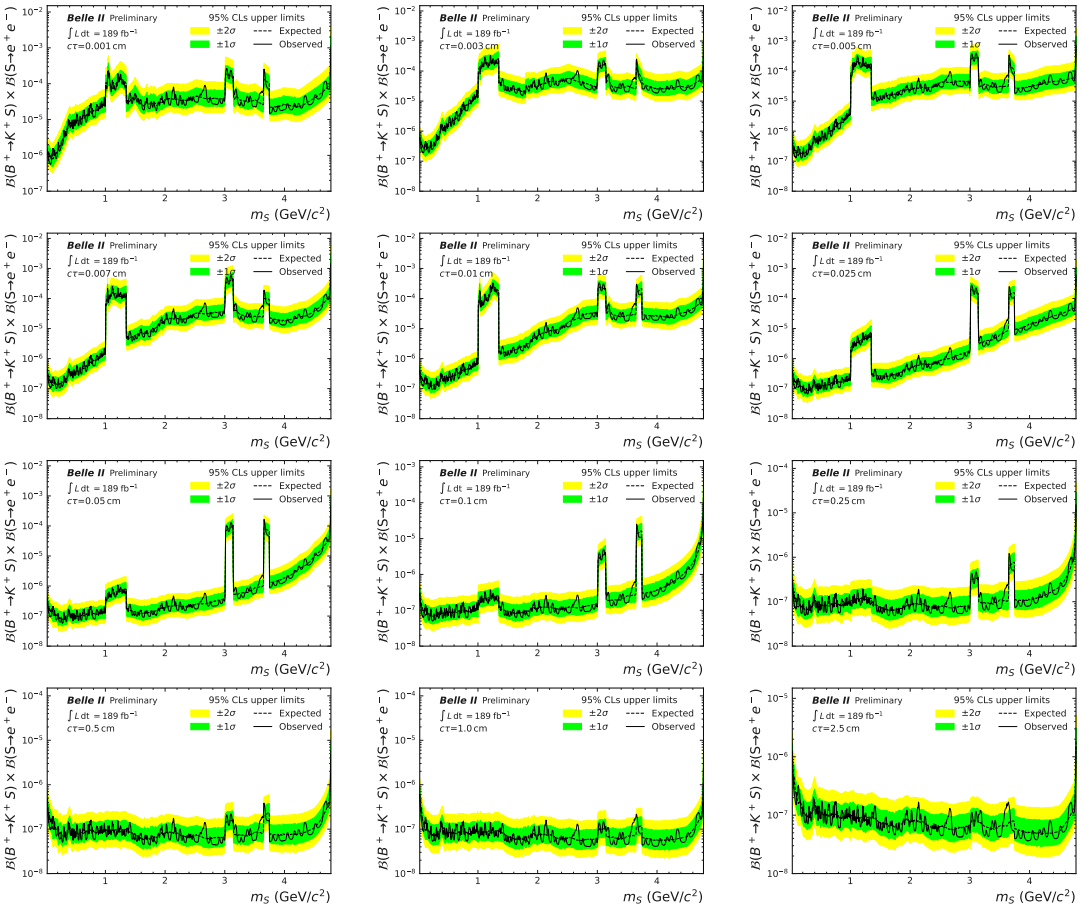

Figure A.1.: Expected and observed limits on the product branching fractions $\mathcal{B}(B^+ \to K^+S) \times \mathcal{B}(S \to e^+e^-)$ for lifetimes $0.001 < c\tau < 2.5$ cm.

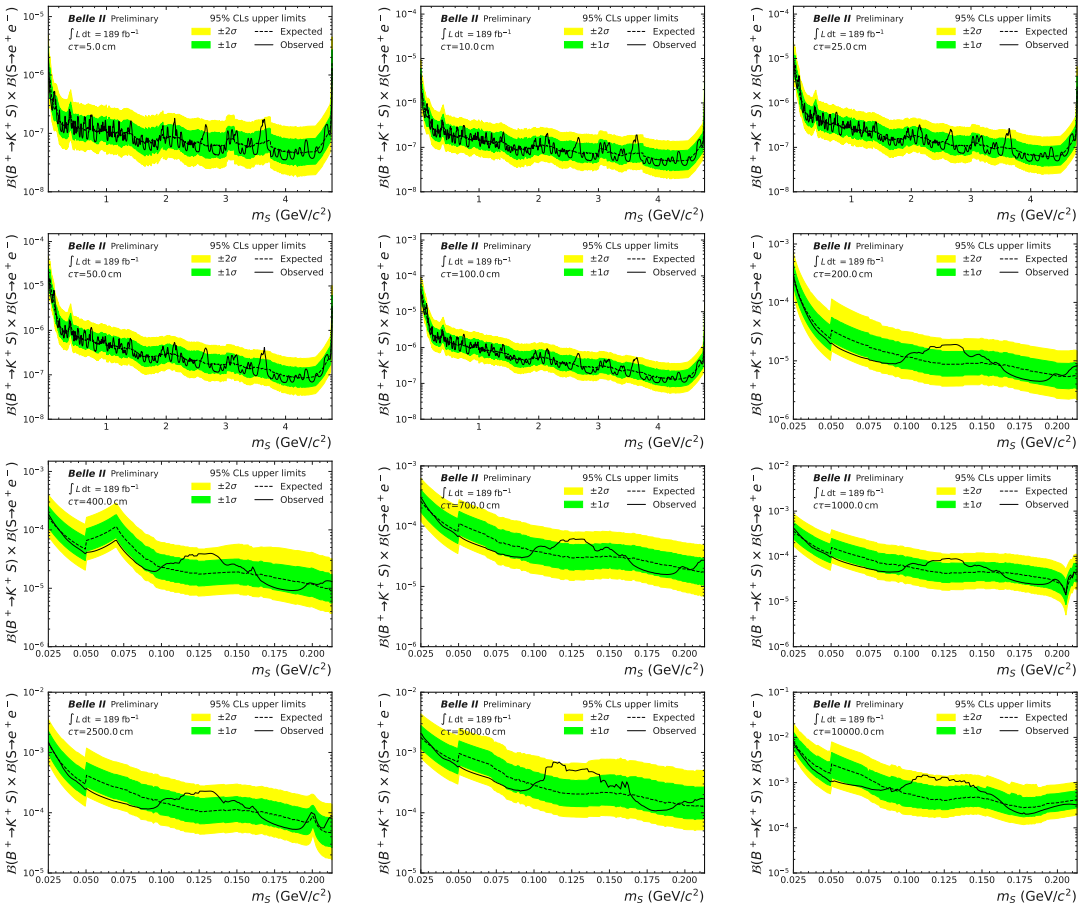

Figure A.2.: Expected and observed limits on the product of branching fractions $\mathcal{B}(B^+ \to K^+S) \times \mathcal{B}(S \to e^+e^-)$ for lifetimes $5 < c\tau < 10000$ cm.

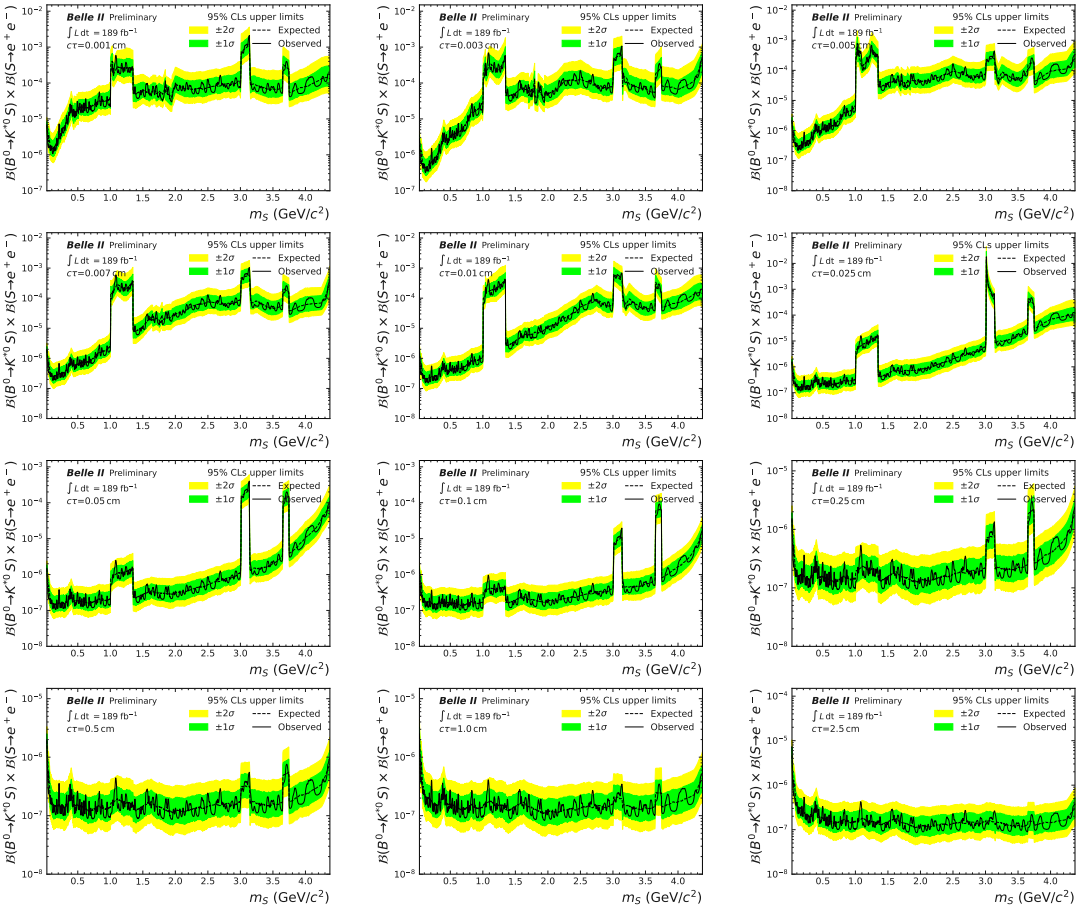

Figure A.3.: Expected and observed limits on the product of branching fractions $\mathcal{B}(B^0 \to K^{*0}(\to K^+\pi^-)S) \times \mathcal{B}(S \to e^+e^-)$ for lifetimes $0.001 < c\tau < 2.5$ cm.

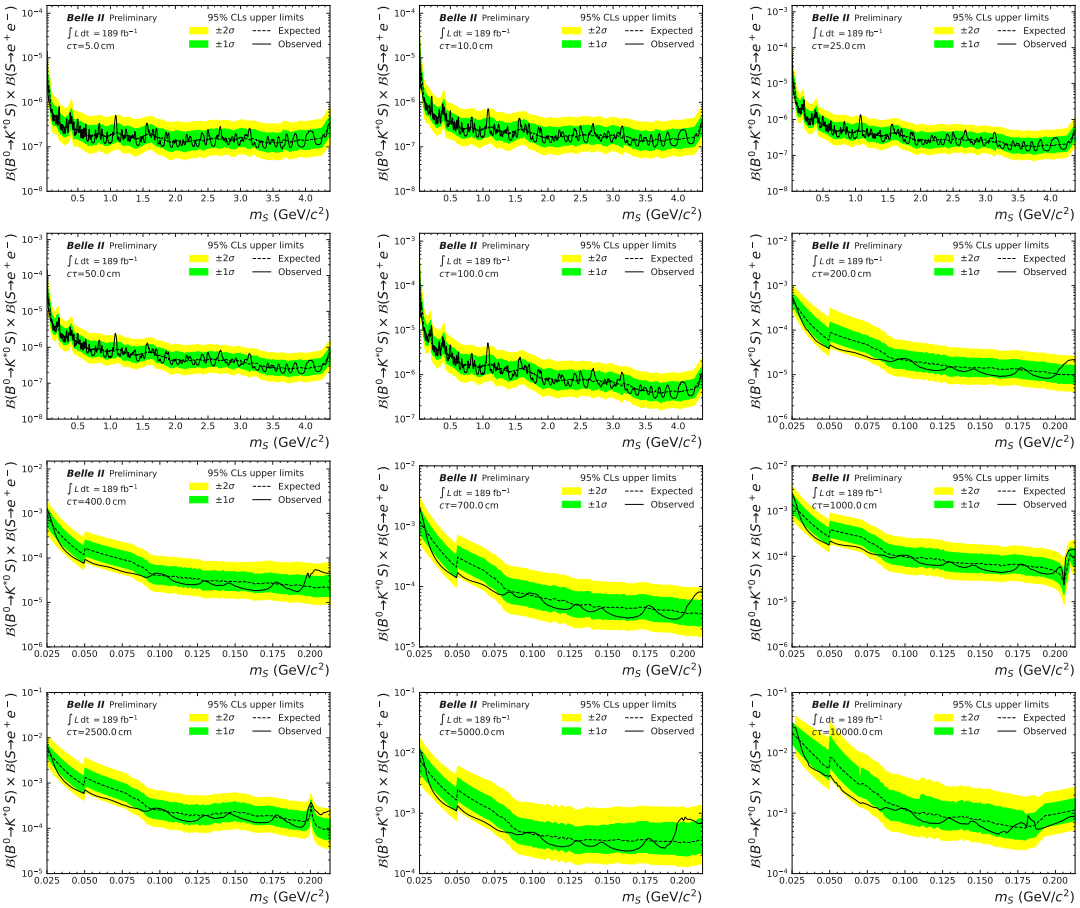

Figure A.4.: Expected and observed limits on the product of branching fractions $\mathcal{B}(B^0 \to K^{*0}(\to K^+\pi^-)S) \times \mathcal{B}(S \to e^+e^-)$ for lifetimes $5 < c\tau < 10000$ cm.

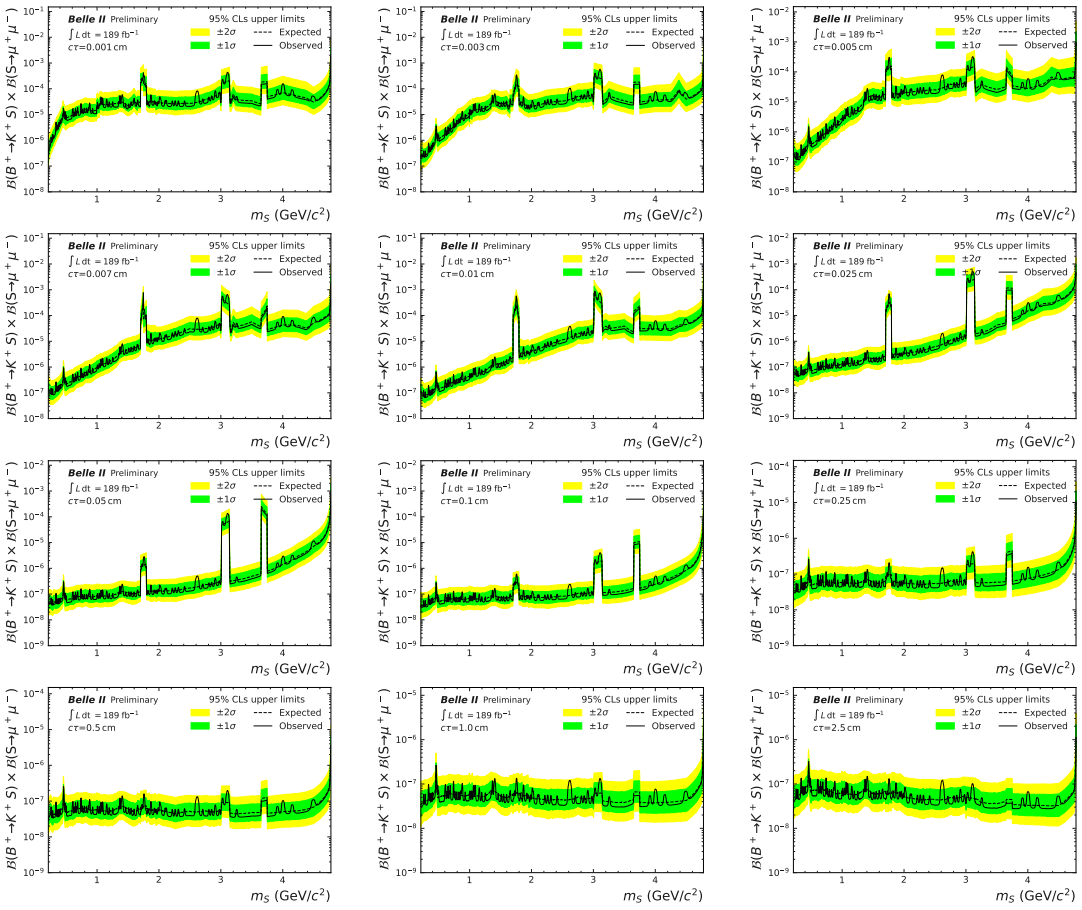

Figure A.5.: Expected and observed limits on the product of branching fractions $\mathcal{B}(B^+ \to K^+ S) \times \mathcal{B}(S \to \mu^+\mu^-)$ for lifetimes $0.001 < c\tau < 2.5$ cm.

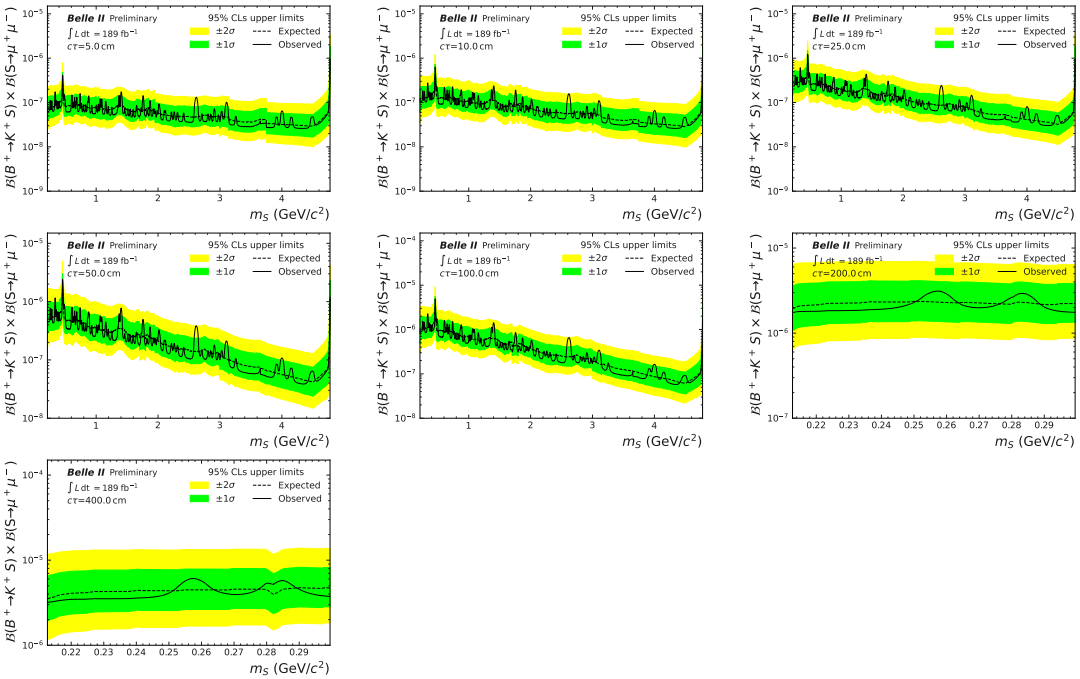

Figure A.6.: Expected and observed limits on the product of branching fractions $\mathcal{B}(B^+ \to K^+ S) \times \mathcal{B}(S \to \mu^+\mu^-)$ for lifetimes $5 < c\tau < 400$ cm.

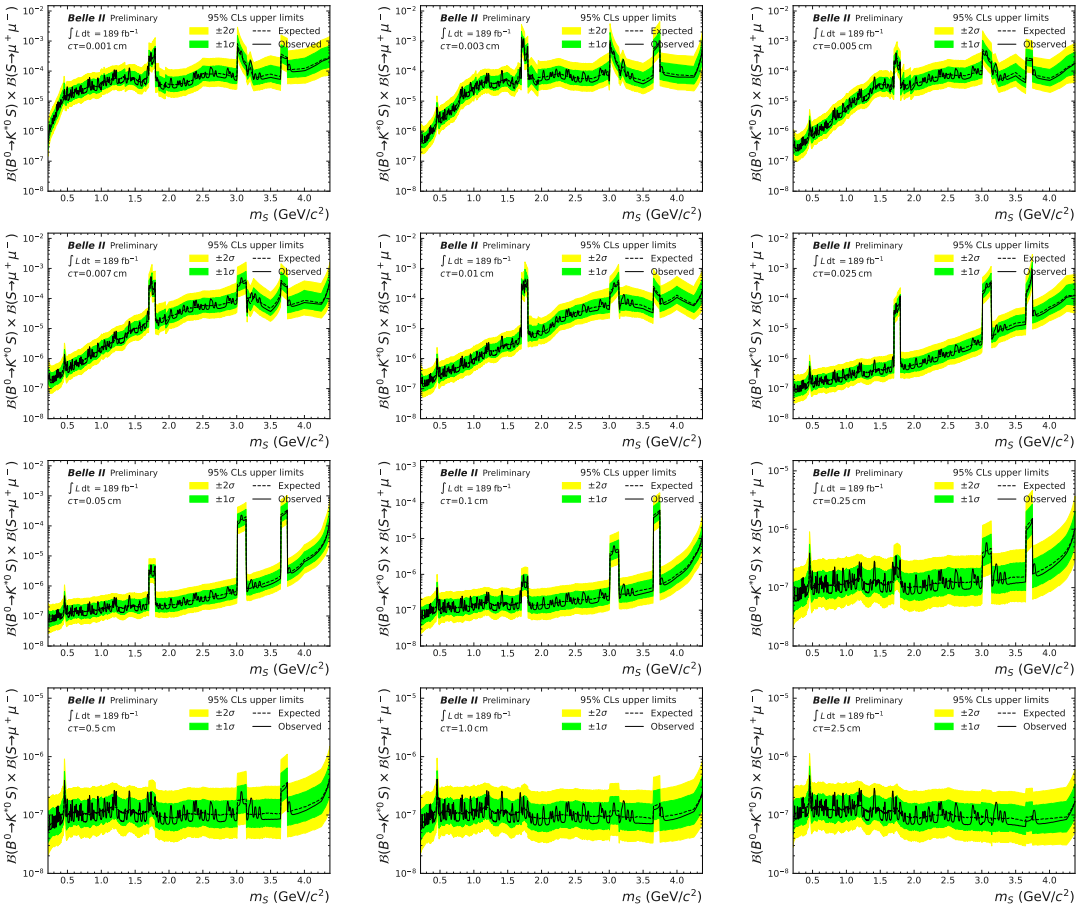

Figure A.7.: Expected and observed limits on the product of branching fractions $\mathcal{B}(B^0 \to K^{*0}(\to K^+\pi^-)S) \times \mathcal{B}(S \to \mu^+\mu^-)$ for lifetimes $0.001 < c\tau < 2.5$ cm.

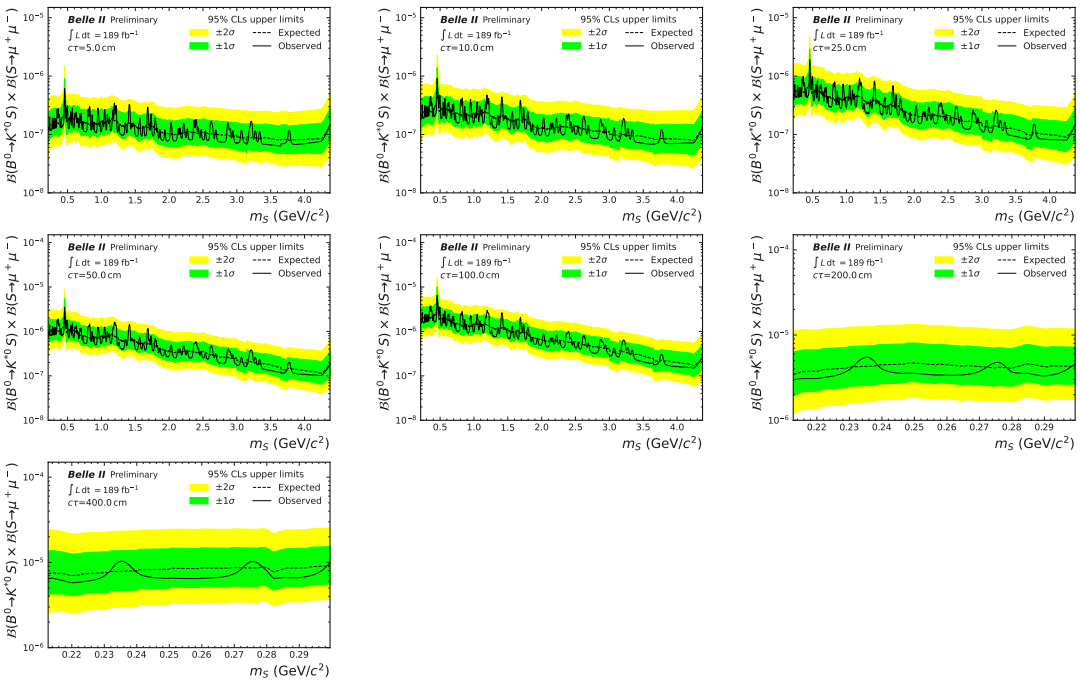

Figure A.8.: Expected and observed limits on the product of branching fractions $\mathcal{B}(B^0 \to K^{*0}(\to K^+\pi^-)S) \times \mathcal{B}(S \to \mu^+\mu^-)$ for lifetimes $5 < c\tau < 400$ cm.

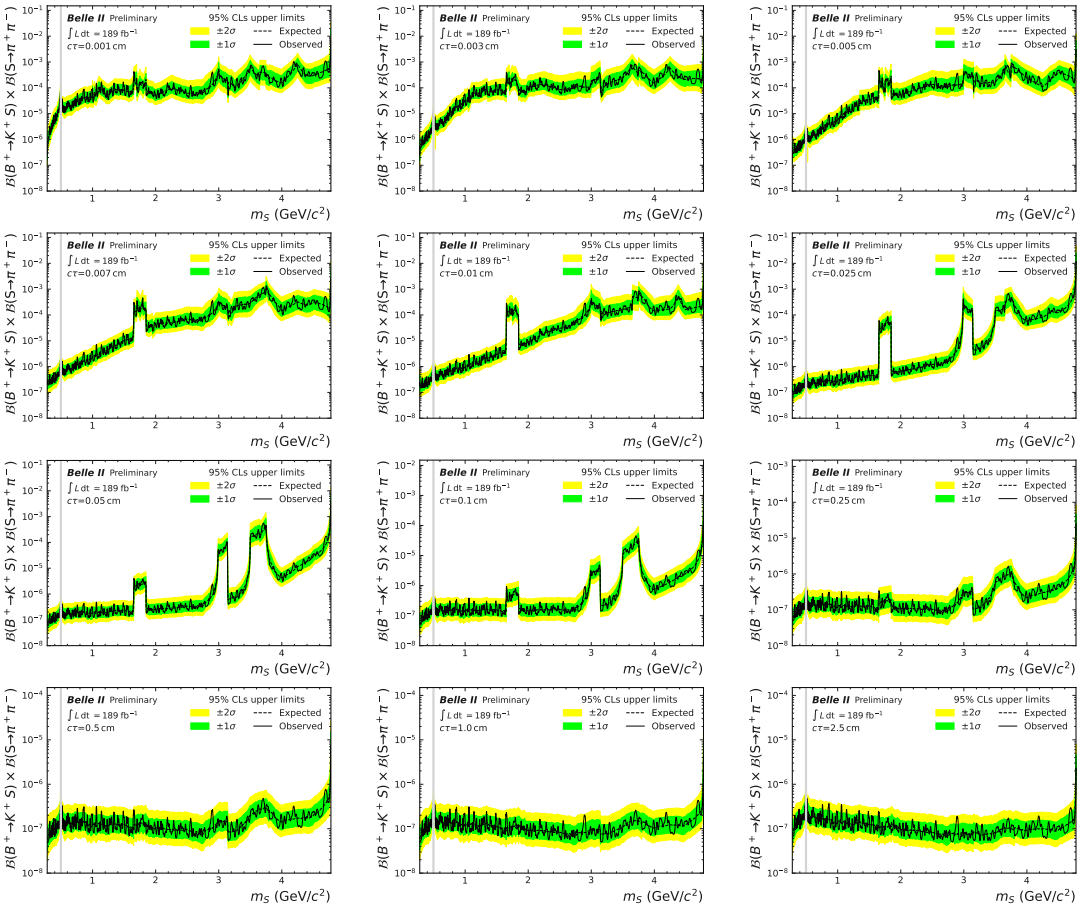

Figure A.9.: Expected and observed limits on the product of branching fractions $\mathcal{B}(B^+ \to K^+S) \times \mathcal{B}(S \to \pi^+\pi^-)$ for lifetimes $0.001 < c\tau < 2.5$ cm. The region corresponding to the fully-vetoed K^0_S for $S \to \pi^+\pi^-$ is marked in gray.

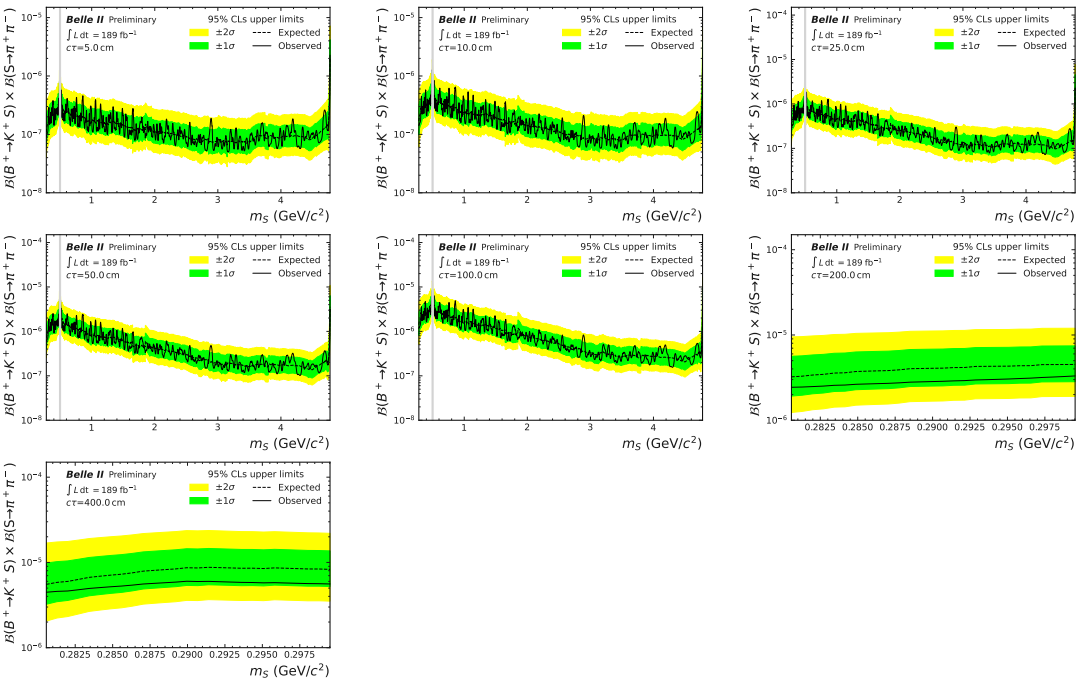

Figure A.10.: Expected and observed limits on the product of branching fractions $\mathcal{B}(B^+ \to K^+ S) \times \mathcal{B}(S \to \pi^+\pi^-)$ for lifetimes $5 < c\tau < 400$ cm. The region corresponding to the fully-vetoed K^0_S for $S \to \pi^+\pi^-$ is marked in gray.

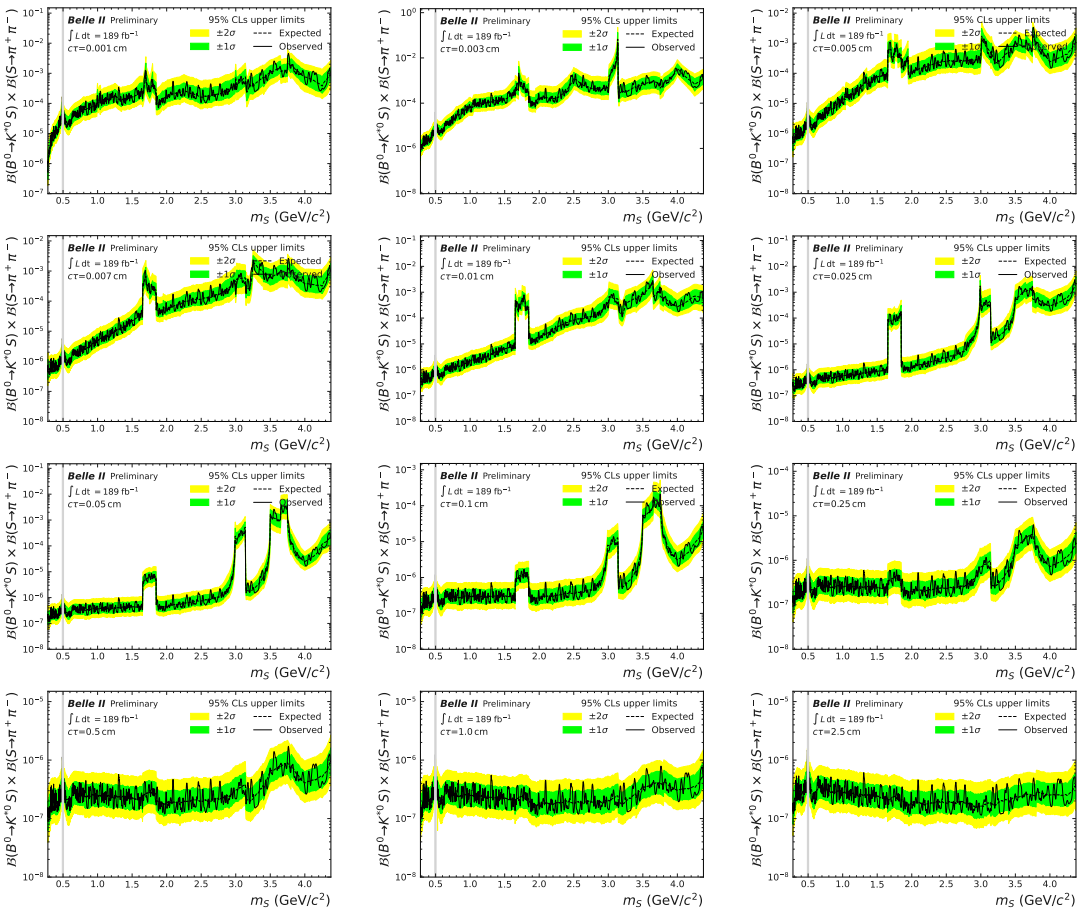

Figure A.11.: Expected and observed limits on the product of branching fractions $\mathcal{B}(B^0 \to K^{*0}(\to K^+\pi^-)S) \times \mathcal{B}(S \to \pi^+\pi^-)$ for lifetimes $0.001 < c\tau < 2.5$ cm. The region corresponding to the fully-vetoed K^0_S for $S \to \pi^+\pi^-$ is marked in gray.

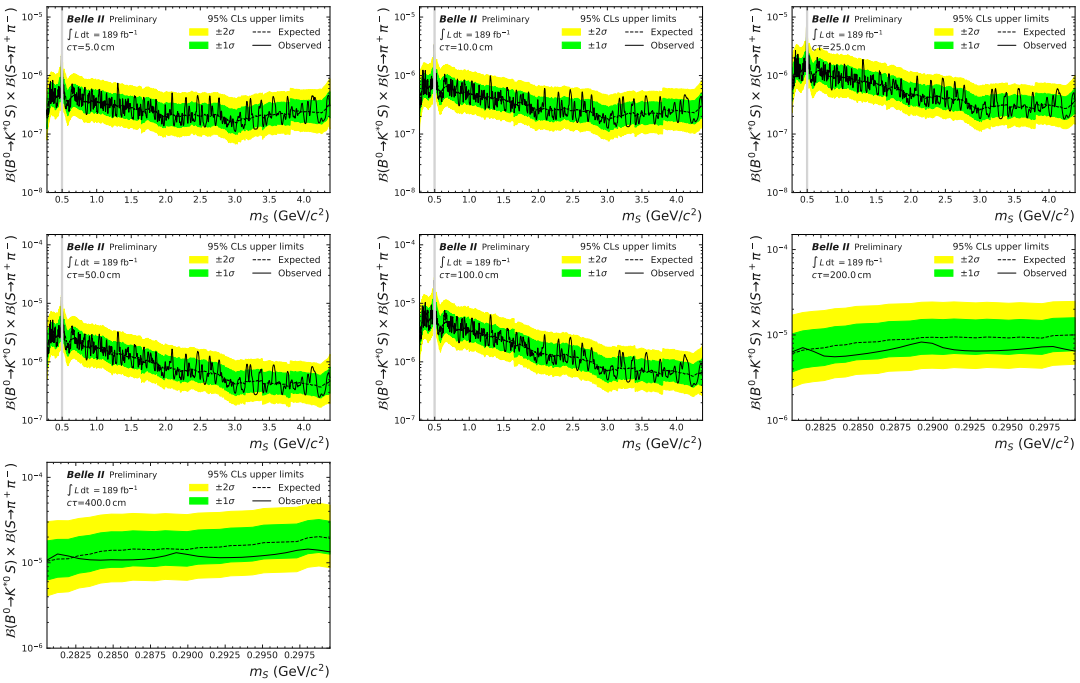

Figure A.12.: Expected and observed limits on the product of branching fractions $\mathcal{B}(B^0 \to K^{*0}(\to K^+\pi^-)S) \times \mathcal{B}(S \to \pi^+\pi^-)$ for lifetimes $5 < c\tau < 400$ cm. The region corresponding to the fully-vetoed K^0_S for $S \to \pi^+\pi^-$ is marked in gray.

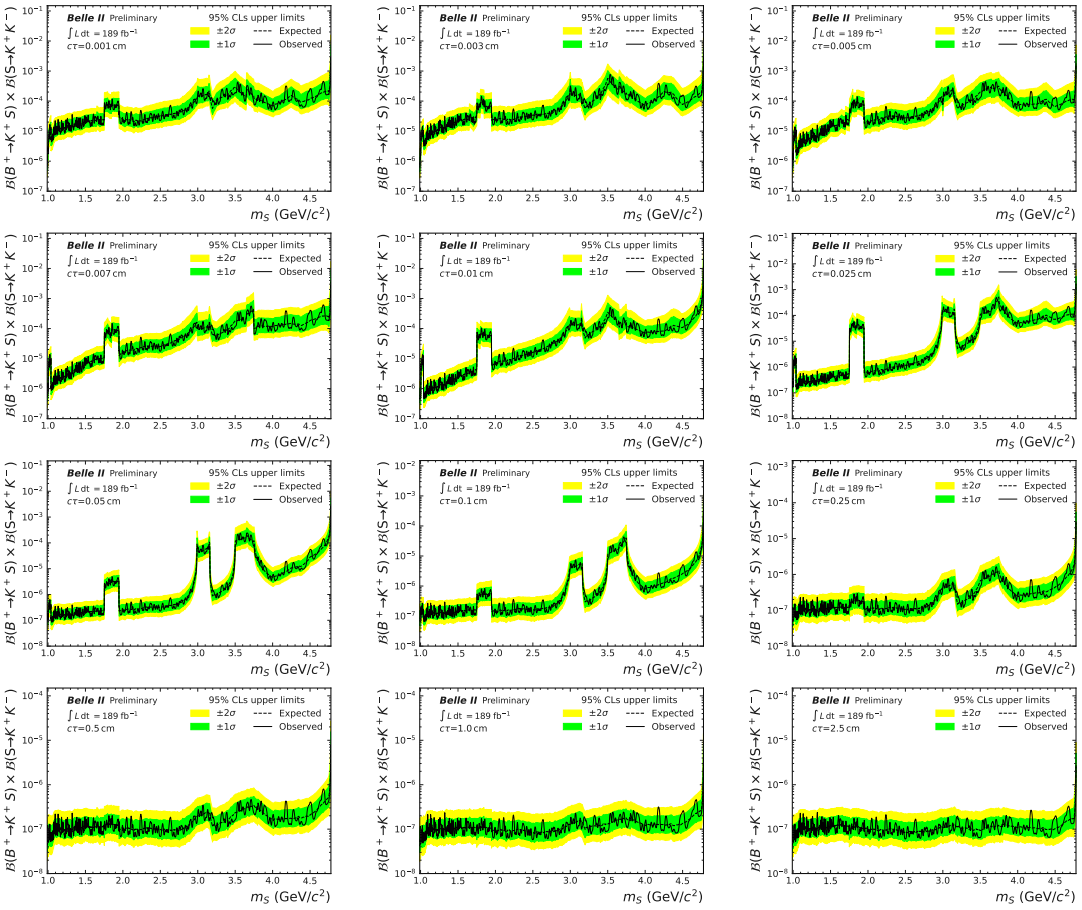

Figure A.13.: Expected and observed limits on the product of branching fractions $\mathcal{B}(B^+ \to KS) \times \mathcal{B}(S \to K^+K^-)$ for lifetimes $0.001 < c\tau < 2.5$ cm.

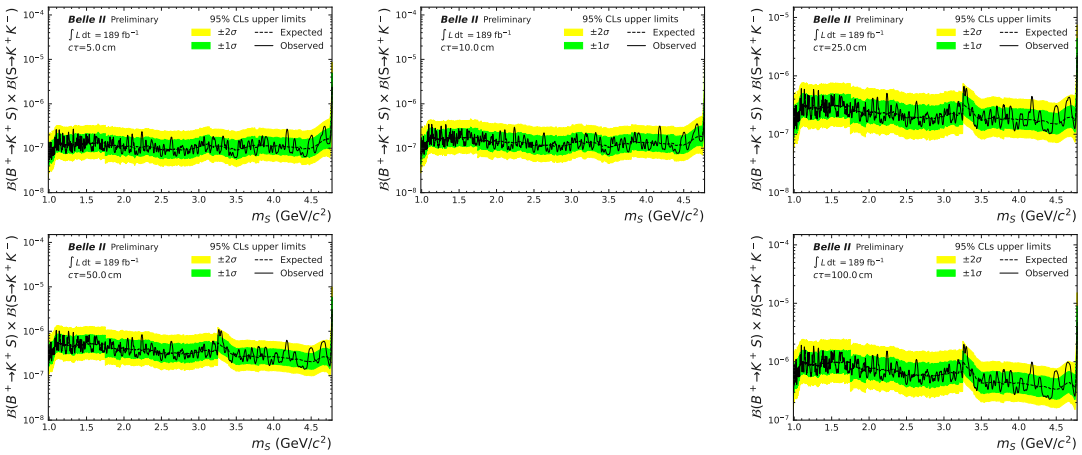

Figure A.14.: Expected and observed limits on the product of branching fractions $\mathcal{B}(B^+ \to KS) \times \mathcal{B}(S \to K^+K^-)$ for lifetimes $5 < c\tau < 100$ cm.

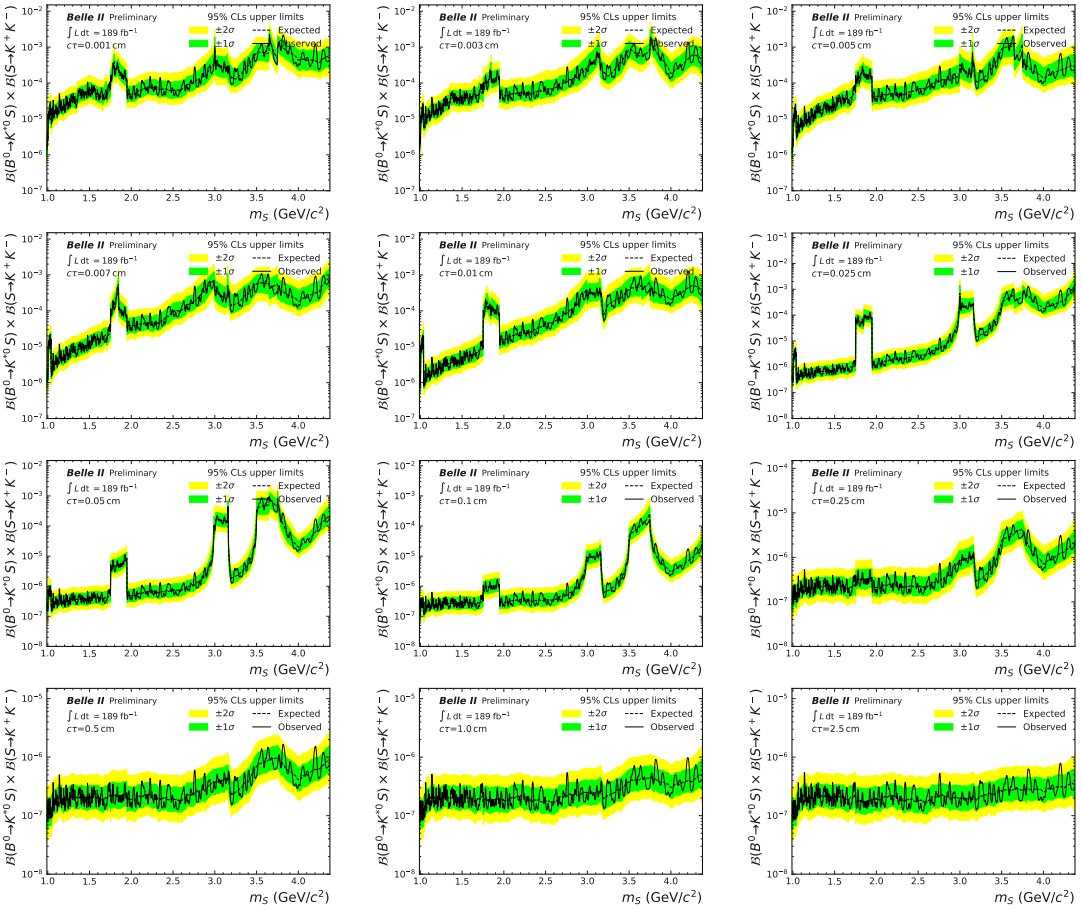

Figure A.15.: Expected and observed limits on the product of branching fractions $\mathcal{B}(B^0 \to K^{*0}(\to K^+\pi^-)S) \times \mathcal{B}(S \to K^+K^-)$ for lifetimes $0.001 < c\tau < 2.5$ cm.

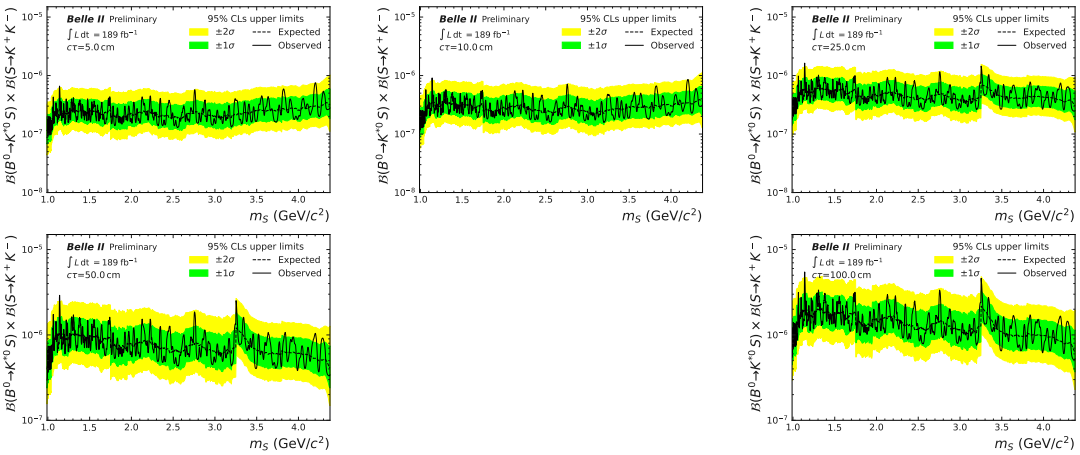

Figure A.16.: Expected and observed limits on the product of branching fractions $\mathcal{B}(B^0 \to K^{*0}(\to K^+\pi^-)S) \times \mathcal{B}(S \to K^+K^-)$ for lifetimes $5 < c\tau < 100$ cm.

Disclaimer

Data analyses in high energy physics such as the search presented in this doctoral thesis are only possible due to a collaborative effort of many involved people. The particle beams that are collided to study physics process are provided by the SuperKEKB accelerator collaboration. The Belle II detector was built and is operated by the Belle II collaboration. This collaboration also develops and maintains the software that is used to record and process the experimental data as well as implements toolkits to simulate various physics processes. The author of this thesis is part of the Belle II collaboration since 2019 and performed all studies detailed in this thesis except for the following:

- The simulation of background processes introduced in Sec. 4.3.1.
- The choice of parameters used in the bremsstrahlung correction detailed in Sec. 5.2.
- The determination of correction factors and associated uncertainties on:
  - The particle identification described in Sec. 7.2.1.
  - The track momentum described in Sec. 7.2.2.
- The determination of the systematic uncertainties due to:
  - The track finding efficiency for charged particles originating from close to the interaction point described in Sec. 9.1.1.
  - The number of B-meson pairs including the absolute value and the fractions of charged and neutral pairs described in Sec. 9.1.4.
  - The track momentum resolution detailed in Sec.9.1.6.

An article summarising the results of the search was written by Prof. Dr. Torben Ferber and the author of the thesis and submitted to PRL and as preprint [176] on June 05, 2023.

Acknowledgements

I would like to thank Elisabeth, as well as Inga and Hartmut for their unconditional support and confidence in me. Without them paving the way, this thesis would not have been possible.

On an academic level, my first thanks goes to Torben, who provides just the right amount of guidance, accepted to be a little more strict from time to time, is dedicated to seek opportunities for his mentees, and works hard to make high energy physics more fair for everybody.

A similar thanks is extended again to Oleg for his ongoing support and boosting my early scientific career.

I feel gracious to have spent my time working on this project alongside many awesome people, among them Anselm, Cyrille, Filippo, Henrikas, Marcel, Munira, Sam and Sally.

Many thanks go to the people who provided feedback to earlier versions of this manuscript: Arthur, Alberto, Cyrille, Eldar, Filippo, Henrikas, Marcel, Munira and Sam.

I would not have arrived at this point without the friendship, support, and occasional drink with Chris, Clemens, Felix, Jonas, Lars and Tobi.

I also want to express my gratitude to the Belle II group at ETP for making me feel welcome.

Declaration

I hereby declare upon oath that I have not submitted a dissertation with the same research topic to an academic higher education institution that had been already accepted or evaluated as insufficient in an earlier doctoral procedure.

Ich versichere, dass ich keine Dissertation mit dem gleichen Forschungsthema schon einmal in einem fruheren Promotionsverfahren an einer wissenschaftlichen Hochschule eingereicht habe, die angenommen oder als ungenugend beurteilt worden ist.

I hereby declare upon oath that I have written the present dissertation independently and have not used further resources and aids than those stated in the dissertation.

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

I hereby declare upon oath that the dissertation submitted in electronic form and the printed bound copy of the dissertation submitted for archiving are identical.

Ich versichere an Eides statt, dass die in elektronischer Form eingereichte Dissertation und das zur Archivierung eingereichte gedruckte gebundene Exemplar der Dissertation identisch sind.

Sascha Simon Dreyer
Hamburg, 07 July 2023