

Searching for beyond the Standard Model physics using direct and indirect methods at LHCb

Samuel C. P. Hall

Imperial College London

Department of High Energy Physics

HZP Comp

A thesis to be submitted to Imperial College London
for the degree of Doctor of Philosophy

Abstract

It is known that the Standard Model of particle physics is incomplete in its description of nature at a fundamental level. For example, the Standard Model can neither incorporate dark matter nor explain the matter-dominated nature of the Universe. This thesis presents three analyses undertaken using data collected by the LHCb detector. Each analysis searches for indications of physics beyond the Standard Model in different decays of B mesons, using different techniques. Notably, two analyses look for indications of new physics using indirect methods, and one uses a direct approach.

The first analysis shows evidence for the rare decay $B^+ \rightarrow D_s^+ \phi$ with greater than 3σ significance; this also constitutes the first evidence for a fully-hadronic annihilation-type decay of a B^+ meson. A measurement of the branching fraction of the decay $B^+ \rightarrow D_s^+ \phi$ is seen to be higher than, but still compatible with, Standard Model predictions. The CP -asymmetry of the decay is also measured, and its value is precisely in line with the Standard Model expectations.

The second analysis claims the first observations of the decays $B^+ \rightarrow K^+ \pi^+ \pi^- \mu^+ \mu^-$ and $B^+ \rightarrow \phi K^+ \mu^+ \mu^-$ which are both flavour changing neutral currents, forbidden at leading order in the Standard Model. Branching fractions of both these decays are measured, and for the high statistics channel $B^+ \rightarrow K^+ \pi^+ \pi^- \mu^+ \mu^-$ the differential branching fraction, as a function of the invariant mass squared of the dimuon system, is also presented.

These first two analyses both constitute indirect searches for physics beyond the scope of the Standard Model, where the observables are sensitive to contributions from new physics entering at loop-level. In contrast, the third analysis presented in this thesis involves the direct search for a new dark boson, χ , which is a messenger particle between a dark sector and the Standard Model particles. Using a frequentist technique, the dimuon component of candidates of the decay $B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$ for an excess consistent with $\chi \rightarrow \mu^+ \mu^-$.

Chapter 2

The Standard Model and beyond

“Before beginning a Hunt, it is wise to ask someone what you are looking for before you begin looking for it.”

Winnie the Pooh, A.A. Milne

This thesis contains the work undertaken in three analyses; each of which concerns a different area of interest in high energy physics. The following chapter aims to motivate each analysis in turn after introducing the Standard Model of particle physics.

Firstly, the formulation of the SM will be outlined, with particular detail paid to the flavour sector. Various successes of the SM will then be discussed before going on to identify its shortcomings using arguments from both experiment and theory. These shortcomings will then be used to motivate the three analyses: a search for the decay $B^+ \rightarrow D_s^+ \phi$ (Chap. 5); a search for the two related decays $B^+ \rightarrow K^+ \pi^+ \pi^- \mu^+ \mu^-$ and $B^+ \rightarrow \phi K^+ \mu^+ \mu^-$ (Chap. 6); a search for dark sector particles in $B^0 \rightarrow K^*(892)^0$ (Chap. 7). Theory specific to each of these analyses will be detailed in the relevant chapter.

2.1 The Standard Model

The behaviour of fundamental particles and forces are described by the SM of particle physics, which was concocted in the 1970s, when the Higgs mechanism was incorporated into Glashow’s electroweak theory by Salam and Weinberg. The theory prescribes a treatment as to how fundamental particles interact via three of the four fundamental forces, namely: the strong, weak and electromagnetic forces.

Mathematically, the SM is a locally gauge invariant quantum field theory. It inhabits a space-time with a global Poincaré symmetry that obeys a local $SU_C(3) \otimes SU_L(2) \otimes U_Y(1)$

Perhaps say whose framework
you are following -

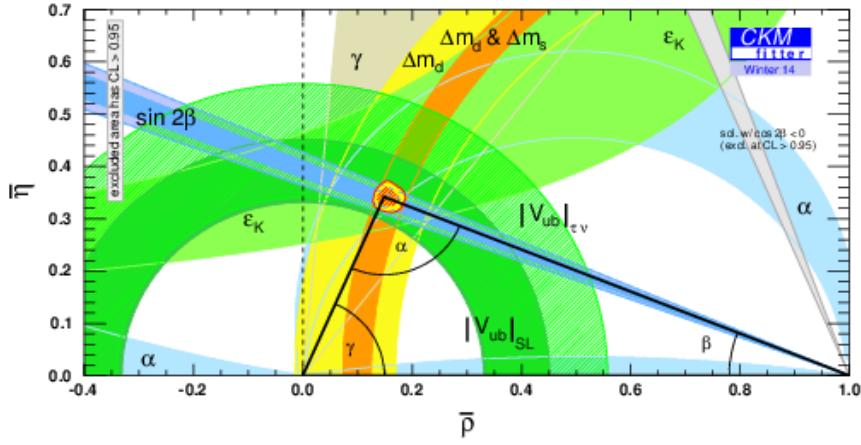


Figure 2.3: Diagram of the UT with coloured bands indicating various constraints on side lengths, angles and position of the apex, which is taken from the CKMfitter group in Ref. [25]. The constraints on V_{ub} from the combination of inclusive and exclusive modes ($|V_{ub}|_{SL}$) is given separately to a value obtained using $\mathcal{B}(B^+ \rightarrow \tau^+ \nu_\tau)$, ($|V_{ub}|_{\tau\nu}$).

measurements. A measurement from the LHCb experiment uses the baryonic decay $\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$ calculated a value of $|V_{ub}|$ to be $(3.27 \pm 0.23) \times 10^{-3}$ [22]. This is an exclusive measurement, and is in agreement with other exclusive measurements.

Another method to access the CKM matrix parameter $|V_{ub}|$ is via the annihilation-type decay $B^+ \rightarrow \tau^+ \nu_\tau$. Measurements from both the BaBar and Belle experiments of $\mathcal{B}(B^+ \rightarrow \tau^+ \nu_\tau)$ [23, 24] suffer from small statistics, but are found to be in better agreement with values of $|V_{ub}|$ determined using inclusive measurements than exclusive. Searching for the decay $B^+ \rightarrow \tau^+ \nu_\tau$ is not viable at LHCb; instead, decays of the same topologies can be searched for. The decay $B^+ \rightarrow D_s^+ \phi$ is also an annihilation-type decay in which V_{ub} appears in the amplitude; an analysis of this decay is described in Chap. 5.

Current measurements of angles and side lengths of the UT, from Ref. [25], are shown in Fig. 2.3. This figure also shows global V_{ub} measurements from the semi-leptonic and $B^+ \rightarrow \tau^+ \nu_\tau$ modes alongside one another.

Unnatural NP models with parameters that differ wildly in magnitude tend to lead to parameters or processes that must cancel to absurdly high precision in order to agree with experimental observations. These precise cancellations are known as *fine tuning*. In the SM, quantum loop corrections to the Higgs mass are of the order 10^{19} for $m_H \simeq 125$ GeV [8, 9]. This means that the cancellations required to result in a Higgs mass comparable to the masses of the weak vector bosons must be exact to 17 orders of magnitude. This instance of fine tuning is known as the *hierarchy problem*. A solution for the hierarchy problem is to introduce NP particles, whose contributions to loop level processes reduce the magnitude of fine tuning required to a level deemed acceptable. SUSY immediately solves the hierarchy problem because for every SM particle that contributes to the Higgs mass, a SUSY particle

Explain
for
non-LHCb
people

Chapter 3

The LHCb experiment

The following chapter first introduces the LHC and the LHCb detector before briefly describing data collection and processing. Other important features, such as particle identification and triggering, of the LHCb experiment will also be outlined.

3.1 The LHC

The LHC is a superconducting synchrotron which can simultaneously accelerate beams of proton bunches in opposite directions. Physically, the LHC is located at CERN, near Geneva, Switzerland; it is 27 km in circumference and spans the Franco-Swiss border at a depth of about 100 m. Protons are supplied to the LHC from the Super Proton Synchrotron (SPS) with an energy of 450 GeV, they can then be accelerated and collided with a centre-of-mass energy of up to 14 TeV. In the years 2011 and 2012 collisions operated with a centre-of-mass energy was 7 and 8 TeV respectively. Once the desired energy is reached the beams are collided at four interaction points. The LHCb detector is situated at one of them [32]. Collisions of proton bunches occur every 50 ns reaching luminosities of up to $7 \times 10^{32} \text{ cm}^2\text{s}^{-1}$, however the beams entering LHCb must be luminosity levelled, to $3(4) \times 10^{32} \text{ cm}^2\text{s}^{-1}$ in 2011(2012), in order to reduce detector occupancy. These high energy collisions produce a vast number of $b\bar{b}$ pairs which subsequently hadronize into b hadrons. It is the prospect of studying these bound states of b (and other heavy flavour) quarks that has motivated the design of the LHCb detector.

3.2 The LHCb detector

Before introducing the LHCb detector, it is helpful to first define the Cartesian coordinate system around which the LHCb detector is built. The z direction is defined by the LHC

collected in 2012. In total, the data collected in 2011 and 2012 is known as Run-1 data.

Even the ~~much~~ much reduced, HLT2 output rate of 5 kHz is a vast amount of data for an analyst to sift through in a timely manner. To improve the speed to access data, additional selections are applied to the dataset biannually which further categorize each event. This is known as stripping. Stripped datasets are the only ones accessible to analysts, which makes the process of retrieving data of interest fast. Stripping selections in this thesis vary, and will be described when appropriate.

Reliable analysis of real data would not be possible without selections of simulated events. These allows collaborator's access to pure samples of specific, requested decays to aid their research. This can be for the evaluation of efficiencies, understanding effects in data, or making analysis decisions without compromising blinded data. These events are generated in two independent phases: generation and simulation. Proton-proton collisions are generated using PYTHIA [36] with a specific LHCb configuration [37], and subsequent hadronic decays are handled by the EVTGEN [38] package. The simulation phase is designed to mimic the LHCb detector's response to particles, this is done with GEANT4 [39] as described in Ref. [40]. Simulated events after the hadronization stage and before detector modelling are known as *generator level* simulation.

- ④ Summarise detector performance
conditions etc Handwritten
- ④ Explain simulations in detail
- ④ Reference for later chapter

Chapter 4

Multivariate selection techniques

The analyses detailed in Chap. 5, Chap. 6 and Chap. 7 make prodigious use of multivariate techniques to reduce combinatorial backgrounds. Combinatorial backgrounds are formed from random combinations of tracks which appear to form a vertex, pass selection criteria, and satisfy relevant PID assignments. To remove these backgrounds Multivariate Analysis (MVA) techniques can be employed. A multivariate discriminator exploits correlations between weakly discriminating variables to produce a single, more separating, classifier.

MVA techniques used in HEP tend to be supervised learning algorithms, whereby a selection of events are given input, and an algorithm produces a response based on how best to separate them. Input into the algorithms to separate a background are: a sample of the signal and background candidates that must be separated, and a set of variables to be used to do so. Samples of events are split in two; some are used for training the MVA, and the remaining are used for testing it. The input, or training, variables define an n -dimensional space populated by the input samples. The algorithm then classifies regions in this n -dimensional space as signal- or background-like; such that an arbitrary event placed somewhere in the space would also be classified based on the point it inhabits. The Boosted Decision Tree (BDT) algorithm is used throughout this thesis because it can handle a weighted training sample, including negative weights, and can exploit non-linear correlations between variables [41, 42].

A BDT is composed of a combination of numerous Decision Trees (DTs), each of which is a classifier in its own right — albeit a weak one — being able to distinguish between high density regions of signal and background populations.

Training a DT begins with a single parent node populated by the whole training sample, which inhabits the parameter space defined by the variables x_i , whose true distribution is $f(x_i)$. The sample on the parent node is split by selecting a cut based on maximizing some figure-of-merit. Child nodes are then split, and the process is repeated until there is no possible improvement in the separation of signal and background. The definition of

improvement is usually related to the signal and background purity of a node:

$$p_{\text{sig}} = (1 - \varepsilon_{\text{bkg}}) \quad p_{\text{bkg}} = (1 - \varepsilon_{\text{sig}}). \quad (4.1)$$

*defn Es,
dark eqns*

figure?

The final child nodes, or leaves, are each associated with signal or background depending on the purity of the sample which populates it. Each leaf therefore maps out an area in n -dimensions, and is classified as a signal or background leaf depending on the purity of the training sample enclosed by that area. The hypothesised category, as output by the DT, $h(x_i)$, will ideally be equal to $f(x_i)$, but in reality there will be events which are misclassified. A figure-of-merit which is often used to determine the cut used at each node is the G_{ini} index, which is defined as

$$G_{ini} = 2p_{\text{bkg}}p_{\text{sig}} = 2(1 - \varepsilon_{\text{sig}})(1 - \varepsilon_{\text{bkg}}) = \frac{2sb}{(s + b)^2}, \quad (4.2)$$

where s and b are the weighted sum of signal and background candidates, respectively, after a given cut.

Decision Trees have the advantages over other machine learning algorithms — such as neural nets — of being able to deal with weighted training samples, and being insensitive to variables with very little separation power because the G_{ini} index never identifies a cut on them as being profitable. However, DTs are sensitive to statistical fluctuations in the training sample. To negate this problem DTs can be *Boosted* using any one of a number of algorithms. The procedure of boosting removes the power that statistical fluctuations has over the final BDT.

A different boosting method is used to train the BDTs in each analyses in this thesis. The algorithms used are outlined in the remainder of this chapter.

4.1 The bagging algorithm

Bootstrap aggregating, or bagging¹, is a method of boosting whereby the effects of statistical fluctuations are negated by making many independent DTs and using them to make a decision based on the average response. Training n DTs, it is possible to define three errors, namely: square of the error of a single estimator

$$\epsilon_t(x_i) = (f(x_i) - h_t(x_i))^2 \quad (4.3)$$

¹ This section is based on Ref. [43].

Chapter 5

Search for the decay $B^+ \rightarrow D_s^+ \phi$

5.1 Introduction

In the SM the decay $B^+ \rightarrow D_s^+ \phi$ proceeds via the annihilation of the constituent b and u quarks of a B^+ meson forming a virtual W^+ boson from the CC interaction, the processes is suppressed by the CKM matrix element V_{ub} ¹. To achieve the final state, the W^+ decays into a $c\bar{s}$ pair and an additional $s\bar{s}$ pair must be created from the QCD field. This is the only diagram that can perpetuate such a decay at tree-level, because the initial state quarks are all different to those in the final state. A Feynman diagram of the decay $B^+ \rightarrow D_s^+ \phi$ is shown in Fig. 5.1, where the final state mesons can be formed in the way indicated, or the $s\bar{s}$ pair from the QCD field can form the ϕ , although this is colour-suppressed. Also, the gluon that forms the $s\bar{s}$ pair can originate from any of the initial or final state quarks. This analysis was published in Ref. [1].

Annihilation decays of B^+ mesons are rare in the SM due to the magnitude of $|V_{ub}| \sim 4 \times 10^{-3}$. In fact, no fully hadronic decays proceeding via annihilation-type diagrams have yet been observed.

Predictions for the branching fraction $\mathcal{B}(B^+ \rightarrow D_s^+ \phi)$ are calculated using the OPE defined by the effective Hamiltonian [47–50]:

$$\mathcal{H}_{\text{eff}} = -4 \frac{G_F}{\sqrt{2}} V_{ub} V_{cs}^* [C_1(\Lambda) \mathcal{O}_1 + C_2(\Lambda) \mathcal{O}_2] \quad (5.1)$$

where

$$\begin{aligned} \mathcal{O}_1 &= (\bar{b}\gamma_\mu P_L u)(\bar{c}\gamma_\mu P_L s) \\ \mathcal{O}_2 &= (\bar{b}\gamma_\mu P_L s)(\bar{c}\gamma_\mu P_L u). \end{aligned} \quad (5.2)$$

¹ All mentions of the ϕ meson refers to the $\phi(1020)$.

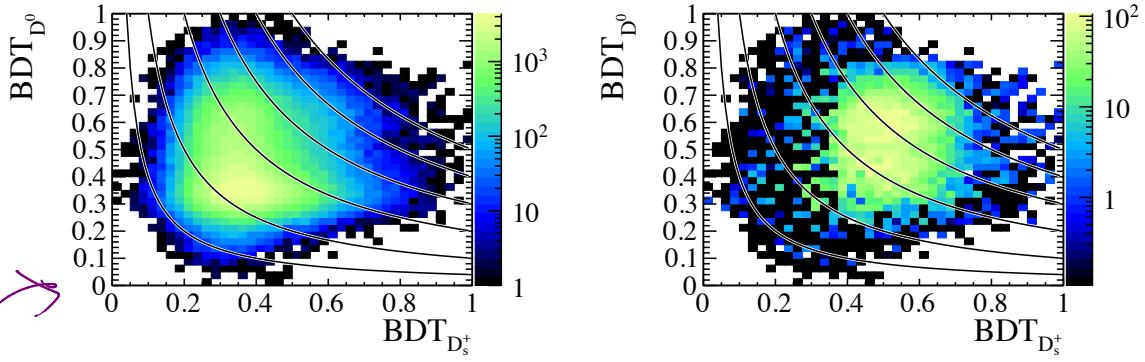


Figure 5.2: An example distribution of the response for a BDT_{D^0} against a $\text{BDT}_{D_s^+}$ for a sample of (left) $B^+ \rightarrow D_s^+ \bar{D}^0$ candidates where the background dominates and occupies the lower left of the plots, and (right) the same candidates after sWeighting has been applied where the signal peak is observed further towards the upper right. The lines overlaid on the plot show the boundaries for the cuts of $\text{BDT}_{D_s^+} \times \text{BDT}_{D^0} > 0.01, 0.04, 0.10, 0.20, 0.30, 0.40$, and 0.50 . It is seen that using the product of the BDT discriminants is more effective for rare decays where a tighter cut will be needed, in this region candidates are selected if their response is particularly positive, at the expense of the other meson.

of using a BDT to identify each meson is also used in Ref. [53], which measures branching fraction ratios of various $B \rightarrow DD$ decays. The bagging boosting technique used gives a response between zero and one. Therefore, it is natural to cut on the product of the two BDT responses, $\text{BDT}_{D_s^+} \times \text{BDT}_\phi > X$, as opposed to $\text{BDT}_{D_s^+} > X_1$ and $\text{BDT}_\phi > X_2$. Cutting on the product of the BDTs improves the performance of the selection, because a very strong D_s^+ candidate (for example) will be selected at the expense of a slightly weaker ϕ selection, this is particularly true for tighter cuts. Figure 5.2 shows the effect of cutting on $\text{BDT}_{D_s^+} \times \text{BDT}_{D^0}$ in the normalization channel $B^+ \rightarrow D_s^+ \bar{D}^0$.

The $D_s^+ \rightarrow K^+ K^- \pi^+$ BDT was trained using $D_s^+ \rightarrow K^+ K^- \pi^+$ decays from data taken from the high statistics channel $\bar{B}_s^0 \rightarrow D_s^+ \pi^-$. The signal sample of D_s^+ decays came from the $D_s^+ \rightarrow K^+ K^- \pi^+$ candidates that fell within 3σ of the known D_s^+ mass, the sample was background subtracted using the sWeighting technique [54] on the B^+ mass spectrum. Background data was taken from candidates falling in the upper mass sideband of the B^+ , and either sideband of the D_s^+ . Similarly, for the ϕ BDT, the signal sample was sWeighted and the background comes from the ϕ mass sidebands; but the sample is taken from the high statistics $B_s^0 \rightarrow J/\psi \phi$ mode.

In total, there are five kinematic and geometric training variables for the parent meson. For the daughter tracks there are a total of 23 variables, including kinematic, geometric and PID variables. Since the BDT was trained using data, it is possible to use PID variables that are poorly described in simulation. A summary of all training variables is given in Table 5.2.

The cut for the BDT was optimized using the metric $S/\sqrt{S + B}$, In this case, the number

Table 5.2: List of training variables used in the D_s^+ and ϕ BDTs. Each BDT uses five variables associated with the parent particle and 23 variables from each daughter track.

Particle		Variable
D_s^+, ϕ	Kinematic variables	p, p_T
	Geometric variables	$\chi_{\text{vtx}}^2, \chi_{\text{IP}}^2, \chi_{\text{FD}}^2$
Tracks	Kinematic variables	p, p_T
	Geometric variables	$\min(\chi_{\text{IP}}^2)$
	Track variables	4 variables characterizing the track quality
	PID variables	16 variables containing PID information, such as <code>isMuon</code> and DLL variables from the RICH detectors

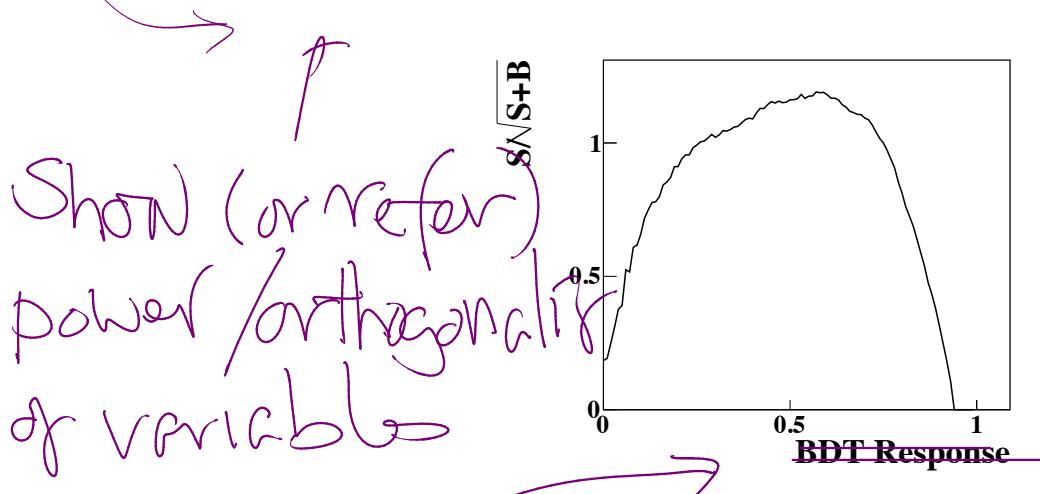


Figure 5.3: Value of the figure of merit $S/\sqrt{S + B}$ is shown as a function of the BDT response, $\text{BDT}_{D_s^+} \times \text{BDT}_\phi$. The maximum value of the figure of merit is 0.57, which is chosen as the final BDT cut.

of signal events, S , was estimated from the yield from the decay $B_s^0 \rightarrow D_s^- \pi^+$, according to:

$$S = \frac{\mathcal{B}(B^+ \rightarrow D_s^+ \phi)}{\mathcal{B}(\bar{B}_s^0 \rightarrow D_s^+ \pi^-)} \frac{\varepsilon_{\text{gen}}(B^+ \rightarrow D_s^+ \phi)}{\varepsilon_{\text{gen}}(\bar{B}_s^0 \rightarrow D_s^+ \pi^-)} \frac{f_d}{f_s} N(\bar{B}_s^0 \rightarrow D_s^+ \pi^-), \quad (5.5)$$

where

f_s/f_d quantifies the fraction of B_s^0 mesons produced relative to B^0 mesons. The generator level efficiency, ε_{gen} , is the efficiency introduced by the acceptance region of the LHCb detector, and the necessity that all daughter particles must travel through the detector. Background yield for a given cut is estimated as:

$$B = c \cdot N_c(B_s^0 \rightarrow D_s^- \pi^+) \cdot N_c(B_s^0 \rightarrow J/\psi \phi), \quad (5.6)$$

where N_c indicates the yield of the combinatoric background for the indicated decay, and c is a constant scaled such that $N_c(B_s^0 \rightarrow D_s^- \pi^+) \cdot N_c(B_s^0 \rightarrow J/\psi \phi) = N_c(B^+ \rightarrow D_s^+ \phi)$ with no BDT cut. The optimization procedure results in the optimal cut as $\text{BDT}_{D_s^+} \times \text{BDT}_\phi > 0.57$, as is shown in Fig. 5.3.

Both the signal and normalization channels have a $D_s^+ \rightarrow K^+ K^- \pi^+$, and therefore efficiencies from the D^+ and A_c^+ vetoes cancel to a large extent. Given that these cuts are very efficient, the difference is negligible, and the efficiency ratio is assumed to be unity for the calculation. The only differences between the two modes are due the decay kinematics, since the ϕ is lighter than the D^0 . There is a difference of 1.5 % between $\varepsilon_{\text{veto}}$ for the decays $\bar{B}_s^0 \rightarrow D_s^+ \pi^-$ and $B^+ \rightarrow D_s^+ \bar{D}^0$, and since the mass of the ϕ is nearer the D^0 mass than the pion mass, a systematic uncertainty of 1 % is assigned.

There are other systematic uncertainties that affected the selection. The BDT cut is assigned an uncertainty of 3 %, which is due to the sizes of the validation samples used to calculate the efficiencies. Mass windows around the D_s^+ and ϕ lead to a 3 % systematic uncertainty. Also, the low statistics of the simulation samples used to deduce efficiencies led to a 3 % uncertainty.

The mass fits introduced systematic uncertainties from each component. If the parameters describing the signal shape are allowed to float the fit results in a yield which is 5 % higher than the nominal fit. This is assigned as a systematic uncertainty.

The total uncertainty from the background shape is 5 %; and is estimated by making changes to the background model. By removing either the $B^+ \rightarrow D_s^{*+} \phi$ or $\bar{B}_s^0 \rightarrow D_s^{*+} K^{*0} K^-$ components, the yield changes by only 1 %. Changing the constraints on **A/B** and **C/D** for $\bar{B}_s^0 \rightarrow D_s^+ K^{*0} K^-$ by a factor of 2 results in a 1 % change in signal yield. The combinatorial background is estimated by allowing the slope to float free, this led to an approximate 3.5 % systematic uncertainty.

Contributions from all sources of systematic uncertainties are summarized in Table 5.6. The dominant systematic uncertainty — discounting the uncertainties on the branching fraction of the normalization channel — is from the mass fits, which is unsurprising considering the complexity of the fit, and the treatment of the backgrounds. Regardless of the constraints that are, or are not, included in the fit: the lowest significance obtained is still greater than 3σ . Therefore the significance is quoted as greater than 3σ .

Doorbells feature principle to constrain effect of BDT cut value

5.5 Direct CP asymmetry

The CP asymmetry is defined in Eq. 5.4, but this must be modified to account for background in the sample:

$$\mathcal{A}_{CP}(B^+ \rightarrow D_s^+ \phi) = \frac{N(B^- \rightarrow D_s^- \phi) - N(B^+ \rightarrow D_s^+ \phi)}{N(B^- \rightarrow D_s^- \phi) + N(B^+ \rightarrow D_s^+ \phi) - N_{\text{bkg}}}. \quad (5.10)$$

Here, the yields, N , refer to the decay with the indicated charge, and N_{bkg} denotes the amount of background contaminating the signal.

Chapter 6

Search for the decays

$B^+ \rightarrow K^+ \pi^+ \pi^- \mu^+ \mu^-$ and

$B^+ \rightarrow \phi K^+ \mu^+ \mu^-$

6.1 Introduction

The decays $B^+ \rightarrow K^+ \pi^+ \pi^- \mu^+ \mu^-$ and $B^+ \rightarrow \phi K^+ \mu^+ \mu^-$ both are $b \rightarrow s \mu^+ \mu^-$ FCNC transitions, which are forbidden at tree-level in the SM¹. Therefore, these processes are sensitive to virtual NP particles contributing to the decay amplitude in loops. The analysis described in the following chapter makes a measurement of the branching fractions of both decays and the differential branching fraction of $B^+ \rightarrow K^+ \pi^+ \pi^- \mu^+ \mu^-$ in bins of q^2 , where q^2 is the invariant mass of the dimuon system squared. This analysis was published in Ref. [2].

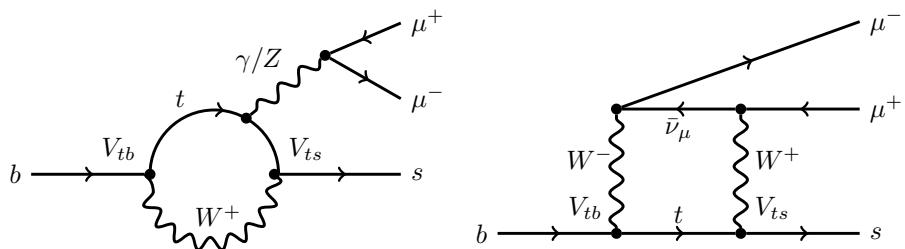


Figure 6.1: Schematic Feynman diagrams for the operators \mathcal{O}_7 , \mathcal{O}_9 , and \mathcal{O}_{10} which are most sensitive to the $b \rightarrow s \mu^+ \mu^-$ FCNC. The propagators are the (left) photonic and Z penguin diagram, and the (right) W^+ -mediated box diagram. Operator \mathcal{O}_7 describes the photonic penguin diagram; while \mathcal{O}_9 and \mathcal{O}_{10} are the vector and axial-vector parts of both the Z and W^+ diagrams.

¹All mentions of the ϕ refer implicitly to the $\phi(1020)$ meson.

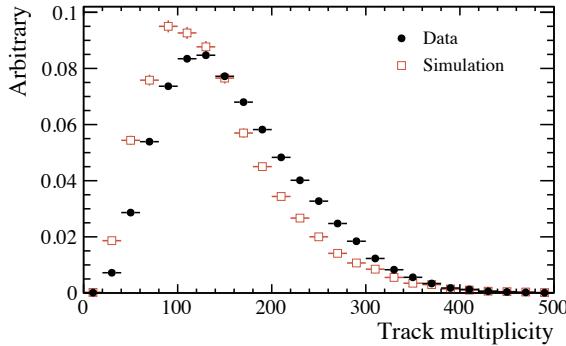


Figure 6.7: Distributions of track multiplicity for (black circles) data and (red squares) simulation. Simulated events are known to mis-model the track multiplicity, having a lower average number of tracks per event.

state where the discrepancy comes from

of this resampling technique, it is observed that the simulated PID distributions that have been resampled matches data distributions much better than the raw distributions; the differences that remain are accounted for in the systematic uncertainties. Muon PID distributions are sufficiently well modelled in simulation that correcting them is unnecessary.

Tracking efficiency varies depending upon the regions of the detector through which the particle passes, and the modelling of the detector in the simulation behaves differently to in actuality. To correct for this, each candidate is weighted based on the relative tracking efficiency between data and simulation, this is dependent upon p and η . The same is true for the response of the `isMuon` variable. Figure 6.9 shows how the tracking efficiency and `isMuon` criteria are corrected for throughout the detector volume. After all reweighting all the aforementioned variables in simulated events, the BDT distributions are seen to be in agreement, this is shown in Fig. 6.10.

Once the simulation has been corrected, the total efficiency, ε_{tot} was calculated for each normalization and signal mode using simulated events. The value for ε_{tot} is calculated to be $\varepsilon_{\text{gen}} \times \varepsilon_{\text{reco\&sel}} \times \varepsilon_{\text{trig}}$, where: ε_{gen} is the generator selection efficiency; $\varepsilon_{\text{reco\&sel}}$ is the reconstruction and selection efficiency; $\varepsilon_{\text{trig}}$ trigger efficiency. The generator efficiency defines the probability that a B^+ decays into daughter particles which all pass through the LHCb detector acceptance, and is approximately 15 % for each signal and normalization channel.

Since efficiency calculations require reliable simulated samples of decays, accurate physics models must, or should, be used. This raises a dilemma deciding how to calculate the efficiency of the signal decays $B^+ \rightarrow K^+\pi^+\pi^-\mu^+\mu^-$ and $B^+ \rightarrow \phi K^+\mu^+\mu^-$, because no physics models exist for them. For the decay $B^+ \rightarrow K^+\pi^+\pi^-\mu^+\mu^-$, an appropriate choice was the physics model for $B^+ \rightarrow K_1(1270)^+\mu^+\mu^-$ from Ref. [70] and $\theta_{K_1} = 34^\circ$, because this was assumed to be a dominant contribution. As there is no available physics model

Chapter 7

Search for the decay of a dark sector particle $\chi \rightarrow \mu^+ \mu^-$ in $B^0 \rightarrow K^{*0} \mu^+ \mu^-$

7.1 Introduction

This chapter describes a procedure for searching for a dark boson, χ , of unknown mass and lifetime¹. A frequentist method is applied to the dimuon distribution of $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ candidates, to search for an excess of events above the SM background, consistent with a χ decaying into a pair of muons. Lifetime information is added by splitting candidates into two bins of decay time: those which are prompt, and the χ vertex is the same as the K^{*0} vertex; and those which are displaced.

Chapter 2 explains that the SM cannot explain the numerous experimental observations of DM. Little is known about dark matter, except that it interacts gravitationally, and does not interact with electromagnetic radiation to any significant extent. A possible extension to the SM is to introduce a dark sector, which can contain a rich variety of distinct particles operating through forces that are hitherto unknown. Dark sector particles would be gauge singlet states with respect to the SM, and only be able to communicate with known particles via weakly interacting messenger particles through one of four *portals*: the vector, axion, Higgs, and neutrino portals [73]. Interaction terms for messengers in each of these portals are given in Table 7.1.

Theories involving dark sectors are tremendously attractive because it is relatively easy to construct a complex theory that explains various unexplained phenomena. ~~Yet, these can have little impact on the SM observables since the interaction between the two sectors can be extremely weak.~~

The Higgs portal has a scalar messenger particle which can mix with the SM Higgs. There

¹ Throughout this chapter, the symbol χ shall denote a general dark boson and all references to a K^{*0} will be implicitly referencing the $K^*(892)^0$; unless explicitly stated otherwise.

Table 7.1: A summary of portals through which a new dark boson could operate, as given in Ref. [73]. Terms are defined as: $F_{\mu\nu}$ is the field strength tensor of the photon; $F'_{\mu\nu}$ is the dark photon field; ϵ characterizes mixing between the SM and the dark photon; f_χ is scale at which Peccei-Quinn global $U(1)$ symmetry is spontaneously broken; $G_{\mu\nu}$ is the gluon field strength tensor; S is a dark scalar field with coupling strengths μ and λ to the Higgs field; and the sterile neutrino couples to a H with a strength Y_N .

Portal	Particles	Operator(s)	
Vector	Dark photons	$-\frac{\epsilon}{2 \cos \theta_W} F_{\mu\nu} F'^{\mu\nu}$	
Axion	Pseudoscalars	$\frac{\chi}{f_\chi} F_{\mu\nu} \tilde{F}^{\mu\nu}$	$\frac{\chi}{f_\chi} G_{i\mu\nu} \tilde{G}_i^{\mu\nu}$
Higgs	Dark scalars		$(\mu\chi + \lambda\chi^2) H^\dagger H$
Neutrino	Sterile neutrinos		$Y_N \ell H \chi$

are a number of models which incorporate a scalar messenger particle that interacts with the Higgs. One class of models incorporate an *inflaton*, which is the quanta of the hypothesised inflaton field responsible for the inflationary period of the Universe; beginning around $t = 10^{-36}$ s. Figure 7.1 shows the allowed parameter space of the mixing angle between the DM and Higgs, θ , as a function of mass. It is possible that inflatons are light, in the range $270 < m_\chi < 10^4$ MeV [74], and might therefore be accessible in the decay $B^0 \rightarrow K^{*0} \chi$. These models also help to solve other problems, such as the BAU [75, 76].

Chapter 2 introduces the idea of PQ symmetry breaking leading to an axion which resolves the strong CP problem. Unlike other dark boson portals, the axion portal introduces a term in the Lagrangian which couples messenger axions to fermions directly. In order for the axion portal to couple to a dark sector containing TeV-scale DM, the messenger particles are predicted to have a mass in the range $360 < m_\chi < 800$ MeV and a decay constant in the range $1 \lesssim f_\chi \lesssim 3$ TeV [78]. Figure 7.2 shows a Feynman diagram of how the decay $B^0 \rightarrow K^{*0} \chi$ might proceed; it shows an FCNC where the χ results from a coupling to a t quark. Therefore, searching for evidence of a $B^0 \rightarrow K^{*0} \chi$ where $\chi \rightarrow \mu^+ \mu^-$ is particularly sensitive to portals which couple strongly to mass, as is the case for axions.

It is known that SM would be a gauged field theory, and it is therefore reasonable to assume that the dark sector is also gauged, but under a different group. If this were to be the case, the SM $U_Y(1)$ generator could kinetically mix with the generator of the dark $U(1)$ group, giving rise to a particle, often called a *dark photon*, interacting through the vector portal.

In principle, the following analysis is sensitive to any dark sector particle. Practically, other experiments have searched directly for dark bosons with mass-independent couplings using much larger data samples. For example, the NA48/2 collaboration has searched for a dark photon directly in the decay $\pi^0 \rightarrow \gamma A'$ [79], and the BaBar collaboration have searched for evidence in the decay $e^- e^- \rightarrow \gamma A'$ [80]. The coupling here is mass independent,

Search for the decay of a dark sector particle $\chi \rightarrow \mu^+ \mu^-$ in $B^0 \rightarrow K^{*0} \mu^+ \mu^-$

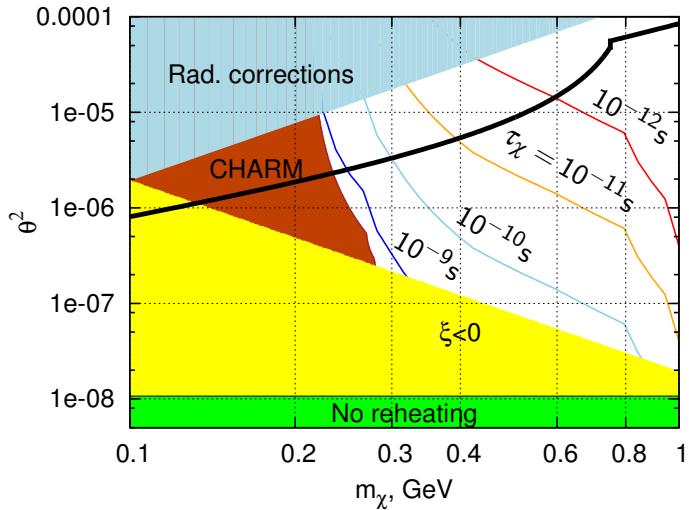


Figure 7.1: Allowed and excluded regions of the Higgs-inflaton mixing parameter-squared as a function of inflaton mass, taken from Ref. [77].

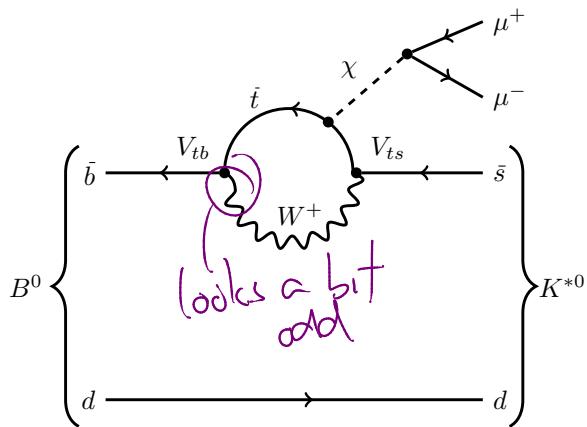


Figure 7.2: Feynman diagram showing the decay $B^0 \rightarrow K^{*0} \chi$, and $\chi \rightarrow \mu^+ \mu^-$. Depending on the portal through which the χ acts, it couples directly to the muons, or need to mix with a SM Higgs, Z , or photon.

Search for the decay of a dark sector particle $\chi \rightarrow \mu^+ \mu^-$ in $B^0 \rightarrow K^{*0} \mu^+ \mu^-$

and therefore this search is less sensitive to dark photons in comparison to these direct production searches.

SUSY could also be restricted, at energies reached thus far, to a dark sector. It is known that SUSY is a broken symmetry and thus must have associated Goldstone particles: a fermionic *goldstino* and associated super-partners called *sgoldstinos*, which are scalar and pseudoscalar. In some models, the sgoldstinos are the messenger particles between the SM and the dark SUSY sector. After SUSY breaking, the goldstino becomes the longitudinal component of the gravitino, and the sgoldstinos are massless particles, which gain mass from corrections at higher orders. Then, after electroweak symmetry breaking, these sgoldstinos interact with SM fermions via Yukawa-like interactions, but suppressed by the SUSY breaking parameter, F , [81]. This is interesting because although the sgoldstino masses are unknown, a measurement of its coupling to fermions would give access to F and the scale of SUSY breaking since $\sqrt{F} \sim \Lambda_{\text{SUSY}}$. The suppression of the coupling between the sgoldstinos and fermions means that the larger the scale Λ_{SUSY} , the longer the lifetime of the sgoldstinos.

Naïvely, one might expect that in the case that χ is a scalar or pseudoscalar, the decay $B^0 \rightarrow K^+ \mu^+ \mu^-$ to be more sensitive than $B^0 \rightarrow K^{*0} \mu^+ \mu^-$. The latter mode for a spin-0 χ requires orbital angular momentum of one, because the B^0 is a pseudoscalar and the K^{*0} is a vector, and therefore leads to some suppression due to a barrier factor. However, this suppression is only significant at high dimuon masses, close to threshold, because of there is plenty of phasespace in the decay $B^0 \rightarrow K^{*0} \chi$. A further complication of using the decay $B^+ \rightarrow K^+ \mu^+ \mu^-$ for this analysis is the lack of a good quality B^0 decay vertex. Decay rate predictions for a χ operating through the axion portal for decays of the type $B \rightarrow K \chi$ to be [82]:

$$\Gamma(B \rightarrow K \chi) = \Gamma_0 \frac{\lambda_K (m_B^2 - m_K^2)^2}{m_B^6} [f_0(m_\chi^2)] \quad (7.1)$$

$$\Gamma(B \rightarrow K^* \chi) = \Gamma_0 \frac{\lambda_{K^*}^3}{m_B^6} [A_0(m_\chi^2)] \quad (7.2)$$

where the phasespace factor is

$$\lambda_\kappa = \left[(m_B^2 - m_\chi^2 - m_\kappa^2)^2 - 4m_\chi^2 m_\kappa^2 \right]^{\frac{1}{2}}, \quad (7.3)$$

form factors are denoted as f_0 A_0 , and Γ_0 is a constant. Figure 7.3 shows that the phasespace factor is the dominant factor in the shape for all the decays, and that searching for a new scalar or vector particle in the decays $B^0 \rightarrow K^{*0} \chi$ and $B^+ \rightarrow K^+ \chi$ is similarly sensitive for $m_\chi \lesssim 2000$ MeV, but at high masses $B^+ \rightarrow K^+ \chi$ is more sensitive.

The following chapter introduces the analysis strategy, an overview of the selection, and

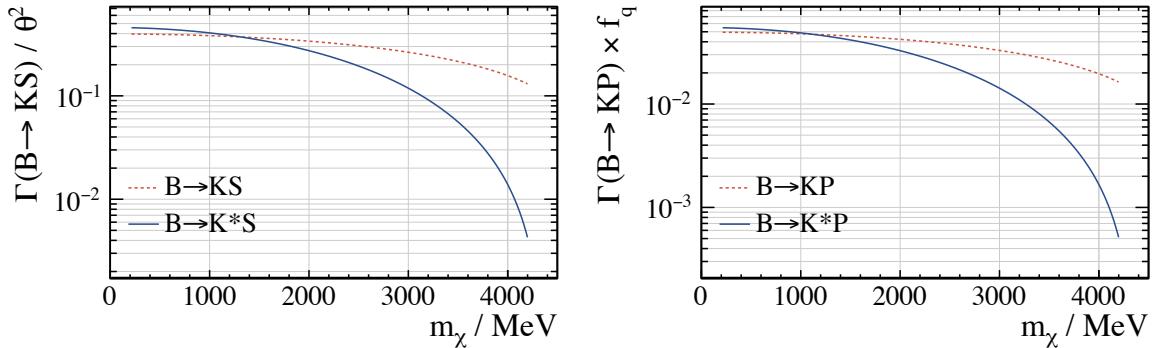


Figure 7.3: Decay rate predictions, from Eq. 7.3, for decays of the form $B \rightarrow KX$, where X is either (left) a scalar (S) or (right) an axial-vector (P). The parameters θ and f_q are parameters of the model in Ref [82]. The shapes of the curves are dominated by the available phasespace, and sensitivity is comparable for $m_\chi \lesssim 2000$ MeV.

how the discrete samples of $B^0 \rightarrow K^{*0} \chi$ are used to parameterize various distributions at all masses. The analysis is performed blindly, some results of the unblinding procedure are given in the final selection along with a calculated p -value.

7.2 Analysis strategy

This analysis is a completely general search for a particle with unknown mass, m_χ , and lifetime, τ_χ . Exhaustive details of the analysis strategy can be found in Ref. [83], but the following section will outline the important points. Broadly, the search involves a frequentist scan of the dimuon invariant mass spectrum, separated into two bins in decay time, from $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ for an excess of events consistent with $\chi \rightarrow \mu^+ \mu^-$. Firstly, an explanation of the search in the mass dimension will be presented, and then extended to deal with the decay time dimension.

7.2.1 Searching in the mass dimension

A scan in mass is performed where the test mass, m_t , is incremented in steps of $\frac{1}{2}\sigma_m$, where σ_m is the local mass resolution defined at m_t . In this analysis, σ_m lies in the range $1 \lesssim \sigma_m \lesssim 6$ MeV. At each m_t a signal and background regions are defined as

$$|m - m_t| < 2\sigma_m \quad (7.4)$$

and

$$3\sigma_m < |m - m_t| < (2x + 3)\sigma_m, \quad (7.5)$$

respectively. The width of the sideband region with respect to the signal region is defined by x . Using two sideband regions, one either side of the signal means the background contribution to the signal region can be estimated by assuming that the background is, on average, linear over the range of interest. A gap of $1\sigma_m$ separates the signal and background regions to allow for leakage of signal candidates beyond $2\sigma_m$.

At each mass point in the search the number of events in the signal region, n_s , and the number of events in the background region, n_b , are counted. If the background is, on average, linear in the local mass region, and if there is no signal contribution: $\langle n_s \rangle = \frac{1}{x} \langle n_b \rangle$. However, if there is evidence of signal, then $n_s = s + b$, where s is the number of signal events, and b is the background component in the signal region, as estimated from the sideband regions. Thus, one can construct a likelihood:

$$\mathcal{L}(n_s, n_b | s, b) = \mathcal{P}(n_s, s + b) \cdot \mathcal{P}(n_b, xb), \quad (7.6)$$

where $\mathcal{P}(n, \lambda)$ is the probability of observing n from a Poisson distribution parameterized by λ . This is simply the likelihood of the background estimate from the sidebands fluctuating to the observed number of events in the signal region.

The likelihood given in Eq. 7.6 assumes that the background is, on average, exactly linear, and that the sideband region is always factor of x larger than the signal region. In reality, this is not precisely true, and local deviations mean that the actual scaling factor is an unknown y , with an uncertainty σ_y , which is accounted for by modifying the likelihood function to be

$$\mathcal{L}(n_s, n_b, x | s, b, y) = \mathcal{P}(n_s, s + b) \cdot \mathcal{P}(n_b, yb) \cdot \mathcal{G}(y, x, \sigma_y). \quad (7.7)$$

Here, $\mathcal{G}(n, \mu, \sigma)$ is the probability of observing n given a Gaussian distribution with a mean, μ , and standard deviation, σ . This modification allows the uncertainty on the background shape to be immediately accounted for in the method, meaning that no additional systematic uncertainty is required.

The profile likelihood, Λ , is defined as

$$\Lambda(s | n_s, n_b) = \frac{\mathcal{L}(s, \hat{b}(s), \hat{y}(s) | n_s, n_b, x)}{\mathcal{L}(\hat{s}, \hat{b}, \hat{y} | n_s, n_b, x)}, \quad (7.8)$$

where \hat{s} , \hat{b} , and \hat{y} are chosen to maximize the likelihood; the functions $\hat{b}(s)$ and $\hat{y}(s)$ maximize the likelihood for a given s . The function $-2 \ln (\Lambda(s | n_s, n_b))$ behaves like a χ^2 distribution with one degree of freedom.

7.2.2 Searching in the lifetime dimension

Sensitivity to the lifetime of the χ , τ_χ , is introduced by splitting the data at each test mass into two regions: a prompt and a displaced region, defined by $\tau < 3\sigma_\tau$ and $\tau > 3\sigma_\tau$ respectively; where σ_τ is defined to be the local decay time resolution. The joint likelihood is simply the product of the two individual likelihoods:

$$\mathcal{L}(n_s^p, n_b^p, n_s^d, n_b^d, x | s^p, b^p, y^p, s^d, b^d, y^d) = \mathcal{L}(n_s^p, n_b^p, x | s^p, b^p, y^p) \times \mathcal{L}(n_s^d, n_b^d, x | s^d, b^d, y^d), \quad (7.9)$$

where superscripts p and d denote the prompt and displaced regions, respectively. The combined profile likelihood is:

$$\Lambda_{\text{Tot}} = \Lambda(s^p | n_s^p, n_b^p) \cdot \Lambda(s^d | n_s^d, n_b^d), \quad (7.10)$$

and $-2 \ln(\Lambda_{\text{Tot}})$ behaves as a χ^2 distribution with two degrees of freedom. From this information a p -value for a given mass m_t can be calculated quantifying how incompatible the signal region is with the null hypothesis of zero signal.

The information supplied by the addition of two bins in the lifetime dimension is approximately optimal for all χ lifetimes, except for when $\tau \sim 3\sigma_\tau$. In this case, it is marginally more optimal to include shape information of the background distribution from the sidebands; but this introduces significantly more complications to the analysis. Therefore background shape information is not used.

7.2.3 Calculation of p -value

At each m_t , a p -value is calculated using the profile likelihood of the joint likelihood given in Eq. 7.9. Because there are $\mathcal{O}(1000)$ test masses, each of which is not completely independent the look-elsewhere effect must be accounted for. To do this the minimum local p -value is translated into a global p -value an ensemble of toys.

After the p -value at each m_t is calculated, the region in which the lowest p -value consistent with zero signal will be isolated and removed from the sample; leaving data which is entirely background. This is assuming that there is only one NP particle that can be observed in this analysis. The remaining background-like distribution is then turned into a Probability Density Function (PDF) — where the region that is removed is interpolated across — from which toy datasets can be generated. This procedure is undertaken later, and described more fully then.

Since this method requires the generation of $\mathcal{O}(10^7)$ datasets to probe down to 5σ , if the minimum local p -value corresponds to less than this, the asymptotic formula will be

Search for the decay of a dark sector particle $\chi \rightarrow \mu^+ \mu^-$ in $B^0 \rightarrow K^{*0} \mu^+ \mu^-$

used [84]. The asymptotic formula uses a few toy datasets to set an upper bound on the conversion of local to global minimum p -value. As shown in Ref. [85], this bound is given by

$$P_{\text{global}} \leq P(\chi_s^2 > c) + \langle N(c) \rangle, \quad (7.11)$$

where s is the number of degrees of freedom of the χ^2 , in this case $s = 2$. The function $N(c)$ returns the number of *upcrossings* of a dataset above the level of c , where an upcrossing is defined to be when the p -value crosses from below the value of c , to above it. The value of c for a given p -value is its significance squared. Clearly, for large p -values, the ability to obtain a reliable value of $N(c)$ will be highly dependent on the toy datasets generated, and therefore an approximation is needed. Reference ?? shows that Eq. 7.11 can be approximated to be

$$P_{\text{global}} \leq P(\chi_s^2 > c) + \langle N(c_0) \rangle \exp\left(-\frac{c - c_0}{2}\right) \cdot \left(\frac{c}{c_0}\right)^{\frac{(s-1)}{2}}, \quad (7.12)$$

where c_0 is the number of upcrossings for a small value of c , optimally $c = s - 1$. Therefore, once $\langle N(c_0) \rangle$ has been determined using a few hundred toy datasets, the asymptotic boundary is trivial to calculate. This can be used to check the final conversion, and used to calculate the global p -value should a greater than 5σ be observed.

7.2.4 Limit setting

Upper limits will be set as a function of m_χ and τ_χ . This requires further modification to the likelihood function to account for the relationship between the number of signal events in the prompt and displaced regions for a given value of decay time. Additional Gaussian terms are added: one to account for the uncertainty in the fraction of signal events that are observed in the two lifetime regions, and another to account for the uncertainty in the efficiency ratio with respect to the normalization channel. The resulting likelihood for a given m_χ and τ_χ is

$$\begin{aligned} \mathcal{L}(n_s^p, n_b^p, n_s^d, n_b^d, x, \tau | \dots) &= \mathcal{L}(n_s^d, n_b^d, x | \varepsilon s f, b^p, y^p) \\ &\times \mathcal{L}(n_s^p, n_b^p, x | \varepsilon s(1-f), b^d, y^d) \\ &\times \mathcal{G}(f, \bar{f}(\tau_\chi), \sigma(f)) \times \mathcal{G}(\varepsilon, \bar{\varepsilon}(\tau_\chi), \sigma(\varepsilon)), \end{aligned} \quad (7.13)$$

where the fraction of signal events in the prompt region is given by f , which has an expected value from simulation \bar{f} and an uncertainty $\sigma(f)$. The same nomenclature is used for the efficiency measured relative to the normalization channel, ε .

The normalization channel used will be the SM decay $B^0 \rightarrow K^{*0} \mu^+ \mu^-$, restricted to the region $1.1 < q^2 < 6.0 \text{ GeV}^2$. This range is chosen to minimize theoretical uncertainties at

Table 7.2: Selected particles properties for mesons which decay into a dimuon pair final state, and the branching fractions for the relevant decays [3]. Central values of mass and width are given in MeV.

Meson (X)	Mass	Width	$\mathcal{B}(B^0 \rightarrow K^{*0}X)$	$\mathcal{B}(X \rightarrow \mu^+ \mu^-)$	$\mathcal{B}_{\text{tot}}(B^0 \rightarrow K^{*0} \mu^+ \mu^-)$
η	547.9	0.001	$(1.59 \pm 0.10) \times 10^{-5}$	$(5.8 \pm 0.8) \times 10^{-6}$	$(9.2 \pm 1.4) \times 10^{-11}$
ρ	775.3	147.8	$(3.9 \pm 1.3) \times 10^{-6}$	$(4.55 \pm 0.28) \times 10^{-5}$	$(1.8 \pm 0.6) \times 10^{-10}$
ω	782.7	8.5	$(2.0 \pm 0.5) \times 10^{-6}$	$(9.0 \pm 3.1) \times 10^{-5}$	$(1.8 \pm 0.8) \times 10^{-10}$
ϕ	1019.5	4.3	$(1.00 \pm 0.05) \times 10^{-5}$	$(2.87 \pm 0.19) \times 10^{-4}$	$(2.9 \pm 0.2) \times 10^{-9}$
D^0	1864.8	10.1	$(4.2 \pm 0.6) \times 10^{-5}$	$< 6.2 \times 10^{-9}$	$< 2.6 \times 10^{-14}$
J/ψ	3096.9	0.093	$(5.96 \pm 0.03) \times 10^{-2}$	$(1.32 \pm 0.06) \times 10^{-3}$	$(7.9 \pm 0.4) \times 10^{-5}$
$\psi(2S)$	3686.1	0.299	$(7.9 \pm 0.9) \times 10^{-3}$	$(6.0 \pm 0.4) \times 10^{-4}$	$(4.7 \pm 0.6) \times 10^{-6}$

low and high q^2 and reduce experimental uncertainties by removing the region around the ϕ meson which is centred at $q^2 \simeq 1.04 \text{ GeV}^2$.

7.2.5 Resonant backgrounds

In the absence of resonances, the dimuon background from the SM decay $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ is expected to be locally-linear to within 1% over the entire mass range. However, the inclusion of resonances can lead to significant departures from linearity, dependent upon the resonance's width (Γ), magnitude, and the value of x . Table 7.2 lists a number of mesons that decay to a dimuon final state and could contribute as background to the decay $B^0 \rightarrow K^{*0} \mu^+ \mu^-$. Wide resonances, $\Gamma \gtrsim 20\sigma_m$, such as the ρ , are sufficiently wide not to be problematic, even if they dominate the local background distribution. Conversely, narrow resonances where $\Gamma \lesssim 5\sigma_m$ lead to significant deviations from a locally-linear background and must be vetoed. Dimuon decays of the ϕ , J/ψ , and $\psi(2S)$ mesons have the highest branching fractions, and are also among the narrowest resonances; they are therefore vetoed. Intermediate resonances, $5 \lesssim \Gamma \lesssim 20\sigma_m$, are considered on a case-by-case basis since they are only troublesome if they account for a large fraction of the local background. For this analysis, these ranges roughly translate to requiring that resonances with $\Gamma < 25 \text{ MeV}$ will be vetoed, and those with $\Gamma > 125 \text{ MeV}$ shall be ignored. Other resonances in Table 7.2 are broad, and contribute to the dimuon structures at low mass. It is shown in Ref. [83] that the local-linearity approximation is accurate to $\sim 5\%$ in regions where there may be contributions from wide resonances. Below the mass of the J/ψ the values of $x = 5$ and $\sigma_y = 0.05$ are chosen.

Resonances that have a natural width in the intermediate region $5 \lesssim \Gamma \lesssim 20\sigma_m$, are various $c\bar{c}$ states with masses above the mass of the $\psi(2S)$. For example, there is contribution from the $\psi(4160)$ which was observed by LHCb in the decay $B^+ \rightarrow K^+ \mu^+ \mu^-$ [86]. All known charmonium resonances in this region are wide, $\Gamma \sim 70 \text{ MeV}$, and are dealt with by reducing the size of the sidebands and increasing the uncertainty on the background

Find LHCb official justification
not from [83]

or show effects
minimal

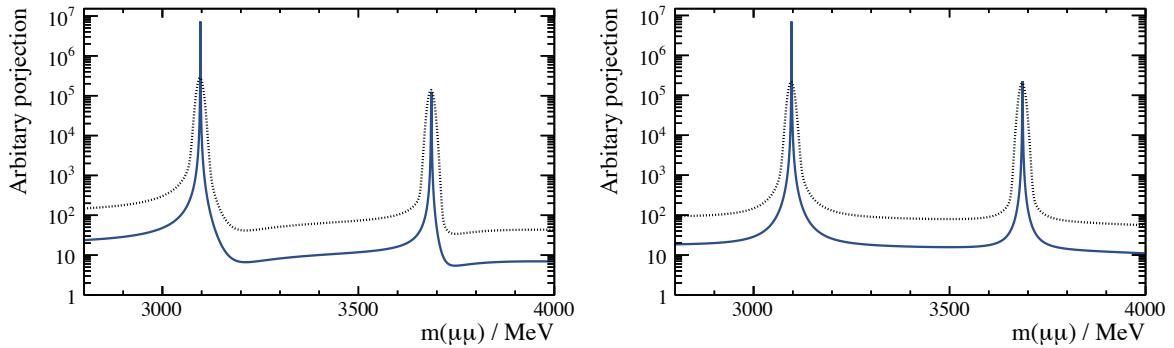


Figure 7.4: Theoretical distributions of the decays $J/\psi \rightarrow \mu^+ \mu^-$ and $\psi(2S) \rightarrow \mu^+ \mu^-$ interfering with a non-resonant dimuon component using a model in Ref. ?? with (left) no phase difference, and (right) maximal phase difference. The solid blue line shows the raw model, and the dotted line is the same distribution which has been convolved with a Gaussian to account for detector resolution effects.

shape, by setting $x = 1$ and $\sigma_y = 0.1$.

Interference effects between $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ and $B^0 \rightarrow K^{*0} \rho$, where $\rho \rightarrow \mu^+ \mu^-$, could lead to non-linearity in the background shape, despite the fact that the total branching fractions of $B^0 \rightarrow K^{*0} \rho (\rightarrow \mu^+ \mu^-)$ decay ~~note~~ is so small that less than one event is expected in the data sample. However, this deviation can only be a maximum of $\sim 5\%$, which is already accounted for in the choice of $\sigma_y = 0.05$.

The contribution from the decay $\phi \rightarrow \mu^+ \mu^-$ to the $K^{*0} \mu^+ \mu^-$ final state is removed by excluding dimuon candidates in the range $1000 < m_{\mu^+ \mu^-} < 1040$ MeV. Vetoed regions for the J/ψ and $\psi(2S)$ are calculated using a theoretical model which can account for the interference effects between the $c\bar{c} \rightarrow \mu^+ \mu^-$ resonances and the non-resonant $\mu^+ \mu^-$ component [87]. The model is used to generate PDFs with different phase differences between the resonant and non-resonant components, which is then convolved with a double Gaussian function to account for detector resolution effects; these are shown in Fig. 7.4. After smearing, the PDFs are then used to calculate the regions around the J/ψ and $\psi(2S)$ where the background is locally linear to better than 5 % for various values of x . The resulting veto regions are very similar to those used in Ref. [61], except that the upper boundary of the $\psi(2S)$ veto is extended to cover the $\psi(3770)$. All the vetoed regions to remove narrow resonances are summarized in Table 7.3.

7.3 Selection

This analysis uses the full 3.0 fb^{-1} of data collected by the LHCb experiment [32] in the years 2011 and 2012. Since the properties of the χ are unknown, a variety of simulated samples of the decay $B^0 \rightarrow K^{*0} \chi$, where $\chi \rightarrow \mu^+ \mu^-$, were generated with a range of m_χ

Search for the decay of a dark sector particle $\chi \rightarrow \mu^+ \mu^-$ in $B^0 \rightarrow K^{*0} \mu^+ \mu^-$

Table 7.3: Summary of the vetoes regions in $m_{\mu^+ \mu^-}$ to remove the contributions from the decays $\phi \rightarrow \mu^+ \mu^-$ and various narrow $\psi \rightarrow \mu^+ \mu^-$ to the dimuon distribution.

Resonance(s)	Vetoed region (MeV)
ϕ	$1000 < m_{\mu^+ \mu^-} < 1040$
J/ψ	$2960 < m_{\mu^+ \mu^-} < 3204$
$\psi(2S), \psi(3770)$	$3614 < m_{\mu^+ \mu^-} < 3875$

Table 7.4: Samples of simulated $B^0 \rightarrow K^{*0} \chi$ generated with given mass and lifetime. A total of 1.5 million events are generated for each sample, but only 150 000 for the samples with a m_χ of 220 and 235 MeV.

τ_χ (ps)		m_χ (MeV)									
10		2500									
100	214	220	235	250	500	800	1000	1500	2000	2500	4000
1000				250					2500		

and τ_χ , a summary of these are shown in Table 7.4. Studies are also performed using simulated events of the decays $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ and $B^0 \rightarrow J/\psi K^{*0}$.

Reconstructed decays of $B^0 \rightarrow K^{*0} \chi$ that must be selected by the L0 triggers for muon, dimuon, or hadronic candidates. Subsequent trigger levels require that the decay has a topology consistent with a B -meson decaying in to a multi-body final state which includes muons. Only **TOS** candidates are used in this analysis for two reasons: firstly, the ratio of trigger efficiency for the SM $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ to that of (possibly displaced) χ mode enters into the limits and, thus, must be precisely determined; and the use of **TIS** events would come with a substantial enhancement of the di- b -hadron backgrounds. The few percent gain in signal efficiency obtained using **TIS** is not worth the increase in these backgrounds di- b -hadron backgrounds.

Since the χ could be long lived, its decay vertex can lie downstream of the B^0 vertex, and could fly far enough to leave acceptance of the VELO before it decays. It is therefore wise to consider different *track types*. For most LHCb analyses, the only appropriate tracks to use are *long* tracks — as have been used implicitly throughout this thesis. Long tracks are fitted using hits in the tracking stations T1–3, the VELO, and can use TT information. But, it is also possible to reconstruct tracks using only some of the tracking stations; for example, a *downstream* track is reconstructed using hits in the TT and tracking stations T1–3, but not the VELO. Figure 7.5 shows the various definitions of track types. The problem with using downstream tracks is that they are not triggered efficiently in HLT2. For example, a simulated χ with a mass of 250 MeV and a lifetime of 100 ps has a reconstruction and stripping efficiency of $\sim 0.9\%$ if the muons are both long tracks, but the equivalent number for downstream tracks is $\sim 2.5\%$; due to the boost

Search for the decay of a dark sector particle $\chi \rightarrow \mu^+ \mu^-$ in $B^0 \rightarrow K^{*0} \mu^+ \mu^-$

Table 7.5: Selection criteria applied to signal candidates in the stripping. Criteria definitions are defined in text. While the B^0 mass is constrained in the fit, the selection makes a cut on the unconstrained mass.

Candidate	Selection criterion			
B^+	$\chi_{\text{vtx}}^2/\text{ndf}$	<	25	state which have LHCb-standard values
	χ_{IP}^2	<	50	
	τ	>	0.2 ps	
	m	\in	[4800, 5800] MeV	
	p_T	>	1000 MeV	
	$\cos \theta_{\text{dir}}$	>	0	
χ	$\chi_{\text{vtx}}^2/\text{ndf}$	<	10	
	χ_{FD}^2	<	25	
	p_T	>	250 MeV	
	DOCA	<	0.2 mm	
	DOCA χ^2	<	25	
Tracks	$\chi_{\text{trk}}^2/\text{ndf}$	<	3	
	$\min(\chi_{\text{IP}}^2)$	>	9	
	\mathcal{P}_{gh}	<	0.3	
K^+, π^+	p_T	>	250 MeV	
	p	>	2000 MeV	
	χ_{IP}^2	>	9	
	\mathcal{P}_K	>	0.1	
	\mathcal{P}_π	>	0.2	
	p_T	>	100 MeV	
	PIDmu	>	-5	
	isMuon		True	

$$\text{DLL}_{K\pi}(\pi^+) > 10.$$

The decay $K_s^0 \rightarrow \pi^+ \pi^-$ has a branching fraction of $(69.20 \pm 0.05) \times 10^{-2}$ [3], and is removed in the preselection by requiring that $|m_{\pi_\mu^+ \pi_\mu^-} - m_{K_s^0}^{\text{PDG}}| < 25$ MeV. This roughly translates to a cut in the dimuon invariant mass spectrum of $436 < m_{\mu^+ \mu^-} < 490$ MeV.

Since the χ can be displaced from the B^0 decay vertex, a potential background for the decay $\chi \rightarrow \mu^+ \mu^-$ is from a $\mu^+ \mu^-$ pair directly from a PV. This is suppressed by requiring that the transverse flight distance (FD_T) of the χ vertex, with respect to the PV, is greater than 0.1 mm.

The high branching fraction of $B^0 \rightarrow J/\psi (\rightarrow \mu^+ \mu^-) K^{*0}$ means that there is also contamination from decays where a hadron is misidentified as a muon, and vice versa. This background can be suppressed by requiring that neither of the K^{*0} daughters satisfy the `isMuon` criteria. Figure 7.6, shows the invariant mass of the hadron-muon pair before and after the requirement that the hadrons fail the `isMuon` criterion. A summary of all preselection cuts is shown in Table 7.6.

Table 7.7: Veto conditions to suppress double and single misidentification of particles. If, under the alternate hypothesis, the χ or K^{*0} candidate mass falls within the range indicated, the candidates are subject to the given PID requirements.

Mass criteria (MeV)	PID requirement
$ m(K^+ K_\pi^-) - m_\phi^{\text{PDG}} $	< 10 $\mathcal{P}_\pi(\pi) > 0.3$ and $\mathcal{P}_K(\pi) < 0.3$
$ m(K_\mu^+ \pi_\mu^-) - m_{D^0}^{\text{PDG}} $	< 25 $\mathcal{P}_\mu(\mu) > 0.3$
$ m(p_\mu \pi_\mu^-) - m_{A^0}^{\text{PDG}} $	< 10 $\mathcal{P}_p(\mu) < 0.3$
$ m(p_\pi K^- \mu^+ \mu^-) - m_{A_b^0}^{\text{PDG}} $	< 50 $\mathcal{P}_p(\pi) < 0.2$

7.3.2 Other backgrounds from particle misidentification

Beyond backgrounds from ~~onia states~~ and the K_s^0 decay there are many mesons that decay into a two-body final state, which could mimic the signal decay given some particle misidentification. A meson that decays via $X \rightarrow hh'$ which is then reconstructed under the incorrect mass hypothesis could pass the selection criteria as either the χ or K^{*0} candidate. This type of contamination is studied by assigning different mass hypotheses to each final state particle and calculating the invariant mass of the $\mu^+ \mu^-$ and $K^+ \pi^-$ candidates. If the mass of one of these objects, after mass reassignment, is seen to peak at the mass of a known particle, then the contamination is removed by applying PID criteria.

Since a selection has been made on the $K^{*0} \rightarrow K^+ \pi^-$ candidate using both PID criteria and constraints on the $K^+ \pi^-$ invariant mass it is expected that there will be little contamination from background sources. To test this, candidate K^{*0} mesons coming from a B^0 candidate with an invariant mass within 80 MeV of the known B^0 mass are assigned different mass hypotheses to check for peaking components in the new $m_{K_h^+ \pi_{h'}^-}$ mass spectrum. The only background that must be removed from this category is from a real $\phi \rightarrow K^+ K^-$ where a kaon in the final state is misidentified as being a pion. If the mass of the $K^+ K_\pi^-$ candidate lies within 10 MeV of the known ϕ mass, the ambiguous pion is subject to the requirements that $\mathcal{P}_\pi < 0.3$ and $\mathcal{P}_K > 0.3$.

Resonances decaying into a pair of hadrons which are mistaken as a pair of muons are more problematic. Weak Decays of mesons can contribute to background, especially $D^0 \rightarrow K^+ \pi^-$ and $A^0 \rightarrow p \pi^-$, which are dealt with in a similar way to the vetoes described in Chap. 5. If the invariant mass of the $K_\mu^+ \pi_\mu^- (p_\mu \pi_\mu^-)$ candidate falls within 25(10) MeV of the nominal $D^0 (A^0)$ mass, then the muons are subject to the requirement that $\mathcal{P}_\mu(K_\mu^+, \pi_\mu^+) > 0.3 (\mathcal{P}_p(p_\mu) < 0.3)$.

Misidentifying the proton as a pion in the decay $\Lambda_b^0 \rightarrow p K^- \mu^+ \mu^-$ can contaminate the selected $B^0 \rightarrow K^{*0} \chi$ candidates. Figure 7.7 shows the invariant mass distribution of the $p_\pi K^- \mu^+ \mu^-$ system for candidates where $\mathcal{P}_p(p_\pi) > 0.2$, a clear peak is observed at the

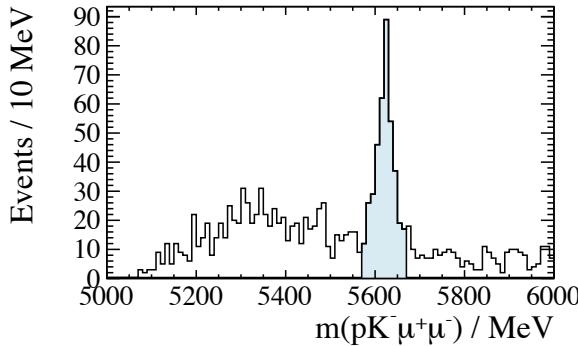


Figure 7.7: Contamination from the decay $\Lambda_b^0 \rightarrow p K^- \mu^+ \mu^-$, where the proton is misidentified as a pion, here a cut of $\mathcal{P}_p(p_\pi)$ has been applied and a clear peak at the known Λ_b^0 mass, 5219.4 MeV, is observed. Candidates are vetoed if they lie within 50 MeV of the Λ_b^0 mass, which is shown as the shaded region.

known mass of the Λ_b^0 . Candidates are removed by if the mass of the Λ_b^0 candidate falls within 50 MeV of the nominal Λ_b^0 mass and satisfies $\mathcal{P}_p(p_\pi) > 0.2$.

The sidebands are used to estimate the level of background in the signal region. Therefore, background contributions are only problematic if they produce a narrow peaking structure in the dimuon mass, because misidentification causes the $m_{\mu^+ \mu^-}$ distribution to be smeared. In general, misidentification is only problematic if the decaying particle has a very narrow natural width, so any remaining misidentification-type backgrounds have a negligible effect in the analysis.

7.3.3 Possible contamination from the $X(1070)$

While searching for potential backgrounds resulting from misidentifying two hadrons as muons, a peak is observed in the invariant mass spectrum of the $K_\mu^+ K_\mu^-$ candidates. This peak was consistent with the $X(1070)$ listed in Ref. [3], which has a mass of (1072 ± 1) MeV with a width of (3.5 ± 0.5) MeV and was observed in the $K_s^0 K_s^0$ distribution from a pion beam interacting with a liquid hydrogen target [88].

Initially the $\mu^+ \mu^-$ pair under the $K^+ K^-$ mass hypothesis appears to have a contribution from a decaying $X(1070)$. Figure 7.9 shows a comparison of simulated $K_s^0 \rightarrow \pi^+ \pi^-$ decays with the observed data near this excess. It is clear that $K_s^0 \rightarrow \pi^+ \pi^-$ decays produce a peak around 1072 MeV under the $K^+ K^-$ hypothesis. There is also a long tail but with low statistics and with a roughly uniform background this tail would not be expected to be visible in the data after the K_s^0 veto in the preselection.

It is clearly not the case that the $X(1070)$ is causing the peak at $m_{K_\mu^+ K_\mu^-} = 1072$ MeV, which is actually due to the decay $K_s^0 \rightarrow \pi^+ \pi^-$. Removing events satisfying $|m_{\pi_\mu^+ \pi_\mu^-} - m_{K_s^0}^{\text{PDG}}| < 25$ MeV removes much of the peak at 1072 MeV, bringing it in line with the background.

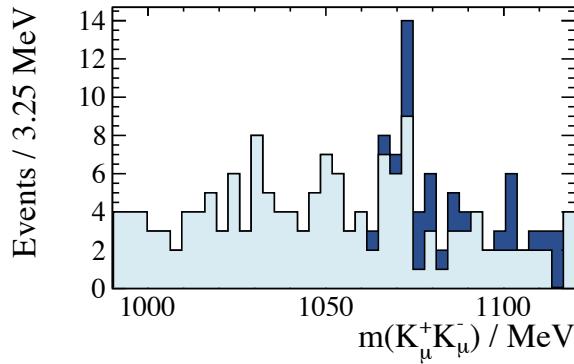


Figure 7.8: Invariant mass distribution of the $K_\mu^+ K_\mu^-$ candidates in data, showing a peak at ~ 1072 MeV in data. The dark and light blue regions show the distribution before and after vetoing the decay $K_s^0 \rightarrow \pi^+ \pi^-$ in the preselection.

Also, the fact that no $\phi \rightarrow K^+ K^-$ is observed in the $m_{K_\mu^+ K_\mu^-}$ spectrum, and that the peak is narrower than the resolution of the LHCb detector, indicate that there this is a false peak and need not be accounted for further.

7.3.4 Combinatorial background and multivariate selection

The data sample is purified from combinatorial background using a multivariate selection technique. Section 4.3 outlines the uBDT algorithm, which trains a BDT whereby events are boosted not only based on misclassification, but also on how uniform the local response of the BDT is for a given set of variables. For the case of this analysis, it is important that the BDT does not bias the sample towards a given mass and lifetime, which makes the uBDT ideal for this analysis.

A uBDT is trained using a signal-proxy from simulated events and background taken from the upper B^0 mass sideband where the B^0 candidate has a mass of over 150 MeV above the nominal B^0 mass. Specifically, the signal-proxy is a concoction of three different simulated samples in which the χ has a mass of: 214, 1000, and 4000 MeV; and each has a lifetime of 100 ps. These samples are chosen to give the uBDT algorithm input the largest range of masses possible; particularly a $m_\chi = 214$ MeV sample was chosen because it is close to threshold and equal to the mass of the P^0 evidenced in the Hyper- CP experiment [20]. It is observed that the response after a uBDT selection does result in a uniform signal efficiency in both the mass and lifetime distributions; this is shown in Fig. 7.10.

The uBDT cut is optimized by maximizing the Punzi figure-of-merit [89], which is defined as

$$\text{Punzi}_{\sigma_p} = \frac{S}{\sqrt{B + \frac{1}{2}\sigma_p}}, \quad (7.14)$$

where S and B are the signal and background yields, respectively; and σ_p is the desired level of observation. The signal and background yields are calculated as follows: S is the

Search for the decay of a dark sector particle $\chi \rightarrow \mu^+ \mu^-$ in $B^0 \rightarrow K^{*0} \mu^+ \mu^-$

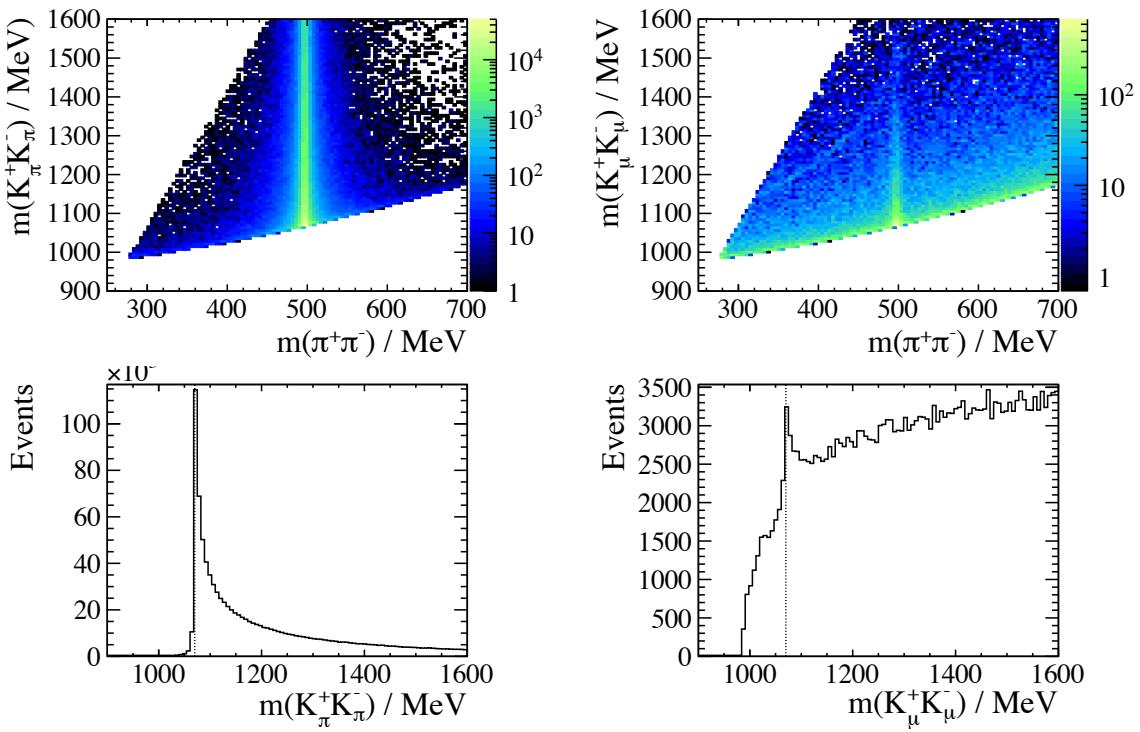


Figure 7.9: A comparison of $K_s^0 \rightarrow \pi\pi$ under different mass hypotheses, for (left) simulated events, and (right) events from data. The (top) plots show the two dimensional distributions of the invariant mass distributions of a $\pi^+\pi^-$ pair and the same candidates in the K^+K^- mass hypothesis, the (bottom) plots show the projections of the K^+K^- systems. Vertical lines in the lower plots indicate 1072 MeV.

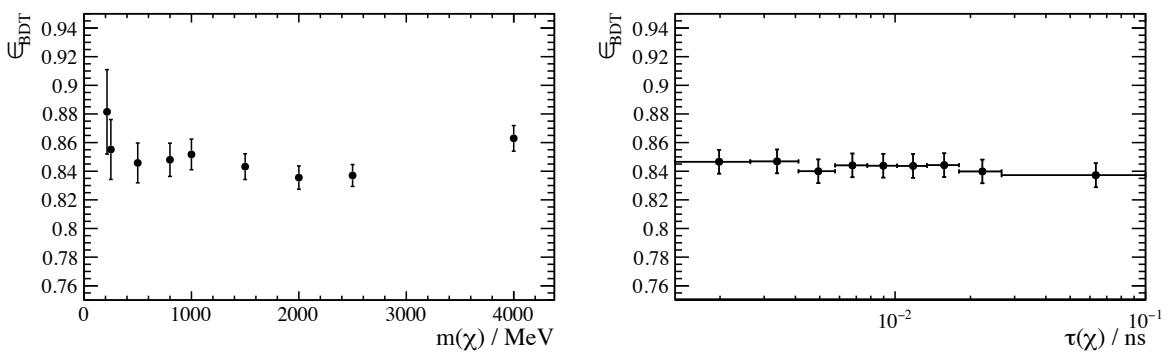


Figure 7.10: Signal efficiencies from simulation for a uBDT cut which is approximately 85 % efficient, the distributions are approximately flat in both the (left) mass and (right) lifetime dimensions.

number of simulated signal events which survive a uBDT cut; B is the background yield as estimated using $B^0 \rightarrow K^{*0} \chi$ candidates that fall outside 80 MeV from the nominal B^0 mass [3]. The invariant mass distributions of these candidates in data are fit to a decaying exponential, and the background yield is taken to be the integral of this exponential within 60 MeV (approximately 3σ) of the known B^0 mass.

The figure-of-merit is considered separately for both the prompt and displaced candidates. The optimal working point varies from sample to sample due to the fact that the background yield depends on (m, τ) . However, the optimal point in the prompt case is approximately $\text{uBDT} > 0.15$ for all samples, a working point that is approximately 85 % efficient by construction. For this efficiency there is an estimated zero background in for a given value of m_χ . This point was calculated using a value of $\sigma_p = 5$, Eq. 7.14, although the result is seen to be insensitive to its exact value.

7.4 Evolution of parameters as a function of mass

Searching for a χ of unknown mass and lifetime requires knowledge of parameters such as: mass resolution, lifetime resolution, and efficiency; at every point on the (m_χ, τ_χ) plane that is searched. This is clearly not possible by generating an arbitrary number of simulated signal events. Instead, the available simulated samples are used to interpolate (and extrapolate) to all masses and lifetimes of interest.

7.4.1 Mass resolution of the χ candidate

The size of the signal and background regions are defined in terms of the local mass resolution, which varies across the whole mass range. To understand the evolution of σ_m , the mass distribution of various signal samples are fitted to a function constructed as the sum of two Gaussian distributions with the same mean, this is known as a double Gaussian. Fitted distributions are used to define the 2σ intervals, and then each point is intersected with a cubic spline. Figure 7.11 shows the resulting function. The mass resolution is observed to be ~ 1 MeV for very low m_χ and quickly increases to a plateau around $2\sigma = 15$ MeV before dropping off again, because the invariant mass of the $K^+ \pi^- \mu^+ \mu^-$ system is constrained to the known B^0 mass.

7.4.2 Lifetime resolution of the χ candidate

Similar to finding the mass resolution as a function of mass, the evolution of the lifetime distribution is obtained by extracting the $3\sigma_\tau$ limits from fitted distributions at known χ mass points. Linear splines are then used to interpolate to all values of m_χ , cubic splines

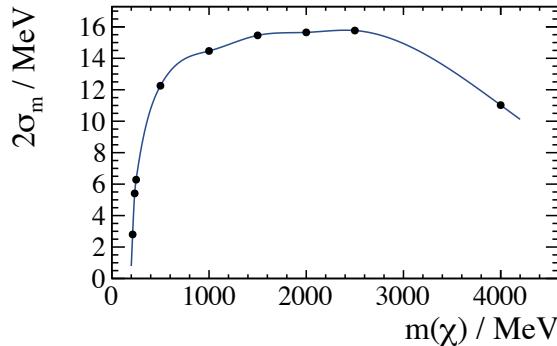


Figure 7.11: Evolution of $2\sigma_m$ as a function of mass. The mass distribution of each simulated signal sample is fit to a distribution constructed of two Gaussian functions with the same mean, the value of $2\sigma_m$ is then extracted by finding the 2σ interval of the fitted distribution. Each black point shows the value of $2\sigma_m$ for a given simulated $B^0 \rightarrow K^{*0} \chi$ sample, and the red line is the cubic spline intersecting each point.

led to massive deviations from the obvious trend, so linear interpolation is used, but for the majority of m_{DM} , the lifetime resolution is flat and this approximation has negligible effect. For each $B^0 \rightarrow K^{*0} \chi$ sample with $m \leq 250$ MeV, $\Delta\tau = \tau_\chi^{\text{meas}} - \tau_\chi^{\text{true}}$ distribution is fit to a double Gaussian function. In samples with $m_\chi < 250$ MeV the $\Delta\tau$ distribution is observed to be significantly distorted from a simple double Gaussian, as shown in Fig. 7.12.

The distortion from a double Gaussian is explained by the narrow opening angle, θ_{open} , between muons for low mass dark bosons. When a χ is produced near the dimuon mass threshold, the daughter particles are produced nearly at rest in the frame of the χ . Therefore, the two muons have a very small opening angle in the lab frame and so the separation of muon hits in the VELO are comparable to the resolution of the VELO strips. This leads to poor spatial resolution of the χ decay vertex. By the time that the two muons are resolved from one another the measured vertex is further downstream than it really is. This effect causes a tail to extend upwards in the $\Delta\tau$ distributions for low m_χ .

The effect of a small opening angle can be seen by comparing the measured and true opening angle distributions for simulated decays of $B^0 \rightarrow K^{*0} \chi$, where $m_\chi = 214$ MeV. Figure 7.13 shows a significant discrepancy between the true and measured opening angles ~~X~~ distributions for $\theta_{\text{open}} \lesssim 0.002$ rad. In the same figure, the evolution of θ_{open} with m_χ at generator level is shown, ~~it~~ it can be seen that when $m_\chi = 250$ MeV, the opening angle is predominantly larger than 0.002 rad.

The discrepancy between real and true opening angle can be verified using data from a decay channel with very high statistics, such as the opening angle of the $K^+ \pi^-$ system in the decay $B^0 \rightarrow K^+ \pi^- \mu^+ \mu^-$ for $m_{K^+ \pi^-} < 640$ MeV. Figure 7.14 shows that the opening angle of the $K^+ \pi^-$ system is very low, and a similar peak at $\theta_{\text{open}} = 0$ when $m_{K^+ \pi^-}$ is near threshold, ($m_{K^+} + m_{\pi^-} = 633.3$ MeV).

Search for the decay of a dark sector particle $\chi \rightarrow \mu^+ \mu^-$ in $B^0 \rightarrow K^{*0} \mu^+ \mu^-$

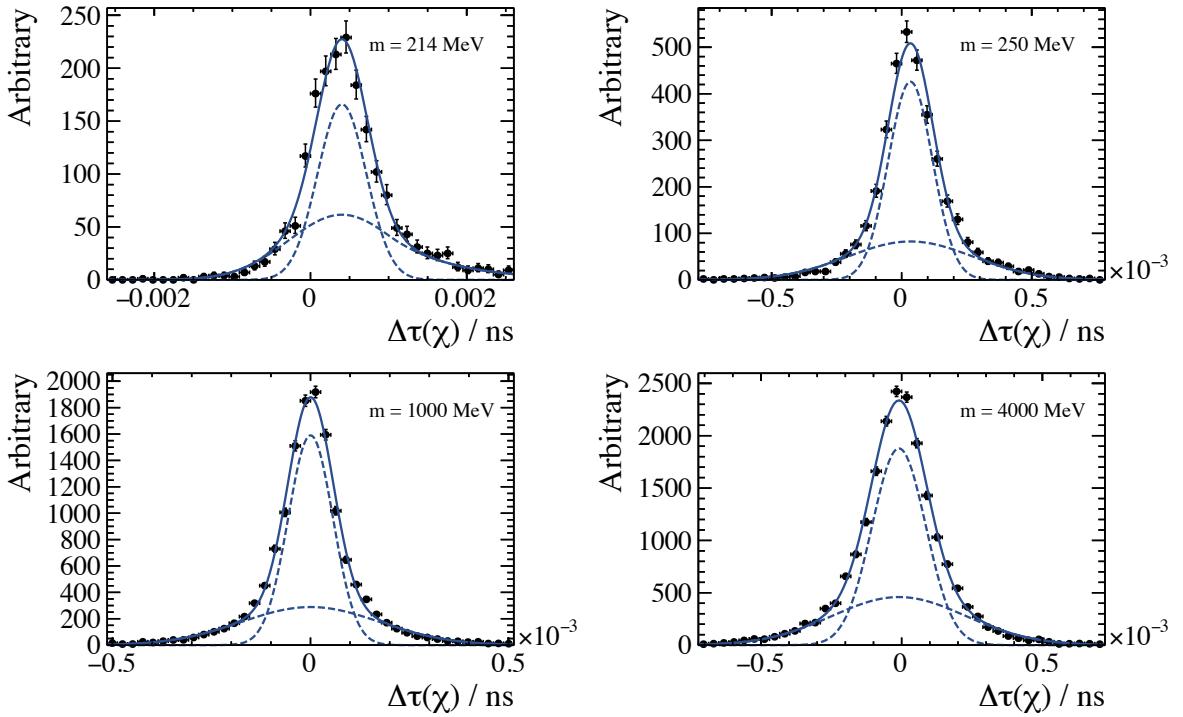


Figure 7.12: Fits to the lifetime resolution parameter, $\Delta\tau$, for individual mass samples. Each fit for $m \geq 250$ MeV is made using a double Gaussian function, and for $m < 250$ MeV the wider Gaussian has an exponential tail on the right-hand side. The solid line shows the total fit, and the dashed lines indicate the two components.

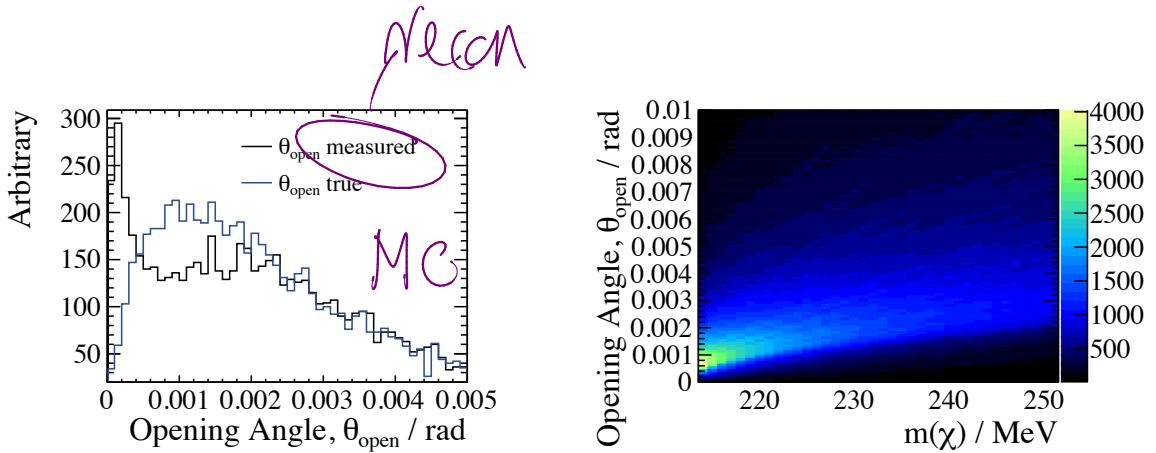


Figure 7.13: Opening angle of the two muons decaying from the χ is highly dependent on m_χ ; for low dimuon masses the opening angle measurement is biased. For $m_\chi = 214$ MeV the difference between generated and reconstructed opening angles are shown, (left) the greatest difference being below $\theta_{\text{open}} = 0.002$. Generator level distributions of opening angles for a range of masses, (right) show that most χ decays have $\theta_{\text{open}} > 0.002$ for $m_\chi \gtrsim 235$ MeV.

Search for the decay of a dark sector particle $\chi \rightarrow \mu^+ \mu^-$ in $B^0 \rightarrow K^{*0} \mu^+ \mu^-$

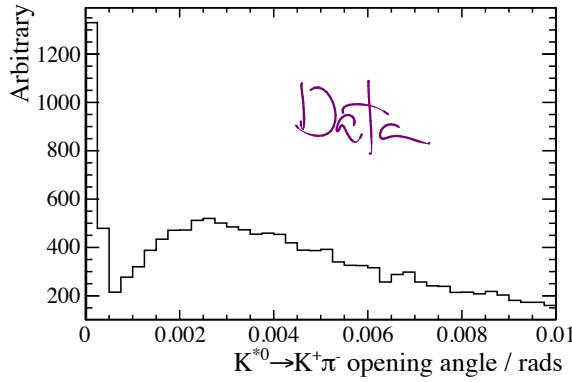


Figure 7.14: Opening angle of the $K^+\pi^-$ system in $B^0 \rightarrow K^+\pi^-\mu^+\mu^-$, where the invariant mass of the $K^+\pi^-$ system is close to threshold.

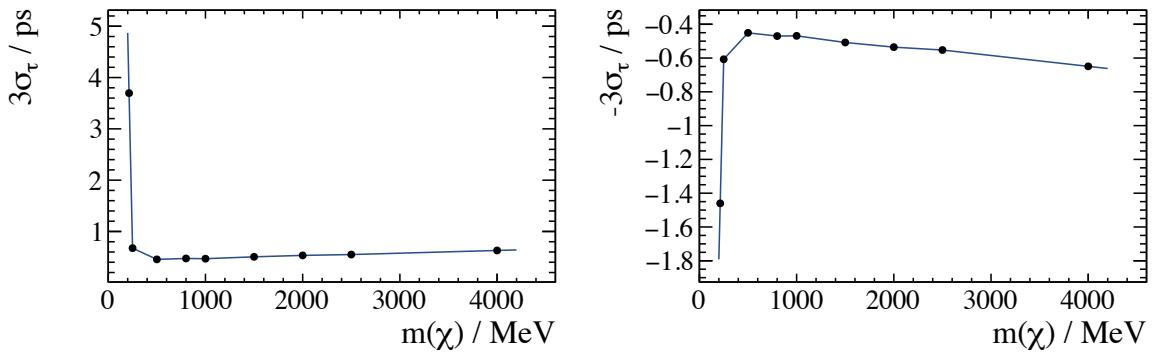


Figure 7.15: Resolutions for the lifetime of a particle as a function of mass in the (upper left) positive and (upper right) negative directions at the 3σ levels, used to define prompt and displaced regions. The black points are from fitted values and the red lines show spline interpolation.

Applying a cut to remove events where θ_{open} is small would be very inefficient for a low mass dark bosons, and therefore the function used to fit $\Delta\tau$ is modified to account for the positive skew. For simulated samples generated with $m_\chi < 250$ MeV the mean of the double Gaussian is allowed to float away from zero, and the wider of the two Gaussian functions is modified to incorporate an exponential tail extending to high $\Delta\tau$. Accounting for this effect allows the $\pm 3\sigma_\tau$ values to be taken directly from the fits in Fig. 7.12, and takes into account the bias in decay time resolution.

The prompt region is defined by $-3\sigma_\tau < \tau_\chi < 3\sigma_\tau$ and the displaced region by $\tau_\chi > 3\sigma_\tau$. Figure 7.15 shows how σ_τ varies with mass, where each point comes from the fits to $\Delta\tau$ described above, and linear spline interpolation is used to access σ_τ at all values of m_t .

Search for the decay of a dark sector particle $\chi \rightarrow \mu^+ \mu^-$ in $B^0 \rightarrow K^{*0} \mu^+ \mu^-$

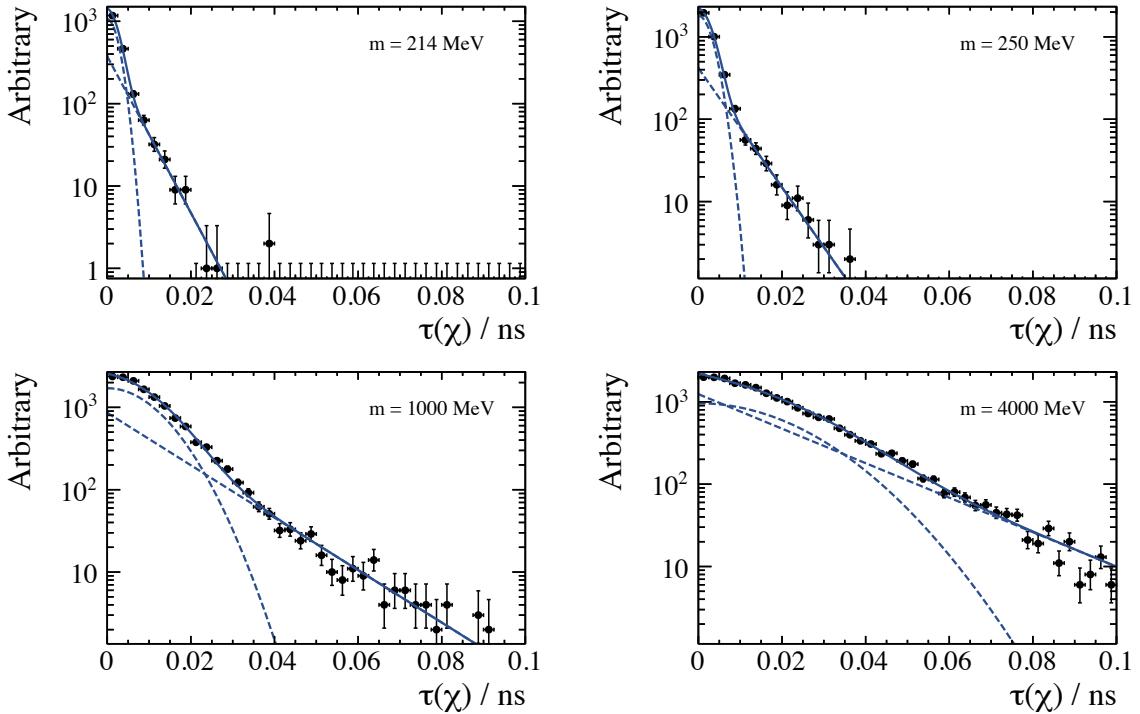


Figure 7.16: Fits, using the function in Eq. 7.15, to the decay time distributions of simulated events produced with the masses: 214, 250, 1000, and 4000 MeV (as indicated) each with $\tau_\chi = 100 \text{ ps}$. The black points are the simulated events, the solid line is the total fit, with dashed lines indicating the Gaussian and exponential components.

7.4.3 Parameterizing the efficiency of the χ selection

Setting limits as a function of mass and lifetime requires knowledge of the efficiency for any arbitrary dark boson, as outlined in Eq. 7.13. To do this, the lifetime distribution for a given mass is fit to a simple function:

$$\mathcal{T}(\tau) = f\mathcal{G}(\mu = 0, \sigma) + (1 - f)\exp(-\alpha\tau), \quad (7.15)$$

which depends upon only a few parameters: a f , σ , and α . There is excellent agreement between τ_χ distributions from simulated samples of $B^0 \rightarrow K^{*0}\chi$ and this simple parameterization, as shown in Fig. 7.16.

The parameters σ , and α evolve smoothly as a function of mass, and ~~as~~ spline interpolation is used to determine their values for arbitrary masses. Linear regression is used to evolve f , because of the large errors on the values of f yielded by these fits. So, σ , α , and f , become functions of mass, therefore \mathcal{T} becomes a function of mass and lifetime: $\mathcal{T}(m, \tau)$. Figure 7.17 shows how the aforementioned parameters evolve as a function of mass.

Once the two-dimensional map of $\mathcal{T}(m, \tau)$ has been produced, it is scaled such that for all

Search for the decay of a dark sector particle $\chi \rightarrow \mu^+ \mu^-$ in $B^0 \rightarrow K^{*0} \mu^+ \mu^-$

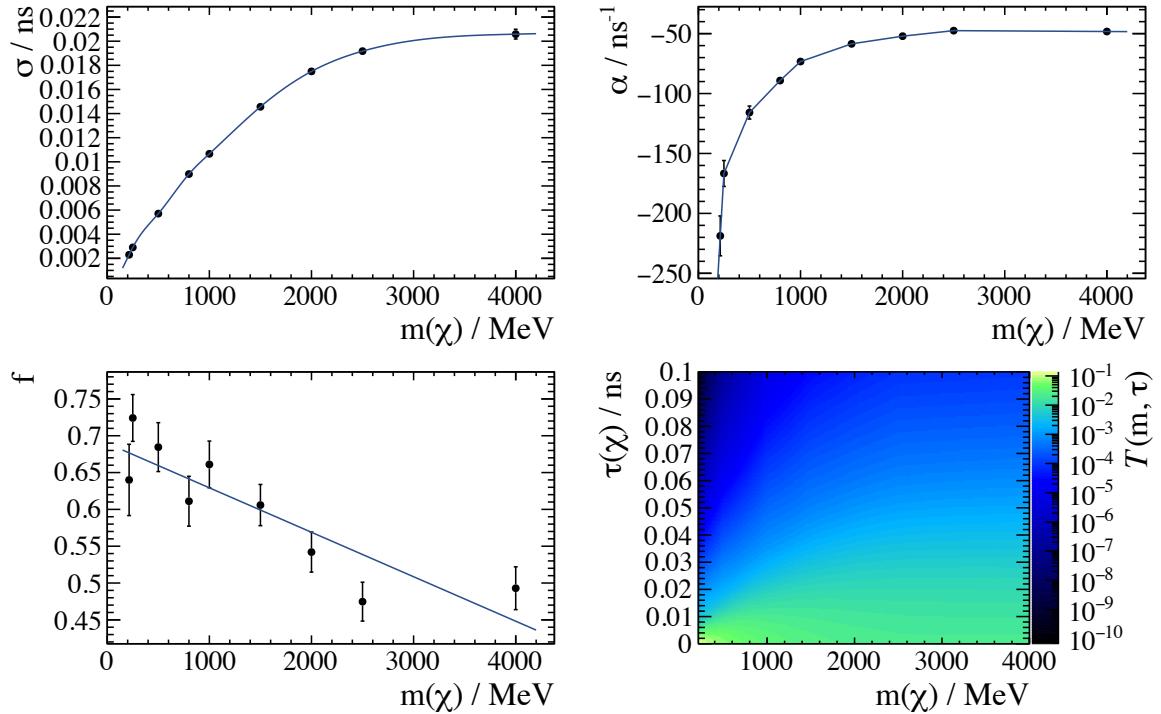


Figure 7.17: Parameters from Eq. 7.15 as a function of mass from simulated events. The parameters are (upper left) σ , (upper right) α , (lower left) f , and (lower right) the two-dimensional projection of the total parameterization. Splines are used to parameterize each shape, except for the parameter f , where a linear fit is used.

values of m ,

$$\mathcal{T}(m, \tau = 0 \text{ ns}) = \frac{\varepsilon_{\tau=0}(B^0 \rightarrow K^{*0} \mu^+ \mu^-)}{\varepsilon_{1.1 < q^2 < 6.0}(B^0 \rightarrow K^{*0} \mu^+ \mu^-)} \mathcal{T}(m, \tau). \quad (7.16)$$

Here, the numerator is q^2 dependent and is calculated using simulated events, and the denominator is the total efficiency for candidates in the indicated q^2 region, which is calculated to be 3.93 %. This efficiency ratio is approximately equal to one for all values of m_χ . Figure 7.18 shows both the efficiency for a prompt dimuon pair — or equivalently a χ — as a function of q^2 , and the rescaled distribution of $\mathcal{T}(m, \tau)$, which shall be called $\mathcal{R}(m, \tau)$.

In summary, $\mathcal{R}(m, \tau)$ is the lifetime distribution for a dark boson where $\tau_\chi = 100 \text{ ps}$, scaled such that at $\tau = 0$ the distribution is equal to $\varepsilon_{B^0 \rightarrow K^{*0} \mu^+ \mu^-}^{\tau=0} / \varepsilon_{B^0 \rightarrow K^{*0} \mu^+ \mu^-}^{1.1 < q^2 < 6.0}$. To assess limits for all values of mass and lifetime initial lifetime distribution must be taken into account by convolving $\mathcal{R}(m, \tau)$ with a decaying exponential to derive an efficiency map:

$$\varepsilon(m, \tau) = \frac{1}{\tau} \int_0^{100 \text{ ps}} \mathcal{R}(m, \tau') \exp\left(-\frac{\tau'}{\tau}\right) d\tau'. \quad (7.17)$$

The upper lifetime acceptance is chosen to be 100 ps, because the efficiency at longer

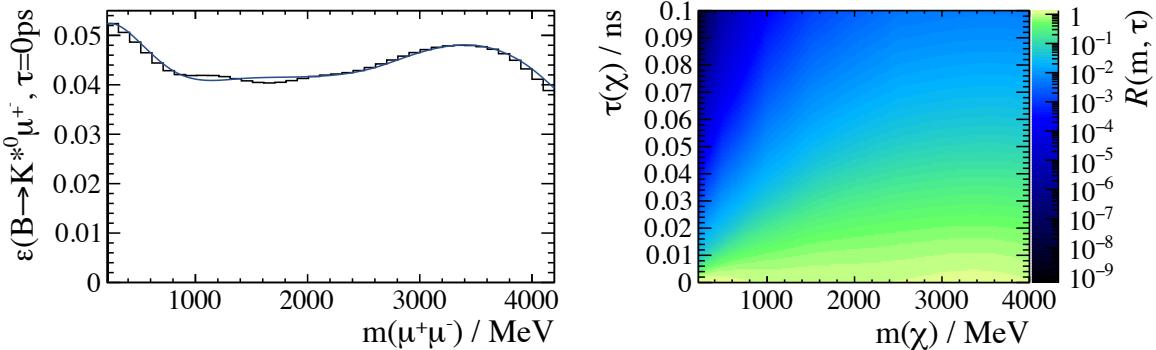


Figure 7.18: Efficiency $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ for prompt muons is shown on the left, in bins, and a spline is used to parameterize the shape, the deviations between 1000 and 2000 MeV are too small to have any significant impact on the limits. This is divided by the efficiency in the normalization region, and the $\tau = 0$ lifetime bin from Fig. 7.17 is set to this value for all masses; the decay time dimension is scaled.

lifetimes, for all values of m_χ , is very poor.

7.5 Unblinding results

This search for a new dark boson is done blindly to ensure that no bias is introduced in the course of the analysis. The data is unblinded in stages to ensure that the selection is behaving as expected on real data.

Selected $B^0 \rightarrow J/\psi K^{*0}$ events are used to check that the selection was not biased based on neither the year, nor the polarity of the LHCb magnet. No bias is observed, in fact efficiencies for the uBDT are observed to be consistent to the 10^{-4} level in all four regions.

The unblinding procedure begins by checking the yield of the normalization channel $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ in the range $1.1 < q^2 < 6.0 \text{ GeV}^2$ and comparing it with the yield from the SM analysis. The yield is taken from an unbinned fit to selected prompt $B^0 \rightarrow K^{*0} \chi$ candidates using a mass model of two Gaussian functions sharing the same mean, and with a power-law tail on the low mass side. A simple exponential models the background component. This yields 527 B^0 candidates, which can be compared to ~ 625 events in the SM selection, where the drop in signal is, again, expected given the search is for a rare process. Together with the drop in signal, comes a drop in background yield, from ~ 630 background events over the full mass range, to only 290. Figure 7.19 shows the B^0 candidate mass spectrum for the normalization channel, and the fitted distribution overlaid.

After unblinding the region $1.1 < q^2 < 6.0 \text{ GeV}^2$, other prompt q^2 regions were also used to confirm that the ratio of BDT selection efficiencies $\varepsilon_{\text{BDT}}^{q^2 \text{ bin}}(B^0 \rightarrow K^{*0} \mu^+ \mu^-) / \varepsilon_{\text{BDT}}(B^0 \rightarrow J/\psi K^{*0})$ are approximately the same in data and simulation. To

Give it
a better
name

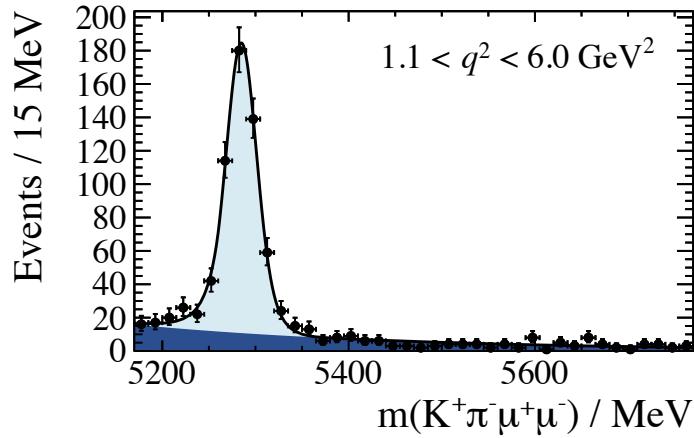


Figure 7.19: Fit to the invariant mass spectrum of the B^0 candidates in selected data in the range $1.1 < q^2 < 6.0 \text{ GeV}^2$. The signal model is the sum of two Gaussian functions with power-law tails on the low-mass side with parameters taken from the analysis described in Ref. [61], the background model is a decaying exponential. This fit results in a signal yield of (527 ± 26) compared to approximately 625 in the SM analysis.

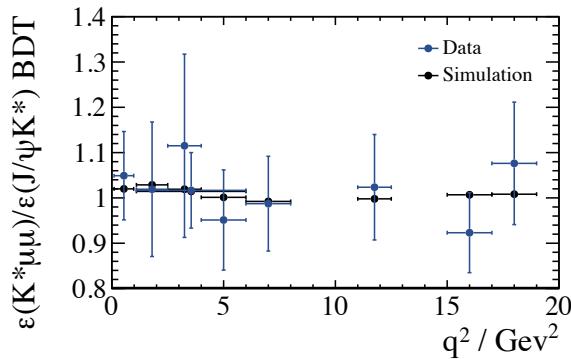


Figure 7.20: Ratio of the efficiencies of the BDT selection for a range of q^2 bins, as used in the SM $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ analysis, with respect to the efficiency for $B^0 \rightarrow J/\psi K^{*0}$. Data and simulation are shown to be in good agreement.

determine these efficiencies, the full selection without the BDT is taken, and a fit performed and the signal yield is extracted. Next, the BDT is applied and a second fit is performed, then take the ratio of the signal yields. A comparison between these numbers in data and simulation is shown in comparison is shown in Fig. 7.20, the two distributions are shown to be in good agreement, centred around unity with about 5–10 % precision.

Using the unblinded distributions it is possible to estimate the amount of combinatorial background that will remain in the final selection. Fitting an exponential to model the background across the signal region allows an approximate number of events in the combinatorial background to be deduced. This number of events can be taken from the upper-mass sideband, and the invariant mass of the dimuon pair can be plotted. With this method, it expected that a maximum of 10(2) events will contribute to the background in

Search for the decay of a dark sector particle $\chi \rightarrow \mu^+ \mu^-$ in $B^0 \rightarrow K^{*0} \mu^+ \mu^-$

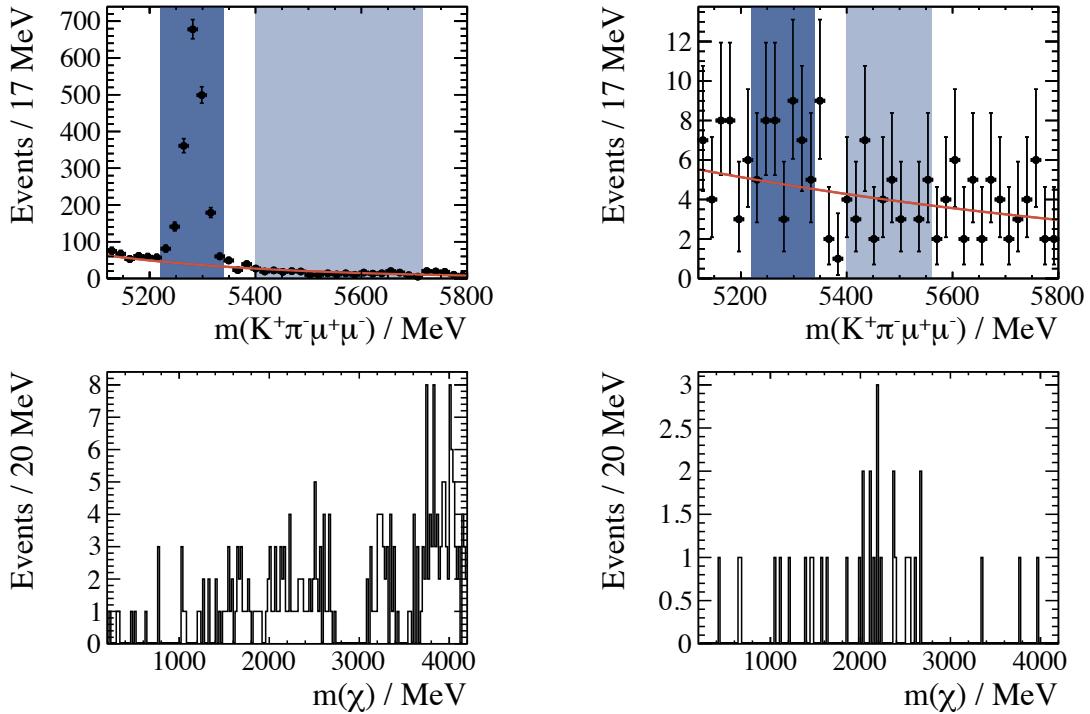


Figure 7.21: Fits to the background component of selected $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ events are shown in the upper plots, where the dark blue regions are not used in the fit, but are used to determine the number of events contributing to the combinatorial background by integrating across the region. The light blue region ~~is~~ covers the same number of background events as in the darker region; these are purely combinatorial background, and the invariant dimuon masses are plotted below, using a bin width that is approximately equal to the signal region in where the resolution is at its worst.

Reword

the prompt(displaced) region at a single test mass, but on average the value is 1.8(0.2) events per bin. Figure 7.21 shows the shape and scale of the combinatorial background, and the fits used to calculate it.

7.5.1 Calculation of p -value

The unblinded distribution in $m_{\mu^+ \mu^-}$ is shown in Fig. 7.22. Using the statistical method described in Sec. 7.2 the following ranges in m_t are scanned:

$$\begin{aligned}
 & 253.4 < m_t < 369.5 \text{ MeV}, \\
 & 574.5 < m_t < 906.5 \text{ MeV}, \\
 & 1136.0 < m_t < 2813.0 \text{ MeV}, \\
 & 3246.5 < m_t < 3576.5 \text{ MeV}, \\
 & 3796.0 < m_t < 4356.0 \text{ MeV}.
 \end{aligned}$$

Search for the decay of a dark sector particle $\chi \rightarrow \mu^+ \mu^-$ in $B^0 \rightarrow K^{*0} \mu^+ \mu^-$

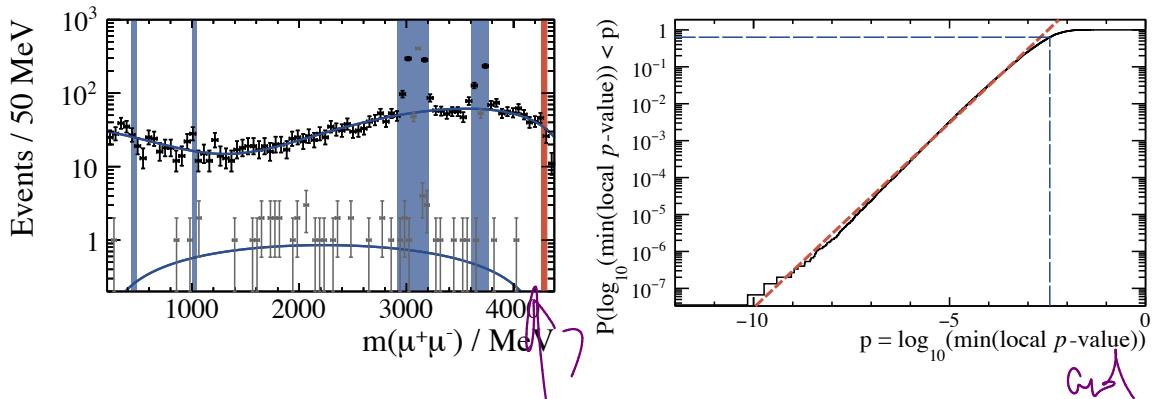


Figure 7.22: Invariant mass distributions of the dimuon pair are shown on the left, prompt and displaced candidates are black and grey points, respectively. Chebychev polynomials fitted to the data are shown as solid blue lines. The red region indicates $|m_t - 4285.0| < 5\sigma_m$ ($x = 1$ in this region). On the right is the result of running the method on 1.5×10^{-7} toy datasets, used to convert the local p -value (3.6×10^{-3}) to a global one (0.63), this conversion is indicated by the blue dashed line. The result of the asymptotic formula is shown by the red dashed line, and is seen to be an excellent approximation of the true distribution.

Values of m_t do not reach threshold boundaries because of the sidebands extending in either direction. The minimum local p -value is found to be 3.6×10^{-3} at $m_t^{\min} = 4285.0$ MeV.

The look-elsewhere effect must be considered. To convert the p -value to a global one, a PDF is fit to the $m_{\mu^+\mu^-}$ distribution of the prompt and displaced $B^0 \rightarrow K^{*0}\chi$ candidates that lie outside of the vetoed regions, and outside of the signal region centred at m_t^{\min} . A fourth(second) order Chebychev polynomial is fit to candidates in the prompt(displaced) region. From these PDFs 1.5×10^7 toy datasets are generated, and the minimum p -value for each one is calculated. Constructing a cumulative distribution of these p -values makes an easy conversion from local to global p -values, this conversion is shown in Fig. 7.22. Shown alongside the cumulative histogram is the asymptotic approximation, which is seen to be in excellent agreement for local p -values less than about 10^{-4} . The histogram converts the local p -value to a global p -value of 0.63, equivalent to a shift in significance of 2.9σ to 0.48σ . These results show no evidence for a new dark boson in the mass ranges given above.

7.5.2 Systematic uncertainties

Systematic uncertainties must be assessed in order to set limits. Sources of systematic uncertainties are from: the ratio of efficiencies $\varepsilon(B^0 \rightarrow K^{*0}\chi)/\varepsilon(B^0 \rightarrow K^{*0}\mu^+\mu^-)$; the definition of prompt and displaced regions; and the uncertainty on the $B^0 \rightarrow K^{*0}\mu^+\mu^-$ branching fraction in the range $1.0 < q^2 < 6.0$ GeV 2 .

The efficiency ratio $\varepsilon(B^0 \rightarrow K^{*0}\chi)/\varepsilon(B^0 \rightarrow K^{*0}\mu^+\mu^-)$ is approximately one, by construction, for each mass at zero lifetime. For larger lifetimes, the efficiency can be checked using

Search for the decay of a dark sector particle $\chi \rightarrow \mu^+ \mu^-$ in $B^0 \rightarrow K^{*0} \mu^+ \mu^-$

data consistent with the decay $B^0 \rightarrow J/\psi K_s^0$. It should be noted that in the displaced region a large uncertainty on the efficiency ratio will translate to a small uncertainty in the limits, because of the low statistics in that region.

Studies undertaken in Ref. [83] show that the defining the boundary that separates the prompt and displaced regions to be $3\sigma_\tau$ is nearly optimal for any value of τ_χ . Only if $\tau_\chi \simeq 3\sigma_\tau$, then there may be some effect on the limits. The total effect of this must be demonstrated.

An important consideration is the spin of the χ . Since the spin of the χ is unknown, there are a range of possible angular distributions that may arise, each having a different detection efficiency. This can be studied using the angular acceptance models used in the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ angular analysis [61].

It has been noted that this analysis relies on the assumption that the background is smoothly varying, and can be approximated as being locally linear. This assumption clearly introduces an uncertainty, but this is already accounted for in the method by the addition of the Gaussian, $\mathcal{G}(y, x, \sigma_y)$, in the likelihood shown in Eq. 7.7. There is no need to add a systematic uncertainty for the chosen value of σ_y , because it is, itself, and ~~uncertainty~~ maximal one.

7.6 Summary

A selection designed to be uniformly efficient in all regions of the dark boson's mass-lifetime parameter space has been presented with a view to embark in a search for a new dark boson. A frequentist method designed to search for an signal in any arbitrary mass spectrum has also been presented and used in the case of the dimuon mass spectrum from $B^0 \rightarrow K^{*0} \mu^+ \mu^-$.

This strategy has extracted a p -value of a particle considering test masses which do not go all the way to the boundaries of various vetoes in the invariant dimuon mass spectrum. It is determined that the maximum deviation of the selected candidates from the null hypothesis of zero signal has a global significance of 0.48σ at $m_t = 4285.0$ MeV. This is consistent with no new particle in the $\mu^+ \mu^-$ distribution over the mass ranges probed. The full analysis will get much closer to these edges and probe the interesting $m_{\mu^+ \mu^-} = 214$ MeV region. The next step in the analysis is to set limits and present them in a model-independent way. Of course, they can be translated to specific models for interpretation.

Figure 7.23 shows the projected sensitivity of this analysis to the inflaton model in Ref. [77]. There is a region excluded by theory, where the model does not satisfy known cosmological constraints. It is predicted that it should be possible to rule out the mass range $250 < m_\chi < 450$ MeV entirely, and come within an order of magnitude of mixing

Search for the decay of a dark sector particle $\chi \rightarrow \mu^+ \mu^-$ in $B^0 \rightarrow K^{*0} \mu^+ \mu^-$

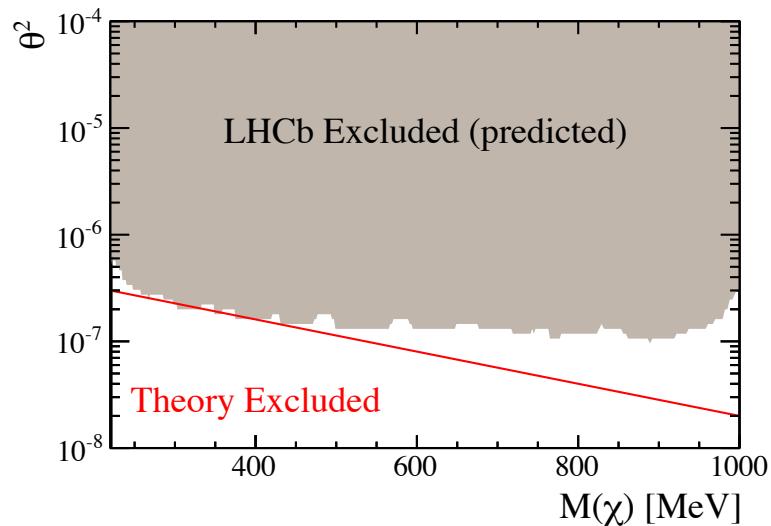


Figure 7.23: Projected exclusions for an inflaton model from Ref. [77], in the mass range $1 < m_\chi < 1000$ MeV. The region below the red line is excluded by theory, since the model fails cosmological constraints in this region. In this mass range, it is expected that this analysis will exclude all but a small area of parameter space for this model.

parameter θ^2 for masses up to 1000 MeV. Assumptions made in making this are that: the lifetime of the inflaton is in the range $1 < \tau_\chi < 1000$ ps, that $\mathcal{B}(B^0 \rightarrow K^{*0}\chi) \simeq 10^{-6}$, and that the χ has the same couplings as the Higgs boson.

Chapter 8

Conclusions

update

This thesis presents three analyses: two have been published, claiming first observations [1, 2], and the third is currently under review; all were undertaken using data collected by the LHCb experiment [32]. Each has the objective of finding evidence of physics BSM in the decays of B mesons. The analysis techniques employed are all slightly different, most obviously the first two analyses are indirect searches for NP, while the third is a direct search for a dark boson of unknown mass and lifetime.

An analysis of the decay $B^+ \rightarrow D_s^+ \phi$ is presented in Chap. 5. First evidence for the decay was seen with a statistical significance of greater than 3σ , this also constitutes the first evidence for a fully hadronic decay via an annihilation-type diagram. The branching fraction measurement made in this analysis is sensitive to NP effects contributing to the decay $B^+ \rightarrow D_s^+ \phi$. This can only propagate via one diagram at tree-level, and this is suppressed by a factor $|V_{ub}|^2$. The element V_{ub} has the largest uncertainties in the CKM matrix and there are historic tensions between values of V_{ub} made using inclusive and exclusive modes.

The branching fraction measurement

$$\mathcal{B}(B^+ \rightarrow D_s^+ \phi) = (1.87^{+1.25}_{-0.73}(\text{stat}) \pm 0.19(\text{syst}) \pm 0.32(\text{norm})) \times 10^{-6}$$

is somewhat higher than SM predictions, which are of order 10^{-7} [47–50], but not incompatible considering large theoretical uncertainties. This value of $\mathcal{B}(B^+ \rightarrow D_s^+ \phi)$ sheds no light on the true value of V_{ub} . Reference [90] asserts that these discrepancies cannot be explained by physics BSM, and is rather due to underestimated uncertainties in either theory or experiment.

Since the decay $B^+ \rightarrow D_s^+ \phi$ is mediated by a W^+ boson, in NP scenarios another charged boson, such as a H^+ from a 2HDM, could contribute to the decay amplitude. These additional processes could alter the branching fraction significantly, which is not observed, or introduce extra phases into the decay, causing the CP -asymmetry to deviate from $\mathcal{A}_{CP} = 0$, as expected in the SM. The value measured, after correcting for production and

detection asymmetries, is

$$\mathcal{A}_{CP}(B^+ \rightarrow D_s^+ \phi) = -(0.01 \pm 0.41(\text{stat}) \pm 0.03(\text{syst})),$$

which is perfectly consistent with SM expectations.

Chapter 6 presents an analysis leading to the first observations of the two decays $B^+ \rightarrow K^+ \pi^+ \pi^- \mu^+ \mu^-$ and $B^+ \rightarrow \phi K^+ \mu^+ \mu^-$. Their branching fractions were measured, as was the differential branching fraction of $B^+ \rightarrow K^+ \pi^+ \pi^- \mu^+ \mu^-$ in bins of q^2 . The integrated branching fractions of these decays are

$$\begin{aligned}\mathcal{B}(B^+ \rightarrow K^+ \pi^+ \pi^- \mu^+ \mu^-) &= (4.36^{+0.29}_{-0.27}(\text{stat}) \pm 0.21(\text{syst}) \pm 0.18(\text{norm})) \times 10^{-7}, \\ \mathcal{B}(B^+ \rightarrow \phi K^+ \mu^+ \mu^-) &= (0.82^{+0.19}_{-0.17}(\text{stat})^{+0.10}_{-0.04}(\text{syst}) \pm 0.27(\text{norm})) \times 10^{-7},\end{aligned}$$

and both have statistical significances of greater than 5σ . The decay $B^+ \rightarrow K^+ \pi^+ \pi^- \mu^+ \mu^-$ has a large branching fraction and could be used for future analyses similar to those of interest in $B^0 \rightarrow K^{*0} \mu^+ \mu^-$, since both are sensitive to the same operators. An angular analysis would help constrain the scalar, pseudoscalar, and tensor amplitudes of the decay, all of which are vanishingly small in the SM. This is made difficult by the number of contributing states to the $K^+ \pi^+ \pi^-$ system, but with more statistics it will be possible to gain a better understanding of strange states that decay into kaons and pions.

Given larger statistics, the additional channels $B^+ \rightarrow K^+ K^- \pi^+ \mu^+ \mu^-$ and $B^+ \rightarrow \pi^+ \pi^+ \pi^- \mu^+ \mu^-$ may be observable, and give access to the ratio of CKM matrix elements V_{td}/V_{ts} . These would be complimentary to the current measurements from B -meson oscillations.

The large uncertainties in the measurement of $\mathcal{B}(B^+ \rightarrow \phi K^+ \mu^+ \mu^-)$ are primarily due to uncertainties propagated from the branching fraction of the normalization channel $B^+ \rightarrow J/\psi \phi K^+$. The paper in Ref. [2] quotes the ratio of branching fractions in order for $\mathcal{B}(B^+ \rightarrow \phi K^+ \mu^+ \mu^-)$ to be calculated given an improved measurement.

Finally, a direct search for a NP particle, χ , belonging to some dark sector is presented. Using $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ candidates, the dimuon invariant mass spectrum is searched for a signal indicative of a dark boson decaying via $\chi \rightarrow \mu^+ \mu^-$. The selection is designed specifically not to bias any corner of the mass-lifetime space that the χ might inhabit, this is done primarily with the aid of the uBoost algorithm. Efficiencies and resolutions are parameterized using discrete simulated samples of $B^0 \rightarrow K^{*0} \chi$ and spline interpolation is used to understand selection and resolution effects for any value of m_χ . A novel frequentist strategy was employed which scans in mass and, at each point, calculates the local p -value that the observed signal is consistent with the null hypothesis of zero signal. One the look-elsewhere effect was accounted for, the significance of the minimum local p -value was equivalent to a significance of 0.48σ at 0.48σ at $m_{\mu^+ \mu^-} = 4285.0$ MeV. This

is consistent with with no new particle. Further studies will push tested masses to the boundaries of vetoed regions. The projected limits for an inflaton model indicate that much of the allowed parameter space will be excluded for $m_\chi < 1000$ MeV. A similar approach could be used in any arbitrary mass spectrum to search for a multitude of particles appearing above a smoothly varying background.

In conclusion, the SM continues in its resilience, seeming to be in agreement to the limit of accuracy that experimental high energy particle physics can reach. There is a complementarity that exists between indirect and direct measurements. Historically, it has often been the case that indications of future discoveries were first anticipated by observations made by indirect experiments. This has been the case because indirect

SQ measurements are sensitive to loop-level processes and therefore virtual particles. This pattern looks set to endure, as precision measurements in the flavour sector continue to play an important role in cornering the nature of NP. As of Run 1 of the LHC, there has been no clear indication of where new physics may lie; but, there are a number of discrepancies between SM predictions and measurements of observables in the flavour sector with significances greater than 3σ . Considering the current landscape, it is perhaps increasingly difficult to interpret these discrepancies, and how they impact various NP models.

Unfortunately, interpretation of precise measurements of B physics observables that can be made experimentally are often made difficult by the form-factor parameterization which must be adopted. It is these QCD effects, particularly the form-factors, that are the dominant sources of theoretical uncertainty. Difficulties in dealing with QCD has been demonstrated by a lack of consistent predictions for $\mathcal{B}(B^+ \rightarrow D_s^+ \phi)$, and the absence of any predictions for either $\mathcal{B}(B^+ \rightarrow K^+ \pi^+ \pi^- \mu^+ \mu^-)$ and $\mathcal{B}(B^+ \rightarrow \phi K^+ \mu^+ \mu^-)$.

The idea that nature is natural, is an attractive one. As such, it is not unreasonable to expect — or at least hope — that NP lives just around the corner. Run-2 of the LHC will collide protons with a centre-of-mass energy of about 14 TeV with the aim to see signals indicative of NP. That being said, there are still areas of parameter space of various theoretical scenarios that are accessible at lower masses to probe. In the absence of direct evidence from the high energy frontier, and the plethora of dark sector models which contain weakly interacting messenger particles, the intensity frontier may be a good place to search.

The interplay between direct and indirect searches in the arena of high energy physics has long been important, and will continue to be so.