CMS Physics Analysis Summary

Contact: cms-pag-conveners-bphysics@cern.ch 2009/07/30

Study of $B_s^0 \to \mu^+\mu^-$ in CMS

The CMS Collaboration

Abstract

The rare leptonic decay $B_s^0 \to \mu^+\mu^-$ has a highly suppressed rate in the Standard Model (SM) and is therefore a potentially sensitive probe of physics beyond the SM. In the minimal supersymmetric extension of the SM the branching fraction for this decay can be enhanced substantially, especially at large $\tan\beta$. In this note we present a Monte Carlo simulation study for the sensitivity of the decay $B_s^0 \to \mu^+\mu^-$ with the CMS detector at the LHC with an integrated luminosity of $1\,\text{fb}^{-1}$ at $\sqrt{s} = 14\,\text{TeV}$. The study is based on the full detector simulation of signal and background events.

1 Introduction

The leptonic decay mode[1] $B_s^0 \to \mu^+\mu^-$has a highly suppressed rate of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.86 \pm 0.15) \times 10^{-9}$ [1] in the Standard Model (SM), since it involves a flavor-changing neutral current $b \to s$ transition and requires an internal quark annihilation within the B meson which further suppresses the decay by $(f_B/m_B)^2 \approx 2 \times 10^{-3}$, where m_B (f_B) is the mass (decay constant) of the B meson. In addition, the decay is helicity suppressed. To date this decay has not been observed; the current best limits by CDF [2] and D0 [3] are still an order of magnitude above the SM expectation.

Since this process is highly suppressed in the SM, it is potentially a sensitive probe of physics beyond the SM (see Figure 1). In the minimal supersymmetric extension of the Standard Model (MSSM) the branching fraction for these decays can be enhanced by orders of magnitude, especially at large $\tan\beta$ [4, 5]. In supersymmetric models with modified minimal flavor violation (MFV) at large $\tan\beta$, the branching fraction can be increased by up to four orders of magnitude [6]. A measurement of both $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$ decays is interesting since they can be enhanced separately even at low $\tan\beta$ in specific models containing leptoquarks [7] and supersymmetric models without R-parity conservation [8].

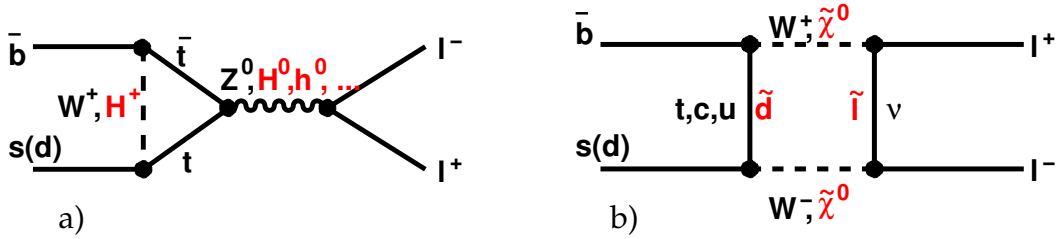

Figure 1: Illustration of the rare decays $B^0_{s(d)} \to \ell^+\ell^-$. In the SM, these decays proceed through $W^\pm$ and Z^0 bosons in Z-penguin (a) and box (b) interactions. The box diagram is suppressed by a factor of $m_W^2/m_t^2 \approx 0.2$ with respect to the Z-penguin diagram.

In the MSSM, the branching fraction enhancement for $B_s^0 \to \mu^+\mu^-$ is proportional to $\tan^6\beta$, which provides a certain sensitivity to $\tan\beta$. Experimental bounds can restrict allowed regions of parameter space, specifically the mass of the charged Higgs boson and $\tan\beta$. Recently, there has been significant interest [9–11] in using the decay mode $B_s^0 \to \mu^+\mu^-$ to "measure" the key parameter $\tan\beta$ of the MSSM and to constrain other extensions of the SM. The determination of $\tan\beta$ is difficult—there is no general technique to measure it at hadron colliders—yet all supersymmetric observables depend on it. It has been shown that with very general assumptions that do not depend on specific models, it is possible to put significant lower (and to a lesser extent also upper) bounds on $\tan\beta$. Since however, based on very general principles, $\tan\beta$ is constrained from above [12], already a lower bound on $\tan\beta$ is arguably tantamount to a measurement.

The present analysis, developed at the CMS experiment [13], uses a relative normalization to the well-measured decay $B^\pm \to J/\psi K^\pm$ to avoid a dependence on the unknown $b\bar{b}$ production cross section and luminosity measurements. Furthermore many systematic errors cancel to first order when deriving the upper limit normalizing to a similar decay channel measured in data.

[1] Charge conjugation is implied throughout this note; exceptions will be clearly spelled out.

2 Event Generation and Monte Carlo Simulation

The Monte Carlo (MC) event samples were generated, simulated and reconstructed in the context of official CMS MC 2007 productions. Depending on the production, the generation was based on PYTHIA V6.227 and PYTHIA 6.409 [14]. All signal and background events are selected from the generic QCD $2 \rightarrow 2$ and $2 \rightarrow 3$ subprocesses and present a mixture of gluon-gluon and quark-antiquark fusion, flavor excitation, and gluon splitting. In all event samples, a generator-level filter required the presence of two muons (or hadrons for rare decays), each with transverse momentum $p_\perp > 2.5\,\text{GeV}$ and in the central part of the detector $-2.5 < \eta < 2.5$.

The background sources that mimic the signal topology can be grouped into three categories. First, $q\bar{q}$ events (where $q = b, c$) with $q \rightarrow \mu X$ (prompt or cascade) decays of both q-hadrons. Second, minimum bias QCD events where a true muon is combined with a hadron misidentified as a muon (punch-through or in-flight decay of a hadron). And finally, rare B^0, B^+, B_s^0 and Λ_b decays, mostly from semileptonic decays.

In the non-peaking dimuon background samples $b\bar{b} \rightarrow \mu^+\mu^- + X$ and $c\bar{c} \rightarrow \mu^+\mu^- + X$, both heavy quarks are forced to decay in a multitude of direct semimuonic decay channels. Semimuonic charm decays after hadronic B decays are not contained in this event sample, as all B mesons are forced to decay semimuonically.

Rare b-hadron decays could potentially lead to sizable background contributions. Two cases can be distinguished: (1) Peaking backgrounds from rare decays, where a heavy particle decays into a pair of hadrons. Examples for these decays include $B_s^0 \rightarrow K^+K^-$, $\Lambda_b \rightarrow pK^-$. (2) Non-peaking backgrounds from rare B^0, B^+, and B_s^0 decays, comprising hadronic, semileptonic, and radiative decays. The invariant dimuon mass distribution for these decays is a continuum with an upper edge at the mass of the decaying particle; the finite momentum resolution could lead to events reconstructed in the $B^0 \rightarrow \mu^+\mu^-$ signal mass window. Because semileptonic decays have branching fractions several orders of magnitude above $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$, this background could be problematic. For each decay channel, events were generated and analyzed without requiring explicit muon identification. The misidentification probability (and muon identification efficiency, when one final state particle is a muon) were applied as weighting factors at the end.

All events were passed through a full GEANT based detector simulation, and reconstructed with the standard CMS software. For the event simulation, no pileup events were mixed in. The event samples were reconstructed with alignment conditions expected after $100\,\text{pb}^{-1}$ of data.

3 Trigger

This analysis is not primarily targeted at the initial very low-luminosity startup period of the LHC but requires about $1.0\,\text{fb}^{-1}$. Therefore the trigger strategy is based on an instantaneous luminosity of at least $10^{32}\,\text{cm}^{-2}\text{s}^{-1}$ as described in Ref. [15].

The Level-1 (L1) muon trigger provides fast identification of muon candidates and an estimate of their transverse momentum $p_\perp$ based on signals from the drift-tubes (DT, $|\eta| < 1.2$), the cathode strip chambers (CSC, $0.9 < |\eta| < 2.4$) and resistive plate chambers (RPC, $|\eta| < 2.1$, in the startup phase $|\eta| < 1.6$). The DT and CSC sub-triggers compare segment slopes in successive layers for their $p_\perp$ estimate, while the RPC sub-trigger is based on predefined hit patterns to classify the muon trajectory. The Global Muon Trigger matches the DT and CSC

candidates with RPC candidates and rejects unconfirmed candidates. The four highest-quality muon candidates with the largest $p_\perp$ are passed on to the global trigger, which sorts them by rank. The rank is determined by $p_\perp$ and quality. In the global trigger, separate threshold requirements could be applied to each of the four muon candidates; other possible requirements on the azimuth or pseudorapidity of single muon candidates are possible. In this analysis the L1 condition requires two L1 muons anywhere in the muon detector with $p_\perp > 3\,\text{GeV}$ each. No isolation or charge requirement is applied.

The high-level trigger (HLT) starts with the level-2 (L2) muon reconstruction. L1 muon candidates serve as seeds for the reconstruction of (stand-alone) tracks in the muon chambers with higher $p_\perp$ resolution compared to L1. A transverse momentum requirement of $p_\perp > 3\,\text{GeV}$ is applied to these L2 muons. In the next step, L2 muons are used to determine regions of interest where tracks in the central tracker are reconstructed and combined with the L2 muons. Both combined muon tracks have to satisfy $p_\perp > 3\,\text{GeV}$. The two muons are fit to a common decay vertex, a good vertex quality is required with $\chi^2 < 10$. The significance of the transverse decay length is required to be above 3 and the angle α between the reconstructed dimuon momentum vector and the vector from the primary to the decay vertex has to fulfill $\cos\alpha > 0.9$. The primary vertex at the HLT is determined with pixel tracks.

4 Muon Reconstruction

In the first step of the off-line global muon reconstruction, a stand-alone muon track is used to define a region of interest, rectangular in $\eta\phi$ space. Here the muon and the tracker tracks are extrapolated to the surface of the last tracker hit (for low-$p_\perp$ muons). The tracker tracks in this region are further selected based on spatial and momentum matching criteria. Finally, a global track fit is performed for all remaining combinations of tracker tracks and stand-alone muon tracks. The global track with the best χ^2 is chosen; there is at most one global muon track for each stand-alone muon.

Hadrons can be misidentified as muons mainly because of two reasons: (1) High-momentum hadrons can traverse the calorimeters without hadronic interaction. (2) Hadrons, in particular charged kaons, decay dominantly into muons, which will be measured in the muon system. In the following we do not distinguish between the two cases. We have used all background event samples to determine whether particles produced as hadrons close to the interaction region have been identified as muons, using the full simulation and reconstruction chain as described in section 2. We extract (conservative) misidentification probabilities for the three charged hadron species $\varepsilon_\pi = 0.60\%$, $\varepsilon_K = 1.10\%$, $\varepsilon_p = 0.25\%$. The misidentification probabilities are used as scaling weights for the rare background contributions, which are dominated by hadrons that have been misidentified.

5 Event Selection for $B_s^0 \to \mu^+\mu^-$

For the offline event selection, variables related to the primary vertex, the muon candidates, and the B_s^0 candidate with its associated secondary vertex are calculated. The primary vertex is determined with the standard algorithm [16] used in CMS. In the following a description of the calculation of all relevant variables is provided. Table 1 summarizes the numerical values for all selection criteria applied on these variables for signal and various background samples. For the figures illustrating the distributions used in the analysis, all previous selection requirements have been applied. In all figures of this section the background is composed of $b\bar{b} \to \mu^+\mu^- + X$. The background from rare decays (peaking and non-peaking) will be discussed in Sec. 7 and

is not included in any form in this section. The most important selection criteria have been optimized in a grid search for best upper limit.

If more than two muon candidates are found, the pair with the smallest $\eta\phi$ separation is chosen. We require $p_\perp > 4\,\text{GeV}$ and $|\eta| < 2.4$ for both muons. Figure 2a illustrates the muon transverse momentum distribution.

The B_s^0 candidates are formed by vertexing the two muon candidates. Figure 2b shows the transverse momentum distribution of the reconstructed B_s^0 candidates. The B_s^0 candidate is required to fulfill $p_\perp > 5\,\text{GeV}$. Figure 2c illustrates the mass resolution obtained on the signal MC event sample. The distribution is fit with two Gaussians, the width $\sigma = 53.0 \pm 1.4\,\text{MeV}$ after all analysis requirements is determined from the second central moment of the double Gaussian. The mass resolution, in particular its strong η-dependence, is limited by an inconsistent treatment in simulation and reconstruction of inhomogeneities in the magnetic field[2].

Signal events are distinguished by two muons originating from the same secondary vertex while the muons in $b\bar{b} \to \mu^+\mu^- + X$ events stem from separate vertices. Vertexing the two muons therefore provides a powerful handle in this background reduction. The transverse momentum vector of the B_s^0 candidate must be close to the direction of the secondary vertex from the primary vertex: the cosine of the opening angle between the two vectors must fulfill $\cos(\alpha) > 0.9985$, corresponding to an angular separation of about 3.1° (see figure 3a). The flight length significance of the B_s^0 candidate is an excellent handle against (prompt) combinatorial background. The significance of the (unsigned) flight length l_{3D} is defined as l_{3D}/σ_{3D}, where σ_{3D} is the error on the flight length. Figure 3a–c illustrate the distributions relevant for vertexing.

In high-$p_\perp$ gluon-splitting events the $b\bar{b}$ quark pair moves closely together due to their boost, and the two decay vertices of the resulting b-hadrons cannot be well separated in all cases, therefore mimicking a common secondary vertex. However, because of the other hadrons in semileptonic decays of both b hadrons, the hadronic activity around the dimuon direction is enhanced compared to the signal decay. This is exploited in isolation requirements. The isolation I, as applied in the searches at the Tevatron, is determined from the B_s^0 candidate transverse momentum and charged tracks with $p_\perp > 0.9\,\text{GeV}$ in a cone with radius $r = 1.0$ around the dimuon direction as follows:

$$I = \frac{p_\perp(B_s^0)}{p_\perp(B_s^0) + \sum_{trk} |p_\perp|},$$

where all track parameters are evaluated at the origin. Figure 3d illustrates the distribution of isolation variable I. In data we will also measure the decay $B_s^0 \to J/\psi\phi$ to validate and (if necessary) reweigh the B_s^0 isolation distribution in MC simulation with that obtained in data.

The efficiency for event selection on the signal and $b\bar{b} \to \mu^+\mu^- + X$ samples is provided in Table 1. The application of all selection requirements leaves no remaining background event, given the limited luminosity of the generated background sample. This does not allow to determine a reliable background estimate. However, the relatively mild correlation to the other selection criteria allows a factorization of the isolation I and χ^2 requirements from the other cuts: Their efficiencies are determined on an event sample where the dimuon mass is $4.8 < m < 6.0\,\text{GeV}$ and the significance of the secondary vertex separation is $l_{3d}/\sigma_{3d} > 7$. The expected event yield from $b\bar{b} \to \mu^+\mu^- + X$ events is then obtained by multiplying the isolation and χ^2

[2]This problem has been fixed in more recent software releases.

efficiencies to the event yield after all the other cuts. This preselection is quite loose to provide enough statistics to allow tight χ^2 or I cuts, but still retains mostly those background events that mimic the signal event signature.

Table 1: Event reduction and efficiency for the offline selection. The events are counted in the mass interval $4.8 < m_{\mu\mu} < 6.0\,\text{GeV}$ and are normalized to a luminosity of $1.0\,\text{fb}^{-1}$. The efficiencies for χ^2 and I, quoted in the middle part of the table, are determined relative to the event sample after the requirements of $4.8 < m_{\mu\mu} < 6.0\,\text{GeV}$ and $l_{3D}/\sigma_{3D} > 7$ (different normalization). The other efficiencies are cumulative. The total event selection efficiency and event yield are provided without and with the assumption of factorization of the χ^2 and I cuts. The errors in the last row are statistical only.

		Signal		$b\bar{b} \to \mu^+\mu^- + X$ background	
Description	Selection Criteria	Events	Efficiency	Events	Efficiency
gen. kinematics	see text	103	—	3.24×10^8	—
L1	see text	51.71	0.503	1.52×10^8	0.469
HLT (w/o mass cut)	see text	17.61	0.171	5.07×10^6	0.016
Good events	rec. candidate, PV	13.39	0.130	4.17×10^6	0.013
Mass cut	$4.8 < m_{\mu\mu} < 6.0\,\text{GeV}$	13.33	0.130	2.00×10^5	6.15×10^{-4}
Pointing angle	$\cos(\alpha) > 0.9985$	9.77	0.095	2.15×10^4	6.63×10^{-5}
Flight distance	$l_{3d}/\sigma_{3d} > 17.0$	5.68	0.055	1979	6.10×10^{-6}
Vertex fit (diff. norm.)	$\chi^2 < 5.0$		0.941		0.411
Isolation (diff. norm.)	$I > 0.850$		0.471		0.018
Total	w/o factorization	2.50	0.024	0.0	0.0
Total	w/ factorization	2.52	0.024	14.97	4.61×10^{-8}
Signal window	$m_{B_s^0} \pm 100\,\text{MeV}$	2.36 ± 0.076	0.023	$2.54^{+0.719}_{-0.560}$	7.82×10^{-9}

The total signal efficiency amounts to $\varepsilon = (2.43 \pm 0.115) \times 10^{-2}$, assuming factorization of the I and χ^2 selection criteria it is $\varepsilon = (2.45 \pm 0.116) \times 10^{-2}$, which is consistent with the former. Both errors are statistical only. For the $b\bar{b} \to \mu^+\mu^- + X$ background sample, the efficiency is determined to be $\varepsilon = (4.61 \pm 0.218) \times 10^{-8}$, assuming factorization of these two criteria (statistical error only).

At this stage the $b\bar{b} \to \mu^+\mu^- + X$ background event yields have been obtained in the full mass window $4.8 < m_{\mu\mu} < 6.0\,\text{GeV}$. For the determination of the final sensitivity only the background yield in the signal window $m_{B_s^0} \pm 100\,\text{MeV}$ is relevant. This reduction factor $f = 0.17$ is determined by loosening the selection cuts to those at the HLT, and then determining the ratio of background events in that window to the total. With a linear background parametrization, f varies only very weakly with the fit parameters.

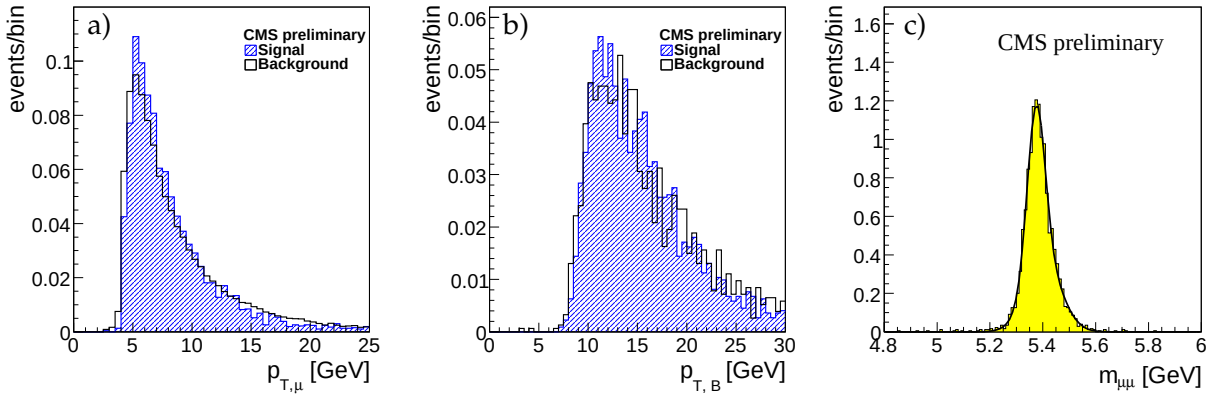

Figure 2: Distributions of a) muon transverse momentum, b) transverse momentum of B_s^0 candidate, c) B_s^0 candidate $m_{\mu\mu}$. All histograms are normalized to unity.

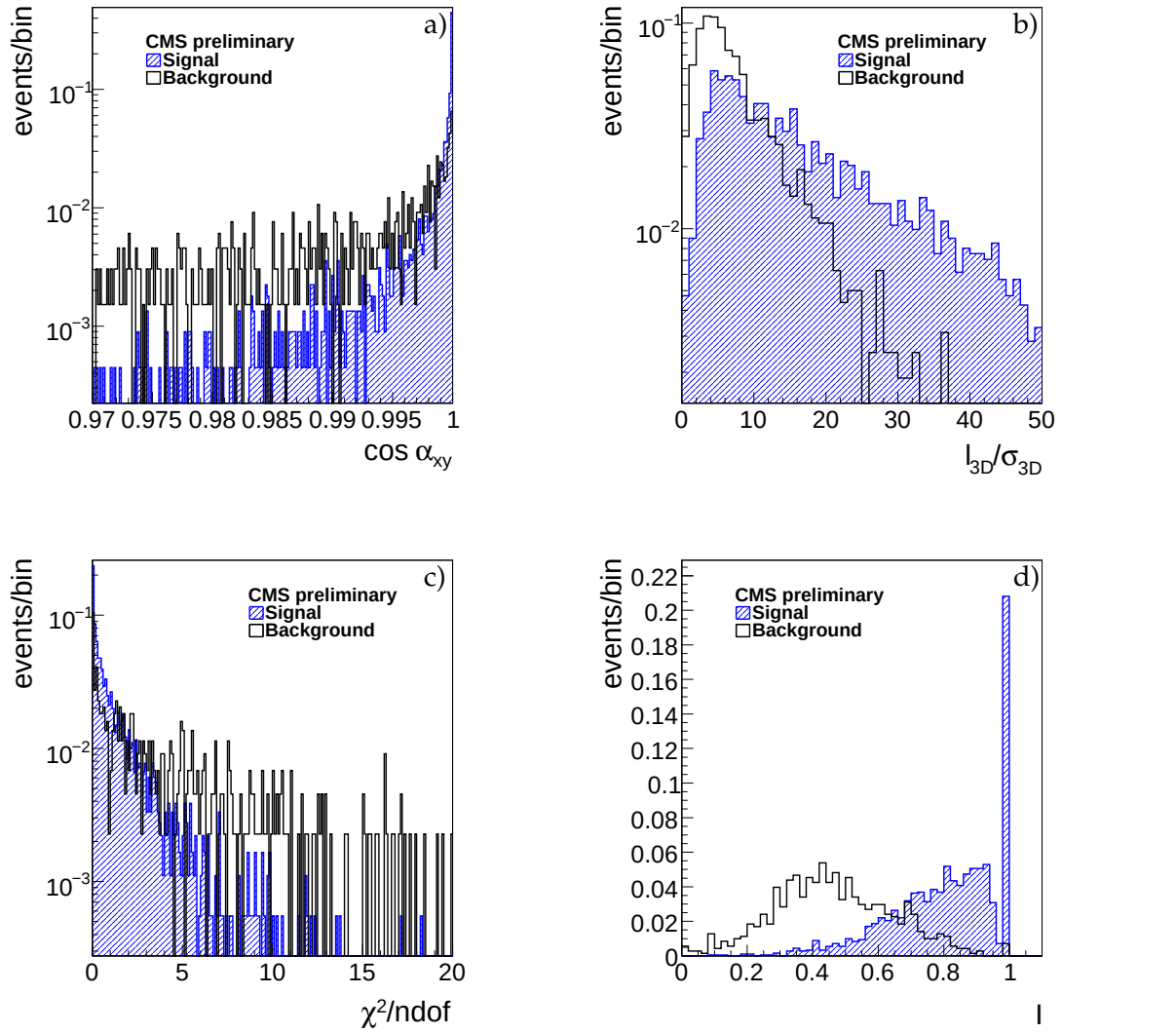

Figure 3: Vertexing and isolation variables for reconstructed B_s^0 candidates with $4.8 < m_{\mu\mu} < 6.0\,\text{GeV}$: a) Pointing angle $\cos\alpha$, b) flight length significance, c) χ^2/dof, d) isolation I. The histograms are normalized to unity.

6 The Normalization Sample $B^{\pm} \to J/\psi K^{\pm}$

To minimize the dependence on the unknown $b\bar{b}$ production cross section and luminosity measurements, the analysis uses a normalization sample with a signature similar to the signal decay $B_s^0 \to \mu^+\mu^-$. The decay $B^{\pm} \to J/\psi K^{\pm}$ has a large and well-measured branching fraction with only one additional track in the final state compared to the signal decay. However, the hadronization of the B^+ meson can be different from the B_s^0 meson, affecting for instance the isolation variable. The large statistics of the $B^{\pm} \to J/\psi K^{\pm}$ sample will allow a detailed comparison of the detector performance and analysis selection efficiencies in data and MC simulation. It will also allow the reweighing of the B^+ transverse momentum spectra so that the MC simulation reproduces the data. The decay $B^{\pm} \to J/\psi K^{\pm}$ is reconstructed using requirements as similar to the signal mode as possible; the B^+ decay vertex is reconstructed using only the two muons and no mass-constraint on the J/ψ mass is applied.

Figure 4 illustrates the combinatorial background after the HLT requirement to be expected from b-hadron decays into J/ψ mesons. While the background is not negligible, it is not expected to pose a significant problem for the extraction of the normalization yield. The background shape is described by an exponential function.

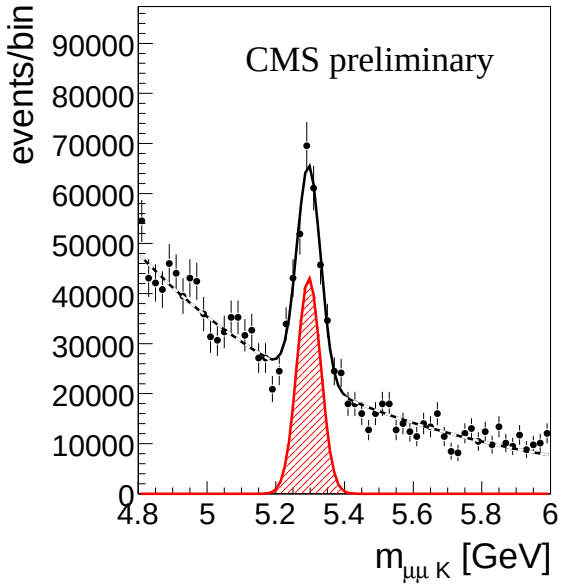

Figure 4: Reconstruction of $B^{\pm} \to J/\psi K^{\pm}$: Signal and background (combinatorial background in $b \to J/\psi X$ sample), normalized to $1.0\,\mathrm{fb}^{-1}$.

The total selection efficiency for normalization signal events is $\varepsilon_{\mathrm{tot,N}} = (1.57 \pm 0.074) \times 10^{-2}$, resulting in a total number $n_N = 3.30 \times 10^4$ of events.

7 Additional Background Studies

A selection of rare decay channels was studied. Figure 5 summarizes the mass distributions for the rare b-hadron decay backgrounds. The background distributions, shown before the application of selection criteria, are absolutely normalized. The signal distribution is normalized to

the same area as the background distribution.

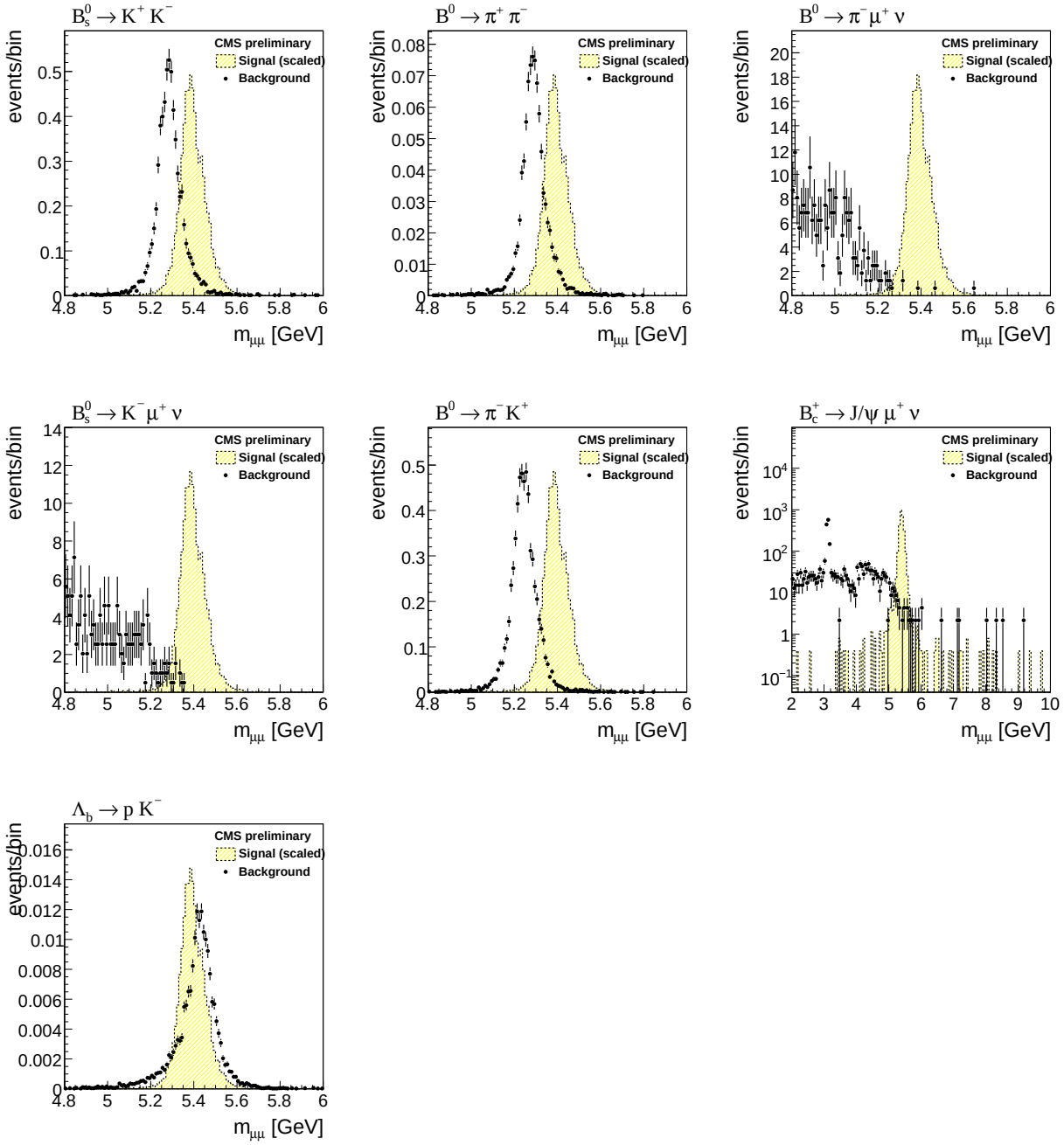

Figure 5: Background $m_{\mu\mu}$ distribution *before* the application of selection criteria (muon identification, in particular) for different channels.

To better investigate the background arising from the combination of one muon with a misidentified hadron, we have performed a generator-level MC study. Minimum bias QCD events,

corresponding to an integrated luminosity of about $2\,\mathrm{nb}^{-1}$, were generated with PYTHIA 6.409 and filtered on the generator level. Events are retained when at least one combination of a muon with a particle h with an invariant mass $4 < m_{\mu h} < 10\,\mathrm{GeV}$ is present; the particles h include muons, pions, kaons, and protons.

At this level, no vertexing or isolation requirements are applied. The contributions from the the $\mu\mu$ pairs and the μh pairs in the mass window $4.8 < m < 6.0\,\mathrm{GeV}$ are similar. We assume that the vertexing and isolation requirements will reduce the μh background similar to the $\mu\mu$ background. Therefore we increase the background yield obtained from the $b\bar{b} \to \mu^+\mu^- + X$ sample by a factor two to also include this background with one misidentified hadron.

8 Systematics

The uncertainty on the muon identification efficiency has no influence on the signal efficiency, as it cancels in the ratio with the normalization sample. It affects the background uncertainty, however. We assume that the muon identification efficiency will be determined with an error of 5%. The hadron misidentification probabilities for the determination of the hadron background have been varied by $\pm 20\%$; the background uncertainty amounts to 7%. Kaon misidentification is the dominant source for this uncertainty.

The tracking efficiency uncertainty is expected to amount to 5%. Since the normalization sample has one kaon track more in the final state, the tracking uncertainty will affect this directly by 5%.

We use the efficiency difference of the vertex χ^2 requirement between MC samples with perfect alignment and with alignment conditions expected after $100\,\mathrm{pb}^{-1}$ of data as an estimate of the uncertainty due to the tracker misalignment. We obtain a signal efficiency uncertainty of 3% and a background uncertainty of 5%.

Because of the limited statistics in the background samples, the selection requirements for the vertex fit χ^2 and isolation are studied independently on an enlarged dataset. The signal efficiency differs by 1% between the factorizing and simultaneous analysis efficiency. In the normalization sample we find a difference of 10%, and for the normalization background 12%. On the background sample, the two efficiencies cannot be compared as no event survives the complete analysis chain. We assume a 20% systematic error for the background yield.

We assume an uncertainty of 5% (relative) for each the L1 and HLT efficiency.

The normalization for this analysis will rely on the measurement of a control sample in data. The largest external uncertainty here is from the ratio of fragmentation probabilities f_s and f_d. The uncertainty amounts to 15%.

The branching fractions of the normalization sample are $\mathcal{B}(B^\pm \to J/\psi K^\pm) = (1.007 \pm 0.035) \times 10^{-3}$ and $\mathcal{B}(J/\psi \to \mu^+\mu^-) = (5.93 \pm 0.06) \times 10^{-2}$. Adding the branching fraction uncertainties quadratically amounts to a total uncertainty of 3.6%.

Combining the systematic error, summarized in Table 2, quadratically with the statistical error, the signal efficiency is known to about 18%, while the background yield uncertainty amounts to about 37%.

Table 2: Summary of systematic uncertainties.

Source	$\Delta\varepsilon_{\text{Signal}}$	$\Delta N_{\text{Background}}$
Muon identification	-	5%
Muon misidentification	-	7%
L1 Efficiency	5%	5%
HLT Efficiency	5%	5%
Misalignment	3%	5%
Kaon tracking efficiency	5%	–
Factorizing selection	1%	20%
f_s/f_u	15%	–
Branching fraction	3.6%	–
Total	18 %	23 %

9 Results

With the event and candidate selection described in Sect. 5 the total cumulative selection efficiency for signal events is $\varepsilon_S = 0.023$ and the background reduction factor is $\varepsilon_B = 7.82 \times 10^{-9}$. With this selection, the first $1.0\,\text{fb}^{-1}$ of integrated luminosity will yield $n_S = 2.36$ signal events and $n_B = 5.07$ background events in the mass window $m_{B_s^0} \pm 100\,\text{MeV}$, where we have combined contributions from $\mu\mu$ pairs and misidentified μh pairs. Additional background events in the mass window arise from rare B-decays as described in Sect. 7. The contribution of these events is $n_B^{\text{rare}} = 1.45$, for a total background expectation of $n_B^{\text{total}} = 6.53$. As described in Sect. 8, we have combined statistical and systematic uncertainties of the background estimate of 37% and for the signal efficiency error of 18%, respectively.

To calculate the upper limit on the number of observed signal events, we follow the Bayesian procedure described as in Ref. [17]. We extract the upper limit at the 90% C.L. from

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) \;=\; \frac{N(n_{obs}, n_B, n_S)}{N(B^\pm \to J/\psi K^\pm)} \frac{f_s}{f_u} \frac{\alpha_{B^+}}{\alpha_{B_s^0}} \frac{\varepsilon^{trig}_{B^+}}{\varepsilon^{trig}_{B_s^0}} \frac{\varepsilon^{ana}_{B^+}}{\varepsilon^{ana}_{B_s^0}} \mathcal{B}(B^\pm \to J/\psi K^\pm)\mathcal{B}(J/\psi \to \mu^+\mu^-),$$

where $\alpha_{B_s^0}$ (α_{B^+}) is the generator-level acceptance for signal (normalization) events, $\varepsilon^{trig}_{B_s^0}$ ($\varepsilon^{trig}_{B^+}$) is the trigger efficiency for signal (normalization) events, $\varepsilon^{ana}_{B_s^0}$ ($\varepsilon^{ana}_{B^+}$) is the analysis efficiency for signal (normalization) events, and $\mathcal{B}(B^\pm \to J/\psi K^\pm) = (1.007 \pm 0.035) \times 10^{-3}$ and $\mathcal{B}(J/\psi \to \mu^+\mu^-) = (5.93 \pm 0.06) \times 10^{-2}$, and finally $f_s = (10.5 \pm 0.9)\%$ and $f_u = (40.2 \pm 0.9)\%$.

Using the event and candidate selection criteria described in Sect. 6 , the total selection efficiency for signal events in the normalization channel is $\varepsilon_{\text{tot,N}} = (1.57 \pm 0.074) \times 10^{-2}$. By normalizing to the number of $B^\pm \to J/\psi K^\pm$ events $n_N = 3.30 \times 10^4$, the resulting upper limit on the branching fraction is given by

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) \leq 1.6 \times 10^{-8}.$$

At 95% C.L. the upper limit on the branching fraction is $\mathcal{B}(B_s^0 \to \mu^+\mu^-) \leq 1.9 \times 10^{-8}$.

While this upper limit is about four times above the SM expectation, it allows already constraints on new physics models with the first $1.0\,\text{fb}^{-1}$ of integrated luminosity. A better de-

termination of the expected background event yield with sidebands and the use of improved selection/analysis methods will further lower the upper limit.

The expected yield of $B_d^0 \to \mu^+\mu^-$ is about ten times smaller, which could, however, be increased by non-MFV physics. Given our mass resolution, we can separate leptonic decays of B_s^0 and B_d^0 mesons nearly at 2σ and could extract a separate upper limit also for the decay $B_d^0 \to \mu^+\mu^-$.

References

[1] M. Artuso *et al.*, "B, D and K decays," Eur. Phys. J. C **57**, 309 (2008).

[2] T. Aaltonen *et al.* [CDF Collaboration], "Search for $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$ Decays with $2\,\mathrm{fb}^{-1}$ of $p\overline{p}$ Collisions," Phys. Rev. Lett. **100**, 101802 (2008).

[3] D0 Collaboration, "A new expected upper limit on $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ using $5\,\mathrm{fb}^{-1}$ of Run II data," Conference Note 5906-CONF, March 2009.

[4] S. R. Choudhury and N. Gaur, "Dileptonic decay of B_s meson in SUSY models with large $\tan\beta$," Phys. Lett. B **451**, 86 (1999).

[5] K. S. Babu and C. F. Kolda, "Higgs-mediated $B_s^0 \to \mu^+\mu^-$ in minimal supersymmetry," Phys. Rev. Lett. **84**, 228 (2000).

[6] C. Bobeth, T. Ewerth, F. Kruger and J. Urban, "Enhancement of $\mathcal{B}(B_d^0 \to \mu^+\mu^-)/\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ in the MSSM with minimal flavour violation and large $\tan\beta$," Phys. Rev. D **66**, 074021 (2002).

[7] S. Davidson, D. C. Bailey and B. A. Campbell, "Model independent constraints on leptoquarks from rare processes," Z. Phys. C **61**, 613 (1994).

[8] R. L. Arnowitt, B. Dutta, T. Kamon and M. Tanaka, "Detection of $B_s^0 \to \mu^+\mu^-$ at the Tevatron Run II and constraints on the SUSY parameter space," Phys. Lett. B **538**, 121 (2002).

[9] G. L. Kane, C. Kolda and J. E. Lennon, "$B_s^0 \to \mu^+\mu^-$ as a probe of $\tan\beta$ at the Tevatron," arXiv:hep-ph/0310042.

[10] S. Baek, "A two-loop contribution to $B_s^0 \to \mu^+\mu^-$ at large $\tan\beta$ in the MSSM," Phys. Lett. B **595**, 461 (2004).

[11] A. Dedes and B. T. Huffman, "Bounding the MSSM Higgs sector from above with the Tevatron's $B_s^0 \to \mu^+\mu^-$," Phys. Lett. B **600**, 261 (2004).

[12] F. Zwirner, "The Quest For Low-Energy Supersymmetry And The Role Of High-Energy $e^+\,e^-$ Colliders," arXiv:hep-ph/9203204.

[13] R. Adolphi *et al.* [CMS Collaboration], JINST **0803**, S08004 (2008) [JINST **3**, S08004 (2008)].

[14] T. Sjostrand, S. Mrenna and P. Skands, JHEP **0605**, 026 (2006) [arXiv:hep-ph/0603175].

[15] CMS Collaboration, "CMS High Level Trigger," CERN/LHCC 2007-021 (2007).

[16] P. Valnaer, L. Barbone, N. De Filippis, T. Speer, O. Buchmuller, F.-P. Schilling, "Impact of CMS Silicon Tracker Misalignment on Track and Vertex Reconstruction," CMS-NOTE-2006-029.

[17] J. Heinrich, C. Blocker, J. Conway, L. Demortier, L. Lyons, G. Punzi and P. K. Sinervo, "Interval estimation in the presence of nuisance parameters. 1. Bayesian approach," arXiv:physics/0409129.