Reply To Comments for $B \to p\overline{p}\mu\nu$

Matthew Tilley, Mark Smith, Mitesh Patel

1 Comments from Patrick Owen

Comment 1:

A measurement of the ratio $R(p\overline{p})$ is therefore of interest and would be the first $|V_{ub}|$ mediated R-ratio to be measured. The observation of the decay $B \to p\overline{p}\mu\nu$ is a necessary

line 38: I dont think this is unambiguously true, R(pi) has technically been measured via Belle's limit of B->pitaunu.

Reply: While I'm not sure if it's technically a ratio measurement, this is true. I have added this to the note.

Comment 2:

the $p\overline{p}\mu$ tracks can be removed with the help of an isolation tool. The tool used is the so-called TupleToolApplypMuIsolation which applies a boosted decision tree (BDT) to the additional tracks in the event [12].

Did you adapt the tool to treat the signal as a ppmu vertex rather than just a pmu vertex?

Reply: We found that it uses the ppmu vertex out of the box. It was coded to treat all p $\overline{p}$ and $\mu^{\pm}$ tracks as signal tracks.

Comment 3:

The normalisation $B^+ \to (J/\psi \to \mu^+\mu^-)K^+$ simulation samples used are Reco16/Sim09b for 2016 simulation and Reco14c/Sim09b for 2012 simulation. The

There isn't any simulation for 2011?

Reply: There is for $B \to J/\psi K$. However I have been using 2012 MC reweighted to 2012+2011 data throughout the analysis. Thought it was best to be consistent here. We will look into this and have ordered some signal 2011 simulation for further efficiency studies.

Comment 4:

$B^+ \to (N^* \to pX)\overline{p}\mu^+\nu_\mu$	Bu_pNstmunu,pX̃TightCut.dec	12813400	2012/2016
$B \to D^0 p\overline{p}\mu\nu$	Bu_D0ppmunu=TightCut.dec	12572000	2012/2016
$B \to (D^0 \to \mu X)p\overline{p}$	Bd_D0ppbar,Xmunu=DecProdCut.dec	11774000	2012/2016

Do these simulation samples also include excited versions of these backgrounds i.e. $B- > N^*N^*munu$ or $D^*ppmunu$?

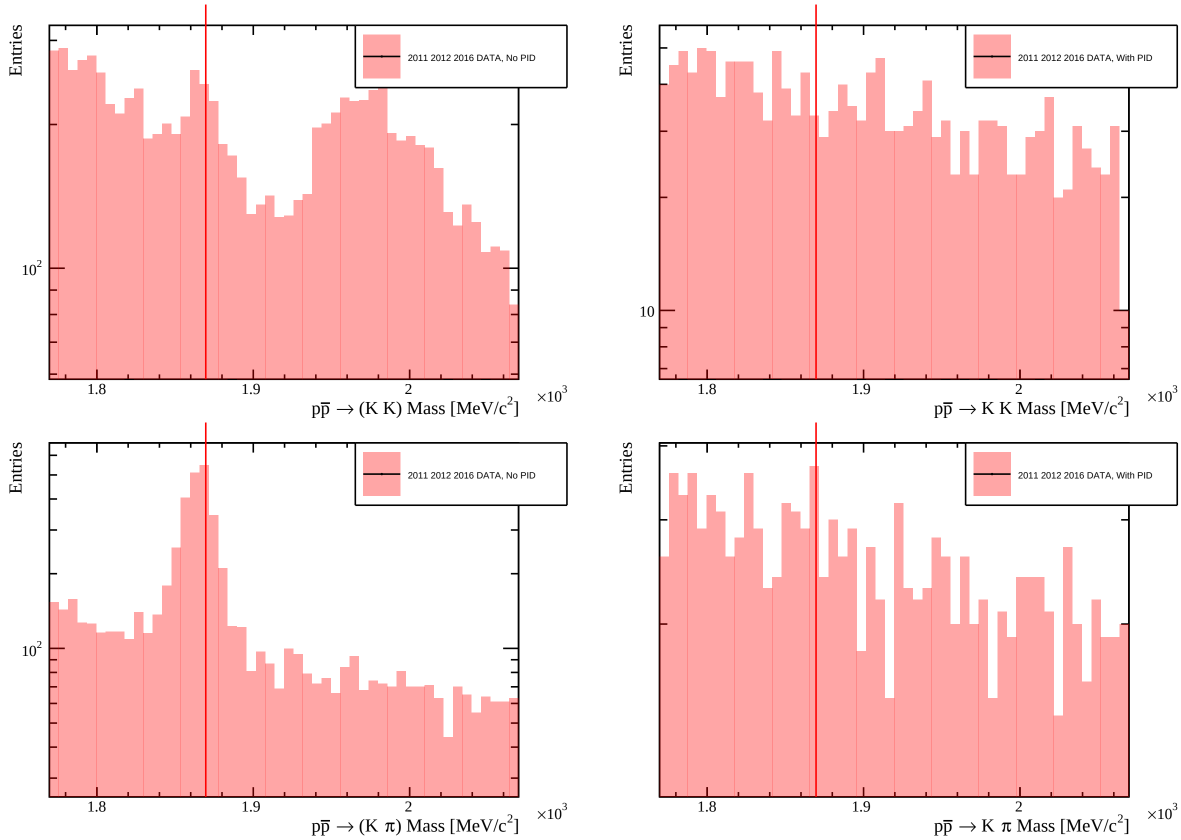
Reply: They do not, the argument for why they are not needed comes from the treatment of $B \to \Lambda_c N^*\mu\nu$ in section 5.2.6. This is taking the largest $|V_{cb}|$ version of the $B \to (X_{bary} \to p)(\overline{X}_{bary} \to \overline{p})\mu\nu$ background. Since the efficiency of this sample is so low and since $B \to N^*p\mu\nu$ is clearly smaller than $B \to \Lambda_c p\mu\nu$ we similarly consider $|V_{ub}|$ versions such as $B \to N^*\overline{N^*}$ to also contribute a negligible amount. It also is worth considering the amount of overall freedom the part-reco shape has, and if an additional component can have much of an effect on the signal yield. This argument has been added to the note and is now subsection 5.2.7 in version 2.

Comment 5:

A data-driven method is used to assess the shape and level of the background coming from misidentified protons and anti-protons. A sample of data is obtained where the particle identification selection is applied to only the proton or the anti-proton. This sample of data is referred to as the fake-p sample. The data is then split into π, K or

Did you check that double mis-ID of the two protons negligible?

Reply: The most obvious place to look for this is in the $D^0 \to K(K/\pi)$ peak. We do see double MisID before our additional PID requirements. However, after the selection it is gone. I've attached a plot giving the protons the kaon hypothesis with a red line for the D0 mass. The peak for D0->K pi is shown below. This has been added to the note as subsection 5.6.



Comment 6:

unknown. The efficiencies of ghost tracks to be in the ProbNN$^{\text{combined}}$ categories is assumed to be approximately zero as the combined ProbNN selection includes $(1 - \text{ProbNN}ghost)$.

What is the U->U efficiency assumed to be in this case?

Reply: It is always chosen such that the column sums to 1. In this case it is 1.

Comment 7:

statistical uncertainty of the PIDCalib samples. The ghost efficiencies are also randomly fluctuated uniformly between zero and the efficiency of the correct particle to be identified in each region. The standard deviation of the sum of the 100 re-sampled weights is added in quadrature with the Poisson uncertainty from the fake-p sample statistics. The muon misID is treated differently and is discussed at the end of this section.

Is the matrix re-inverted every time you fluctuate thees ghost efficiencies or is it just the final weight?

Reply: Yes. The whole unfolding procedure is preformed again each time.

Comment 8:

Decay
$B^+ \to (\Lambda_c^+ \to pX)\overline{p}\mu^+\nu_\mu$
$B^+ \to (\Lambda_c^{+*} \to pX)\overline{p}\mu^+\nu_\mu$
$B^+ \to (\Sigma_c^+ \to pX)\overline{p}\mu^+\nu_\mu$
$B^+ \to (N^* \to pX)\overline{p}\mu^+\nu_\mu$
$B^+ \to (\Delta^+ \to pX)\overline{p}\mu^+\nu_\mu$
$B^{+/0} \to (D^0 \to \mu^+ X)p\overline{p}X$
$B^+ \to D^0 p\overline{p}\mu^+\nu_\mu$

Similar to my comment from before, the anti-proton is always in the ground state here, what about if it is also excited (e.g. N*)? Does that make a difference?

Reply: I expect all the double resonance cases to be smaller than $\Lambda_c N^* \mu\nu$. As I say in response to comment 4.

Comment 9:

Λ_c^+ decay weight	Decay	Decay model
0.02800	$p\ K^-\ \pi^+$	PHSP
0.01065	$p\ K^{*0}$	PHSP
0.00860	$\Delta^{++}\ K^-$	PHSP
0.00414	$\Lambda(1520)^0\pi^+$	PHSP

Table 6: Shouldn't the Lc decay be the more extensive list from table 5 also for this case rather than just pKpi?

Reply: For this sample the list of Λ_c decays is a bit sorter, but the major decays are there. I don't expect this makes a large difference to the corrected mass shapes.

Comment 10:
can have on the signal yield. The simulated $B \to (X_u^* \to pX)\overline{p}\mu\nu X$ cocktail does not include B^0 equivalent decays. However, all of the same N^* resonances contribute to these backgrounds and so the B^+ cocktail covers the B^0 corrected mass shapes.

I guess you don't mean exactly the same N^* resonances, but the isospin equivalent ones right? I would guess that these could decay into different numbers of pions due to isospin rules, maybe its worth a quick mention with reference to the PDG?

Reply: Yes you are right. We have samples in the cocktail with equivalent missing mass to the B^0 decays. I don't expect it to change the overall part-reco shape I've changed the comment in the note to reflect this.

Comment 11:

5.2.6 The Decays $B^+ \to (\Lambda_c^+ \to \overline{p}X)(N_u^* \to pX)\mu\nu$

Ah there's an example of a decay that I was thinking of with more excited particles. What about $B->N*N*munu$? I don't see why this decay is not listed in table 4?

Reply: I have added $B^+ \to (X_{bary} \to \overline{p}X)(\overline{X}_{bary} \to pX)\mu\nu$ decays to the table earlier in the note to avoid confusion. The argument in response to comment 4 is now in the section 5.2.7 on $B^+ \to (\Lambda_c \to \overline{p}X)(N_u^* \to pX)\mu\nu$.

Comment 12:
Another sample rich in combinatorial events is a sample with a high B-vertex χ^2. A value of greater than $3\,\sigma$, corresponding to B-vertex $\chi^2 > 9$ is chosen for this sample. This

I guess you mean > 9/ndf? Otherwise I guess you get a lot of physics background in this sample.

Reply: Yes that is right, note has been updated

Comment 13:

The number of variants of such decays with comparable branching fractions is < 10. This would imply no greater than 85 muon misID events across all bins of $p\bar{p}$ mass. This is considered to be negligible and the muon misID is not included in the nominal data fit.

If I understand correctly, the argument is that the sum of the exclusive branching fractions gives an estimate of the inclusive ones. Would it be better to instead use the inclusive branching fraction $B- > ppX$ of 2.5%?

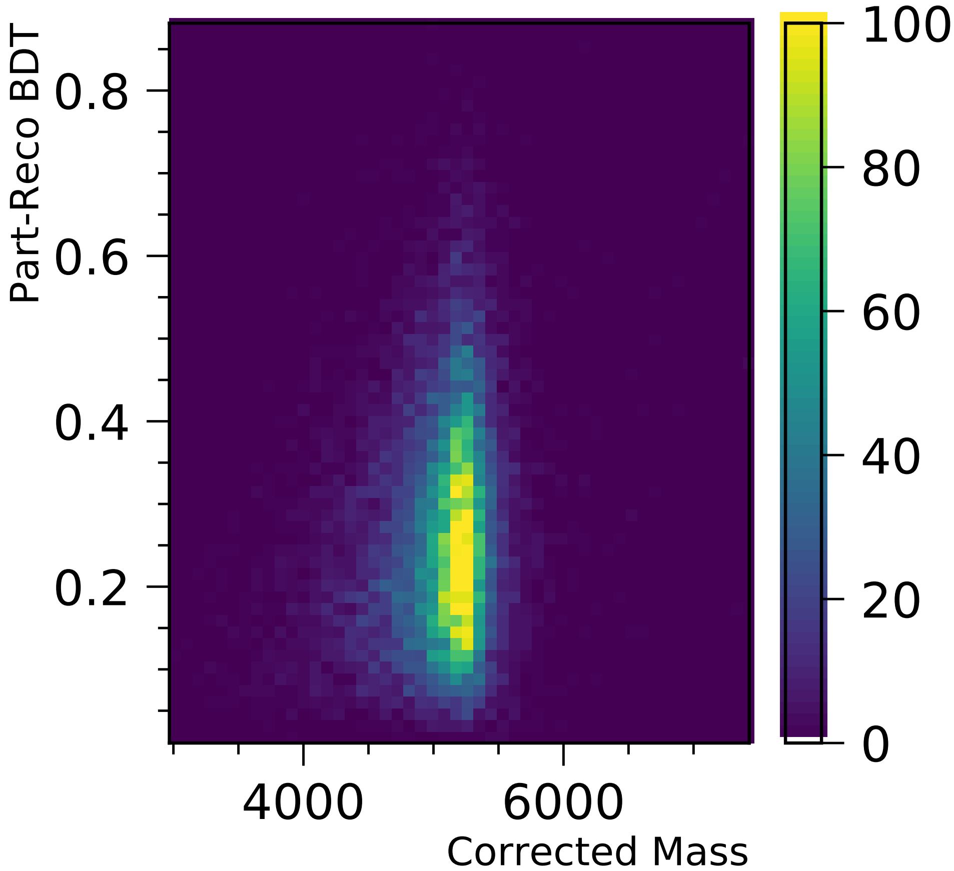
Reply: As far as I understand, the 2.5% allows the protons to come through Λ^0, so I think it would be a large overestimate.

Comment 14:

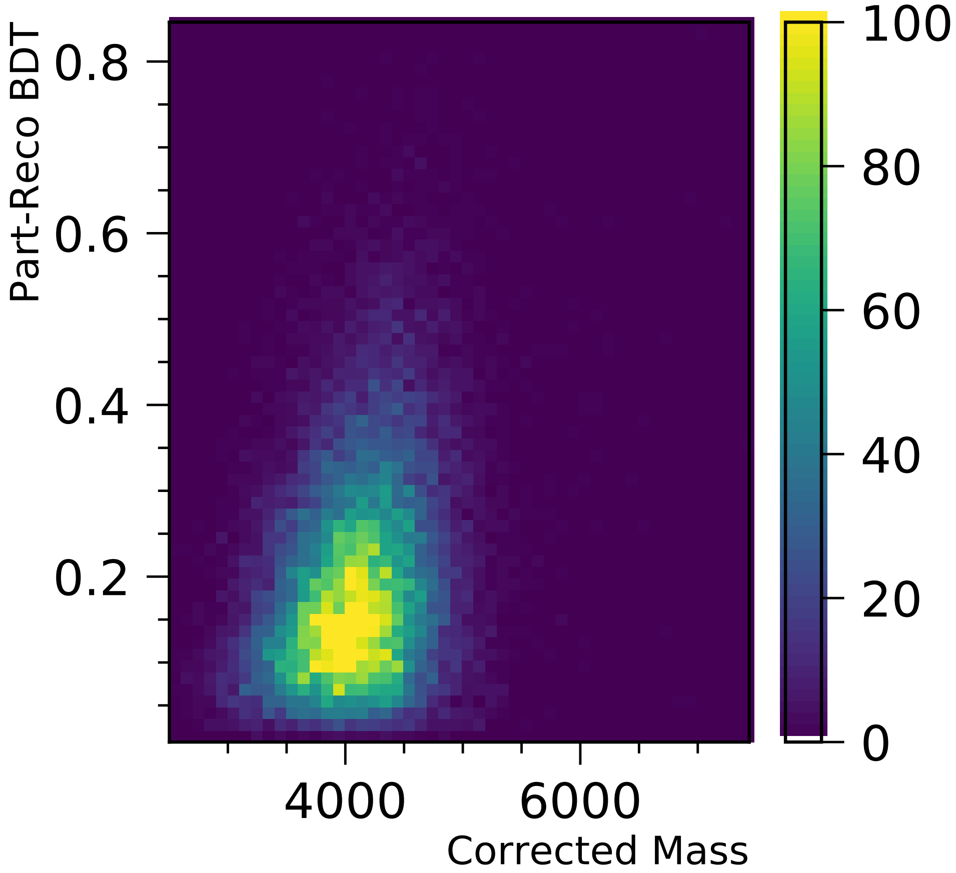
6.2.2 BDT to Remove Partially Reconstructed Backgrounds Without Additional Charged Tracks

How correlated is this BDT with the corrected mass?

Reply: It is quite correlated, though any warping of the corrected mass shapes is included in the optimisation. With the cut point at 0.05 I don't think it's a major concern. Here's a 2D plot of the 2012+2016 signal MC.:-



Here is the same plot with $\Lambda_c \bar{p} \mu \nu$:-



Comment 15:

6.4 Cut Optimisation

I'm somehow missing from this section the cuts of all the BDTs and the respect signal and background efficiencies.

Reply: For training samples as the background
Charged track bdt: 93.2% signal eff. and 80.1% $B \to \Lambda_c\overline{p}\mu\nu$ background rejection
Part-reco bdt: 98.7% signal eff. 18.4% $B \to D^0 p\overline{p}$ background rejection, 6.7% $B \to \Lambda_c\overline{p}\mu\nu$ rejection and 1.9% $B \to N^*\overline{p}\mu\nu$ rejection.
This has been added to a table in the note.

Comment 16:

StdAllNoPIDsProtons particle containers, with no additional selection; and finally, there is simulation that has been re-stripped with the PID selection removed. All efficiencies

This simulation that has been re-stripped is run on generator level MC or the reconstructed MC from the previous step? I would naturally expect the former but the latter seems to be suggested by your definition of the efficiency in the equation below this line.

Reply: The re-stripped MC is reconstructed, I run on the same DSTs for all the tuple types. Except for the local jobs used for the DPC efficiencies. I hope this answers your question.

Comment 17:

$\frac{\epsilon(B\to p\bar{p}\mu\nu)_{m_{p\bar{p}}}}{\epsilon(B\to J/\psi K)}$	Relative efficiency in bins of $p\bar{p}$ mass				
	$1.87-2.0\,\mathrm{GeV}$	$2.0-2.2\,\mathrm{GeV}$	$2.2-2.4\,\mathrm{GeV}$	$2.4-2.6\,\mathrm{GeV}$	$2.6-5.0\,\mathrm{GeV}$
Run 1	1.393 ± 0.119	1.281 ± 0.087	1.242 ± 0.082	1.279 ± 0.084	1.179 ± 0.062
Run 2	1.201 ± 0.107	1.114 ± 0.076	1.131 ± 0.074	1.129 ± 0.074	1.077 ± 0.055

Table 15: Decay product acceptance efficiency for run 1 and run 2. Quoted uncertainties are from simulation statistics.

These DPC rel efficiencies are very far away from unity. Do you understand why they differ so much between signal and normalisation?

Reply: As a check the DPC efficiency using a sample of JpsiK without `LHCbDaughtersinAcceptance` gives the same normalisation efficiency at taking the generator log. I have attached plots of the particle ETA distributions of 2012 and 2016 gen level MC. Also included are the distributions with a cut for the lowest $p\bar{p}$ mass bin. I hope this helps.

The ETA of the muons from $(J/\psi \to \mu\mu)K$ are lower than the protons from $B \to p\bar{p}\mu\nu$. It makes sense that the muons from the J/ψ come out at a wider angle and are more likely to come out of acceptance. This is especially true in the low $p\bar{p}$ mass bin. Within the Jpsi window of ppbar mass 3096 ± 50 the relative efficiency in run 1 is 1.152 ± 0.074 and run 2 is 1.047 ± 0.086. Much closer to 1. The plots include Z-inversion for the generator level MC.

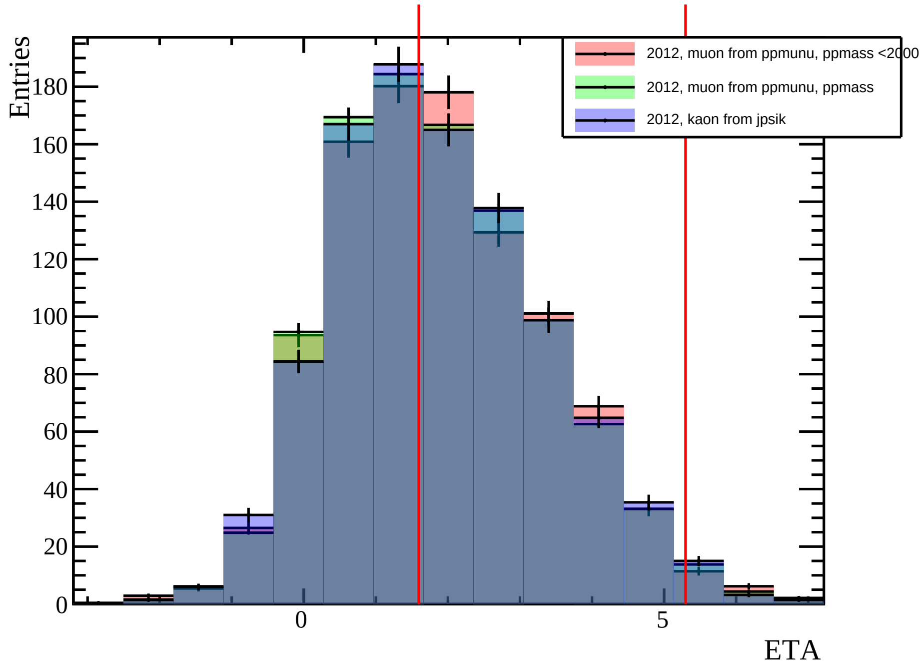


Figure 1: Distribution of 2012 simulation in ETA for the muon in $p\bar{p}\mu\nu$ and the kaon from $B \to J/\psi K$. Red is the muon for ppbar mass$<$ 2 GeV. Green is for ppbar mass in the range 3096 ± 50. Blue is for the kaon. The red lines are the ETA cut points. (Not a major sig/norm difference.)

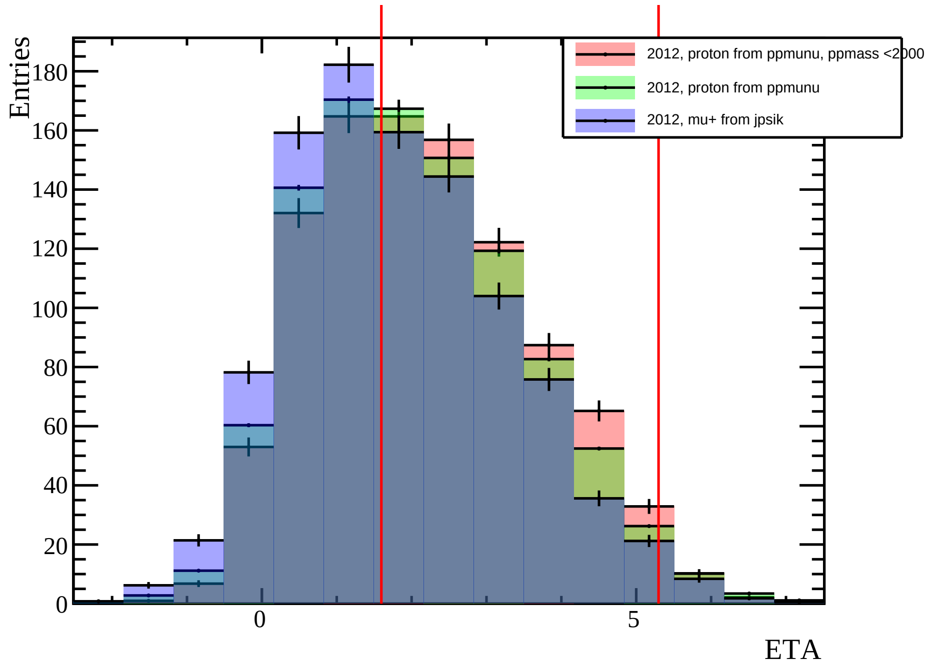


Figure 2: Distribution of 2012 simulation in ETA for the proton in $p\bar{p}\mu\nu$ and the μ^{+} from $B \rightarrow J/\psi K$. Red is the proton for ppbar mass$<$ 2 GeV. Green is for the proton with a ppbar mass in the range 3096 ± 50. Blue is for the kaon. The red lines are the ETA cut points. (Large sig/norm difference at low ETA.)

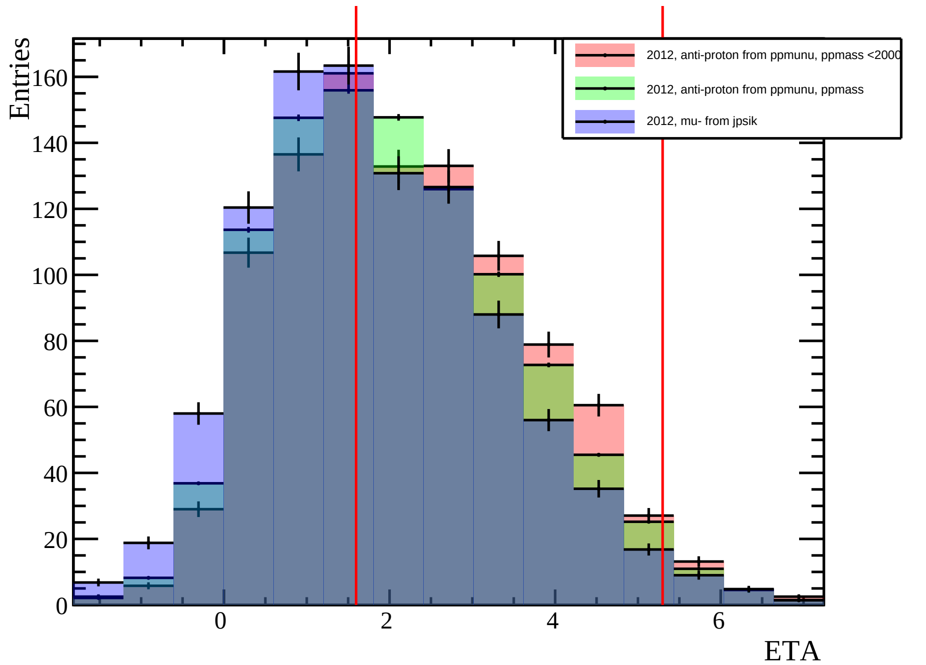


Figure 3: Distribution of 2012 simulation in ETA for the anti-proton in $p\bar{p}\mu\nu$ and the μ^{-} from $B \rightarrow J/\psi K$. Red is the muon for ppbar mass$<$ 2 GeV. Green is for the anti-proton with a ppbar mass in the range 3096 ± 50. Blue is for the μ^{-}. The red lines are the ETA cut points. (Large sig/norm difference at low ETA.)

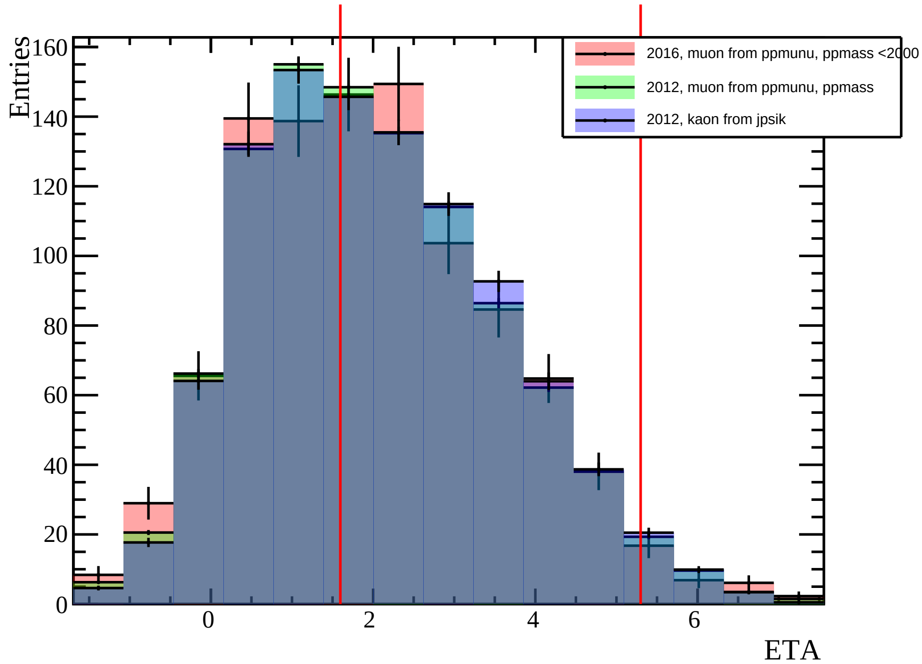


Figure 4: Distribution of 2016 simulation in ETA for the muon in $p\overline{p}\mu\nu$ and the kaon from $B \to J/\psi K$. Red is the muon for ppbar mass$<$ 2 GeV. Green is for ppbar mass in the range 3096 ± 50. Blue is for the kaon. The red lines are the ETA cut points. (Not a major sig/norm difference.)

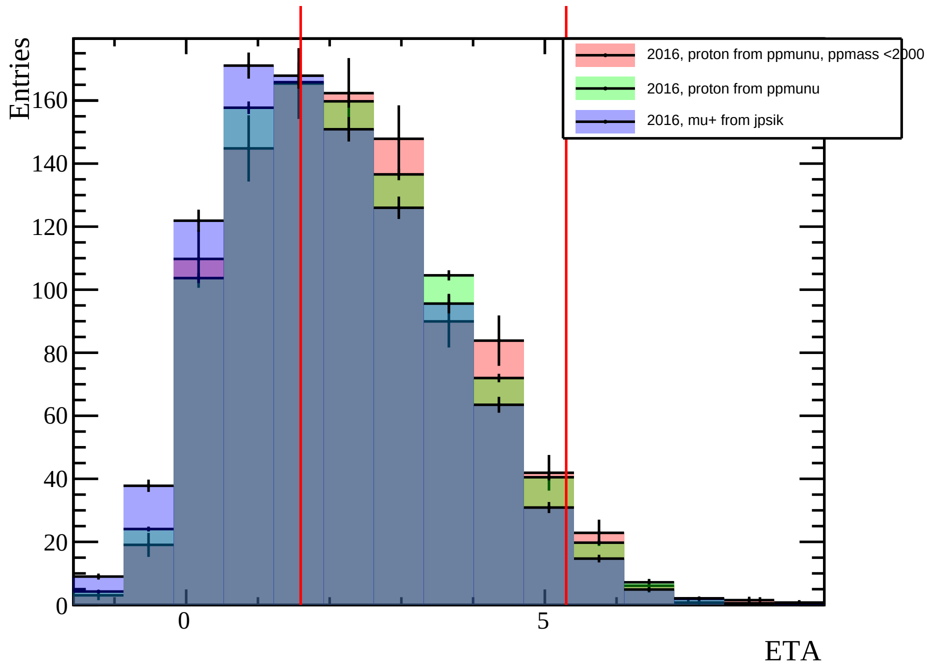


Figure 5: Distribution of 2016 simulation in ETA for the proton in $p\overline{p}\mu\nu$ and the μ^+ from $B \to J/\psi K$. Red is the proton for ppbar mass$<$ 2 GeV. Green is for the proton with a ppbar mass in the range 3096 ± 50. Blue is for the kaon. The red lines are the ETA cut points. (Large sig-norm difference at low ETA.)

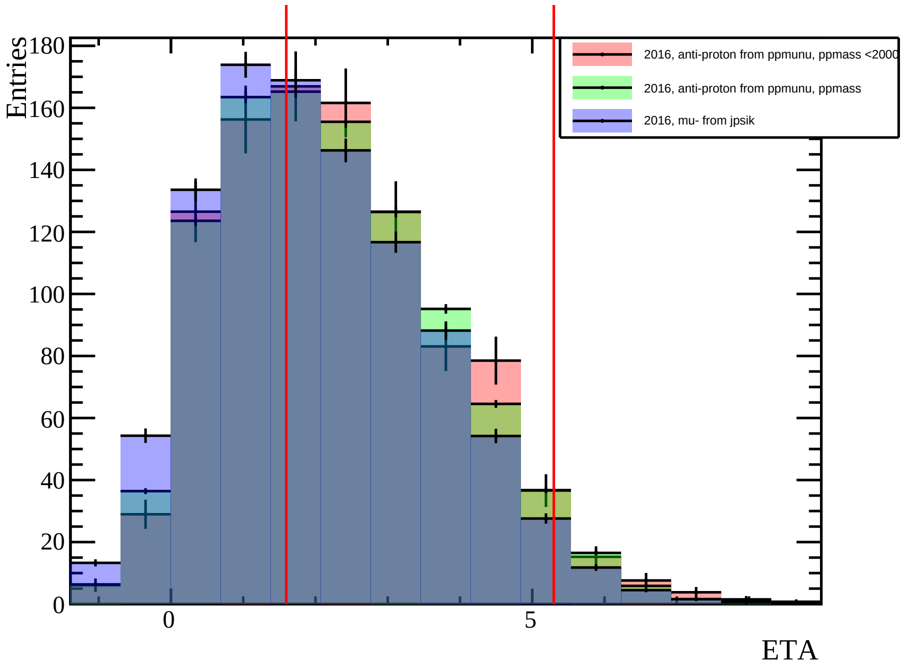


Figure 6: Distribution of 2016 simulation in ETA for the anti-proton in $p\overline{p}\mu\nu$ and the μ^- from $B \to J/\psi K$. Red is the muon for ppbar mass$<$ 2 GeV. Green is for the anti-proton with a ppbar mass in the range 3096 ± 50. Blue is for the μ^-. The red lines are the ETA cut points. (Large sig/norm difference at low ETA.)

Comment 18:

$\frac{\epsilon(B\to p\overline{p}\mu\nu)_{m_{p\overline{p}}}}{\epsilon(B\to J/\psi K)}$	Relative efficiency in bins of $p\overline{p}$ mass				
	$1.87-2.0\,\mathrm{GeV}$	$2.0-2.2\,\mathrm{GeV}$	$2.2-2.4\,\mathrm{GeV}$	$2.4-2.6\,\mathrm{GeV}$	$2.6-5.0\,\mathrm{GeV}$
Run 1	0.319 ± 0.017	0.336 ± 0.02	0.329 ± 0.015	0.327 ± 0.016	0.281 ± 0.037
Run 2	0.371 ± 0.047	0.415 ± 0.066	0.415 ± 0.065	0.391 ± 0.053	0.332 ± 0.077

Table 17: Trigger efficiency for run 1 and run 2. Quoted uncertainties are from simulation statistics.

This is also a very large difference as I would expect the absolute efficiency of the signal to be above 50%. Do you know at which level this efficiency ratio comes from?

Reply: So for the trigger efficiency in simulation, the absolute numbers I get are:

2012
TRIG: [1876, 2000] sig: 0.160633288351 norm: 0.504273460399
TRIG: [2000, 2200] sig: 0.169451706143 norm: 0.504273460399
TRIG: [2200, 2400] sig: 0.166046113995 norm: 0.504273460399
TRIG: [2400, 2600] sig: 0.164988248267 norm: 0.504273460399
TRIG: [2600, 5000] sig: 0.141506369946 norm: 0.504273460399

2016
TRIG: [1876, 2000] sig: 0.170810054315 norm: 0.461017695149
TRIG: [2000, 2200] sig: 0.191145986519 norm: 0.461017695149
TRIG: [2200, 2400] sig: 0.191259435122 norm: 0.461017695149
TRIG: [2400, 2600] sig: 0.180191303713 norm: 0.461017695149
TRIG: [2600, 5000] sig: 0.153170341113 norm: 0.461017695149

So the normalisation is as you'd expect $\approx 50\%$, but the signal is much lower, closer to $15-20\%$.

If I apply only the L0Muon Trigger only I get absolute numbers
TRIG: [1876, 2000] sig: 0.471291671159 norm: 0.790244580858
TRIG: [2000, 2200] sig: 0.461226329001 norm: 0.790244580858
TRIG: [2200, 2400] sig: 0.429743534392 norm: 0.790244580858
TRIG: [2400, 2600] sig: 0.409195879377 norm: 0.790244580858
TRIG: [2600, 5000] sig: 0.332566180417 norm: 0.790244580858

Run 2
TRIG: [1876, 2000] sig: 0.439246909573 norm: 0.724667158964
TRIG: [2000, 2200] sig: 0.432335948644 norm: 0.724667158964
TRIG: [2200, 2400] sig: 0.406753807325 norm: 0.724667158964
TRIG: [2400, 2600] sig: 0.378014218623 norm: 0.724667158964
TRIG: [2600, 5000] sig: 0.314529950039 norm: 0.724667158964

The large difference in L0 muon could just be because two muons are better than one.

Comment 19:

7.5 BDT and Corrected Mass Error Selection

The efficiency of the additional offline selection in corrected mass error, charged track isolation BDT and the part-reco BDT are taken from simulation. The normalisation has a selection efficiency of 50% due to the random splitting of the $B^+ \to (J/\psi \to \mu^+\mu^-)K^+$ sample for reweighting and normalisation. The relative selection efficiency is provided in Tab. 19.

How is the corrected mass error defined for B->J/psi K? Just using one of the muons and ignoring the other one?

Reply: Important to point out here, there is no cut on this for the normalisation channel. In the systematics section I use it for an efficiency check (in that case it is the MCORRERR using all three tracks).

Comment 20:

The PID performance in run 2 seems to be much greater than in run 1. This is the largest difference in efficiency between run 1 and run 2. This difference is exaggerated by the lack of PID selection on the normalisation channel. Changes to the PID efficiencies do not cancel as they do in analyses where the normalisation channel contains the same tracks as the signal in the final state. The algorithm for ProbNN is improved for run 2, this may be the source of the increased efficiency.

A couple of times here there are speculations about what is causing the relative efficiency difference. I think it would be good to follow up on this to find out concretely whats going on here between run I/II.

Reply: The PID does give the largest efficiency difference between run 1 and run 2. It stands out as the most glaring difference. This comes straight from the PIDCalib histograms. We wouldn't necessarily expect it to be the same, and there is not PID on the normalisation so differences do not cancel. Nonetheless, if I trust that the PIDCalib efficiencies are correct then there is not a problem here. I've got the plots below of the histograms that drive this difference from PIDCalib.

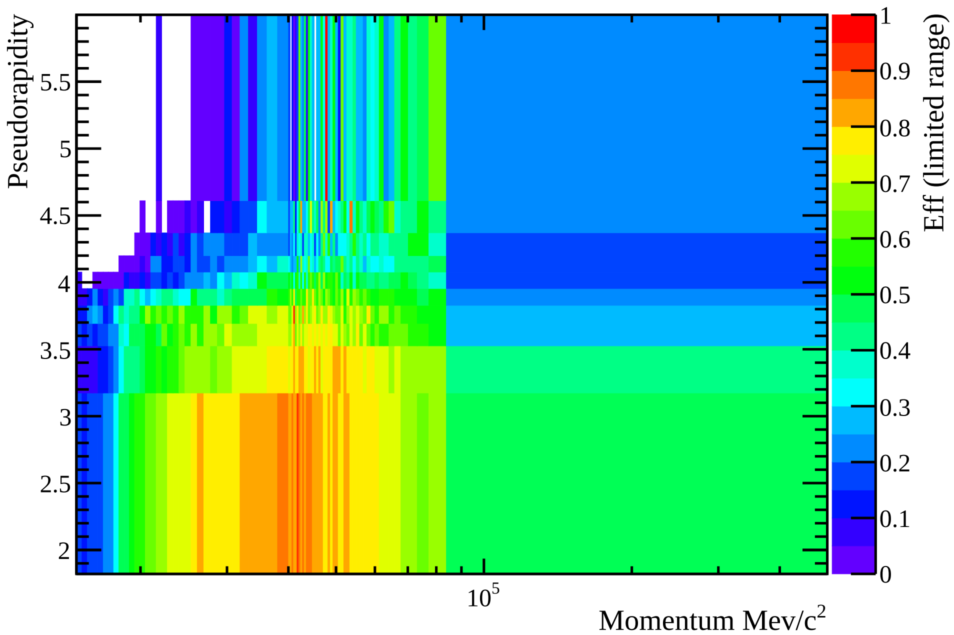


Figure 7: Run 1 proton PIDCalib efficencies in bins of P and ETA. Overall lower in efficency

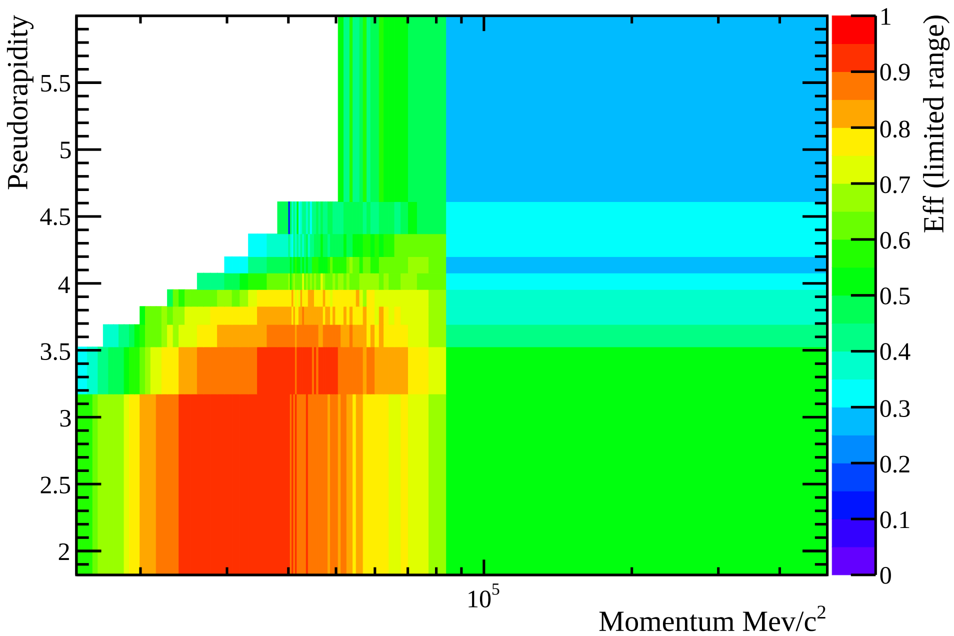


Figure 8: Run 2 proton PIDCalib efficencies in bins of P and ETA. Much higher absolute eff. by a factor of ≈ 1.5 particularly at low ETA.

Comment 21:

The inputs for this weighted efficiency are $\varepsilon_{\mathrm{run}\ 1} = 2.978 \pm 0.029$, $\varepsilon_{2016} = 1.793 \pm 0.179$ and $R_{cs} = 2.0 \pm 0.52$. The propagated uncertainties in each bin along with the systematic uncertainty due to the combination is given in section 10. Efficiencies of combined 2011, 2012 and 2016 simulation, weighted for luminosity are given in Tab. 21.

I know you have a big uncertainty on Rcs but I think it would be more robust to take the ratio of bbar x-sections between 7 and 13 and conservatively assume the systematic uncertainties between the two are uncorrelated.

Reply: I agree, we've changed the analysis and are taking Rcs this way now. Here $R_{cs} = 2.00 \pm 0.35$ instead of 2.0 ± 0.52. This has a very small difference on the final uncertainties in each bin.

Comment 22:

For the backgrounds most similar in average corrected mass, fitting the sum and the difference of two α_n parameters reduces the correlations in the fit. The fractions of each background type are related to the α_n nuisance parameters by

These alpha parameters are shared across the ppbar mass bins or is there one for each bin?

Reply: The fits are totally independent in each bin, no parameter sharing

Comment 23:

described in section 5 is used. A Gaussian constraint is used on the proton misID yield.

Where does this constraint come from? Just the statistical yield?

Reply: Stat yield and the fluctuation of the ghosts etc.

Comment 24:

in the ratio with the normalisation channel. This means that while the systematic uncertainty based on the size of the PID calibration sample is small, the dependence of the PID efficiency on the nTracks distribution, leads to a much larger systematic due to the kinematic reweighting. There is an additional, irreducible systematic uncertainty of

I think if you binned the PID in just kinematics rather than 3D with nTracks, you might have a better control over this. The multiplicty for the PID calibration sample is bound to be the same as you are using Lb->Lcmunu decays for the proton ID.

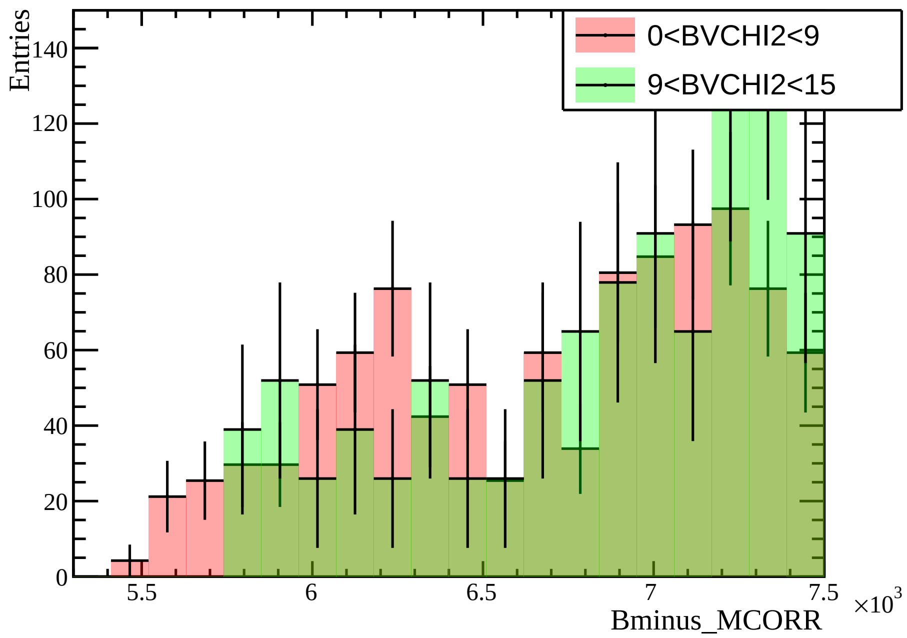
Reply: This is a good point, I don't think it changes the overall picture much though. Perhaps this justifies removing multiplicity reweighing from the strategy. However, nTracks reweighting is also included to ensure we get the correct charged track isolation efficiency.

Comment 25:

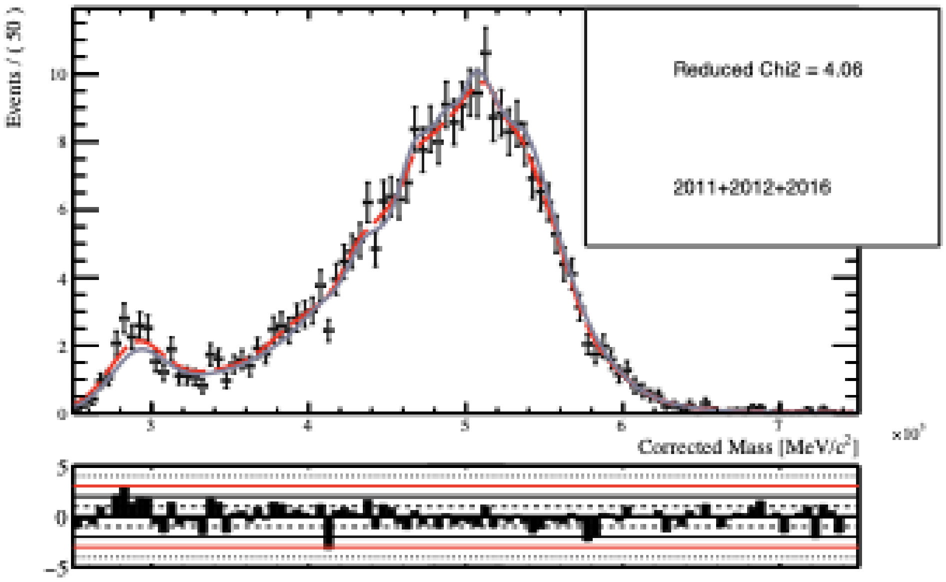
10.2 Uncertainty in the Shape of the Combinatorial Background

Is there any way to look at the events above the B mass to validate the bad vertex sample?

Reply: This was tried and there really doesn't seem to be enough stats in the inverted vertex sample in that region to tell. Regardless, I've attached the corrected mass shape for 2011+2012+2016 data and B mass > 5300. (But with only stripping selection applied). In this case, I cant say these look too different. This has been added to the note.

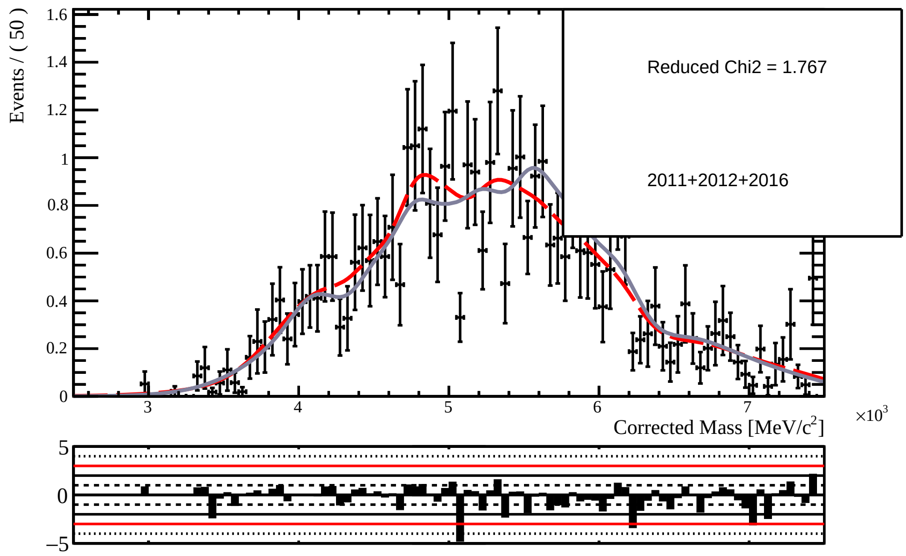


Comment 26:



This is a tiny variation, how does it cause the large 50% systematic in the final bin? (Same question for the smoothing parameter variation).

Reply: What is shown here is an example for the first bin. The an example variation in the last bin looks like the plot bellow (lower stat and a larger fluctuation). This is also 50% of an error on a small number of events in the final bin, it will not affect the overall BF much. By eye you can see that the fluctuations are quite similar to the size of the signal in that bin.



Comment 27:

of fluctuated shapes. The bias on the signal yield is taken as the systematic uncertainty. The uncertainties are presented in Tab. 30 and the pull distributions are shown in Fig.44.

Is there also a bias on the statistical uncertainty as well as the central value?

Reply: There is evidence of a slight over-coverage of uncertainties, pulls width between 0.9 and 1. (implies errors a little too large). This is considered to be acceptable to leave. Note is updated with a comment about this. The pulls also seem to be symmetric.

Comment 28:

10.11 Run 1 and Run 2 Combination Systematic

I really think you shouldn't need this systematic. You know how many J/psiK there are in run1 and run2 with very good precision. Can you not get this number very precisely from that?

Reply: Sorry if I've misunderstood, but this would only give the ratio of number of events after selection. To combine the efficiencies I need the number before, which is why I use the ratio of lumis and cross sections. (Which have an associated error). If I were to use the jpsik events it would require the assumption that the absolute eff for $B \to jpsik$ is correct.

Comment 29:

Fluctuations of the pQCD model are generated by randomly sampling these constants uniformly within $\pm 5\sigma$ deviations of their nominal values.

Do these fluctuations get the pQCD predictions closer to the Belle BF?

Reply: We leave out parameters that do not affect the shape, but affect the level. As a result we are not sensitive to the BF that comes out.

The transition of $B \to p\bar{p}$ is described in the most general way by a vector current and axial-vector current with a factor in front of each of the possible bilinear covariants. The pQCD model comes up with various relations between these 10 numbers giving 6 model parameters. In particular the relations describe the enhancement at low $m(p\bar{p})$ mass - they scale as $\frac{1}{m^6}$ (for 3 hard gluons). This is a general model for $B \to \mathbf{B\overline{B}}$ decays. They therefore fit the decay rates amd angular distributions of many experimentally measured final state decays (eg $B \to p\bar{p}K$, $B \to p\bar{p}\pi$ etc) to extract the 6 parameters. Some parameters are entirely dominated by the BF, others by the shapes of the kinematic variables. We only care about the shapes.

2 Comments from Michel De Cian

Comment 30: l 104: Track-chi2 and ghost probability are not very well reproduced in simulation (this will come back in a question further down). I assume here it would just make the isolation less performant on data than on MC.

Reply: I do not disagree. However, there is a systematic for this.

Comment 31: l 230ff: What simulation version are these? Is it consistent with the signal decay, and no Sim09d?

Reply: signal is 2012-sim09b/2016-sim09c. Norm is 2012-sim09b/2016-sim09b. The sim09a MC was scrapped and no sim09d is used.

Comment 32: Figure 7: I have some troubles understanding the colour scheme of these plots. Are these colour semi-transparent, e.g. I don't see red.

Reply: I've added a comment in the caption to help avoid confusion. The red and blue totally overlap, by not seeing any red it shows that the reweighting is working. (That's why it looks like there is a purple component).

Comment 33: l 342ff: Did you check these statements with MC? It will not give you precise numbers, but should allow you to see if the statement that "the efficiencies of the ghost tracks to be in the ProbNNcombined categories is assumed to be zero" is true, as you can identify ghosts in MC. Also, flucutating eff(U -> pi) up to eff(pi -> pi) seems quite large. Again no constraint from MC possible? Of course it might just not matter in the end, as you will hopefully not have too many ghosts in the sample.

Reply: I agree that we could probably make an argument about ghost efficiency's from MC, we have not done this. We say that since the effect seems to be quite small with a clearly drastic change in the ghost efficiency we have full considered it. I don't think we'll gain much by reducing the size of the fluctuation.

Comment 34: l 364ff: I think normally one should take more systematic uncertainties with PIDCalib into account than just the statistical uncertainty of the samples, I'll mention this again in the systematics section.

Reply: This has been improved. We now use the coarse binned sample that we use for unfolding to get a systematic. This is quite conservative, and the systematic now generated from this can be reduced if necessary, but it doesn't significantly affect the overall result.

Comment 35: l 375ff: Are there really no $Lambda_b$ decays that could play a role? $Lambda_b->pJ/psipi$, with $J/psi->pp$ is too low in BR?

Reply: $Lambda_b->pJ/psipi$ or $Lambda_b->pJ/psi\mu\nu$ already seems low and we would have an additional proton track of the isolation. Also, $J/psi->pp$ is small and vetoed. We have also predicted the level of $Jpsi->ppX$ elsewhere in the note.

Comment 36: l 439: How about $Lambda_b->Lambda_cpipimunu$, with double MisID? Actually I think you don't quote the PID efficiencies so far, maybe it would be good to give some numbers to get a feeling with MisID backgrounds could be dangerous and which not.

Reply: Double misID has been added to the note, section. Here we look at the peak you get from $D^0 \to K(K/\pi)$. We think the large BFs of $B \to D\mu\nu$ decays are most dangerous for double misID. See reply to comment 5 from Patrick. As for PID efficiency from the PIDCalib samples I use for misID unfolding I can say that at the cut point of $\mathrm{ProbNN_{combined}} > 0.45$ the average $K^+K^- \to p\bar{p}$ misID efficiency is 1.6×10^{-5} for run 1 and 1.8×10^{-5} for 2016. This is higher than the efficiency of $\pi \to p$. This has been added to the note in section 5.6.

Comment 37: l 445ff: Can you gain some knowledge from the region of m_{corr} above the B mass, or the visible mass above the end point?

Reply: There is very low stat, but we took a look at the distribution with just stripping cuts. Please see response to Patrick's comment 25 for the plot.

Comment 38: l 496: I assume this would only affect bins below m(pp) = m(J/psi), as missing a particle will shift the mass to the left, which might distribute the events not evenly across the pp mass. Maybe it still does not matter in the end though.

Reply: Yes, we were worried about Jpsi->ppX leaking into the first four bins, however we are confident from this study that it is negligibly small.

Comment 39: l 510: Did you try taking a sample of B -> ppK for the MisID, as kaons normally have a higher change to be misidentified as muons than pions?

Reply: This is a good suggestion, we have not yet used B -> ppK simulation. The decay $B- > (J/\psi \to pp)(K \Rightarrow \mu)$ would be included in the $J/\psi \to pp$ combinatorial study. If this is a concern we can also use our fake-mu sample for an additional systematic. This has been flagged as a yet to be addressed comment in the version 2 of the note.

Comment 40: l 524ff: Are you affected by the AALLSAMEBPV bug for 2016, as I think the 2BodyTopo uses this?

Reply: Yes, I believe it is a bug in the Topo. Attached below are the total relative efficiency with an nPV cut on the signal and normalisation. There is a slight drop in relative efficiency with nPV but the gradient is within 2σ of being flat in each case.

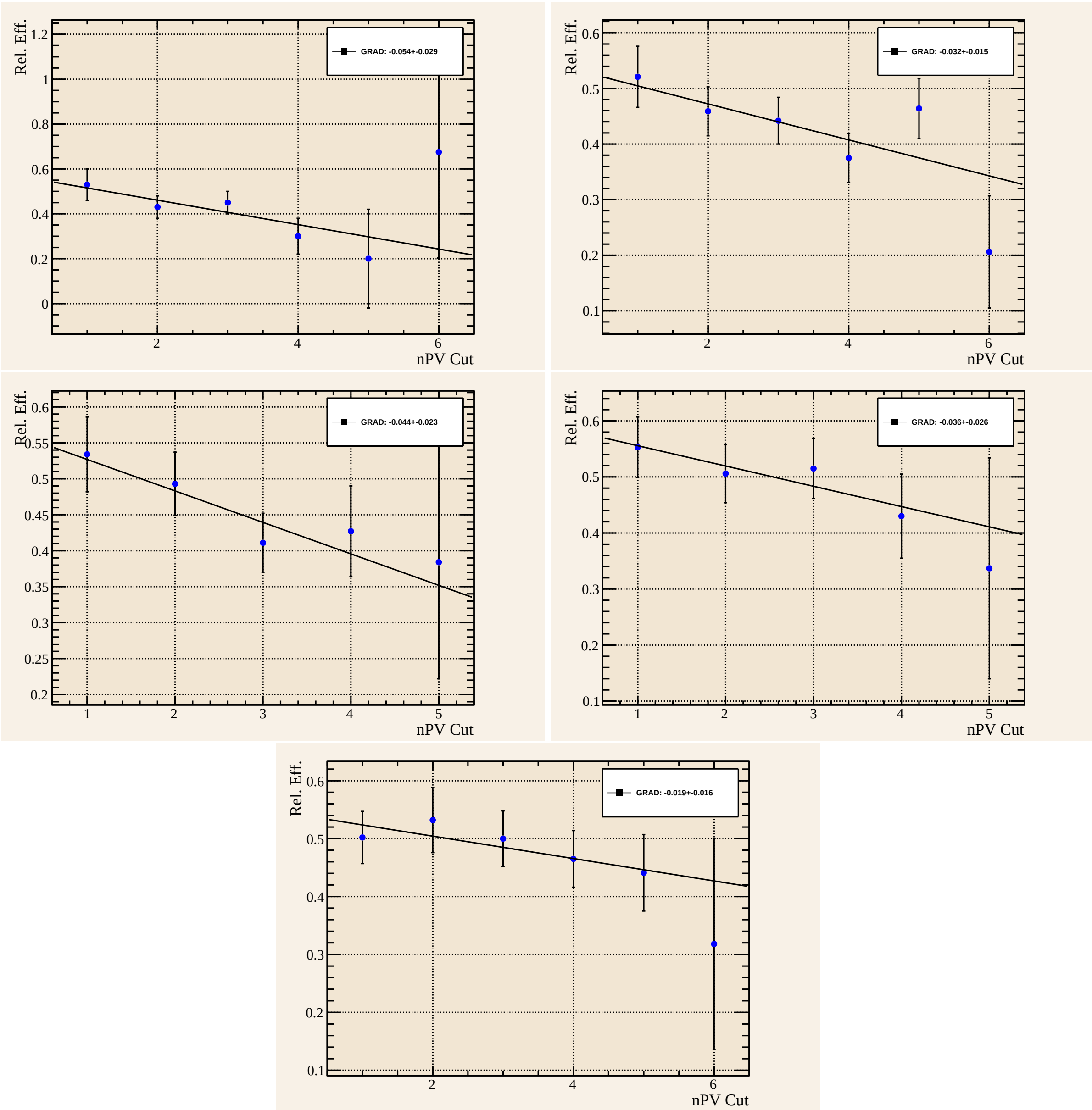


Figure 9: Relative total efficiency of signal/normalisation in 2016 as a function of nPV. Bin $1 \rightarrow 5$.

Since there is still a slight efficiency change with NPV I re-weighted the MC using (below) JpsiK s-weighted data and MC. The difference in the total efficiency in bin $1 \rightarrow 5$ is:
r1 (no bug present): 3.2 %, 2.5 %, 2.2 %, 2.8 %, 2.1 %
r2 (AALLSAMEBPV bug): 2.6 %, 1.8 %, 2.2 %, 1.3 %, 1.5 %
Interesting that re-weighting nPV has a bigger effect in run1 than run2. In both cases it is deemed small enough to ignore.

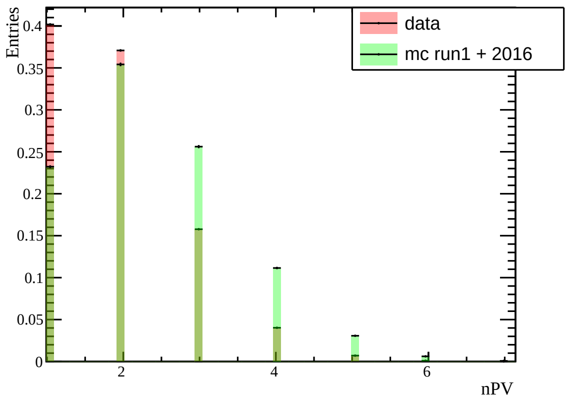


Figure 10: Number of PV's comparison of $B \to J/\psi K$ simulation and s-weighted data. Used for re-weighting.

Comment 41: Table 12: I darkly remember a bug that came into play when applying "not isMuon" at some point during Run II. Did you check that?

Reply: Haven't been able to find this bug yet, we will investigate further. It has been flagged at the start of the version 2 note.

Comment 42: l 564: You don't seem to use any neutral isolation, like cone-isolation, trying to find calo clusters, etc. These have proven to be not useful?

Reply: This is is development for the tau mode, but wasn't deemed necessary for the mu mode.

Comment 43: l 648: Is this equivalent to a more traditional FOM?

Reply: yes, we used toy fits since counting $s/\sqrt{s+b}$ in a region didn't full capture the effect that the changing shapes had (particularly for the way a MCORR error cut improves the peak resolution). So this optimisation should give us the best fractional uncertainty, which is what we are after.

Comment 44: l 677: I am not sure I understand: You generate without DecProdCut and then apply it, and the ratio is the efficiency? At least that would be what I assume. The eta values seem to be a bit different from the usual LHCb ones, do they correspondly exactly to DecProdCut? Also: Why don't you use the reweighted MC to determine the efficiency? I assume it could have an effect (also on the angular distribution).

Reply: I do generate without a DPC and then calculate it in bins of ppbar mass. The cuts I apply offline correspond exactly with the LHCbDauthersinAcceptance cut from here `https://gitlab.cern.ch/lhcb/Gauss/blob/9831c9210565779c2f32174eccaeb3617f897fb0/Gen/Generators/src/component/DaughtersInLHCb.cpp`. Should be 10 mrad < theta < 400. So 5.298 > eta > 1.596. exactly values are used for cal. rounded to 5.3 and 1.6 in the note.

This has been verified by doing the same thing to the JpsiK simulation. I get the same DPC efficiency when applying these ETA cuts as when I look at the generator log with LHCbDauthersinAcceptance applied. See the reply to Patrick's comment 17 for a full explanation of the relative efficiencies and the included ETA plots.

Comment 45: Table 15: I am also puzzled about these very low ratios, even when splitting into pp bins. Is this comparing the DecProdCut efficiency of B -> ppmunu with the DecProdCut efficiency of J/psi K, where the DecProdCut of J/psi K was already applied?

Reply: The reply to Patrick's comment 17 addresses the issue of DPC efficiencies that are >1.

Comment 46: l 707ff: I am not sure the simulation is such a good representation, but I think you asses this later.

Reply: Yes, we do take a systematic for this later on.

Comment 47: l 738ff: Given that you have so much statistics for J/psiK and no PID applied I am surprised you see no J/psi pi.

Reply: I did include a component for a while that fit to zero events. I removed it for simplicity. I think the reason is that while JpsiK still has high stat here, it's actually has a bit lower than usual, there is very tight kinematic selection on JpsiK here (much tighter than normal for consistency with the signal mode). For example, muons have > 15GeV.

Comment 48: l 823: The last bin seems to have quite a high bias. Any explanation for this?

Reply: It is quite biased though we are dealing with very few events in this bin, (and a small error to be biased on). So it doesn't affect the result overall.

Comment 49: Table 25: Maybe I missed it, but how do you combine the systematics between the years? They might have different levels of correlation, or not?

Reply: For this we do assume that these systematic uncertainties are uncorrelated between years.

Comment 50: l 876ff: As mention before, is there anything to learn from MC?

Reply: Something could be learned from simulation of ghost tracks. But, we think we have a conservative treatment of this that doesn't significantly affect the overall result.

Comment 51: l 887ff: Do you change the smoothing parameter once, make toys, get the deviation, and then change it again, make toys, get the deviation, etc. and then take the largest deviation in the end? Or do you fit each toy with a different smoothing parameter and then take the deviation? In the latter case, it could easily be that all deviations average out, but the "true" smoothing parameter is different from the chosen one, which would not be reflected in the systematic uncertainty.

Comment 52: l 895ff: Same question as the one before.

Reply: When generating toys we always you the overall nominal fit to data and then fit a varied model. We operate under the assumption that this is a "good" fit, and so the toys represent the data. For the set of varied models we randomly pick a parameter between 1 and 2. We fit the same nominal MC each time with a varied smoothing parameter. The question we are answering is "does the choice of smoothing parameter bais the signal yield?". If the variations cancel out an result in no bias then we are safe from a systematic (this would imply that the choice of smoothing parameter does not bias the value of the signal yield and we return the same pull distributing as with the nominal toys).

Comment 53: l 909ff: When using J/psi K you only probe one q2 region. Do you see any dependence on q2 of all these quantities on simulation?

Reply: We are restricted to the jpsi peak. But perhaps more relevant would be dependence on the muon/kaon P or PT. Fortunately since we have a model for JpsiK simulation we are much more confident about these simulation distributions than for the signal ppbar mass distributions. I can produce some plots for this if it would be desirable.

Comment 54: l 943ff: I am a bit puzzled about the message here: You say changes in nTrack can cause large changes in PID, but you never quantify it.

Reply: So when I said this I was comparing the effect re-weighting had on PID efficiency with the Selection efficiency. Here are the efficiency changes with and without kinematic re-weighting.
Selection Run1: 1.0%, 0.8%, 0.3%, 0.5%, 0.4%
PID Run1: 3.6%, 3.8%, 3.6%, 3.8%, 3.8%
Selection Run1: 1.3%, 0.8%, 0.6%, 0.1%, 0.6%
PID Run2: 2.3%, 1.9%, 2.2%, 2.0%, 1.9%

It's clear that the PID efficiency changes quite a bit more than for selection (I was thinking about the impact of multiplicity on isolation). However, upon further inspection the largest difference actually seems to come from the trigger efficiencies.
Trigger Run1: 6.6%, 6.9%, 8.1%, 8.2%, 9.8%

Trigger Run2: 5.7%, 5.1%, 4.7%, 4.9%, 5.0%
It makes sense that the trigger efficiency would have some dependence on the multiplicity. This has been added to the note.

Comment 55: l 959ff: Would it not be possible to generate more statistics? Given that is just pure generation (and no simulation), it is normally quite easy to produce large samples.

Reply: Yes, although it takes time to get a lot in the low ppmass bin. Phase-space MC has a small proportion of events in the threshold region. That's why generating enough MC for this takes surprisingly long. Some has been added in version 2, I will generate more as time goes on.

Comment 56: l 954ff: Again the question how you treat correlations between the years?

Reply: We consider the uncertainties to be uncorrelated.

Comment 57: l 981: You essentially take 100 models (i.e. 100 different parameter sets) and then take the maximum and minimum?

Comment 58: Yes

Comment 59: l 986: I think this PID systematic is underestimated: Did you try changing the binning scheme. Especially for regions with steep efficiency changes this can have a O(2%) effect.

Reply: Good suggestion, this has now been done. The much coarser binning used for MisID unfolding was taken as an alternative, the difference in efficiency with it is taken as a systematic. This is quite an extreme change and provides a systemic that is much larger than just taking the stat. uncertainty of the calibration sample. For all three tracks together it has about a 10% effect.

Comment 60: l 998: Could you also do an integrated fit and check the consistency with the calculated branching fraction?

Reply: Another good suggestion, we will look into this.

3 Comments from Lucia Grillo

Comment 61: - line 260-262: Reweighting 2012 MC to match 2011+2012 data looks fine for B kinematics, event multiplicity, nTracks, I am wondering if there are differences between 2011 and 2012 that are not caught by doing this, e.g. something in trigger/reconstruction. I wonder if one could have a look at 2011 (I can't remember if the problem was the MC was not there or there was a bug or something different). Even though in the signal/normalisation probably differences cancel to a large degree

Reply: We do not think the 2011/2012 should be large enough to make a big difference. But we have ordered 2011 simulation to study this.

Comment 62: - line 345: How populated is the "uncategorised tracks" category? (I guess if there are many uncategorised tracks in the highest ppbar mass bin, that would result in a larger systematic)

Reply: For the fraction of uncategorised tracks in each bin for the full 2011+2012+2016 I find in bin $1 \rightarrow 5$ we get:-
11.2 % 11.8 % 13.6 % 13.2 % 18.1 %
So it is true that there are a higher fraction of uncat. tracks in the final bin. However, more importantly the systematic is relative to statistical the uncertainty in the final bin. In the final bin we have very low statistics and fluctuations on the misID shape can then similar in scale to the signal peak. It makes some intuitive sense that we are closer to being systematically dominated in the last $p\overline{p}$ mass bin but not so much in the first two.

Comment 63: - line 457: maybe you could add a couple of sentences about how you contaminated the same-sign sample with mis-identified particles. Related to this, one should make sure that the misID and combinatorial fit models don't have an overlap, but this is not very much relevant here, as you use the high B vertex χ^2 sample instead.

Reply: There is an extent to which we have to accept that the the same-sign proton sample will contain some misID. The interesting point is that even allowing for a significant misID contribution the total number of events after selection is just ≈60, possibly indicating a very low combi background from protons. It is true that the misID and combi shapes do not overlap very much.

Comment 64: - lines 469-470: I guess it is hard to estimate how much the high B vertex chi^2 sample is contaminated by partially reconstructed background because of the low statistics

Reply: Yes, hard to tell, but we believe we are covered by the quite large systematic.

Comment 65: - sec 6.2.1 : how much do you gain by using your BDT to reject B->(Lc->pX)pbarmunu with respect to use just the isolation BDT?

Reply: See reply to Patrick's comment 15. For the Lc background it's a quite a modest 7%, but biggest rejection is of $B \to D^0 p\bar{p}$ decays of ≈20%.

Comment 66: - sec 6.2.2 (fig 19): I think I remember Mark showing different options for the neutral isolation, but I cannot re-find the talk now. Maybe you can just add information if you investigated different options and you decided to go for this BDT.

Reply: This neutral isolation is in the works for the τ mode, but wasn't deemed necessary for the μ mode.

Comment 67: - lines 570-571: you include the same amount of the 3 different background MC cocktails because you expect to have these proportions in data after looking at the efficiencies? (Even though I guess it doesn't make a big difference as long as all components are present in the training sample)

Reply: Yes this is just for training. We could go back and weight them differently relative to each other, but we don't expect this to change the picture much.

Comment 68: - line 747-748: I might be getting confused: you fit separately 2011, 2012 and 2016 normalisation samples, and then you combine the yields? Also because for the signal fits you fit together 2011, 2012 and 2016, and the efficiency ratios are calculated for Run1 and 2016 (or combined)

Reply: Yes, it was chosen to combine 2011+2012+2016 for the signal channel because it's useful to have the additional fit stability. This isn't the case for the norm channel so we fit individually. The efficiencies are calculated separately and then combined based on luminosity and relative cross section for run1/2016.

Comment 69: - line 755: I understand there are no shared parameters among the fits in different ppbar bins, correct?

Reply: Correct, totally independent fits

Comment 70: - sec 10.2: I am not sure you have enough events, but how do your combinatorial background distributions compare to the events with corrected B mass higher than the B mass?how large is the variation of the fitted combinatorial component using the different models you use?

Reply: Yes, we have had a look. Not many events with the full selection applied. But we have made a plot with just stripping cuts. See the response to comment 25.

Comment 71: - lines 975-976: when you fluctuate the pQCD model to generate q^2 distributions compatible with data, do you keep into account the background contributions to your data distribution?

Reply: Sorry that this wasn't made clear. We never take a data distribution, I meant the expected distribution (including threshold enhancement effects ect. not present in phase space MC). We don't expect that we have unfolded the true distribution in data.

Comment 72: A couple of editorial things: - Tab 3 I guess the "pi-like" category is defined like "ProbNNpi>0.3 and !p-like and !K-like", correct? - line 365 a factor 5 instead of 50, right? As the misID line has a pre-scale of 20% (unless I missed a factor 10 somewhere)

Reply: Sorry, this was a typo, now fixed. Pre-scale is in fact 2%.

4 Comments from Suzanne

Comment 73: l.12 Perhaps you can expand with 1-2 sentences on this dibaryon threshold enhancement, where it comes from and why it's interesting. You mention it a few times later in the note and in the conclusion.

Reply: Thank you for the suggestion. A little has been added here about the threshold enhancement effect.

Comment 74: l.34: this is now 3.1 sigma

Reply: Thank you, updated

Comment 75: Table 2: Does it matter that you use only 2012 samples for run 1? I suspect differences between 2011 and 2012 are minor, if any, and most will cancel between the signal and normalisation channel. Perhaps you can show some plots, or add a few sentences. Are the PID algorithms the same?

Reply: Simulation from 2011 has been requested to study potential efficiency differences. As for PID, the calibration sample used is the full run 1 sample.

Comment 76: l. 349: Is it reasonable to assume such a large fluctuation for the uncategorised track efficiencies? It seems very large.

Reply: I agree it is exceptionally large. But fortunately it seems to have little effect on the overall uncertainty, so it is used as a conservative treatment of the effect.

Comment 77: l. 524: Perhaps you could rephrase the title to "Trigger and Stripping selection".

Reply: The note has been changed.

Comment 78: Table 11: The caption should be run 2.

Reply: Fixed in the note.

Comment 79: Table 10-12: I'm probably missing something obvious here, but how is the proton selection "isMuon" in both stripping lines, and also "!isMuon" afterwards?

Reply: this was a typo, it's !isMuon (stupidly used \! instead in the first 2 tables)

Comment 80: Section 9: I was looking for the outcome of the fit parameters here, and eventually found them in the appendix. Could you add a table in this section to make it easier to compare between the different mass bins? Or at least a reference to the appendix.

Reply: I have placed a reference to the appendix in the signal fit section.

Comment 81: Fig. 39: Which mass bin is this? (This would be good also to specify for the other plots in this section.) It seems like the variation is very small, but then the systematics go up to 45.4%. Is this because it's a different bin? I'd be interesting to see the highest mass bin as well, especially since this systematic seems quite large, and I wonder if wonder if the efficiency range you use isn't too broad.

Reply: All the shape variation examples are from the first bin. This has been clarified in the note. See response to comment 26. This has the misID variation plot for the final bin. As you might expect the variation is much larger and the fluctuations in the shapes are now of a similar scale to the small signal yield.

Comment 82: Table 29: Do you know why these effects are so different per mass bin, varying between 0.2% and 31.2% seems a lot for template smoothing.

Reply: It's totally dependent on the shapes and stat fluctuations in the Data and MC. In the final bin the signal stat is so low that the size of the shape fluctuations are of a similar order to the signal yield.

Comment 83: l. 1003: Just out of curiosity, can you see this because all values are scaled by the same numbers in the blinding procedure?

Reply: This plot is totally fake. I just put 2000 events in each bin, it's not using the any output from the fit, just a placeholder until I unbind this plot. (But yes, this will be the key plot to unblind)

Comment 84: Appendix: I missed references to the appendix in the note. Perhaps that's me, if not, please add them.

Reply: Have added references to the appendix.

5 Comments from Abhijit

Comment 85: Abstract: Would be good if you add info about the data sample used.

Reply: Added to the abstract

Comment 86: L27: I might be mistaken, but you mention that due to the over prediction by Geng et al, the reliability on the pQCD is questionable and source of this discrepancy is not known. However, you use their form factor results to re-weight q2 distribution. So you are assuming here that the prediction of the q2 distribution is correct but the prediction of m(ppbar) (which you are not sensitive to) and the helicity angle distributions are not reliable? And that is why you do not reweight helicity angle shape on L308?

Reply: There is a systematic due to large variations on the q^2 model, so we do not trust it absolutely. We can re-weight in helicity angle, however for the purposes of an efficiency systematic we expect it makes very little difference.

Comment 87: L70: Why cut off at 5000 MeV, the phase space limit should be m(ppbar)max = m(B+) - m(mu)   5174 MeV, no?

Reply: I'm not convinced it matters much, (there are very few events up in the range 5.0-5.17). We can move the range if this is a concern.

Comment 88: L187: Has a double K+K- -> ppbar mis-id been considered?

Reply: Yes, we have taken a look now. See answer to Patrick's comment 5 on double misID. We have added this to the version 2 note.

Comment 89: Table 2: You consider here for example, $B+->(Delat+->pX)$ pbar mu+ nu, what about: $B+->(Delta-->pbar)pmu+nu$, or $B+->(Delta+->pX)(Delta-->pbarX)mu+nu$, or $B+->(Delta+->pX)(N*-->pbarX)mu+nu$, etc? Are the shapes similar (due to isospin and Uspin arguments)?

Reply: The $\Delta-$ decays to neutrons so we are safe there. As for the $B \to X_{bary}\overline{X}_{bary}\mu\nu$ see the answer to Patrick's comment 4. This has now been added to the note. We expect these backgrounds to be very small. (Smaller than $B \to \Lambda_C \overline{N}^* \mu\nu$, which has been considered and neglected).

Comment 90: Table 3: Should it be "& !K-like" instead of "or K-like"? And also how where these numbers chosen to define each of the category? (I guess the cut value choice does not matter as long as the samples obtained in each category are mutually exclusive?)

Reply: Yes, I meant !(pi-like or K-like). This is fixed in the note.

Comment 91: L338: "momentum and psuedorapidity", this is referring to track mom and eta of the proton candidate right? Also there is no mention of how N'j is obtained (total number of candidates in each of the category of fake-p sample, right?).

Reply: Yes, N'^j is the observed number of $ph\mu$ candidates in each h category.

Comment 92: L348: so the eff(U->h) is set to zero in the nominal? The range for variation for the systematic seems to be very large for this. (I guess better to be conservative, specially if there is no reasonable estimate of this variation anywhere?)

Reply: yes eff(U->h) is set to zero for the nominal. Yes, the variation is a very conservative choice. But since even this gives little difference we have continued with this strategy.

Comment 93: Figure 12: Not referenced in the text.

Reply: Thank you, this has been fixed.

Comment 94: L399: What about Xi_c or $Omega_c$ -> pX contributions?

Reply: This should be covered by the $B \to \Lambda_c^* \overline{p} \mu \nu$ sample, it is not expected to be significantly different in corrected mass.

Comment 95: L466, 472-475: How the inverted B vertex sample compare to the same sign sample (with full selection) and same sign sample (with full selection except isolation cut)? Are they consistent in shape?

Reply: With full selection, there isn't even enough stat to compare in the same sign. (Which does perhaps indicate a very low level of combinatorial background to begin with.)

Comment 96: L481: You have considered here contribution from Jpsi(->ppbar) with random mu (along with vetoing ccbar -> ppbar contribution), but you could have contributions from nonresonant ppbar candidate + random mu, is this contribution negligible?

Reply: We consider that type of background as part of the combi. shape from the inverted sample. We were concerned specifically about Jpsi(->ppbar) and so we did this study to demonstrate that is was a small contribution.

Comment 97: L497-501: As I understand, the inverted B vertex sample contains two type of combinatorial bkg: 1) (p mu) candidate + random pbar, 2) (ppbar) candidate + random mu. If I am not mistaken the same sign sample tries to emulate the 1st type of combinatorial bkg but not very well the 2nd type. In this case when evaluating the systemics you are setting the contribution of combinatorial bkg from 2nd type to be zero and assume that the shape variation from first type would be a dominant one? It is hard to judge without looking at how the shape of the 2nd type of combinatorial bkg differs from same sign sample, is a comparison possible?

Reply: You've got the correct picture about this, we show the plot in the systematics section. It's exactly this difference in shape (and the effect on the signal yield) that leads to the systematic. We consider that the combinatorial background is small enough that these two handles on it suffice to give a conservative systematic.

Comment 98: L517: Sorry if I have missed it, you haven't considered in the nominal fit but have you in the systematic evaluation?

Reply: Yes that's correct. Nominal uses the inverted vertex sample.

Comment 99: Table 10 and 11: Should be !isMuon cut in p selection column?

Reply: Yes, this is a typo from putting \!, fixed.

Comment 100: Table 12: In the ppbarmu selection (last column), is a cut of B flight direction chi2 > 150 missing?

Reply: Yes, fixed in the note.

Comment 101: Section 6.2: There are two BDTs: 1) for charged additional tracks and 2) neutral additional tracks. To simplify the analysis, has there been an attempt of training a single BDT to beat these two types of bkg? Maybe the correlation with the 4 isolation outputs and the inputs to the 2nd BDT can be beneficial (Although I do not see any correlation between the two BDT outputs in Fig 21).

Reply: This was done at some point early on in the analysis. But it was decided to take separately since there is no correlation expected.

Comment 102: Section 6.4: There are some assumption that are made in the calculation of the yields (L623, L629, L632-639). Have the yields been varied by some appropriate factor (to account for these assumptions) to see what affect they have on the optimised cut points that has been decided?

Reply: It's true that the working point could shift if we change the base assumptions, however we think having this as the nominal set of assumptions is appropriate. There should be no need to run this process again (it takes quite a bit of time) unless there is something wrong with this set of assumptions.

Comment 103: L621: Am I correct in interpreting that there is a simultaneous optimisation of 5 cuts that has been conducted? And the shapes in corrected mass for each of the bkg category has been obtained from MC for 5 given cuts? If so, in line 621, you mention that you integrate over ppbar mass, because you want to avoid additional assumptions about the bkg shapes in ppbar mass bins, but this information you are going to get anyway from the simulation samples right and you do not need to make any assumptions on their shape? Maybe I am missing something here and there might indeed be some additional assumptions required, in any case what has been done here is also fine.

Reply: That's correct. So with phase space MC we know the ppmass distributions are just going to be way off for both sig and background. So the number in each bin will be wrong, it's only the MCORR shape that we use in each bin for the signal fits. We can reweight the signal to an approx model. Performing these toy fits over all bins just seems like the most sensible thing to do in the face of this.

Comment 104: Section 6.5: How where the veto window chosen? And is there a systematic on efficiency by varying this veto window?

Reply: The width of the vetos in the final bin was a relatively arbitrary choice. The number of expected signal events in these regions is so small that it is hard to imagine it having an effect on the overall result.

Comment 105: Table 16: Can see reduced uncertainty in the efficiency ratio in the lowest m(ppbar) bin compared to others. Any explicitly reason for this?

Reply: The uncertainty quoted here is the uncertainty from turning on and off the correction tracking weights. This tracking uncertainty has changed for version 2. From the ETA plots provided earlier in comment 17, it's clear that the first bin has a different distribution in psudorapidity for the protons. With this different ETA distribution the average tracking correction is closer to 1 in the first bin, this may explain the difference. To be clear though, this uncertainty isn't propagated through, but instead the systematic uncertainty is calculated with the change on the full efficiency.

Comment 106: Section 7.4: It seems the IsMuon requirement is included in Stripping and obtained from simulation, why are you not obtaining the efficiency of IsMuon requirement from PIDCalib, is it modelled well in your simulation?

Reply: yes we trust the isMuon from simulation, but not the ProbNNs

Comment 107: Table 18: Although consistent stripping cuts have been chosen for signal for Run1 and Run2, it seems that the ratio is higher for Run 1 than Run 2. Is this coming from the increased signal efficiency of signal in Run 1 or decreased signal efficiency of normalisation in Run 1? For the bin 2.4 to 2.6 GeV the uncertainties seem to be an order of magnitude smaller than other bins.

Reply: Correct, for run 1 the striping eff from simualtion (with simulation stat uncertainty) are:
Norm Strip: [0.11135531 $\pm$ 0.00048697]
Sig1 : [0.13776688 $\pm$ 0.00403892]
Sig2 : [0.13498127 $\pm$ 0.00233636]
Sig3 : [0.15165294 $\pm$ 0.00241632]
Sig4 : [0.16555193 $\pm$ 0.00256176]
Sig5 : [0.22043708 $\pm$ 0.0015534]
For run 2 this is:
Norm Strip: [0.14678713 $\pm$ 0.00024]
Sig1 : [0.15917448 $\pm$ 0.00434424]
Sig2 : [0.15116704 $\pm$ 0.002382]
Sig3 : [0.15872982 $\pm$ 0.00233186]
Sig4 : [0.19072348 $\pm$ 0.00271592]
Sig5 : [0.24560627 $\pm$ 0.00162895]
So to answer your question, the eff increases broadly in run 2, but it is a higher eff. in jpsi that drives the difference.

The table here has an approximate uncertainty by turning the tracking, kinematic and qsq weights on and off. The small diff in bin 4 is by chance. (I intended to only switch on and off the kinematic re-weights, the note has been changed to reflect this).

Comment 108: Table 21: It seems that the efficiency ratio map b/w Run1 and Run2 are significantly different. Do you think it would be interesting to measure the differential branching fraction also as function of sqrts i.e. showing plot 49 for Run 1 and Run 2 separately. I do not if they can serve as cross-checks to each other. I leave it to you.

Reply: We may consider doing this later on. However, the fits to the combined r1 and r2 data are sufficiently complicated that this would require quite a lot of work to get running.

Comment 109: L767: Is the peak position fixed for floated in the fit to data? I do not know if you could actually benefit from a simultaneous fit with common signal shape params in each m(ppbar) bin.

Reply: Fixed from MC, we are keen not to share any parameters between bins, to keep them totally independent. The shape does change slightly between bins, you can see the left hand tail changes slightly.

Comment 110: L782-786: Why not just remove the partially reconstructed bkg with similar shape and include the removed ones as a source of systematic?

Reply: This was done as one point. The trouble was, it was inflating the systematic as all of these changes would be totally correlated. We decided it was best to somehow express this uncertainty in the part-reco shape directly in the likelihood

Comment 111: Section 10.8 and 10.9: Are these systematics really required? In the nominal fit, you are already re-weighting MC with a data-driven procedure which is more accurate than the case that you consider here for systematic. It would make sense to evaluate systematic uncertainty, by using a different control channel for re-weighting other than the normalisation mode, to account for any differences in the kinematics. But this also may not be required.

Reply: I see the argument here. But we consider these systematic uncertainties to be a conservative estimate of the potential effects. These aren't large enough to make much of a difference to the overall uncertainty.