Proposal for the contribution of Imperial CMS Group to the CMS B Parking Effort in 2018

Robert Bainbridge,[1] **Oliver Buchmueller,**[1] **Vincencio Cacchio,**[1] **Sarah Malik,**[1] **and Thomas Strebler**[1]

[1] *High Energy Physics Group, Blackett Laboratory, Imperial College, Prince Consort Road, London, SW7 2AZ, United Kingdom*

ABSTRACT: In this working document we outline a proposal to contribute to the 'B parking' effort of CMS, which intends to collect in 2018 a data set of $O(10^{10})$ B decays. This impressive data set could enable the analysis of $B^{\pm(0)} \to K^{(*)}\ell\ell$ decays in order to perform a significant and competitive measurement of both R_K and R_{K^*}.

Contents

1 Introduction

In this working document we outline a proposal for contributions to the CMS B parking project in 2018. Besides the ongoing contribution to improve the low momentum electron reconstruction, which is key to the success of a competitive measurement of R_K and R_{K^*}, the Imperial CMS group also plans to take a leading role in the analysis to measure the ratio $R_K(q^2 = m^2_{\ell\ell}) = \Gamma(B^\pm \to K^\pm \mu\mu)/\Gamma(B^\pm \to K^\pm ee)$, which originates from charged B decays. In section 2 we outline a high-level step-by-step procedure that eventually leads to the measurement of R_K, while in section 3 we sketch a programme of contributions to the low-momentum electron reconstruction effort.

2 Step-by-Step Analysis Procedure

In order to arrive at a fully commissioned analysis to measure R_K, it is important to perform some auxiliary and cross check measurements along the way. The sequence of these measurements is mainly determined by the availability of a sufficiently large data set of the relevant B decays. Table B shows expected yields for different timelines of partially or fully reconstructable B decays that are relevant for the commissioning of the analysis. In the following sections 2.1 and 2.2 we outline different milestones of the analysis that eventually lead to final measurement of the unitarity ratios. In 2.3 we outline the need for dedicated Monte Carlo production to serve as input for the different analysis steps and in 2.4 we summarise the analysis milestones.

Table 1. *Expected yields of different partially or fully reconstructable B decays. 'N (2018) parked' represents the number of events recorded and parked in the entire run year 2018, while 'N(2018) processed' stands for the number of events that will be processed during the run year (about 5% of all parked events) and are available for immediate analysis. 'N (10^8 trigger)' identifies the number of decays that are contained in a sample of 10^8 triggers, which represents about one average fill. Here we assume a typical B hadron purity of about 80% for single-muon trigger after HLT refinement.*

Mode	N (2018) parked	N(2018) processed	N (10^8 triggers)	$\mathcal{BR}$
Partially reconstructable $B \to D^{*+} l^- \nu$ decay mode				
$B^0 \to D^{*+} l^- \nu \to D^0 \pi^+ l^- \nu$ $\to K^- \pi^+ \pi^+ l^- \nu$	5.5×10^6	2.8×10^5	3.5×10^4	1.1×10^{-3}
Full reconstructable $B \to D\pi$ decay mode				
$B^0 \to D^+ \pi^- \to K^- \pi^+ \pi^+ \pi^-$	1.25×10^6	6×10^4	8.0×10^3	2.5×10^{-4}
$B^\pm \to D^0 \pi^\pm \to K^\pm \pi^\mp \pi^\pm$	4.8×10^4	2.4×10^3	6.1×10^3	1.9×10^{-4}
Main background sample $B \to K^{(*)} J/\psi$ decay mode				
$B^0 \to K^* J/\psi \to K^+ \pi^- \ell^+ \ell^-$	2.6×10^5	1.3×10^4	1.7×10^3	5.24×10^{-5}
$B^\pm \to K^\pm J/\psi \to K^\pm \ell^+ \ell^-$	3.1×10^5	1.6×10^4	2.0×10^3	6.12×10^{-5}
Signal sample $B \to K^{(*)} \ell^+ \ell^-$ non resonante decay mode				
$B^0 \to K^+ \pi^- \ell^+ \ell^-$	3290	165	21	6.6×10^{-7}
$B^\pm \to K^\pm \ell^+ \ell^-$	2250	113	15	4.51×10^{-7}

2.1 Measurement of B purity in data

A direct measurement of the B purity of our parked data stream is essential to ensure that the data that are written to tape possess a sufficiently large component of B events. In order to perform this measurement directly in data, we propose to collected a sample of about 10^8 triggers during a dedicated fill. This fill will taken during the early physics phase of the 2018 run campaign and it is important is that the data are immediately being processed to ensure that they can be used to measure the B purity and to aid the analysis commissioning. This has now been agreed on with management and we expect these data to become available in late May or early June.

With about 4.4×10^4 $D^{*+}l^-$ and several times 10^3 $D\pi$ decays a precise measurement of the B purity, both in partially reconstructed as well as fully reconstructed decays, should be straightforward. This precise measurement would also serve as reference for the data quality monitoring in the course of the run year.

2.2 Analysis commissioning with $B^0 \to J/\psi(l^+l^-)K^\mp\pi^\pm$ and $B^\pm \to J/\psi(l^+l^-)K^\pm$ events

Using control samples that exhibit the same decay topologies like our signal samples but have significantly larger branching fractions would allow us to establish the full analysis strategy in the course of 2018 data taking by using data from normal processing. The most natural control samples would be $B^0 \to J/\psi(l^+l^-)K^\pm\pi^\mp$ and $B^\pm \to J/\psi(l^+l^-)K^\pm$, which correspond to a branching faction that is about 100 times larger than that of the non-resonant signal events $B^0 \to K^*\ell^+\ell^-$ and $B^\pm \to K^\pm\ell^+\ell^-$. The processing of one average fill will yield about 10^8 triggers and, as show in table B, it would also contain about 2000 events of $B^0 \to J/\psi(l^+l^-)K^\pm\pi^\mp$ and $B^\pm \to J/\psi(l^+l^-)K^\pm$. This is about the same number of events that we expected to collect in the entire run year for our signal events. Therefore, this data sample would not only serve a precise determination of the B purity but would also enable us to start commissioning the analysis using its natural control samples (i.e. the $J/\psi(l^+l^-)$ decay mode). Furthermore, it was agreed with management that about every month we will get access to fully processed fill, which in the course of the run year will represent about 5% of the entire parked data sample. As outlined in table B, by the end of the run year we should have access to about a few times 10^4 $B^0 \to J/\psi(l^+l^-)K^\pm\pi^\mp$ and $B^\pm \to J/\psi(l^+l^-)K^\pm$ events, which will be sufficient to fully commission the analysis and demonstrate that we are able to measure $R_{K^{(*)}}(q^2 = m^2_{J/\psi}) = 1$ as expected. It should be note that at $q^2 = m^2_{J/\psi}$ no New Physics contribution are expected and, thus, showing its consistency with unity is a critical test of the analysis chain.

2.3 Dedicated Monte Carlo Production for Relevant B Decays

In order to commission the analysis strategy, large statistics of simulated events need to be made available before the parked data are reconstructed. Those Monte Carlo samples would be used in particular to optimise the selections to apply on the reconstructed objects, which could potentially use machine learning algorithms. In order to have those samples quickly available, those Monte Carlo samples are to be produced directly by the Imperial CMS group using the CMS computing infrastructure.

To reproduce the data sample recorded using muon triggers, all of those generated samples will correspond to the production of a pair of B hadrons together with a muon, potentially coming from the decay of one of the B's. A specific decay of one of the B hadron is subsequently enforced to get enough statistics even for decay modes associated with low branching ratios. The generated samples will focus on the $B^0 \to K^+\pi^-\ell^+\ell^-$ and $B^\pm \to K^\pm\ell^+\ell^-$ decays to optimise the analysis strategy for the R_K measurements. Additional processes used for auxiliary measurements, such as the $B^0 \to D^+\pi^- \to K^-\pi^+\pi^+\pi^-$ and $B^\pm \to D^0\pi^\pm \to K^\pm\pi^\mp\pi^\pm$ to be used for purity measurements, will also be considered.

A step-by-step procedure for this Monte Carlo production is detailed in Appendix ??.

2.4 Summary of Analysis Milestones

Based on the discussion in the previous sections, the analysis contribution of the Imperial CMS group will focus on charged B decays with the final goal to lead the measurement of $R_K(q^2 = m^2_{\ell\ell}) = \Gamma(B^\pm \to K^\pm\mu\mu)/\Gamma(B^\pm \to K^\pm ee)$. The following analysis milestones will be important to meet in order to deliver a timely and competitive measurement of this important lepton universality ratio.

- Using the data of the first processed fill, we are planning to measure the B purity using fully reconstructed the colour-favoured $B^\pm \to D^0\pi^\pm \to K^\pm\pi^\mp\pi^\pm$ decays. This final state is not only ideally suited for this task but, with a final state very similar to the signal sample (i.e. $K^\mp\pi^\pm\pi^\pm$ vs. $K^\pm l^\pm l^\mp$) this decay also represents and excellent test ground to establish the basics of the final analysis chain. The time scale for this is **Summer of 2018**.

- Using about 5% of the parked data that are planned to be directly processed during the course of the run year, we are planning to fully commission the analysis chain and to demonstrate that using $B^\pm \to J/\psi(l^+l^-)K^\pm$ decays the $R_K(q^2 = m^2_{J/\psi}) = 1$. This will be the last step in the analysis commissioning chain, which, if successful, will trigger the timely processing of the full parked data set. The time scale for this is **End of 2018**.

- Assuming that the previous milestones are met successfully, we will proceed to measure $R_K(q^2 = m^2_{\ell\ell}) = \Gamma(B^\pm \to K^\pm\mu\mu)/\Gamma(B^\pm \to K^\pm ee)$. The time scale for this is **Spring of 2019**.

It should be noted, that at any give milestone it might turn out that a competitive measurement of R_K is impossible and, thus, this would be a natural point to revisit the priorities and usefulness of the Imperial contribution to B parking effort.

3 Low-momentum electron reconstruction

One of the most crucial experimental aspects of the $R_{K^{(*)}}$ measurements is the ability to identify low-p_T electrons down to $p_T \gtrsim 1$ GeV. The following subsections outline the performance limitations of the current electron reconstruction algorithms, the consequences for the $R_{K^{(*)}}$ measurements, the proposed strategy for improvements, the primary datasets and simulated event samples and/or skims required to tune and commission the updated algorithms, and computing considerations.

3.1 Nominal performance and consequences for the $R_{K^{(*)}}$ measurements

Studies based on the simulated pair production of charged B hadrons, one of which is decayed inclusively ("tag-side") and the other is forced to decay via $B^{\pm} \to K^{\pm}\ell^{+}\ell^{-}$ ("signal-side"), demonstrate that the acceptance times efficiency $\mathcal{A}\epsilon$ obtained with the current electron reconstuction algorithm severely limits the ability to accurately measure $R_{K^{(*)}}$.

Figure 1 (Left) shows the generator-level p_{T} distributions for the daughter particles from the $B^{\pm} \to K^{\pm}\ell^{+}\ell^{-}$ decay. The p_{T} distributions are very soft, with those for the kaon and subleading lepton peaking at ≈1 GeV. Figure 1 (Right) shows the efficiency to reconstruct electrons as a function of the generator-level p_{T}, as obtained with the default electron reconstruction algorithm. The efficiency is determined to be essentially zero for the region $0 < p_{\mathrm{T}} < 2$ GeV because of explicit thresholds applied during the early seeding steps. The efficency is in the range 0.2–0.8 for the region $2 < p_{\mathrm{T}} < 7$ GeV, beyond which the plateau is reached.

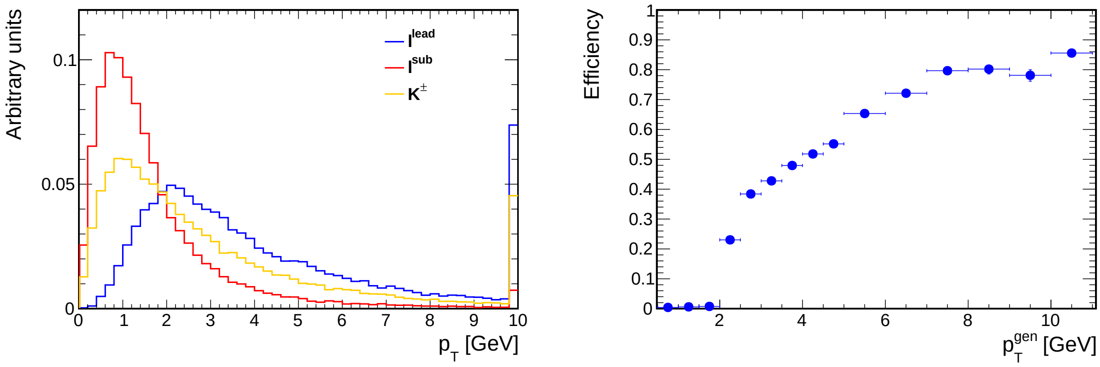

Figure 1. *(Left) Normalised distributions of the generator-level p_{T} for the daughter particles from the $B^{\pm} \to K^{\pm}\ell^{+}\ell^{-}$ decay; namely the kaon ($K^{\pm}$), and the leading (ℓ^{lead}) and subleading (ℓ^{sub}) leptons. (Right) Efficiency to reconstruct electrons as a function of the generator-level p_{T}, as obtained with the default electron reconstruction algorithm.*

Table 2 shows both the $\mathcal{A}$ values for both $B^{\pm} \to K^{\pm}\ell^{+}\ell^{-}$ and $B^{0} \to K^{*}\ell^{+}\ell^{-}$ decays, as well as the $\mathcal{A}\epsilon$ values for both $B^{\pm} \to K^{\pm}e^{+}e^{-}$ and $B^{0} \to K^{*}e^{+}e^{-}$ decays, for a set of minimum tranverse momentum requirements, $p_{\mathrm{T}}^{\mathrm{min}}$, applied to both daughter leptons, while considering a fixed p_{T} requirement for the kaon ($p_{\mathrm{T}} > 0.5$ GeV). The requirement $|\eta| < 2.5$ is applied to all daughter particles. The $\mathcal{A}$ values are comparable for $B^{\pm} \to K^{\pm}\ell^{+}\ell^{-}$ and $B^{0} \to K^{*}\ell^{+}\ell^{-}$ and depend strongly on the transverse momentum threshold $p_{\mathrm{T}}^{\mathrm{min}}$. Acceptances of 8–11% can be raised to 27–44% by reducing $p_{\mathrm{T}}^{\mathrm{min}}$ from the default value of 2 GeV to 0.7 GeV (the latter threshold is the minimum p_{T} required for an electron to reach the ECAL barrel). When folding in the electron reconstruction efficiencies, the $\mathcal{A}\epsilon$ values increase slowly with decreasing $p_{\mathrm{T}}^{\mathrm{min}}$ and do not exceed ≈5% below 2 GeV with the current electron reconstruction algorithm. However, a factor 2–5 increase in $\mathcal{A}\epsilon$ can be achieved with even moderate improvements in efficiencies when combined with the lowering of $p_{\mathrm{T}}^{\mathrm{min}}$.

Table 2. *The $\mathcal{A}$ values for both $B^{\pm} \to K^{\pm}\ell^{+}\ell^{-}$ and $B^{0} \to K^{*}\ell^{+}\ell^{-}$ decays, as well as the $\mathcal{A}\epsilon$ values for both $B^{\pm} \to K^{\pm}e^{+}e^{-}$ and $B^{0} \to K^{*}e^{+}e^{-}$ decays, for a set of minimum tranverse momentum requirements, $p_{\mathrm{T}}^{\mathrm{min}}$, applied to both daughter leptons, while considering a fixed p_{T} requirement for the kaon ($p_{\mathrm{T}} > 0.5$ GeV). The requirement $|\eta| < 2.5$ is applied to all daughter particles.*

$p_{\mathrm{T}}^{\mathrm{min}}$	$\mathcal{A}$		$\mathcal{A}\epsilon$	
[GeV]	$B^{\pm} \to K^{\pm}\ell^{+}\ell^{-}$	$B^{0} \to K^{*}\ell^{+}\ell^{-}$	$B^{\pm} \to K^{\pm}e^{+}e^{-}$	$B^{0} \to K^{*}e^{+}e^{-}$
5.0	0.02	0.02	0.01	0.01
2.0	0.11	0.08	0.05	0.04
1.0	0.32	0.19	0.05	0.04
0.7	0.44	0.27	0.05	0.04

3.2 Baseline strategy for improving $\mathcal{A}\epsilon$

Significant improvements in $\mathcal{A}\epsilon$ are possible, with minimal intervention, through the re-optimisation of existing thresholds and parameters used in the reconstruction chain. This statement is supported by experts within the EGM POG and is based on the fact that the current "tracker-driven" algorithm does not fully exploit the tracking performance in the very low p_{T} regime and, further, several parameters and multivariate techniques have not been tuned to reflect changes in the LHC beam parameters and detector configuration betweens LHC Runs 1 and 2.

Studies are already underway to explore possible performance gains at the tracker-driven seeding step, which involve the consideration of low-p_{T} tracks (i.e. below 2 GeV) as inputs to the seeding step, as well as the retraining of a BDT that improves efficiencies in the region $2 < p_{\mathrm{T}} < 50$ GeV (and potentially below). Similar studies are expected at later stages in the electron reconstruction chain. Further, the electron identification criteria will also be revisited for the low-p_{T} regime. Finally, more involved studies and developments may be used if deemed necessary, although our baseline strategy is to minimise intervention. All studies are being carried out in close collaboration with the EGM POG community.

3.3 Commissioning of low-p_{T} electrons

The following subsections outline a minimal set of primary datasets and simulated event samples that can be used to optimise the electron reconstruction and identification algorithms, measure reconstruction and identification efficiencies, and determine the associated data-to-simulation scale factors.

3.3.1 Simulated samples

Optimisation and efficiency studies are ongoing, based on the following privately produced samples: charged B meson pair production, with $B_{\mathrm{tag}} \to \mu_{\mathrm{tag}}\mathrm{X}$ and $B_{\mathrm{sig}}^{\pm} \to K^{\pm}\ell\ell$, with $\ell = \mathrm{e}, \mu$ and $p_{\mathrm{T}}(\mu_{\mathrm{tag}}) > 7$ GeV; neutral B meson pair production, with $B_{\mathrm{tag}} \to \mu_{\mathrm{tag}}\mathrm{X}$ and $B_{\mathrm{sig}}^{0} \to K^{*0}\ell\ell$, with $\ell = \mathrm{e}, \mu$ and $p_{\mathrm{T}}(\mu_{\mathrm{tag}}) > 5$ GeV. Samples for the prompt and nonprompt production of $\mathrm{J}/\psi \to \ell\ell$ may also be produced.

3.3.2 Data sample of asymmetric $\gamma \to \mathrm{ee}$ events

A large sample of low-p_{T} electrons can be obtained from converted photons resulting from interactions with the beam pipe and inner tracking structures. A collection of `reco::Conversion` objects is available in the `AOD` data tier, the production of which is based on the following inputs: the "general" and "conversion" `reco::Track` collections, the latter of which produced using dedicated steps that supplement the default tracking sequences; and the "ECAL- and tracker-driven" `reco::GsfElectrons` collections.

Preliminary studies demonstrate that a large, relatively pure sample $\gamma \to \mathrm{ee}$ events that contain a low-p_{T} electron, in the range 1–20 GeV, can be obtained through a minimal set of selection criteria comprising the presence of an leading electron with $p_{\mathrm{T}} > 20$ GeV and a two-track vertex consistent with the conversion topology. An analysis of 0.6 fb^{-1} of data from 2017 era F of the `SingleElectron` primary dataset yields $\approx 5 \times 10^4$ low-p_{T} conversions, with $\approx 3 \times 10^4$ events satisfying a tighter p_{T} requirement on the leading electron as well as the trigger requirement `HLT_Ele32_WPTight_Gsf OR HLT_Ele27_WPTight_Gsf`. The `RAW` data tier for this sample is now available on disk and is being used to tune the reconstruction of low-p_{T} electrons.

3.3.3 An unbiased data sample of $\mathrm{J}/\psi \to \mathrm{ee}$ events

A large unbiased sample of $\mathrm{J}/\psi \to \mathrm{ee}$ events can be obtained from the 2017 dataset by relying on J/ψ production from pileup events. The cross section times branching fraction $\sigma_{\mathrm{J}/\psi}\mathcal{B}(\mathrm{J}/\psi \to \mu\mu)$, over the range $6.5 < p_{\mathrm{T}}(\mathrm{J}/\psi) < 30$ GeV and $|y(\mathrm{J}/\psi)| < 2.4$, is ≈ 100 nb at $\sqrt{s} = 7$ TeV [?]. The $\sigma\mathcal{B}$ is assumed to scale to ≈ 1000 nb for $p_{\mathrm{T}}(\mathrm{J}/\psi) > 3$ GeV (Fig. 3 [?]). A further factor two increase is assumed to account for the parton luminosity ratio between $\sqrt{s} = 7$ and 13 TeV. Hence the number $\mathcal{N}$ of $\mathrm{J}/\psi(\to ee)$ events available during the 2017 dataset is given by

$$\begin{aligned} \mathcal{N} &= R_{\mathrm{HLT}} \times t_{\mathrm{LHC}} \times n_{\mathrm{PU}} \times \sigma_{\mathrm{J}/\psi}\mathcal{B}(\mathrm{J}/\psi \to \mu\mu)\,/\,\sigma_{\mathrm{MB}} \\ &= R_{\mathrm{HLT}}\ [\mathrm{Hz}] \times 6 \times 10^6\ [\mathrm{s}] \times 38 \times 2\ [\mu\mathrm{b}]\,/\,80\ [\mathrm{mb}] \\ &= 5700 \times R_{\mathrm{HLT}}\ [\mathrm{Hz}] \end{aligned} \tag{3.1}$$

where R_{HLT}, t_{LHC}, n_{PU}, and σ_{MB} are the HLT trigger rate, total live time of the 2017 data-taking period, average number of pileup events, and the minimum bias cross section, respectively. Estimates for each variable are given above, and their product yields 5700 $\mathrm{J}/\psi \to \mathrm{ee}$ events per Hz of trigger rate.

A skim of events based on the `DoubleMuon` primary dataset could be used to collect events containing at least one high-p_{T} muon (e.g. $\mathrm{Z} \to \mu\mu$) that allows the identification of the primary vertex. A tag-and-probe approach can by used with electron-track pairs that form a vertex consistent with a pileup interaction and an invariant mass consistent with the J/ψ. The `HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8_v4` trigger and the `DoubleMuon` PD provide rates of ≈ 40 and ≈ 100 Hz, respectively, providing a sample of $\mathcal{O}(10^5)$ events for efficiency measurements. A skim of $\mathrm{Z} \to \mu\mu$ events based on the `RAW` data tier can be used to tune the reconstruction of low-p_{T} electrons. Alternatively, the skim for photon conversions, described above, based on the `RAW` data tier from the

SingleElectron primary dataset and the trigger requirement HLT_Ele32_WPTight_Gsf OR HLT_Ele27_WPTight_Gsf could also be used.

3.4 Computing constraints

- The tuning of the electron reconstruction and identification algorithms will require access to the RAW data tier.

- The studies related to reconstruction and identification of low-p_{T} electrons are expected to take several months and thus the integration of algorithm changes into CMSSW are expected to occur towards the end of 2018.

- At a minimum, any developments related to low-p_{T} electrons need only be included in a (patched) CMSSW release dedicated to the reconstruction of the parked dataset.

- The effect of algorithm changes on event size at the RECO, AOD, and miniAOD data tiers, as well as the additional CPU load, will be assessed in due course. Preliminary studies indicate a minimal overhead at the RECO data tier.

- Any high-level analysis that uses low-p_{T} electrons (e.g. the $R_{K^{(*)}}$ measurements) will rely on an "extended" miniAOD data tier that contains a reco::GsfElectron collection (and associated collections) down to $\approx$1 GeV.

3.5 Contributions from Imperial

The contributions from Imperial will focus on improving the $\mathcal{A}\epsilon$ to low-p_{T} electrons according the strategy outlined in Sec. 3.2, using both simulated samples (Sec. 3.3.1) and $\mathrm{J}/\psi \to$ ee events from pileup (Sec. 3.3.3).

4 Summary

In this working document we have outlined a high-level contributions to the B parking effort in CMS, that focus on the low-momentum electron reconstruction as well as the measurement of lepton universality ratio $R_K(q^2 = m_{\ell\ell}^2) = \Gamma(B^\pm \to K^\pm \mu\mu)/\Gamma(B^\pm \to K^\pm ee)$. For the analysis effort we have defined critical milestones towards a timely and competitive measurement of this important quantity. As this is a high-risk-high-gain project, it is possible that at any of these milestones it turns out that a competitive measurement is impossible, which in turn would imply that we revisit our priorities and general involvement in this project.

A Fully Reconstruction of colour-favoured $B^{\pm} \to D^0\pi^{\pm} \to K^{\pm}\pi^{\mp}\pi^{\pm}$ Decays

In this section we outline a strategy to fully reconstruct the colour-favoured $B^+ \to D^0\pi^+ \to K^+\pi^-\pi^+$ decay, where the electrical charge of the K meson matches the one of the charged B meson[1]. While this outlined step-by-step reconstruction procedure is based on past experience, it is imperative to test each step carefully in the Monte Carlo to understand if this is indeed the best approach to reconstruct $B^+ \to D^0\pi^+ \to K^+\pi^-\pi^+$ decays in CMS. The required Monte Carlo sample for this test is currently in production by Thomas and should be available very soon.

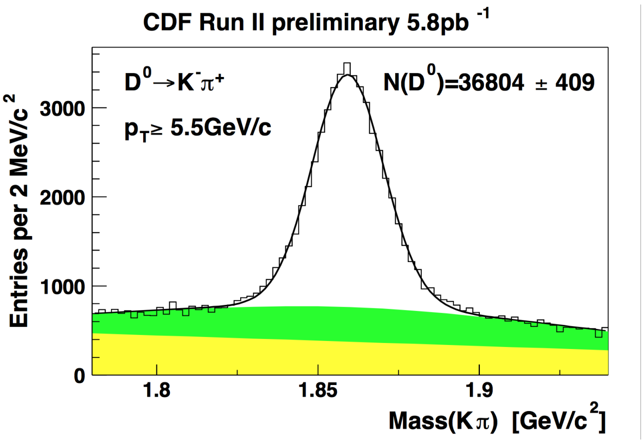

Figure 2. *$M_{K^+\pi^-}$ invariant mass distribution from CDF. The typical mass resolution is around 15 MeV here.*

A.1 Step one: Identify the flavour/electrical charge of the Probe B meson using the trigger muon

As the data sample will be triggered by the single muon trigger, it is possible to identify the electrical charge of the B meson, assuming it originates from a B^+B^- production. Therefore, if the trigger muon possess a negative charge, we assume the probe B to be a B^+ and continue the reconstruction under this hypothesis.

A.2 Step two: Reconstruction of a $D^0 \to K^+\pi^-$ candidate

The next step in the reconstruction chain is to identify the best $D^0 \to K^+\pi^-$ candidate in the event. The following criteria can help to accomplish this:

i) Identify opposite-sign pairs of tracks that stem from a common vertex.

ii) Assume that the positive charged track is a Kaon and apply the corresponding mass hypothesis.

[1] The the colour-favoured decay of a negatively charged B meson is: $B^- \to \bar{D}^0\pi^- \to K^-\pi^+\pi^-$ decay. In the following we will use B^+ as reference but the analog will hold also for B^- with the K meson matching the electrical charge of the charged B meson in question

iii) Calculate the invariant mass $M_{K^+\pi^-}$ and define the mass difference $M_D - M_{K^+\pi^-} = 1.864\ GeV - M_{K^+\pi^-}$. An example of a $M_{K^+\pi^-}$ distribution from CDF is shown in Figure 2.

iv) Chose the candidate in the event that minimises $1.864\ GeV - M_{K^+\pi^-}$. This will be the D^0 candidate used for the B^+ reconstruction step.

A.3 Step three: Reconstruction of a $B^+ \to D^0\pi^+ \to K^+\pi^-\pi^+$ candidate

The final step in the reconstruction chain is to identify the best $B^+ \to D^0\pi^+ \to K^+\pi^-\pi^+$ candidate using the D^0 candidate defined in Step two. The following criteria can help to accomplish this:

i) Identify all positively charged tracks that for with the D^0 candidate a common vertex. The simplest procedure would be to search for common vertex of the $K^+\pi^-$ pair with another positively charged track. Alternatively, the $K^+\pi^-$ pair that represents the D^0 candidate could be refitted using the D^0 mass hypothesis, which could improve the resolution. As a first step, it seems advisable to focus on the simpler "three track" common vertex procedure.

ii) Use all pairs of $D^0\pi^+$ that form a common vertex and calculate the invariant mass $M_{K^+\pi^-\pi^+}$. Define the mass difference: $M_{B^+} - M_{K^+\pi^-\pi^+} = 5.279\ GeV - M_{K^+\pi^-\pi^+}$.

ii) Chose the candidate in the event that minimises $5.279\ GeV - M_{K^+\pi^-\pi^+}$ and plot the distribution of $M_{K^+\pi^-\pi^+}$.

B Potential PhD Thesis Subject: Measurements of Branching Fraction Ratios and CP-asymmetries in Suppressed $B^+ \to (K^-\pi^+)_D\ \pi^+$ and $B^+ \to (K^-\pi^+)_D\ K^+$ decays

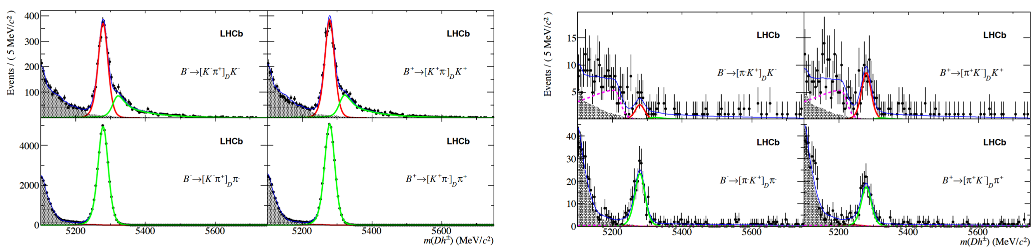

Figure 3. *Invariant mass distributions of selected $B \to (K\pi)_D\ h$ candidates from LHCb. The left four plots show B^- and B^+ candidates for the favoured decay mode, while the four plots to the right show the suppressed decay mode. In the top plots, the bachelor track is identified to be a Kaon, while in the lower plots it is reconstructed as a pion. The dark (red) curve represents the $B \to (K\pi)_D\ K$ events and the light (green) curve stands for $B \to (K\pi)_D\ \pi$ The shaded contribution are partially reconstructed events and the combinatorial component.*

In appendix C we have outlined how to reconstruct the colour-favoured $B^+ \to D^0\pi^+ \to K^+\pi^-\pi^+$ decay. While this decay is important to perform a significant measurement of the

B purity using the first reconstructed data, its suppressed counter-part $B^+ \to (K^-\pi^+)_D\ \pi^+$ has more interesting physics to offer. In fact, the suppressed decays $B^+ \to (K^-\pi^+)_D\ \pi^+$ and $B^+ \to (K^-\pi^+)_D\ K^+$ are sensitive to the CKM angle γ, which is a critical quantity in the unitarity triangle.

Table 3. *Summary of the expected yields of important B modes that can be collected with 2 kHz of parking in 2018. For the 2018 we assume* $\sec_{\mathrm{LHC}} = 7.8 \times 10^6$ *[5] of LHC running and B hadron purity of* $P_{\mathrm{HLT}} = 0.8$ *after a refinement of the L1 seed at the HLT (see Section* **??***). Note that for the yield of* $B^0 \to K^*\ell^+\ell^-$ *decays the factor* $\frac{2}{3}$ *for the decay of* $K^*(893) \to K^\pm\pi^\mp$*, which is required to fully reconstruct this decay, is already included. Note that the numbers in this table do not include the reconstruction efficiency.*

Mode	N_{2018}	f_B [6]	$\mathcal{B}$
Suppressed and favoured $B^+ \to D^0h^+$ decays [h=K,π]			
$B^+ \to (K^+\pi^-)_D^{fav}\ \pi^+$	8.9×10^5	0.4	1.78×10^{-4} [7]
$B^+ \to (K^-\pi^+)_D^{sup}\ \pi^+$	3140	0.4	6.3×10^{-7} [7]
$B^+ \to (K^+\pi^-)_D^{fav}\ K^+$	$\approx 8 \times 10^4$	0.4	$\approx 1.5 \times 10^{-5}$ [7]
$B^+ \to (K^-\pi^+)_D^{sup}\ K^+$	≈ 1500	0.4	$\approx 3 \times 10^{-7}$ [7]
$R_K^{(*)}$			
$B^0 \to K^*\ell^+\ell^-$	3290	0.4	$\frac{2}{3} \times 9.9 \times 10^{-7}$ [7]
$B^\pm \to K^\pm\ell^+\ell^-$	2250	0.4	4.51×10^{-7} [8]

As outlined in [1], a powerful strategy to measure the angle γ in tree-level processes is to study CP-violating observables in the decays $B^+ \to D\ h^+$ where D indicates a neutral charm meson which decays in a mode common to both D^0 and $\bar{D}^0$ states, and h, the bachelor hadron, is either a kaon or a pion. In the case of $B^+ \to D\ h^+$, interference occurs between the suppressed $b \to u\bar{c}s$ and favoured $b \to c\bar{u}s$ decay paths, and similarly for the charge conjugate decay.

Measuring the suppressed decays of $B^+ \to (K^-\pi^+)_D^{sup}\ h^+$, with $h = \pi, K$ and normalising it to the favoured decay scenarios yields the following ratios:

$$R_{Dh} = \frac{B^+ \to (K^-\pi^+)_D^{sup}\ h^+}{B^+ \to (K^+\pi^-)_D^{fav}\ h^+}.$$

An attempt to measure these ratios was carried out in [2–4] but due to limited statistic, especially in the $B^+ \to D\ K^+$ final state, the results are not very significant yet. The latest result is is from LHCb [2] and the corresponding distributions are shown in Figure 3. While a few hundred events of the suppressed decay $B^+ \to (K^-\pi^+)_D^{sup}$ are observed, the suppressed decay with the kaon as bachelor track in the final state only shows a very weak signal.

As shown in Table B, the expected yields before acceptance and reconstruction efficiency for the suppressed $B \to (K\pi)_D\ h$ decays are similar to the one of $B^0 \to K^*\ell^+\ell^-$ and $B^\pm \to K^\pm\ell^+\ell^-$. However, in contrast to the signal events of the $R_K^{(*)}$ measurements, the $B \to (K\pi)_D\ h$ final state does not require an improvement of the low-momentum electron reconstruction and already with state-of-the-art CMS reconstruction tools it should be possible to reconstruct a competitive data set of the suppressed $B \to (K\pi)_D\ h$ decays, especially in the $h = \pi$ scenario.

Therefore, the measurement of R_{Dh} would provide a natural fall-back option for a PhD thesis in case the requirement improvements in the electron reconstruction for a significant $R_K^{(*)}$ cannot be accomplished. It would also be a straightforward continuation of the B purity measurement outlined in appendix C using the favoured decay.

C Procedure for generic Monte Carlo samples generation using the CMS software and computing infrastructure

CAVEAT: The steps described here are to be taken as a generic example to produce a sample using 2017 conditions down to the MINIAOD format. Some tuning of the number of threads for instance may be required depending on the specificity of each sample, on the statistics to generate and on the availability of the Grid sites. Those are inspired by the steps outlined in the following twiki `https://twiki.cern.ch/twiki/bin/viewauth/CMS/BPHMonteCarloContactInfo#Recipe_of_Private_MC_production`

C.1 Step one: GEN-SIM step

A so-called gen-fragment file, corresponding to the process to generate, is needed for this step. Examples of such files are available here `https://github.com/oozcelik/Fragments/blob/master/` If new processes not available here need to be generated, a new gen-fragment needs to be developed by the user. Help can be provided for this by the BPH MC contact O. Ozcelik (ozlem.ozcelik@cern.ch). The gen-fragment can potentially refer to a decay file, such as the ones available here `https://github.com/oozcelik/GeneratorInterface-EvtGenInterface`

The setup to run the GEN-SIM step can be installed following the instructions below

cmsrel CMSSW_9_3_6
cd CMSSW_9_3_6/src
cmsenv
mkdir -p Configuration/GenProduction/python
put the gen fragment in this directory
scram b -j 9
cmsDriver.py Configuration/GenProduction/python/BToKee_13TeV-pythia8-evtgen_cfi.py –fileout file:BToKee_GEN-SIM.root –mc –eventcontent RAWSIM –datatier GEN-SIM –conditions 93X_mc2017_realistic_v3 –beamspot Realistic25ns13TeVEarly2017Collision –step GEN,SIM –nThreads 1 –geometry DB:Extended –era Run2_2017 –python_filename step1_BToKee_GEN-SIM_cfg.py –no_exec –customise Configuration/DataProcessing/Utils.addMonitoring -n 50

Configuration/GenProduction/python/BToKee_13TeV-pythia8-evtgen_cfi.py is to be replaced by the proper gen-fragment file. BToKee_GEN-SIM.root and step1_BToKee_GEN-SIM_cfg.py are arbitrary names for the output root files and the python script to be generated. The python script can then be run using the standard crab submission on the grid.

C.2 Step two: PUMIX step

A different release is a priori needed to run this step. The following instructions can be followed to generate the corresponding python script.

cmsrel CMSSW_9_4_4
cd CMSSW_9_4_4/src
cmsenv
scram b -j 9
cmsDriver.py –mc –eventcontent PREMIXRAW –datatier GEN-SIM-RAW –conditions 94X_mc2017_realistic_v12 –step DIGIPREMIX_S2,DATAMIX,L1,DIGI2RAW,HLT:2e34v40 –nThreads 4 –datamix PreMix –era Run2_2017 –filein file:BToKee_GEN-SIM.root –fileout file:BToKee_PUMix.root –python_filename step2_BToKee_PUMix_cfg.py –pileup_input /store/mc/RunIISum
10_gun/GEN-SIM-DIGI-RAW/MC_v2_94X_mc2017_realistic_v9-v1/30042/98009154-F2CD-E711-A4E1-FA163EC18760.root –no_exec -n -1

The python script can then be run using the standard crab submission on the grid.

C.3 Step three: AODSIM step

The same release as for the previous step can be used. The following instructions can be followed to generate the corresponding python script.

cmsDriver.py –filein file:BToKee_PUMix.root –fileout file:BToKee_AODSIM.root –mc –eventcontent AODSIM runUnscheduled –datatier AODSIM –conditions 94X_mc2017_realistic_v12 –step RAW2DIGI,RECO,RECOSIM,EI –nThreads 4 –era Run2_2017 –python_filename step3_BToKee_AODSIM_cfg.py –no_exec –customise Configuration/DataProcessing/Utils.addMonitoring -n -1

The python script can then be run using the standard crab submission on the grid.

C.4 Step four: MINIAODSIM step

The same release as for the previous step can be used. The following instructions can be followed to generate the corresponding python script.

cmsDriver.py –mc –eventcontent MINIAODSIM –runUnscheduled –datatier MINIAOD-SIM –conditions 94X_mc2017_realistic_v12 –step PAT –era Run2_2017 –filein file:BToKee_AODSIM.root –fileout file:BToKee_MINIAODSIM.root –python_filename step4_BToKee_MINIAODSIM_cfg.py –no_exec -n -1

The python script can then be run using the standard crab submission on the grid.

C.5 Step five: NANOAOD step

Dedicated plugins have been developed to store the relevant variables needed to perform the analysis, within the generic NANOAOD format recently adopted by CMS. The corresponding code is maintained on the following git repository https://github.com/ICBPHCMS/cmssw

cmsrel CMSSW_9_4_6_patch1
cd CMSSW_9_4_6_patch1/src
cmsenv
git cms-merge-topic cms-nanoAOD:master100Xbase
git cms-merge-topic ICBPHCMS:NanoAOD_BPH_101X
scram b -j 4
cmsDriver.py test94X -s NANO –mc –eventcontent NANOAODSIM –datatier NANOAODSIM –filein /store/mc/RunIIFall17MiniAOD/TTToSemiLeptonic_TuneCP5_PSweights_13TeV-powheg-pythia8/MINIAODSIM/94X_mc2017_realistic_v10-v1/60000/A0D71AEE-13E1-E711-B3C9-FA163E629498.root –no_exec –conditions auto:phase1_2017_realistic -n 1000 –era Run2_2017,run2_nanoAOD_94XMiniAODv1

The python script can then be run using the standard crab submission on the grid.

References

[1] D. Atwood, I. Dunietz, A. Soni, Phys. Rev. Lett. **78**, 3257 (1997), hep-ph/9612433

[2] R. Aaij, et al. (LHCb), Phys. Lett. **B712**, 203 (2012), [Erratum: Phys. Lett.B713,351(2012)], 1203.3662

[3] T. Aaltonen, et al. (CDF), Phys. Rev. **D84**, 091504 (2011), 1108.5765

[4] M. Saigo, et al. (Belle), Phys. Rev. Lett. **94**, 091601 (2005), hep-ex/0412025

[5] https://indico.cern.ch/event/688591/contributions/2826324/attachments/1580667/2501056/computing_exo_workshop_jan_2018.pdf

[6] http://www.slac.stanford.edu/xorg/hfag/osc/summer_2017/

[7] http://pdg.lbl.gov/2017/listings/rpp2017-list-B-zero.pdf

[8] http://pdg.lbl.gov/2017/listings/rpp2017-list-B-plus-minus.pdf