$B_s \to \mu\mu\gamma$ in the high-q^2 region

Camille Normand (they/them)
On behalf of the LHCb collaboration

Laboratoire d'Annecy-le-Vieux de Physique des Particules
Università degli Studi di Cagliari

16th February, 2023
NePsi 23 - Università di Pisa

Université degli Studi di Cagliari

Istituto Nazionale di Fisica Nucleare

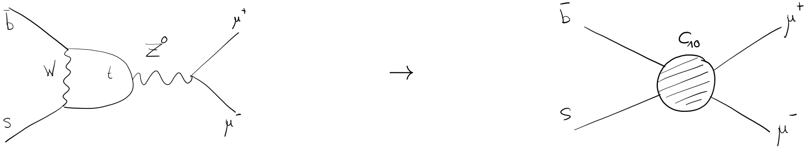
UNIVERSITÉ SAVOIE MONT BLANC

Table of Contents

$B_s \to \mu\mu\gamma$ vs $B_s^0 \to \mu^+\mu^-$

- $B_s \to \mu\mu\gamma$: radiative counterpart to $B_s^0 \to \mu^+\mu^-$.
  Additional photon lifts the helicity suppression of $B_s^0 \to \mu^+\mu^-$, compensating for the additional QED interaction.

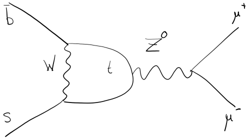
- Sensitive to more Wilson coefficients of interest in $b \to s\ell\ell$ transitions than $B_s^0 \to \mu^+\mu^-$ (C_9 and C_{10} in the high-q^2 region).



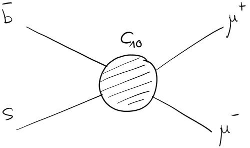
$B_s \to \mu\mu\gamma$ vs $B_s^0 \to \mu^+\mu^-$

- $B_s \to \mu\mu\gamma$: radiative counterpart to $B_s^0 \to \mu^+\mu^-$.
  Additional photon lifts the helicity suppression of $B_s^0 \to \mu^+\mu^-$, compensating for the additional QED interaction.

- Sensitive to more Wilson coefficients of interest in $b \to s\ell\ell$ transitions than $B_s^0 \to \mu^+\mu^-$ (C_9 and C_{10} in the high-q^2 region).



$\rightarrow$



- Larger theoretical uncertainties from $B_s \to \gamma$ form factors + charmonium resonances (cf. L.Vittorio's talk).
  $\longrightarrow$ Not as theoretically clean as $B_s^0 \to \mu^+\mu^-$.

ISR and FSR

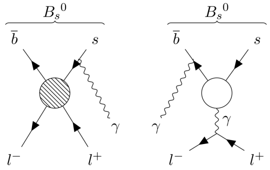


Figure: $B_s \to \mu\mu\gamma$ Initial State Radiation (ISR), hatched circle is C_9/C_{10}, empty circle is C_7.

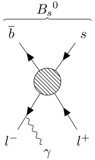


Figure: $B_s \to \mu\mu\gamma$ Final State Radiation (FSR), hatched circle is C_{10}.

Differential branching ratio

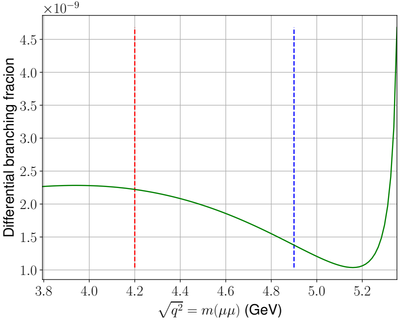


Figure: $B_s \to \mu\mu\gamma$ differential branching fraction. Blue vertical line is the $B_s^0 \to \mu^+\mu^-$ analysis lower mass, red is the objective of this new analysis. Theoretical prediction from flavio.

- Integrated branching fraction $\approx 10^{-10}$

Experimental limit

- LHCb performed a measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction in 2021[1], setting an upper-limit on $B_s \to \mu\mu\gamma$:

$$\mathcal{B}(B_s \to \mu\mu\gamma | [4.9; 6.0]\ \text{GeV}) < 2.0 \times 10^{-9} \tag{1}$$

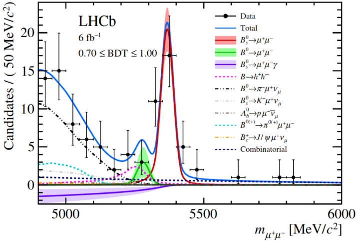


Figure: Fit to the dimuon invariant mass in the recent LHCb analysis.

[1] Aaij et al., arXiv: 2108.09283 (hep-ex), 2021.

Analysis outline

- This analysis aims at measuring the branching fraction in the high-q^2 region $[4.2, 6.0]$GeV, **without reconstructing the photon.**[2]
  $\longrightarrow$ Increase in efficiency.
  $\longrightarrow$ Re-use the same dataset and normalization as $B_s^0 \to \mu^+\mu^-$.
  $\longrightarrow$ Not looking for a mass peak, but for a *shoulder*.

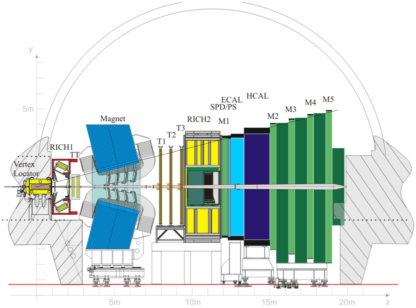


Figure: LHCb experiment during Runs 1 & 2[3].

[2] Dettori, Guadagnoli, and Reboud, arXiv: 1610.00629 (hep-ph), 2017.
[3] Aaij et al., arXiv: 1412.6352 (hep-ex), 2015.

Quick summary

- Theoretical interest: $B_s \to \mu\mu\gamma$, mediated by FCNCs, is sensitive to different WCs than $B_s^0 \to \mu^+\mu^-$: probing different NPs.

- Experimental challenges: Rare decay, which needs a lot of statistics (good candidate for LHC measurements) + large background.

- Specificity: only use the dimuon information, as the $B_s^0 \to \mu^+\mu^-$ analysis.

Table of Contents

Outline

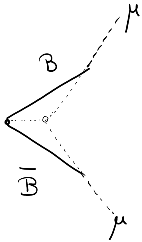
- ✓ Dataset: Run 2 *pp* collisions, 13 TeV, Dimuon samples
- ✓ Trigger and stripping: require dimuon with $m(\mu\mu) > 4.2\,\text{GeV}$.
- ✓ Preselections to "clean" the sample: cuts with very high signal efficiency.
- ✓ BDT: MultiVariate Classifier trained to discriminate signal and combinatorial background.
- ✓ Study of the physics backgrounds: Estimation from MC.
- ✓ Mass fit with all combined components

Selection

- Preselection:
  - On the B_s^0 candidate:
    - $m(B_s^0) \in [4.2, 6.0]$ GeV
    - vertex quality
  - On the muon candidates:
    - Distance of closest approach $< 0.3\,\text{mm}$
    - Track isolated from the rest of the event
    - Tracking quality
    - Tight Particle Identification (PID) requirement.

BDT

- At that point, the data are dominated by "combinatorial background", meaning events with two muons coming from different/successive b-decays.



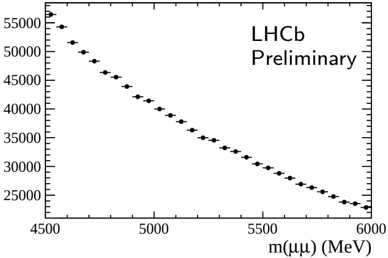
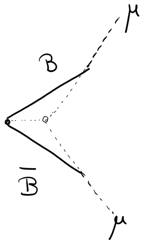


Figure: Distribution of data before BDT

BDT

- At that point, the data are dominated by "combinatorial background", meaning events with two muons coming from different/successive b-decays.



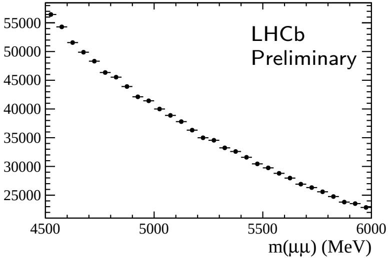
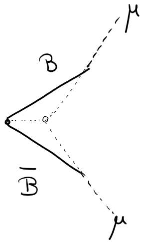


Figure: Distribution of data before BDT

- Number of decay candidates for Run 2 before BDT : $\sim$ 1M vs 10-20 signal events.

BDT

- At that point, the data are dominated by "combinatorial background", meaning events with two muons coming from different/successive b-decays.



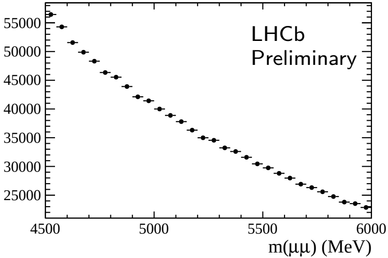


Figure: Distribution of data before BDT

- Number of decay candidates for Run 2 before BDT : $\sim$ 1M vs 10-20 signal events.
- A common way at LHCb to reduce this type of background is by using MultiVariate Analysis (MVA) methods, such as a Boosted Decision Tree (BDT).

What is a BDT for ?

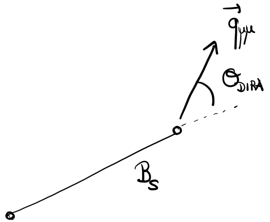
- A BDT contains a training phase, and a testing phase (similar to Neural Networks).

What is a BDT for ?

- A BDT contains a training phase, and a testing phase (similar to Neural Networks).
- Two main inputs:
  - Two training samples labeled as "signal" and "background".
    - Signal: $B_s \to \mu\mu\gamma$ MC in [4.2;6.0] GeV range.
    - Proxy for background: data in [4.2;6.0] GeV range.
  - Input variables that are discriminating between the two training samples.

Choosing the input variables

- Find variables with highest discriminative power between signal and background.



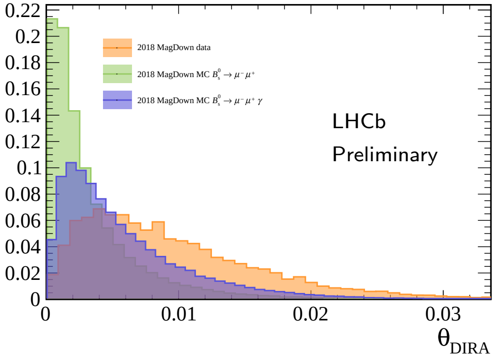


Figure: Distribution of the direction angle for $B_s \to \mu\mu\gamma$ (signal), $B_s^0 \to \mu^+\mu^-$, and data (combinatorial).

Input variables

- Geometric criteria
  - Direction angle
  - Impact parameter of the muons w.r.t. the vertex
  - Distance of closest approach of the muons
- Isolation variables
- Vertex fitting quality

BDT

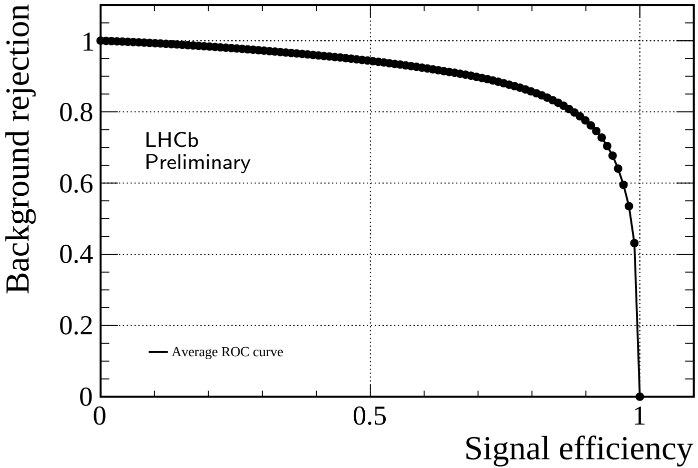


Figure: Average ROC curves - 800 Trees, maxDepth 2, no pruning, 10 Folds

BDT efficiency on signal

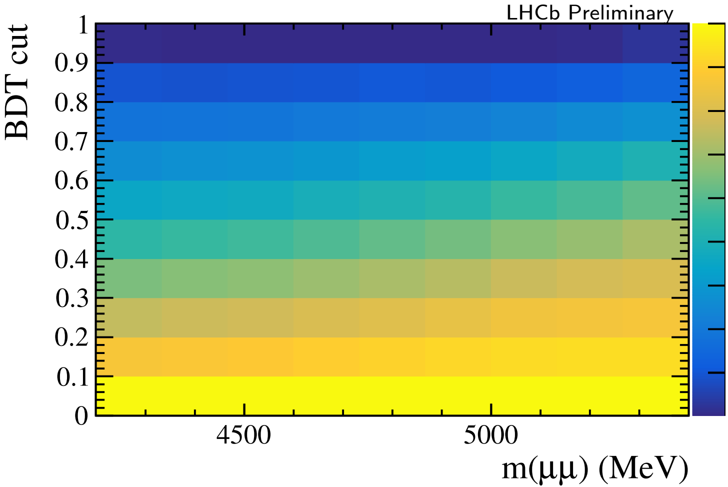


Figure: BDT efficiency vs invariant dimuon mass on signal

Mass shape of our signal

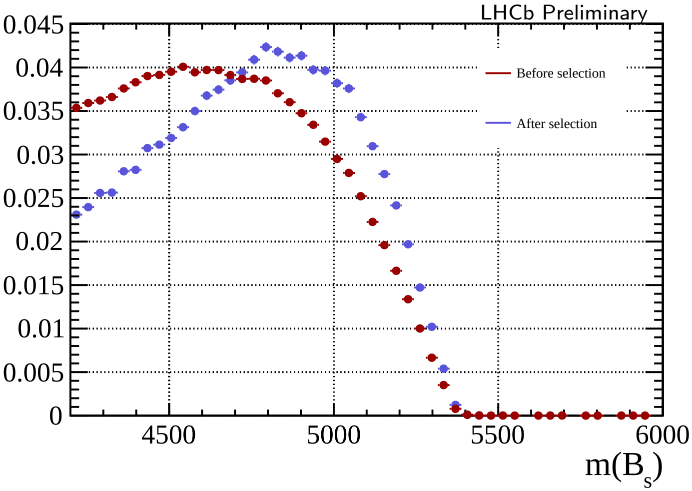


Figure: Mass shape of $B_s \to \mu\mu\gamma$ before and after selection.

BDT efficiency on data

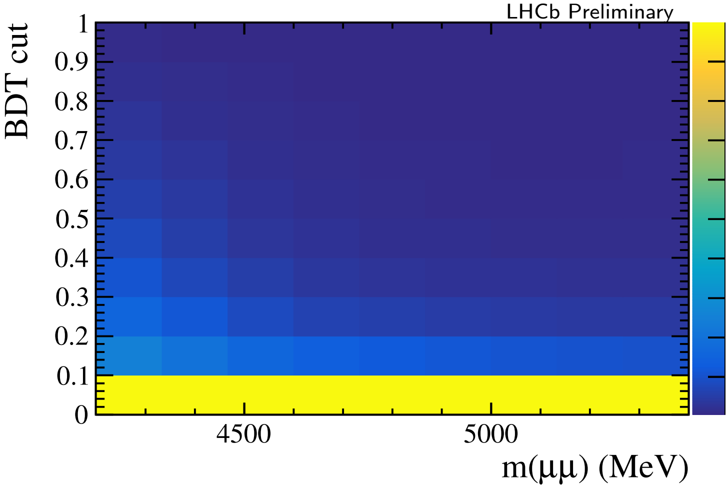


Figure: BDT efficiency vs invariant dimuon mass on data (= combinatorial)

BDT efficiency on $B^0 \to \pi^+\mu^-\nu$

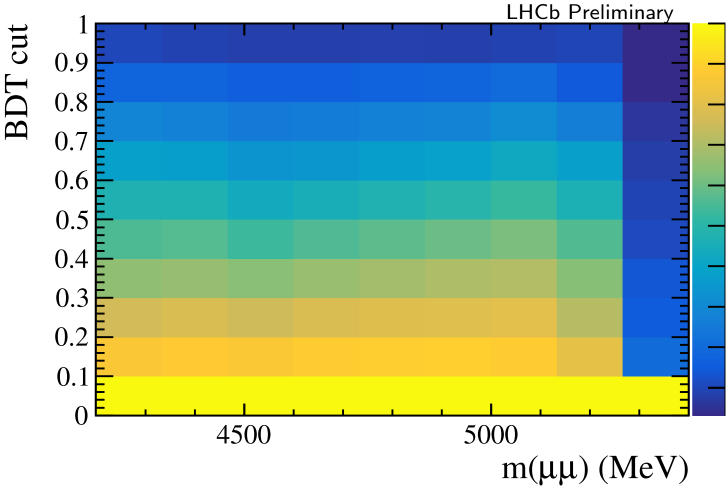


Figure: BDT efficiency vs invariant dimuon on $B^0 \to \pi^+\mu^-\nu$.

Backgrounds

- First set: backgrounds above 4.9 GeV

no mis ID	Branching fraction
$B_s^0 \to \mu^+\mu^-$	3.66×10^{-9}
$B^+ \to \pi^+\mu\mu$	1.83×10^{-8}
$B^0 \to \pi^0\mu\mu$	8.60×10^{-9}
$B^0 \to \mu\mu$	1.60×10^{-10}
$B_c^+ \to J/\psi\mu\nu$	$1.95 \times 10^{-2} \times 0.05$

one mis ID	Branching fraction
$B^0 \to \pi\mu\nu$	1.44×10^{-4}
$B_s^0 \to K\mu\nu$	1.42×10^{-4}
$\Lambda_b \to p\mu\nu$	4.1×10^{-4}
two mis ID	
$B^0 \to K\pi$	1.96×10^{-5}
$B_s^0 \to KK$	2.59×10^{-5}
$B_s^0 \to K\pi$	5.7×10^{-6}
$B^0 \to \pi\pi$	5.12×10^{-6}

Backgrounds

- Second group: backgrounds appearing lower in mass ($< 4.9\,\text{GeV}$)

$B \to M_s \mu\mu$	Branching fraction
$B^0 \to K^0_{\mathrm{S}} \mu\mu$	4.35×10^{-7}
$B^0 \to K^* \mu\mu$	9.60×10^{-7}
$B^0_s \to \phi \mu\mu$	3.39×10^{-7}
$B^+ \to K^+ \mu\mu$	9.40×10^{-7}
$B^+ \to K^* \mu\mu$	8.40×10^{-7}

WIP

- Start to implement mass fit:
  - Study the sensitivity to signal: Decide on the final BDT cut.
- Some other backgrounds to study:
  - double misID'd $B^0 \to \pi^+\pi^-(K^0) = 4.94 \times 10^{-5}$.
  - one mis ID with excited unflavored mesons: $B^0 \to (\rho \to \pi\pi)\mu\nu = 2.94 \times 10^{-4}$.

Question time !

Thank you for your attention !

[1] R. Aaij et al. "LHCb Detector Performance". In: *Int. J. Mod. Phys. A* 30.07 (2015), p. 1530022. DOI: 10.1142/S0217751X15300227. arXiv: 1412.6352 [hep-ex].

[2] R. Aaij et al. "Measurement of the $B_s^0 \to \mu^+\mu^-$ decay properties and search for the $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ decays". In: (Aug. 2021). arXiv: 2108.09283 [hep-ex].

[3] F. Dettori, D. Guadagnoli, and M. Reboud. "$B_s^0 \to \mu^+\mu^-\gamma$ from $B_s^0 \to \mu^+\mu^-$". In: *Phys. Lett.* B768 (2017), pp. 163–167. DOI: 10.1016/j.physletb.2017.02.048. arXiv: 1610.00629 [hep-ph].

Comment on cascade decays

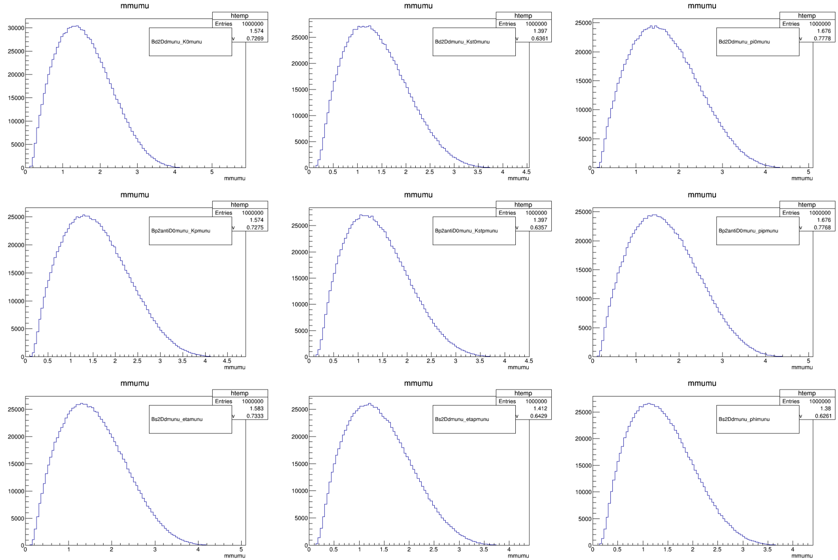
- Cascades $B \to (D \to X\mu\nu)\mu\nu$;

  - ☐ $B_{u,d} \to (D \to \pi\mu\nu)\mu\nu$

  - ☐ $B_{u,d} \to (D \to K\mu\nu)\mu\nu$

  - ☐ $B_s^0 \to (D \to H\mu\nu)\mu\nu$ with $H = \eta, \phi$

Mass distribution of cascade decays, RapidSim



How do we measure the branching fraction ?

- The number of events is given by

$$N(B_s \to \mu\mu\gamma) = \sigma(pp \to b\bar{b}) \times \mathcal{L} \times f(b \to B_s) \times \epsilon_{B_s \to \mu\mu\gamma} \times \mathcal{B}(B_s \to \mu\mu\gamma) \ .$$

- We can do the same with another channel, the *normalization* channel:

$$N(B^+ \to J/\psi K^+) = \sigma(pp \to b\bar{b}) \times \mathcal{L} \times f(b \to B^+) \times \epsilon_{B^+ \to J/\psi K^+} \times \mathcal{B}(B^+ \to J/\psi K^+) \ .$$

How do we measure the branching fraction ?

- By simply taking the ratio

$$\mathcal{B}(\text{sig}) = \frac{f_{\text{norm}}}{f_{\text{sig}}} \frac{\epsilon_{\text{norm}}}{\epsilon_{\text{sig}}} \frac{N_{\text{sig}}}{N_{\text{norm}}} \mathcal{B}_{norm} \ ,$$

with :

  - f_{i} : hadronisation fractions,
  - ϵ_{i} : total efficiencies,
  - N_{i} : numbers of measured candidates,
  - $\mathcal{B}_i$: theoretical branching fractions.

What are the efficiencies ?

- We want to “disentangle” our signal from everything else (especially for this type of analysis with a very low branching fraction).
  - We apply “requirements” on the variables of the candidates to reduce backgrounds while maximizing the number of signal candidates that pass these cuts.
  - For each cut, we have an efficiency that is:

$$\epsilon_{\text{sig}} = \frac{\text{Number of events after the cut}}{\text{Number of events before the cut}}$$

What are these requirements ?

- These requirements are applied through several steps:
  - Trigger selections: the event contains at least two muons,
  - Stripping line: the two muons have an invariant mass less than 1.2 GeV away from the B_s mass,
  - Reconstruction/selection: various cuts on the reconstruction quality, track fitting quality etc,
  - PID selections: pions, kaons (and to a less extent protons) might be misidentified as muons, and we want to reduce the number of misidentified decays.
- Efficiencies on signal are computed using MC simulation of the decay (generation + interaction with the detector)

DIRA for different mass bins

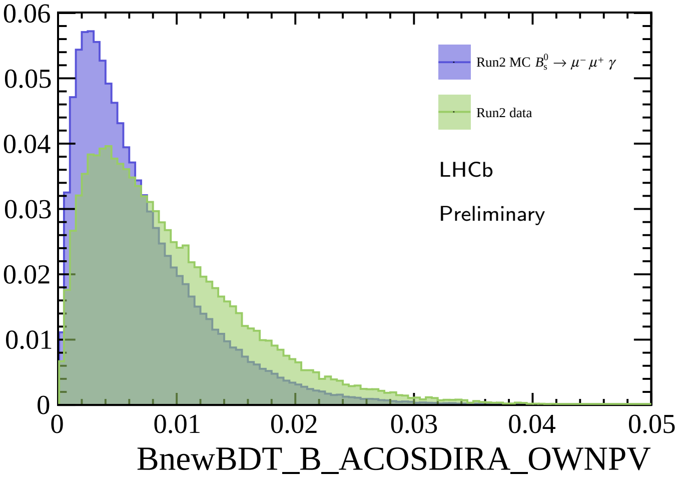


Figure: Distribution of the direction angle for $B_s \to \mu\mu\gamma$ (signal), and data (combinatorial), in [4.5,4.9] GeV.

DIRA for different mass bins

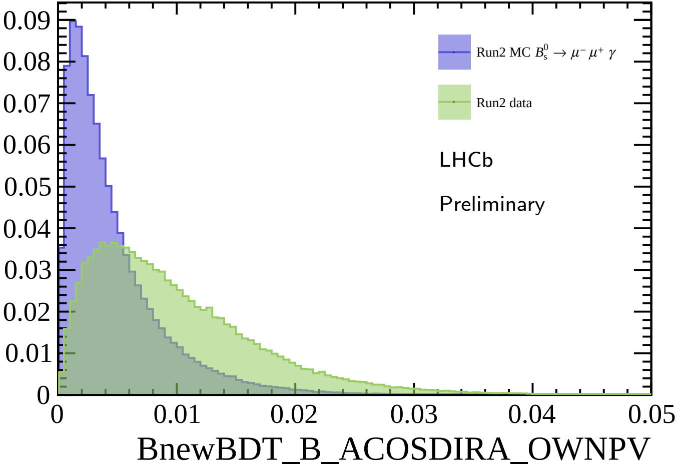


Figure: Distribution of the direction angle for $B_s \to \mu\mu\gamma$ (signal), and data (combinatorial), in [4.9,6.0] GeV.

BDT > Comparison between the two BDTs on $B_s^0 \to \mu^+\mu^-$

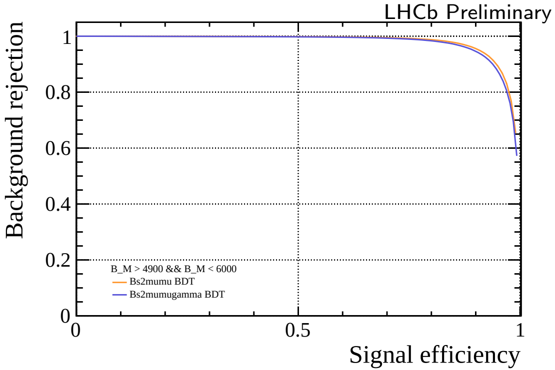


Figure: ROC curves on $B_s^0 \to \mu^+\mu^-$ of (blue) the $B_s \to \mu\mu\gamma$ specific BDT vs (orange) the $B_s \to \mu\mu$ analysis BDT, in the [4.9,6] GeV mass range.

$B_s \to \mu\mu\gamma$ analysis

- Cascades:
  - ☐ $B^0 \to (D^- \to K^0\mu^-\nu)\mu^+\nu = 8.76\% \times 2.31\%$
  - ☐ $B^0 \to (D^- \to K^{*0}\mu^-\nu)\mu^+\nu = 5\% \times 2.31\%$
  - ☐ $B^0 \to (D^- \to \pi^0\mu^-\nu)\mu^+\nu = 0.35\% \times 2.31\%$
  - ☐ $B^+ \to (\bar{D}^0 \to K^+\mu^-\nu)\mu^+\nu = 3.41\% \times 2.35\%$
  - ☐ $B^+ \to (\bar{D}^0 \to K^{*+}\mu^-\nu)\mu^+\nu = 1.89\% \times 2.35\%$
  - ☐ $B^+ \to (\bar{D}^0 \to \pi^+\mu^-\nu)\mu^+\nu = 0.26\% \times 2.35\%$
  - ☐ $B_s \to (D_s^- \to \phi\mu^-\nu)\mu^+\nu = 1.9\% \times 8.1\%$
  - ☐ $B_s \to (D_s^- \to \eta\mu^-\nu)\mu^+\nu = 2.4\% \times 8.1\%$
  - ☐ $B_s \to (D_s^- \to \eta'\mu^-\nu)\mu^+\nu = 1.1\% \times 8.1\%$

Mcorr

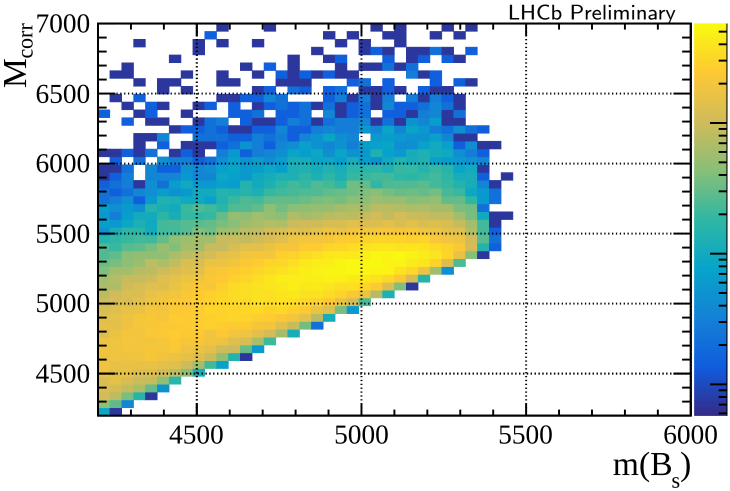


Figure: B_M vs $MCORR$ $B_s \to \mu\mu\gamma$

Mcorr

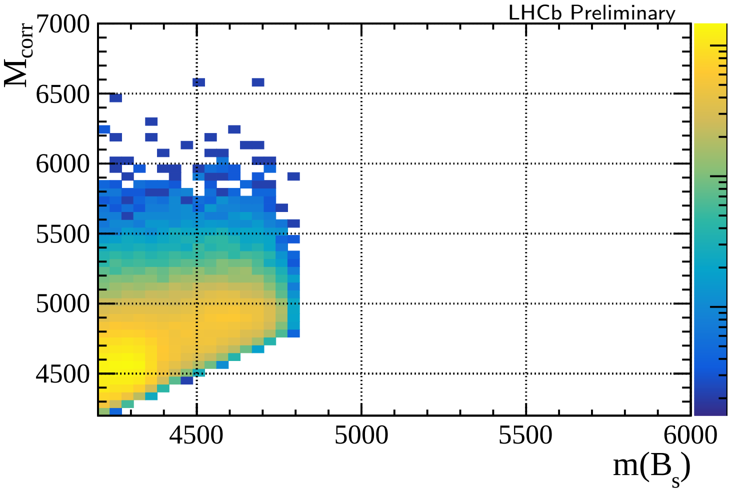


Figure: B_M vs $MCORR$ $B \rightarrow K\mu\mu$