Test $\tau - e$ universality at LHCb

Ching-Hua Li

Supervisors: Dorothea vom Bruch & Olivier Leroy

Aix Marseille Univ, CNRS/IN2P3, CPPM, Marseille, France

27th October, 2025

Outline

Introduction

- Standard model and lepton flavor universality (LFU)
- Current status of LFU measurement
- Introduction of LHCb

Methodology

- Data flow for candidate reconstruction
- $R(D^*)_{\tau/e}$ measurement
- Kinematic reconstruction and fitting variables

Background study

- Fake D^{*-}/ D^0 background
- D^{**} background

Works done and ongoing

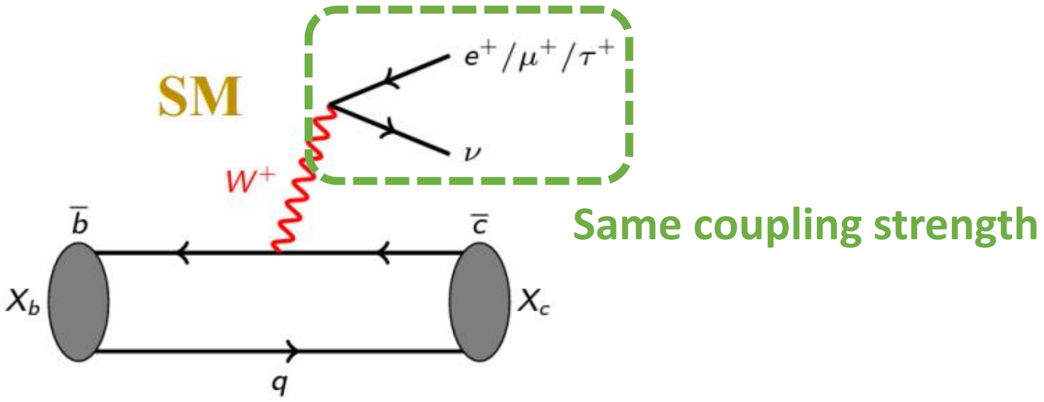
Outlook

Introduction

Lepton Flavour Universality (LFU) in the SM

What is LFU ?

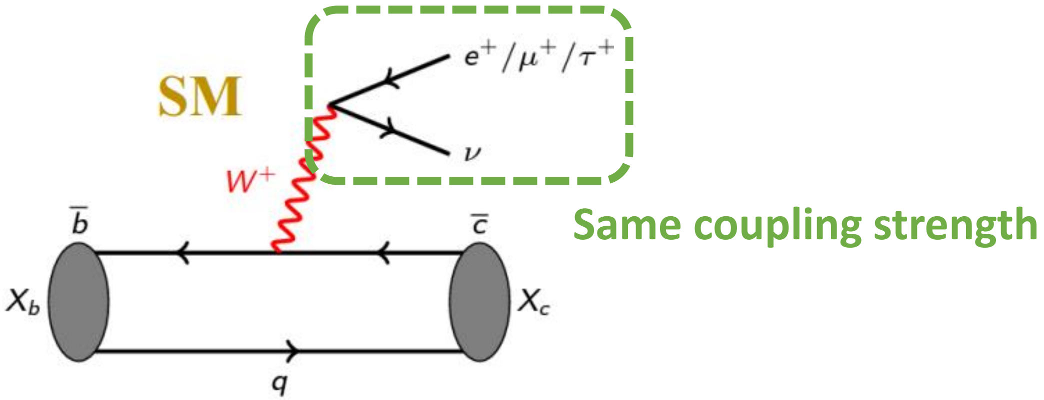
In Standard Model (SM), electroweak couplings to each lepton generation are **identical.**



Lepton Flavour Universality (LFU) in the SM

What is LFU ?

In Standard Model (SM), electroweak couplings to each lepton generation are **identical.**

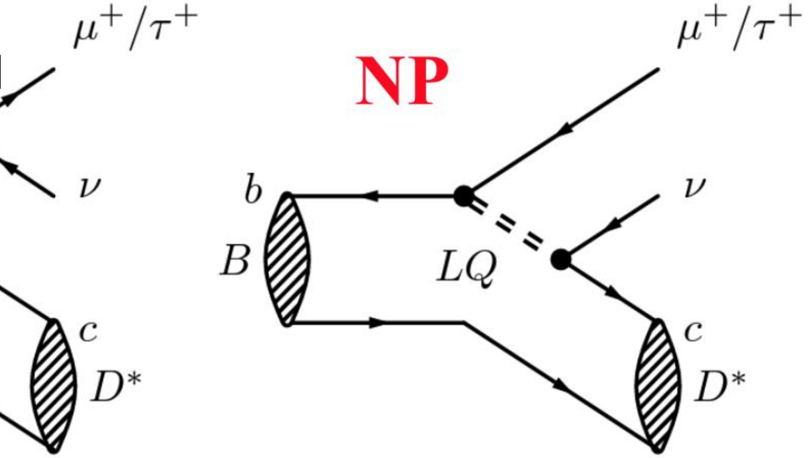


How can LFU be violated ?

The couplings could be affected by New Physics (NP) contributions.

e.g. Leptoquarks, two Higgs doublet, non-universal left-right model

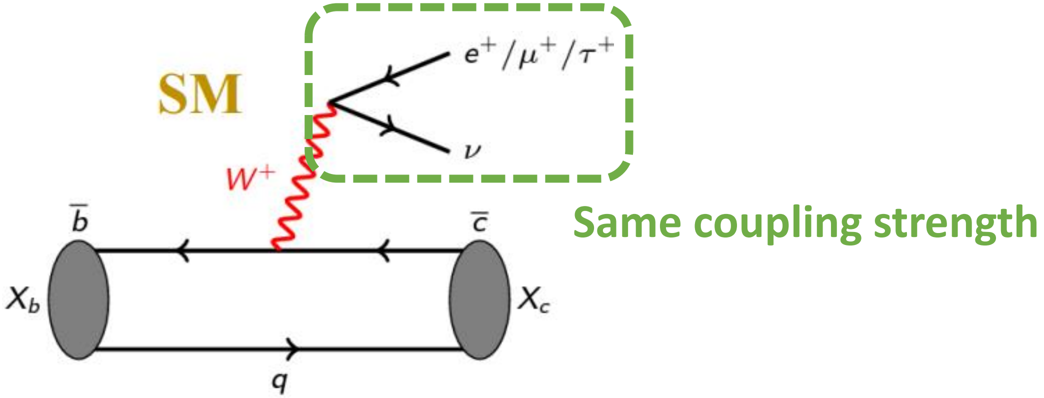
[PRD 85,094025 (2012) , PRD 110, 055023 (2024)]



Lepton Flavour Universality (LFU) in the SM

What is LFU ?

In Standard Model (SM), electroweak couplings to each lepton generation are **identical.**



How can LFU be violated ?

The couplings could be affected by New Physics (NP) contributions.

e.g. Leptoquarks, two Higgs doublet, non-universal left-right model

How do people test LFU ?

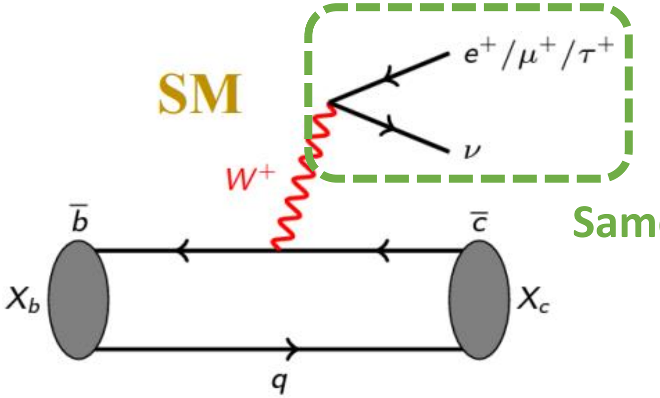
We can test LFU by measuring the ratio of decay rates with different leptonic final state,

e.g. measuring $R(X_c) = \frac{BF(X_b \to X_c l v)}{BF(X_b \to X_c l' v)}$, $X_{b(c)}$ is a meson or baryon containing a *b* or *c* quark

Lepton Flavour Universality (LFU) in the SM

What is LFU ?

In Standard Model (SM), electroweak couplings to each lepton generation are **identical.**



How can LFU be violated ?

The couplings could be affected by New Physics (NP) contributions.

e.g. Leptoqu

How do p

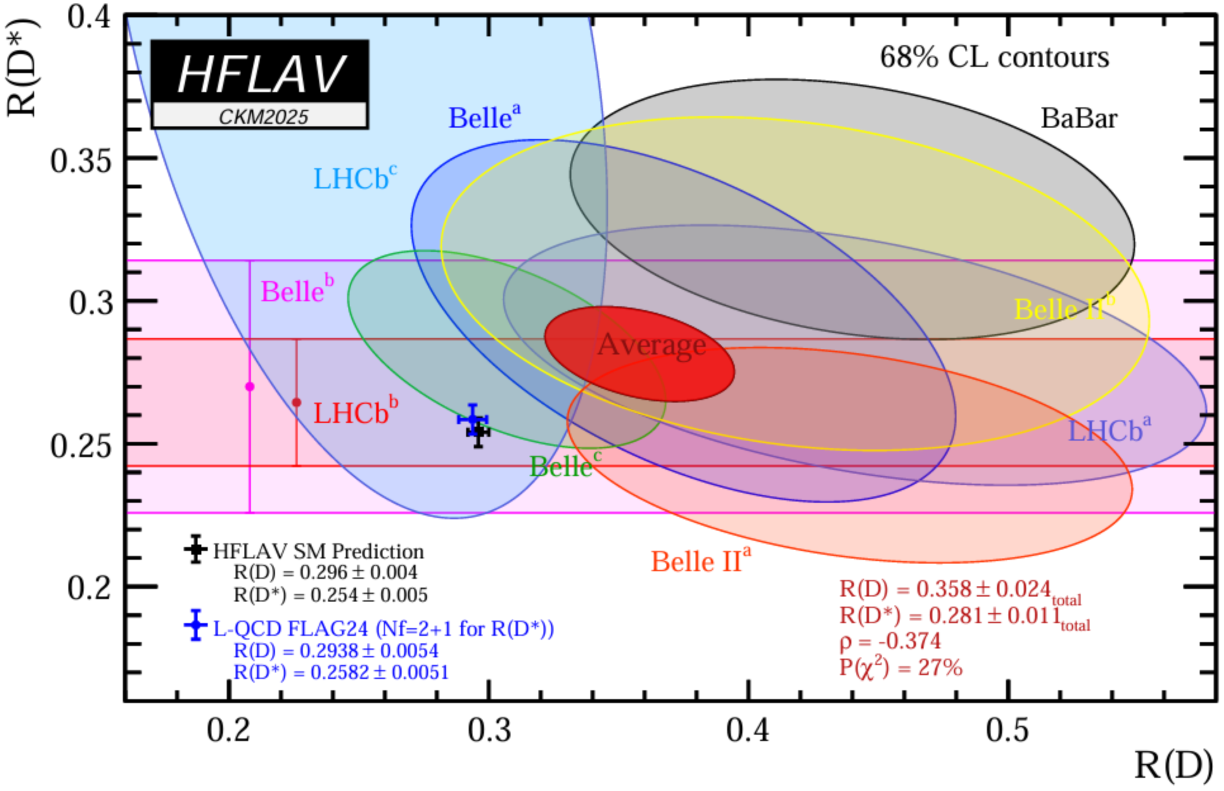
We can test

e.g. measuring $R(X_c) = \frac{}{BF(X_b \to X_c l' v)}$, $X_{b(c)}$ is a meson or baryon containing a b or c quark

My PhD work mainly focuses on the measurement of $R(D^*)_{\tau/e} = \frac{BF(B^0 \to D^{*-}\tau^+ v_\tau)}{BF(B^0 \to D^{*-} e^+ v_e)}$ at LHCb

Experimental and theoretical status for $R(D^{(*)})$

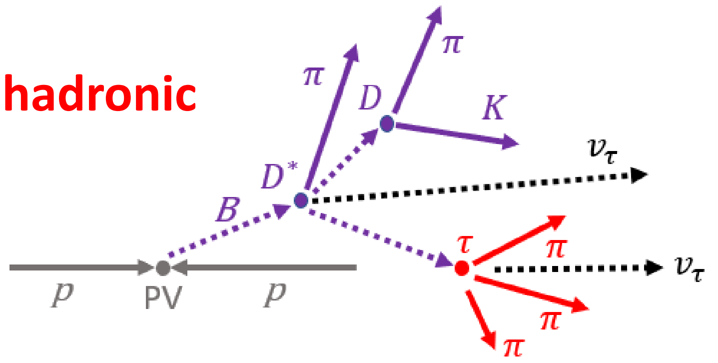
[Link of HFLAV $R(D^{(*)})$ average in 2025]



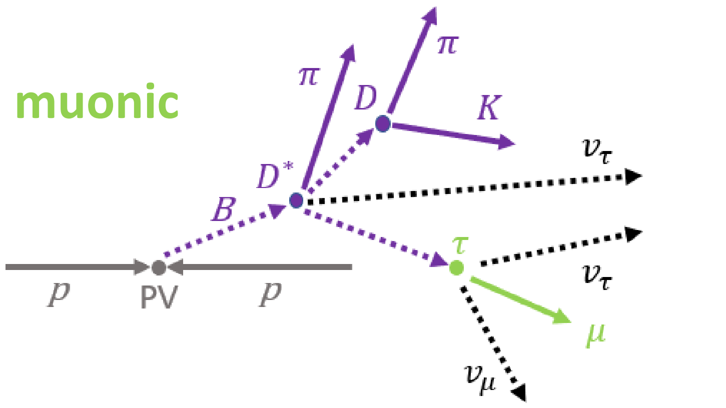
3.8 σ tension in $R(D^{(*)})$ measurement with SM prediction

Currently, we have several $R(D^{(*)-})_{\tau/\mu}$ measurements using **hadronic** and **muonic** τ decays in LHCb.

hadronic

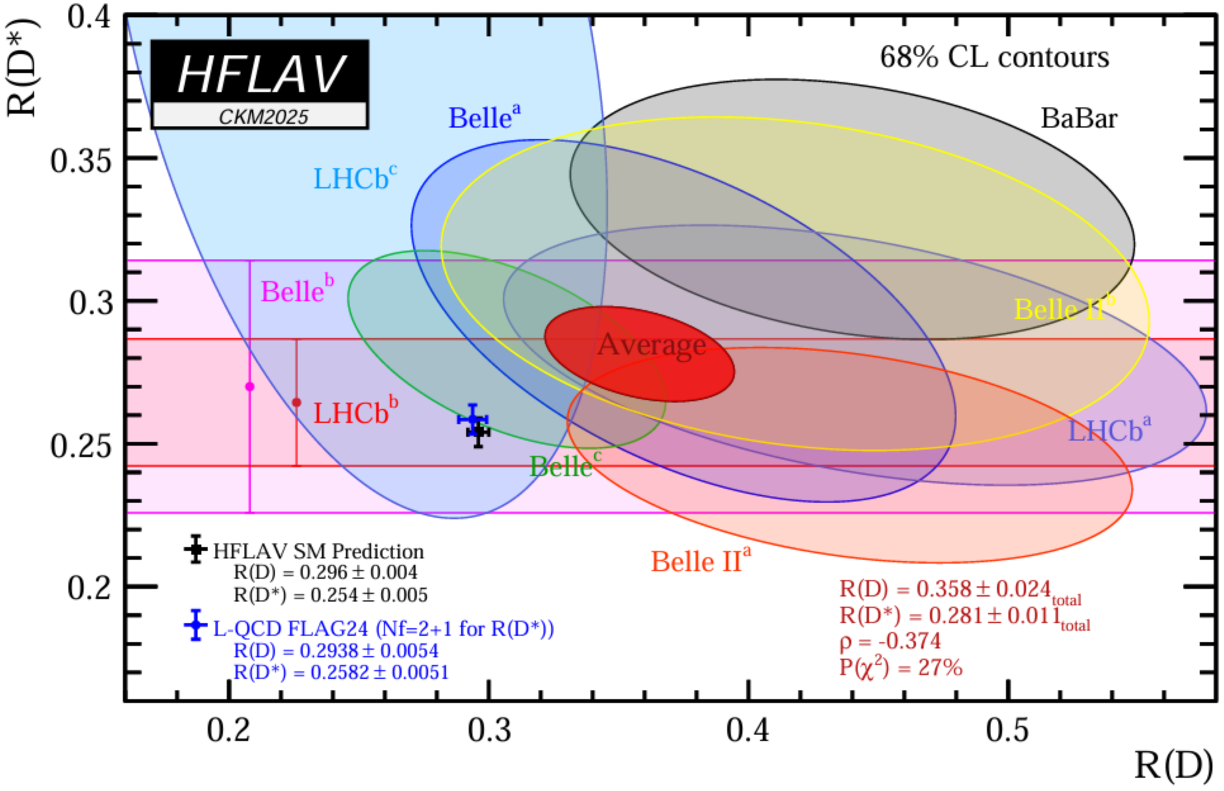


muonic



Experimental and theoretical status for $R(D^{(*)})$

[Link of HFLAV $R(D^{(*)})$ average in 2025]

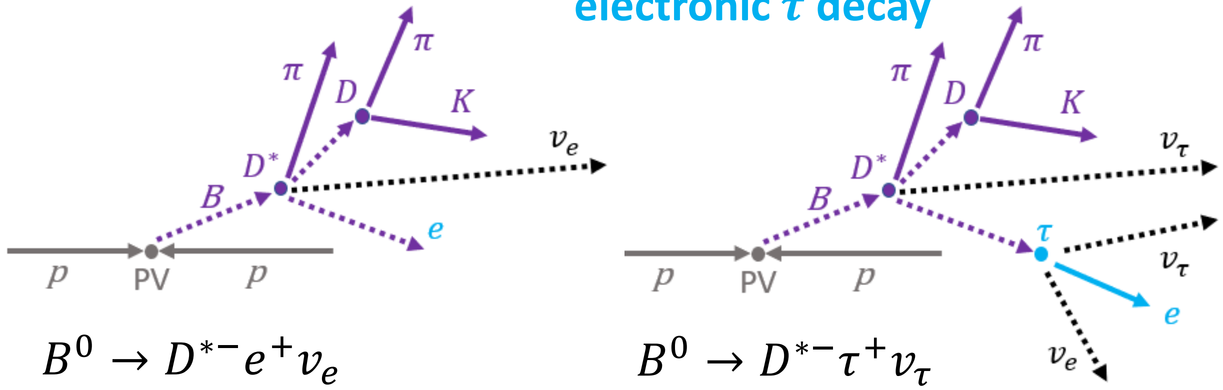


3.8 σ tension in $R(D^{(*)})$ measurement with SM prediction

Currently, we have several $R(D^{(*)-})_{\tau/\mu}$ measurements using **hadronic** and **muonic** τ decays in LHCb.

In our analysis, we study lepton universality by measuring the **electronic** modes

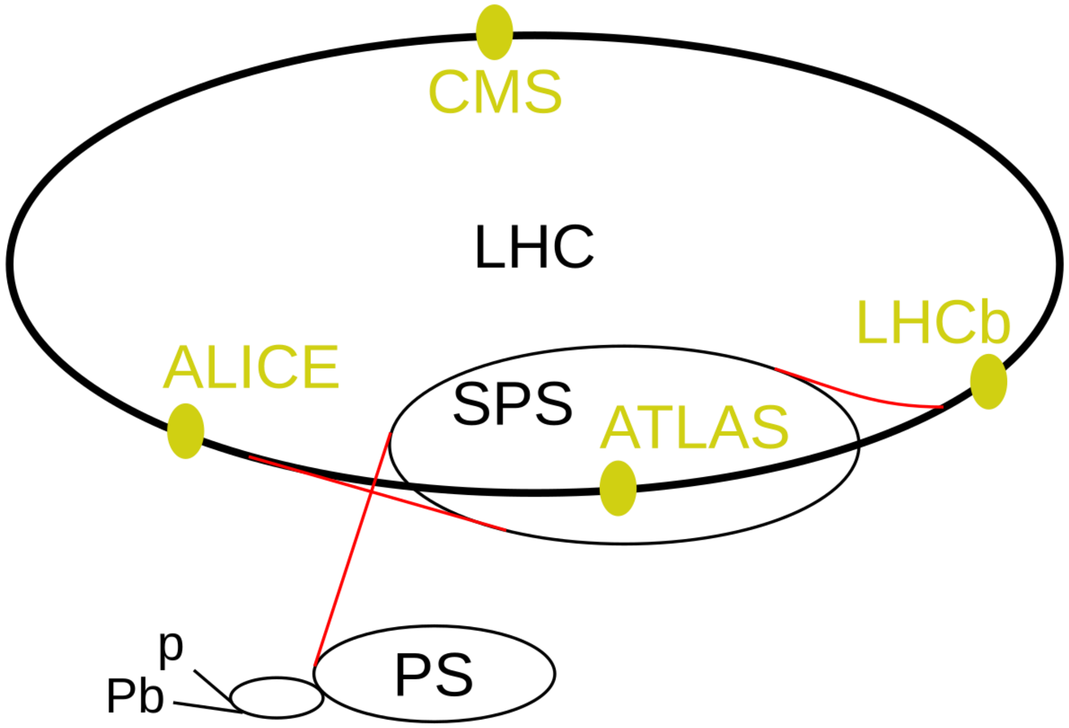
electronic τ decay



It's the first time people test $\tau - e$ universality at LHCb !

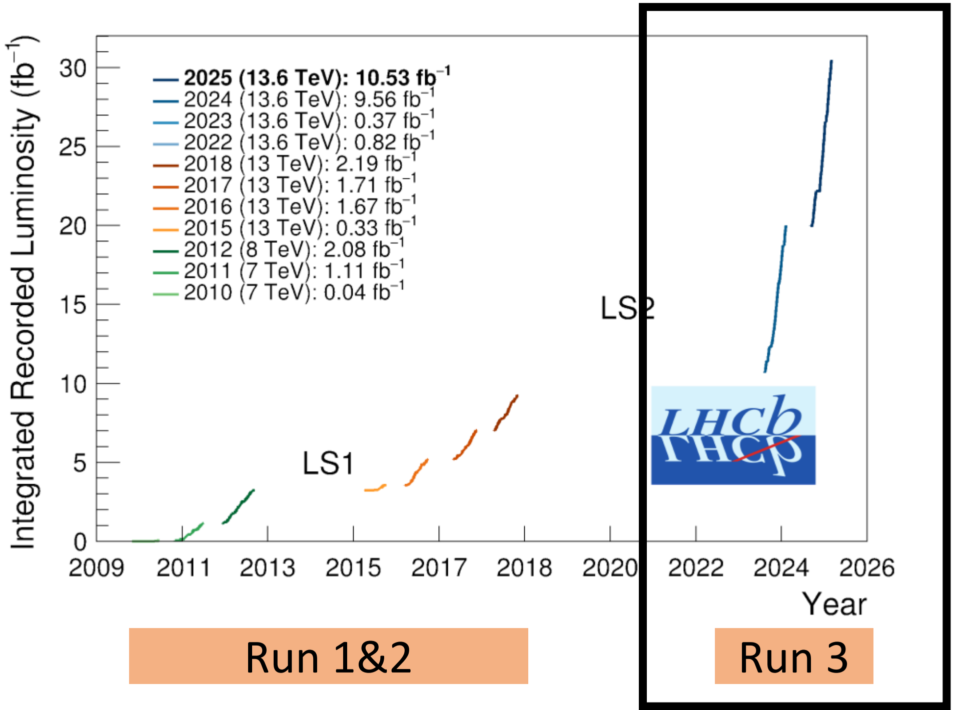
LHCb experiment

- Collect the $pp \to b\bar{b}$ data at a center-of-mass of 13 TeV.



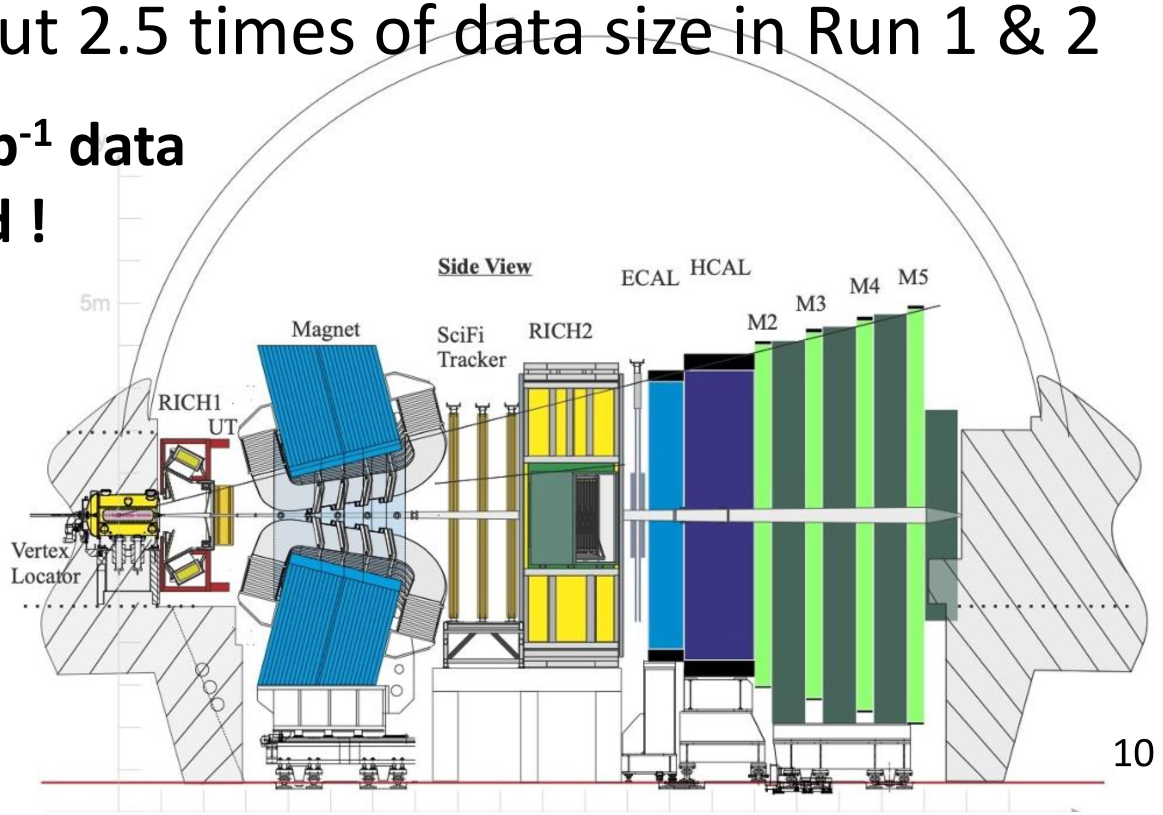
LHCb experiment

- Collect the $pp \to b\bar{b}$ data at a center-of-mass of 13 TeV.
- Single-arm spectrometer covering most of the region where $\sigma(pp \to b\bar{b}X)$ is maximal.
- Aimed to take 25 fb^{-1} of data in Run 3, about 2.5 times of data size in Run 1 & 2



Run 1&2 Run 3

Around 20 fb^{-1} data are collected !



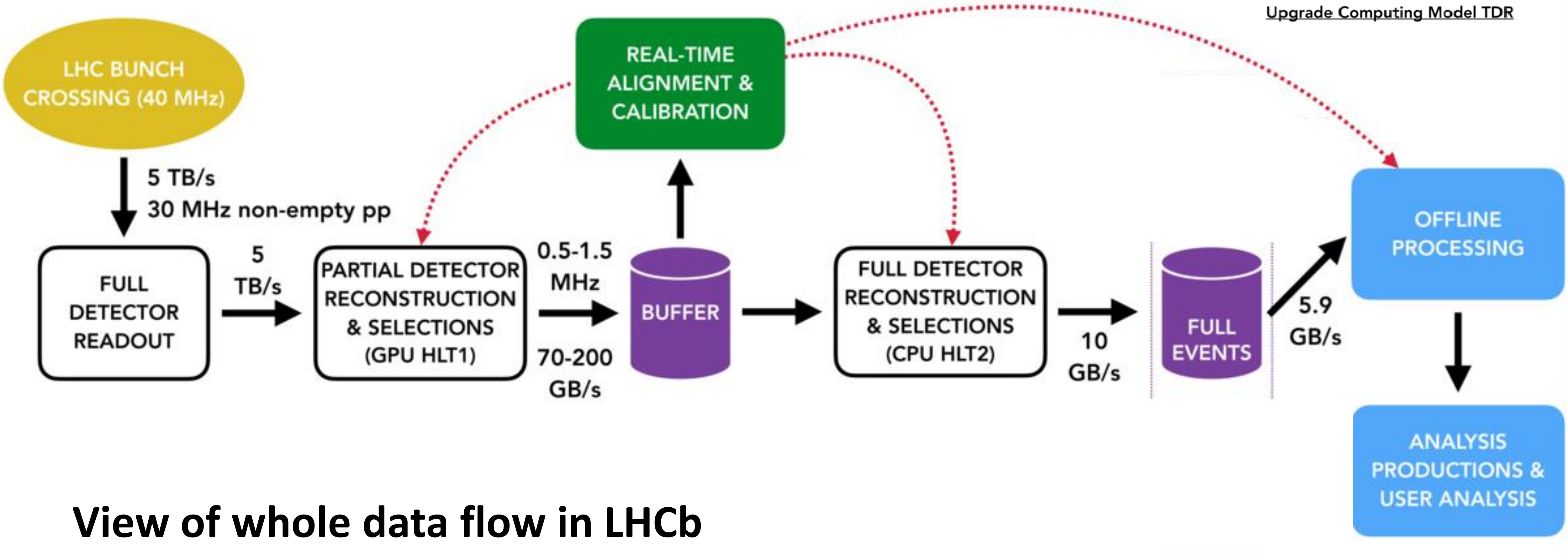
Methodology

Data flow for $B^0 \rightarrow D^{*-} l^+ v$ reconstruction

All numbers related to the dataflow are taken from the LHCb

Upgrade Trigger and Online TDR

Upgrade Computing Model TDR



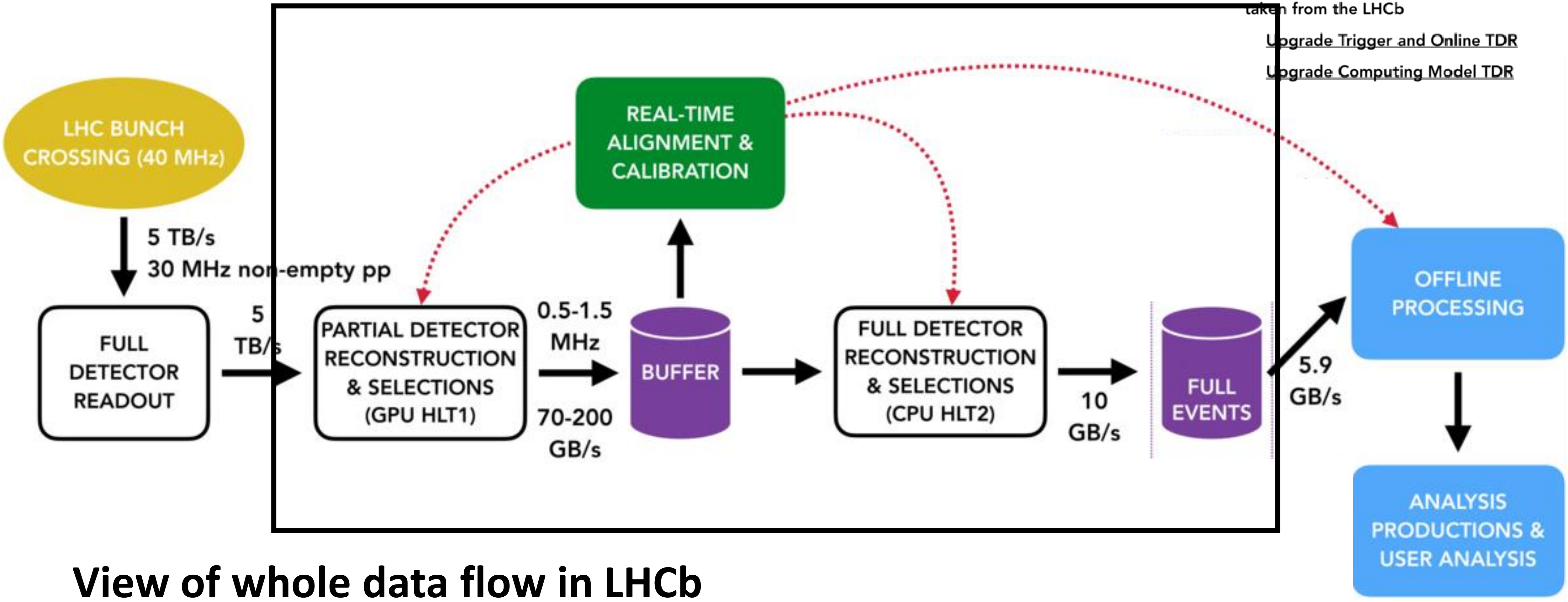
View of whole data flow in LHCb

Data flow for $B^0 \rightarrow D^{*-}l^+v$ reconstruction

All numbers related to the dataflow are taken from the LHCb

Upgrade Trigger and Online TDR

Upgrade Computing Model TDR

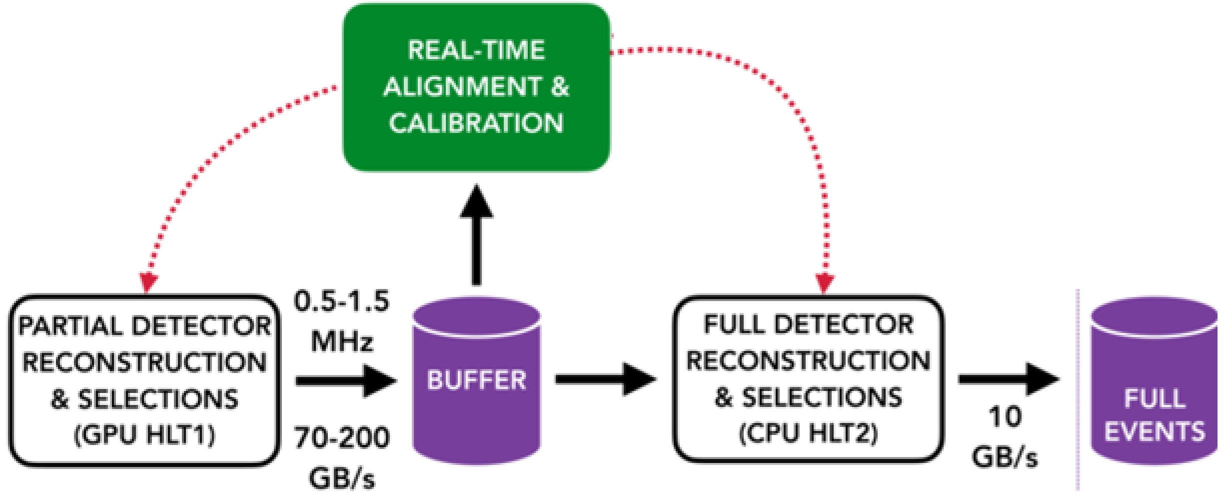


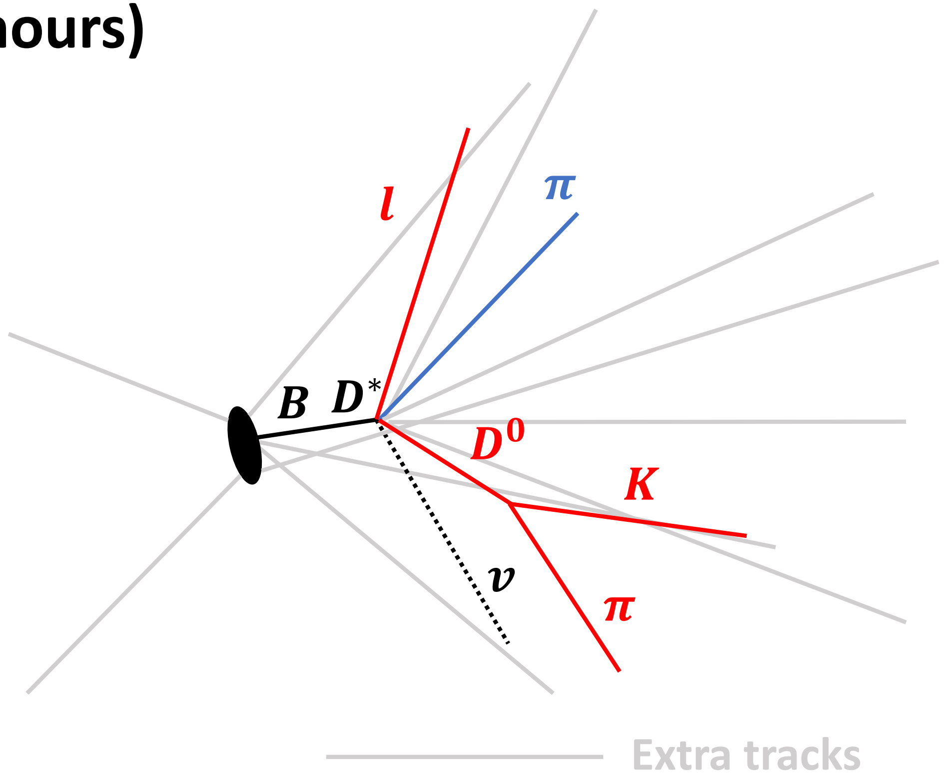
View of whole data flow in LHCb

We start from the online reconstruction

Data flow for $B^0 \rightarrow D^{*-} l^+ v$ reconstruction

During online reconstruction (~ few seconds to hours)





We store the following particles into tape

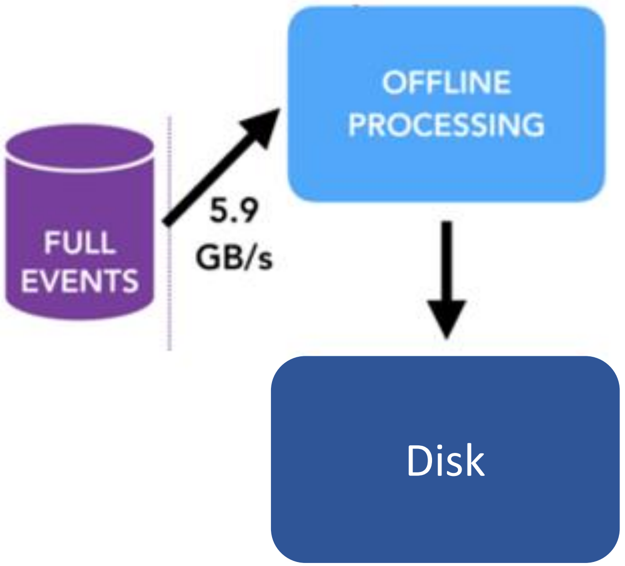
Ground-state particles $D^0 l$ are reconstructed & selected

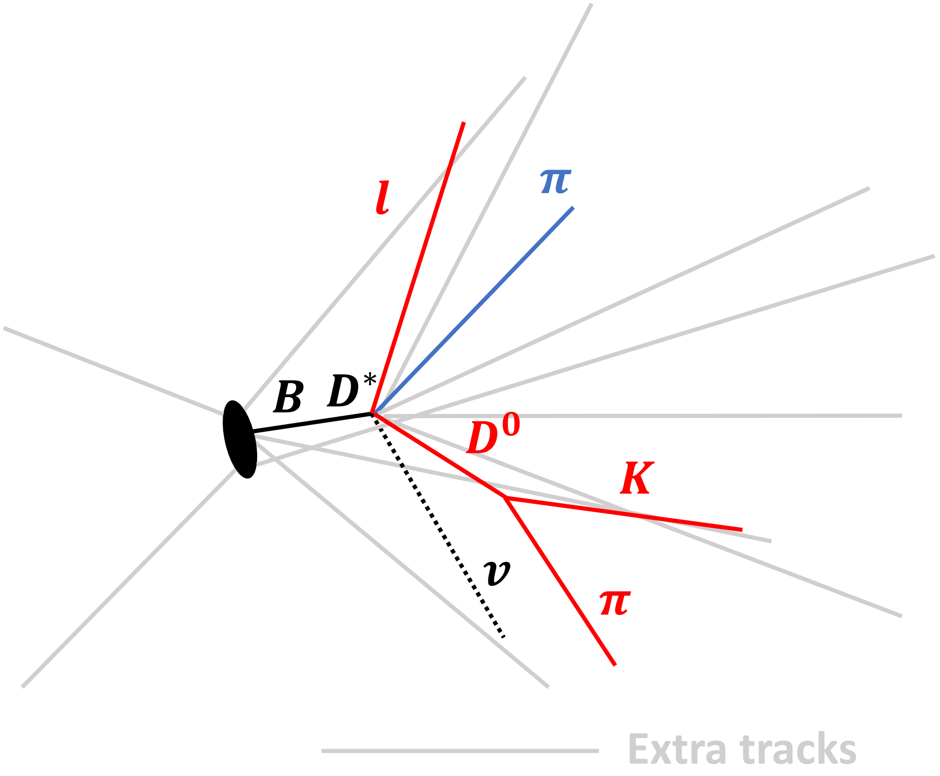
All extra tracks including π from $D^{*-} (\rightarrow D^0 \pi^-)$

Noted : We can't detect neutrino in LHCb

Data flow for $B^0 \to D^{*-} l^+ v$ reconstruction

During offline processing (~ few hours, days)

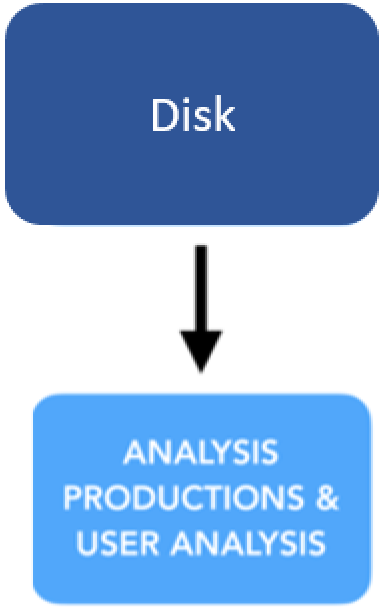


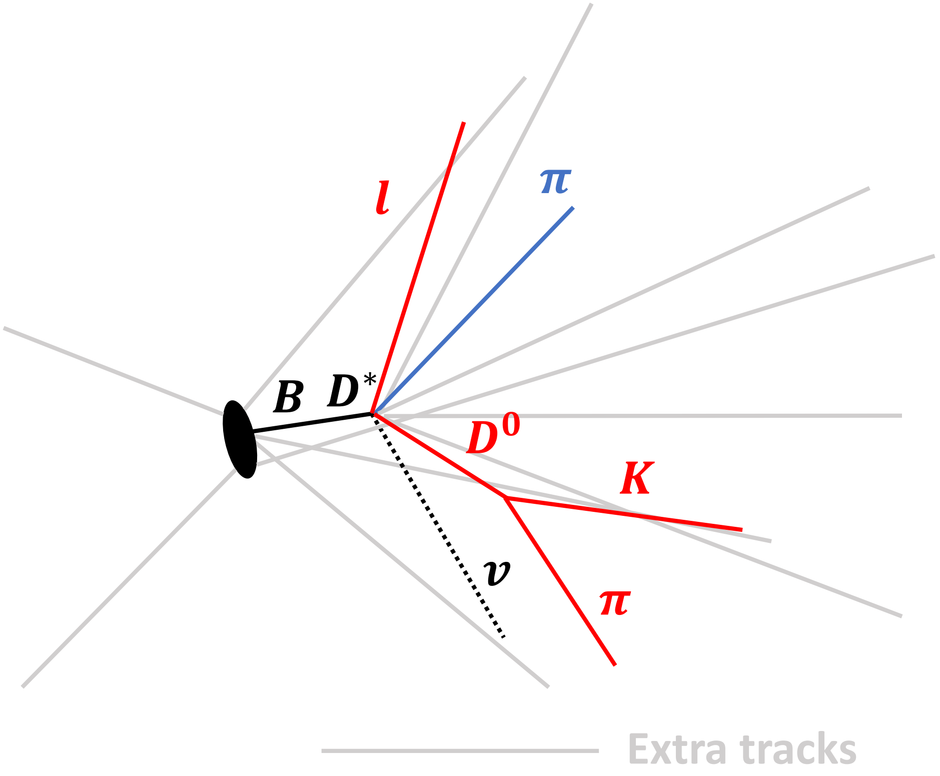


- We pair **ground-state particles** and extra tracks (including π) to get the B candidates
- Select and store the extra tracks based on the MVA score (detailed on p.17) to the disks

Data flow for $B^0 \to D^{*-} l^+ v$ reconstruction

At the tuple making stage (~ few hours, days)

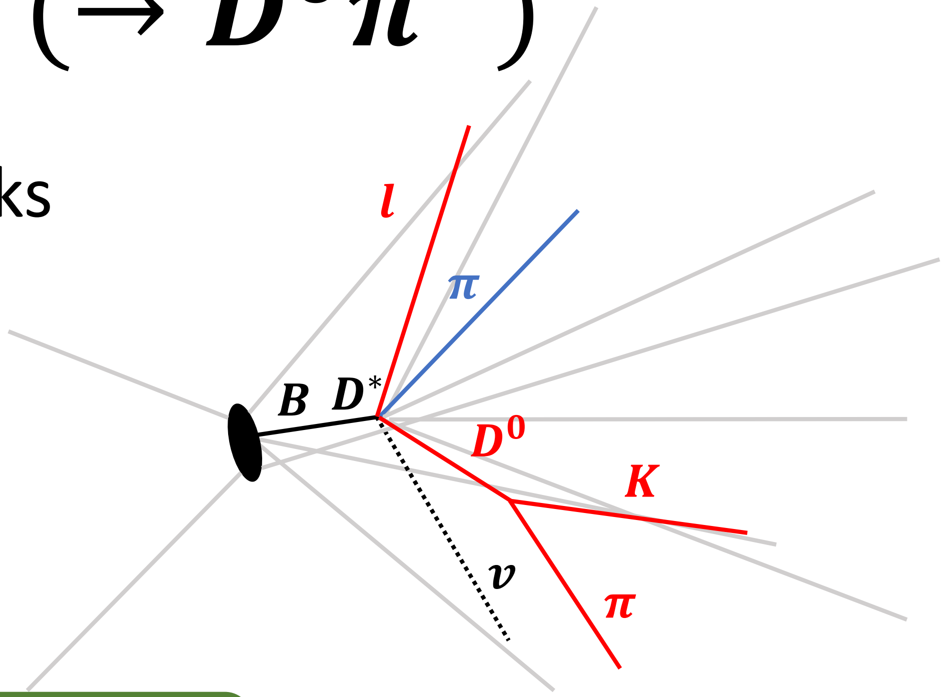




- At the final stage, we reconstruct the B^0 candidates with $\boldsymbol{D^0 l}$ and **extra tracks** to make tuples which includes various physics quantities.
- The B^0 is **signal** candidate if it is reconstructed with $\boldsymbol{D^0 l}$ and $\boldsymbol{\pi}$, otherwise **background**

Finding charged π from $D^{*-}(\rightarrow D^0\pi^-)$

We trained a model using **XGBoost** to identify tracks that likely originate from the B hadron decay.



After the online trigger reconstruction

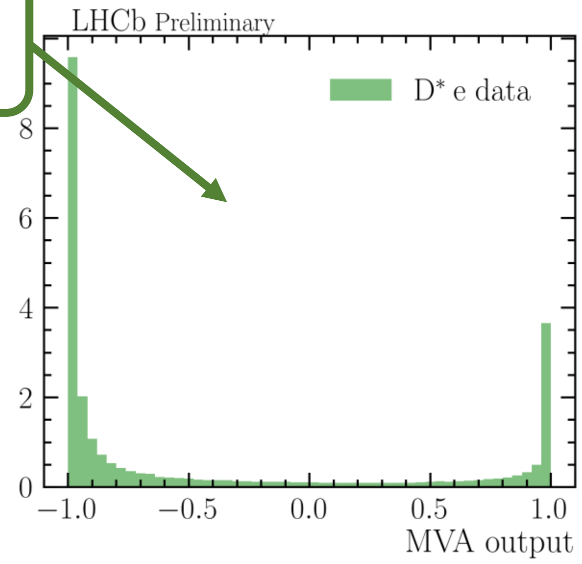
Signal child
Ex : $D^0 l$

Extra pion

Parent particle
Ex : B^0

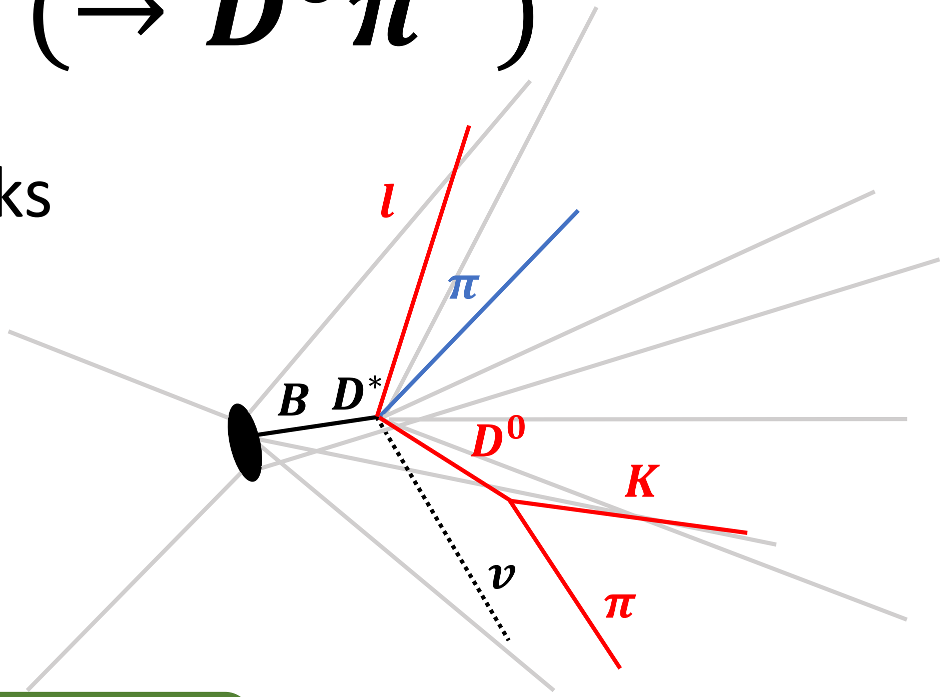
Physics quantities of each parent particle

Acquire the isolation MVA



Finding charged π from $D^{*-}(\rightarrow D^0\pi^-)$

We trained a model using **XGBoost** to identify tracks that likely originate from the B hadron decay.

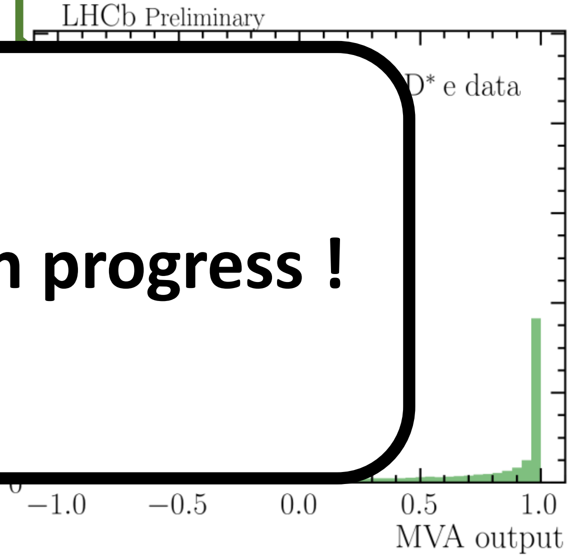


After the online trigger reconstruction

Signal child
Ex

Extra pion

Acquire the



The preparation of publication for this tool is in progress !

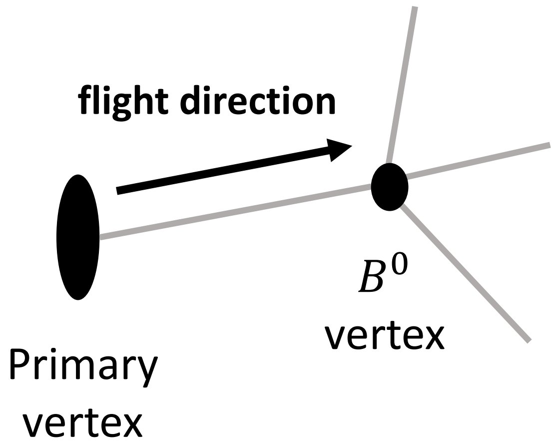
Kinematic reconstruction

Due to the presence of neutrino and we can't detect it in LHCb, we reconstruct B-meson kinematics with the following procedure

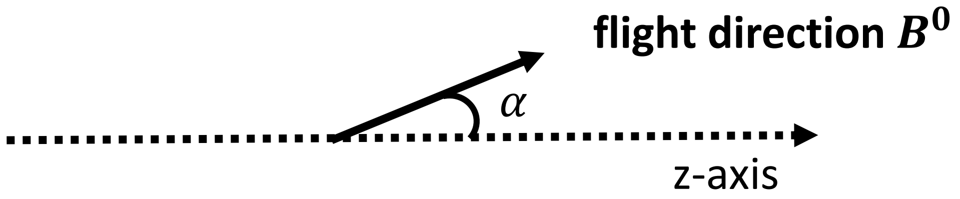
Correct reconstructed $(p_{B^0})_z$ with true m_{B^0}

$$(p_{B^0})_z \approx \frac{m_{B^0,PDG}}{m_{B^0,reco}} \times (p_{B^0,reco})_z.$$

Derive $|\vec{p}_{B^0}|$ using the flight direction of B^0 obtained from the positions of primary vertex and B^0 vertex



$|\vec{p}_{B^0}| = (p_{B^0})_z / \cos\alpha$, where α is angle between flight direction and the z-axis



$R(D^*)_{\tau/e}$ determination

$$R(D^*)_{\tau/e} \equiv \frac{BR(B^0 \to D^{*-}\tau^+\nu_\tau)}{BR(B^0 \to D^{*-}e^+\nu_e)}$$,where $\tau^+ \to e^+\nu_e\bar{\nu_\tau}$

$$= \frac{N(B^0 \to D^{*-}\tau^+\nu_\tau)}{N(B^0 \to D^{*-}e^+\nu_e)} \times \frac{\epsilon(B^0 \to D^{*-}e^+\nu_e)}{\epsilon(B^0 \to D^{*-}\tau^+\nu_\tau)}$$

Extracted from the simulated sample

$R(D^*)_{\tau/e}$ determination

$$R(D^*)_{\tau/e} \equiv \frac{BR(B^0 \to D^{*-}\tau^+\nu_\tau)}{BR(B^0 \to D^{*-}e^+\nu_e)}$$,where $\tau^+ \to e^+\nu_e\bar{\nu_\tau}$

$$= \frac{N(B^0 \to D^{*-}\tau^+\nu_\tau)}{N(B^0 \to D^{*-}e^+\nu_e)} \times \frac{\epsilon(B^0 \to D^{*-}e^+\nu_e)}{\epsilon(B^0 \to D^{*-}\tau^+\nu_\tau)}$$

Extracted from the simulated sample

Extracted from the fitter

We use the strategy used in Run 1 & 2 $R(D^*)_{\mu/e}$ and $R(D^*)_{\tau/\mu}$ analysis to extract $R(D^*)_{\tau/e}$ value

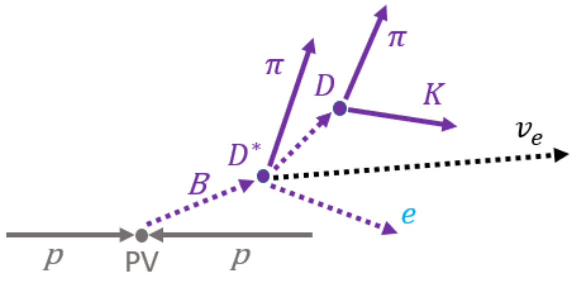
- Use the same fitting variables (introduced on the next page)
- Include statistical uncertainties in the templates of fitting variables for both signal and background with Barlow and Beeston method
- Perform template fit on data to extract $R(D^*)_{\tau/e}$

Fitting variables

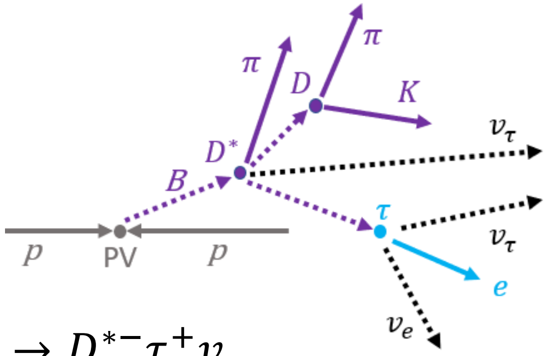
The fitting variables are defined as

- $M^2_{missing} = (p_{B^0} - p_{D^*} - p_e)^2$
- $q^2 = (p_{B^0} - p_{D^*})^2$
- E_e^*: Electron energy in B rest frame,

which are derived from modified $\vec{p}_{B^0}$

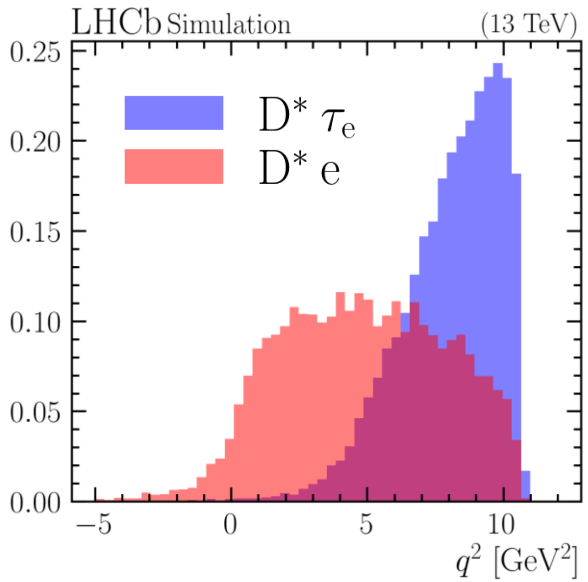


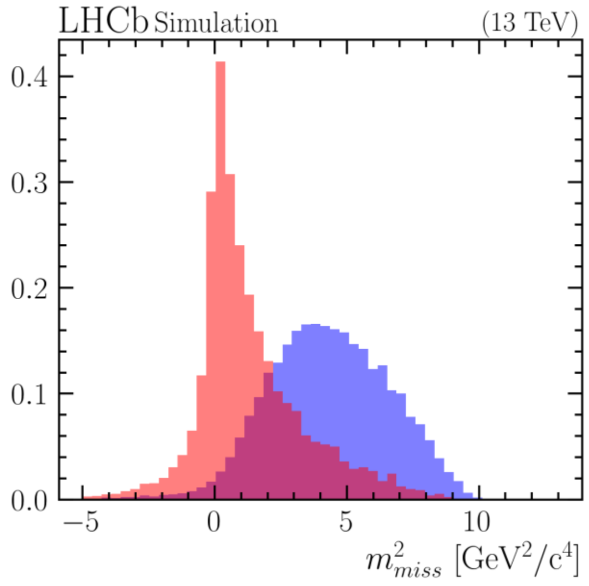
$B^0 \rightarrow D^{*-} e^+ \nu_e$

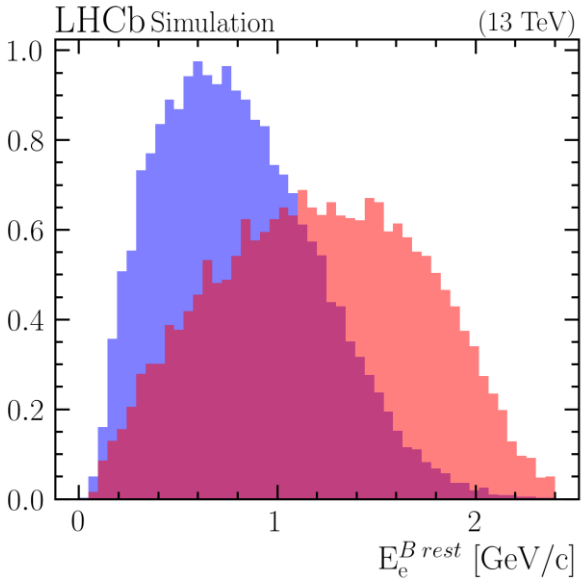


$B^0 \rightarrow D^{*-} \tau^+ \nu_\tau$

Noted : The electron is softer







***$D^*\tau$* and *D^*e* sample are well separated !**

Background study

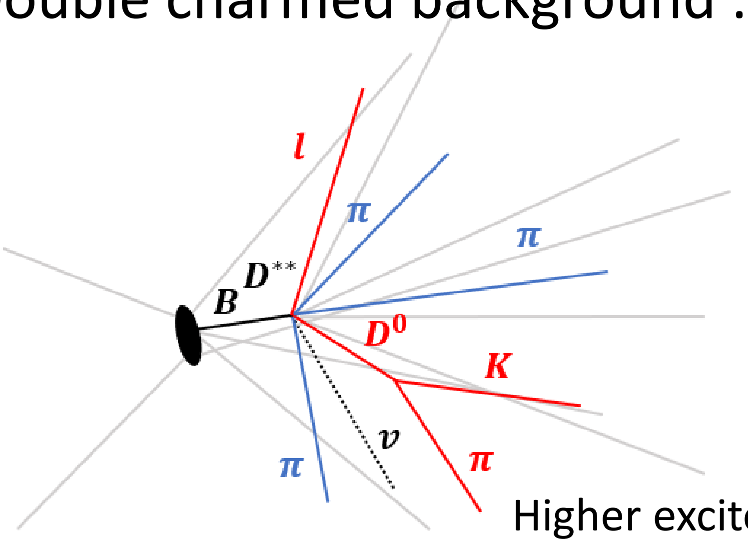
Source of backgrounds in this analysis

Combinatorial background

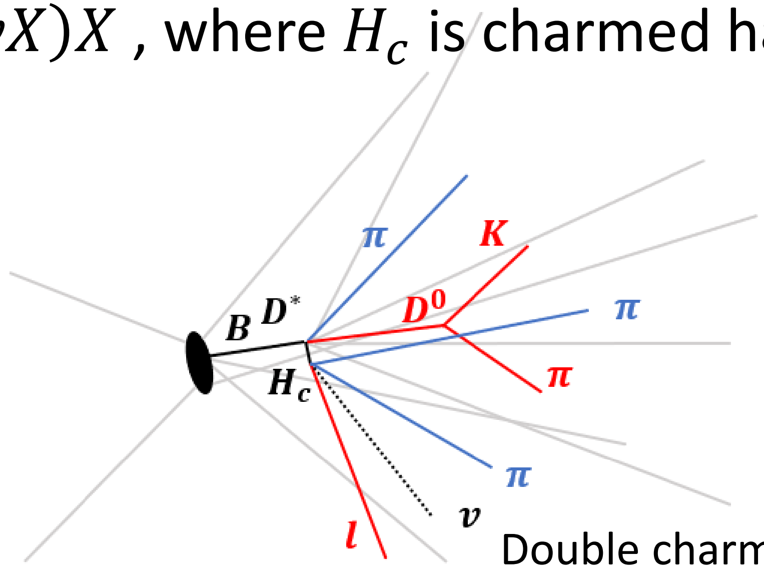
- Wrong sign combination
- e misID background : mis-identified $K\ or\ \pi \rightarrow e$ background

Background from other decay

- Higher excited background : $\bar{B} \rightarrow D^{**}(\rightarrow D^{*}n\pi)l\nu, (n \geq 1)$, where D^{**} is higher excited charmed state
- Double charmed background : $\bar{B} \rightarrow D^{*}H_c(\rightarrow l\nu X)X$, where H_c is charmed hadron



Higher excited background



Double charmed background

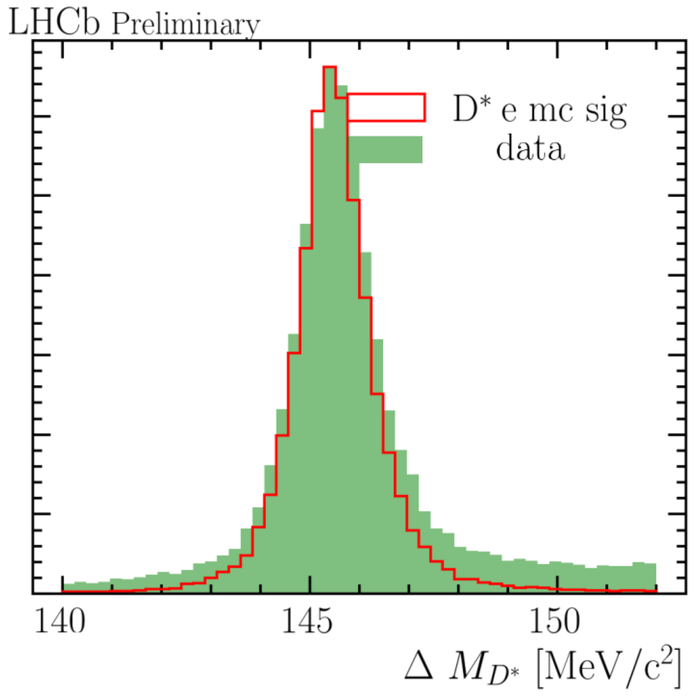
Combinatorial background suppression

The reconstruction of $D^{*-}(\rightarrow D^0(\rightarrow K^-\pi^+)\pi^-)$ is necessary.

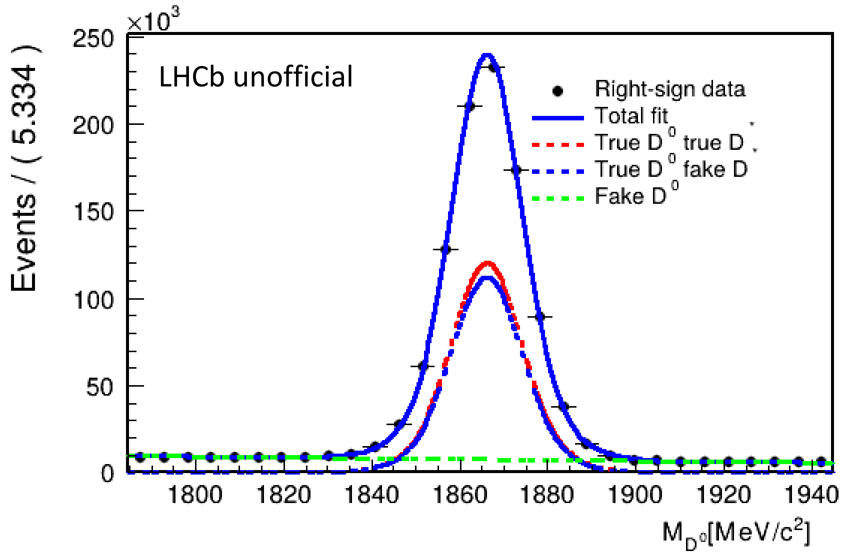
After the trigger and offline selections, there are still lots of background coming from wrong D^0- π combination and fake D^0.

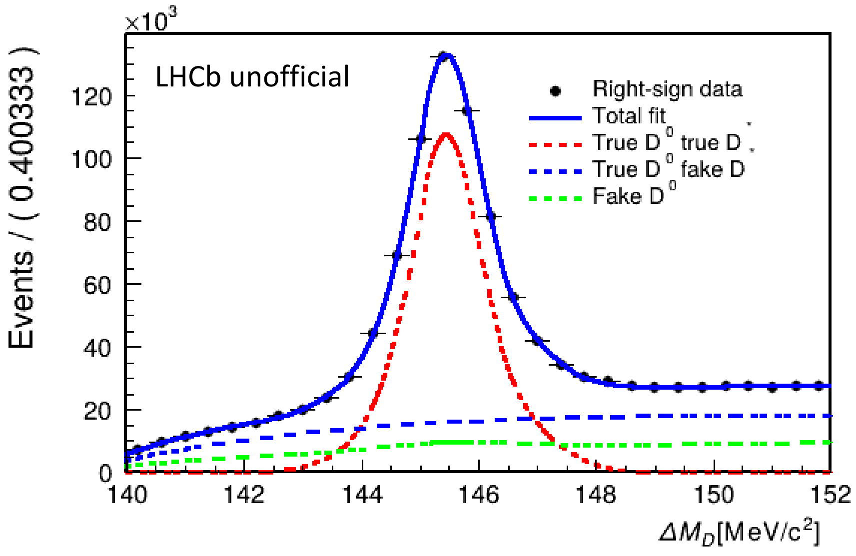
To project out their contribution, $_{s}\boldsymbol{Weight}$ **approach is used**

by performing 2D $M_{D^0} \times \Delta M_{D^*}$ fit on 3 different components



ΔM_{D^*} distribution in data with offline selection, where $\Delta M_{D^*} \equiv |M_{D^*} - M_{D^0}|$



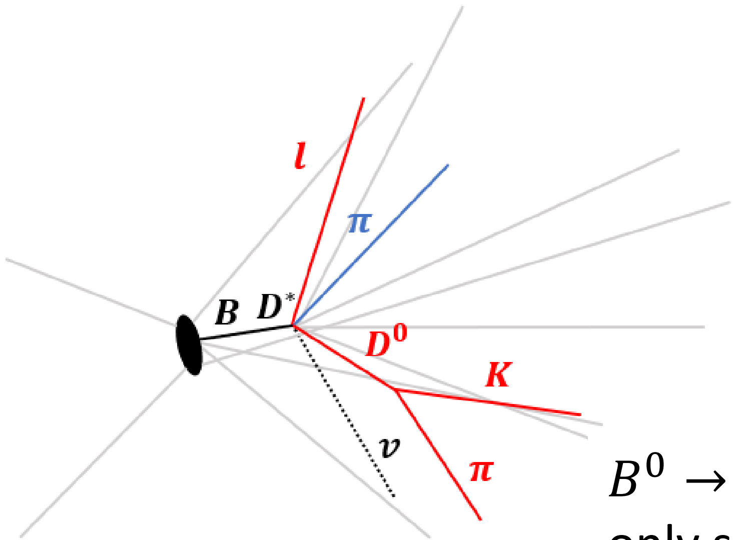


D^{**} background suppression

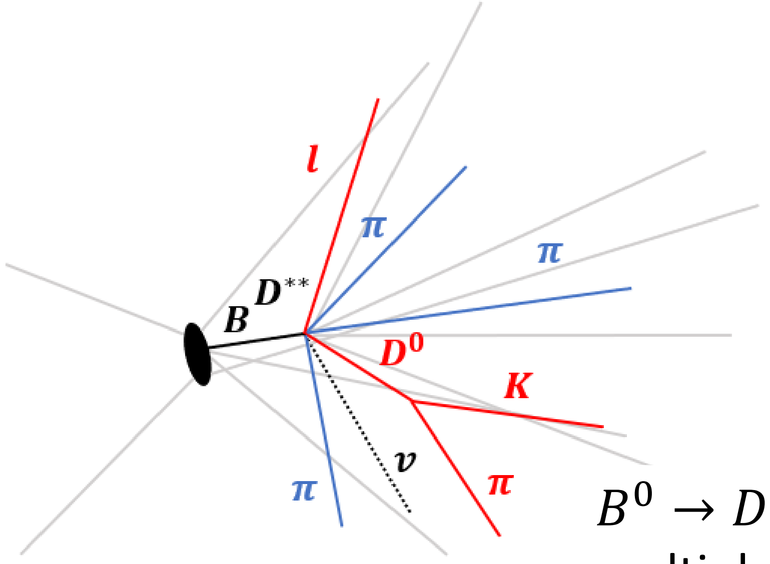
Dominated semi-leptonic background in Run 1 & 2 $R(D^*)_{\tau/\mu}$ analysis :

$\bar{B} \to D^{**}(\to D^* n\pi) l v, (n \geq 1)$ **and** $\bar{B} \to D^* H_c(\to l v X) X$

The charged isolation MVA is used to reject this background. We remove any events with multiple tracks having high MVA score.



$B^0 \to D^{*-} l^+ v$ signal : only single high MVA π

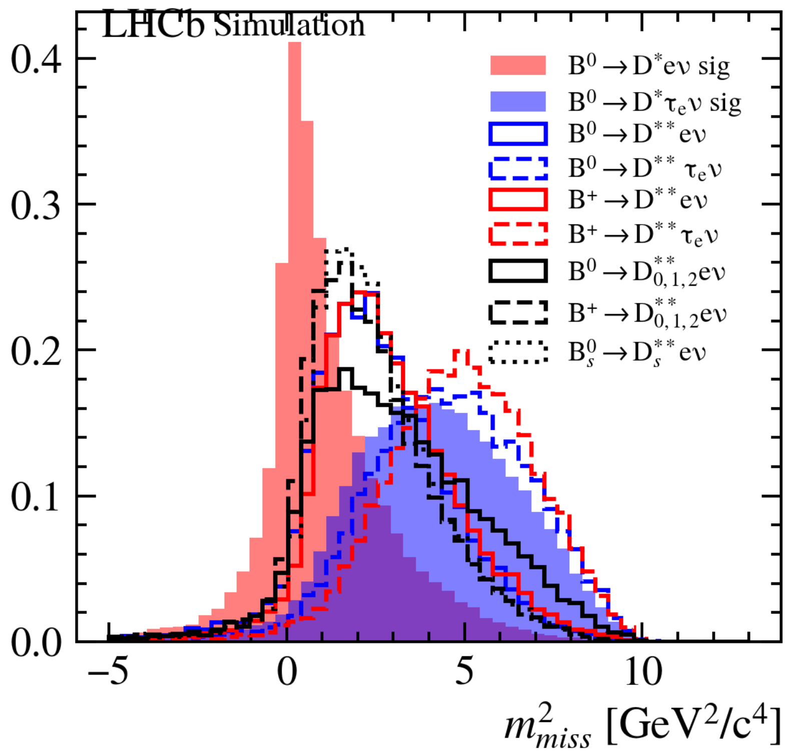


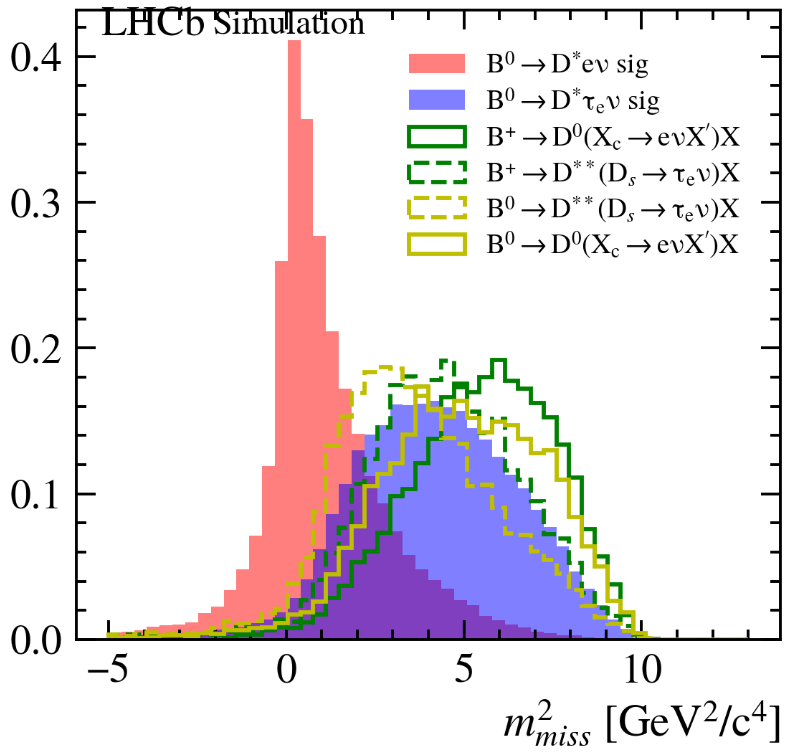
$B^0 \to D^{**-}(\to D^* n\pi) l^+ v$ bkg : multiple high MVA $\pi's$

After the background suppression, the retention rate are around **(10-50) %** across different background channels in the simulation samples.

Kinematic distribution in background sample

$M^2_{miss} = (p_{B^0} - p_{D^*} - p_e)^2$ distributions in different background and signal channels





We can use the difference in **shapes** and **correlations** between different fitting variables to distinguish them !

In this analysis we model the templates of $B^0 \to D^{*-} e(\tau)^+ \nu$ signal and D^{**} background with **simulated sample**

PhD work: completed and ongoing

Completed work :

- Prepared scripts for candidates reconstruction and tuple making.
- Candidate selections study.
- Isolation MVA : model training and data validation.
- Combinatorial and D^{**} background suppression.

Works still ongoing :

- Complete efficiency study
- Data-driven corrections will be applied on simulated sample
- Finalise fitter study with all background components
- Write thesis

Outlook

After my PhD finished, this analysis is still ongoing

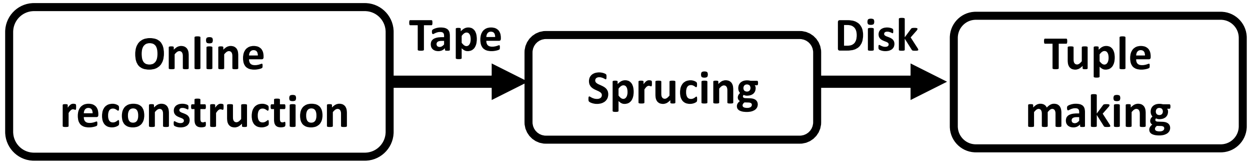
- This measurement is planned to use full Run 3 data (including 2026 data)
- The estimation of systematic error will be studied
- The consistency study for data in different years will be done

After all finished

- This will be the first $R(D^*)$ measurement at LHCb with electron-only decays
- Provide extra input for world average $R(D^*)$ value

Backup

Data flow for $B^0 \to D^{*-} l^+ v$ reconstruction



Online reconstruction (HLT1 – HLT2)

- Select and store ground-state particles - $D^0 l$
- Store all the extra tracks (e.g. $\pi^{\pm}$ from $D^{*-}(\to D^0\pi^-)$).

Sprucing

- Pair ground-state particles with extra tracks, select and store the extra tracks with MVA output (detailed on next page) to meet the bandwidth requirement on disk.

Tuple making

- We pair the ground-state particles stored in HLT2 and extra tracks stored in sprucing stage to make tuples which includes various physics quantities

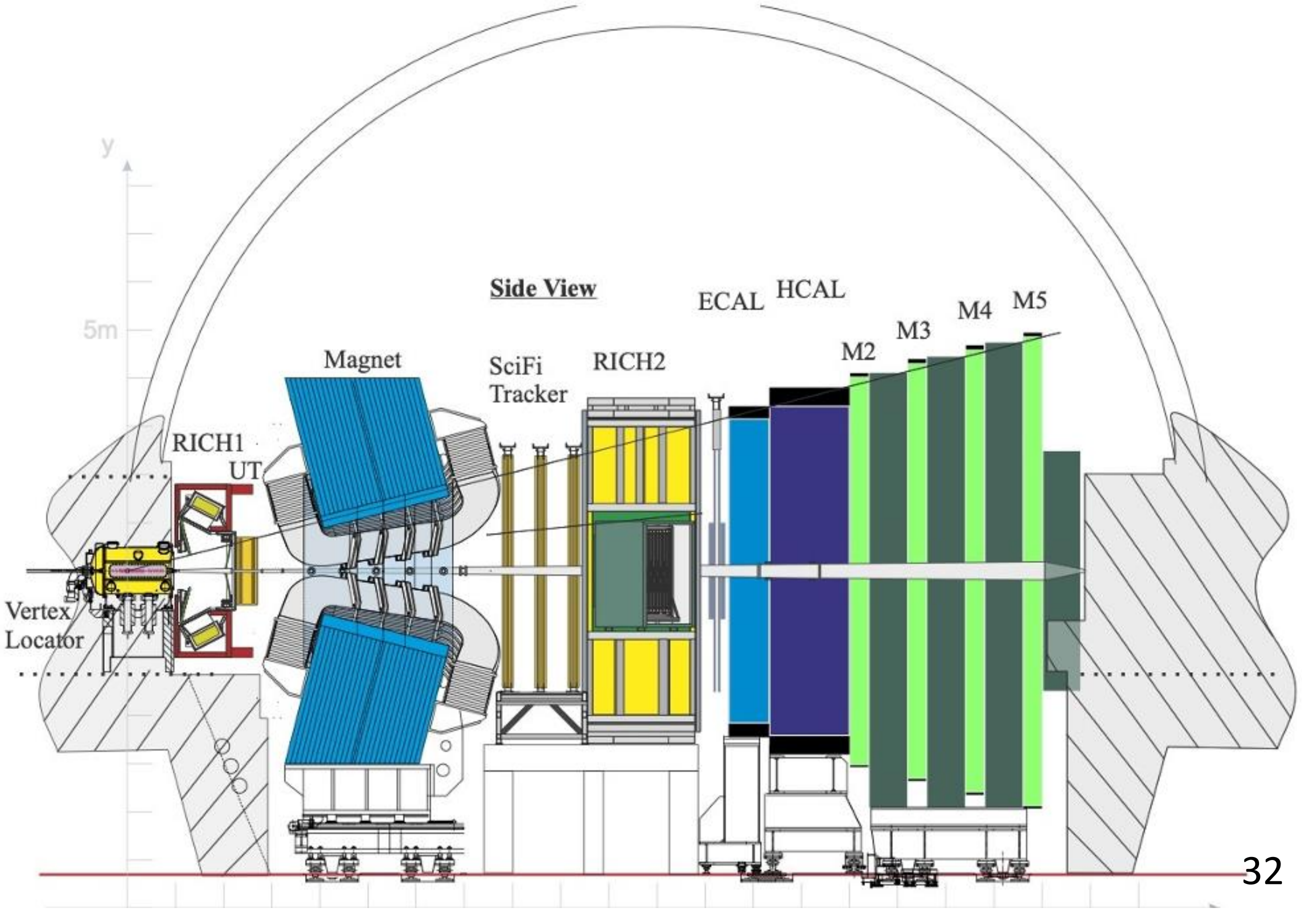
LHCb experiment

- Collect the $pp \to b\bar{b}$ data at a central-of-mass of 13 TeV
- Single-arm spectrometer covering most of the region where $\sigma(pp \to b\bar{b}X)$ is maximal

Detector performance

Tracking :
$\delta p/p \sim (0.5 - 0.8)\%$ at 3 – 100 $\mathrm{GeV/c}$
LHCb-PUB-2024-009

Particle identification:
e ID $\in 95\ \%,\ 1\ \%\ \pi \to e$
K ID $\in 90\ \%,\ 3\ \%\ \pi \to K$
PoS LHCP2024 (2025) 149



Candidate selections in Hlt2

In Hlt2, $D^0(\to K\pi)$ and e candidates are selected and reconstructed as B^-.

Hlt2SLB_BuToD0ENu_D0ToKPi line is used for signal candidate selection.

Table 2: Selections applied in Hlt2 line.

B^+ selections	
Mass (MeV/c^2)	$1500 < M < 10000$
χ^2_{vertex}/ndf	< 9
χ^2_{FD}	> 20
DIRA	> 0.999
DOCA (mm)	< 0.3
$z^{D^0}_{end\,vtx} - z^{B^{*+}}_{end\,vtx}$ (mm)	$>$ -5
D^0 selections	
Mass (MeV/c^2)	$1784.84 < M < 1944.84$
χ^2_{DOCA}	< 5
χ^2_{FD}	> 25
χ^2_{vertex}/ndf	< 6
DIRA	> 0.99

$K^\pm, \pi^\pm$ selections	
p (GeV/c)	> 15
p_T (MeV/c)	> 600
PID K for $K(\pi)$	> 0 (< 2)
$min.\ \chi^2_{IP}$	> 10
$e^\pm$ selections	
p (GeV/c)	> 3
p_T (MeV/c)	> 300
PID e	> 0
$min.\ \chi^2_{IP}$	> 9

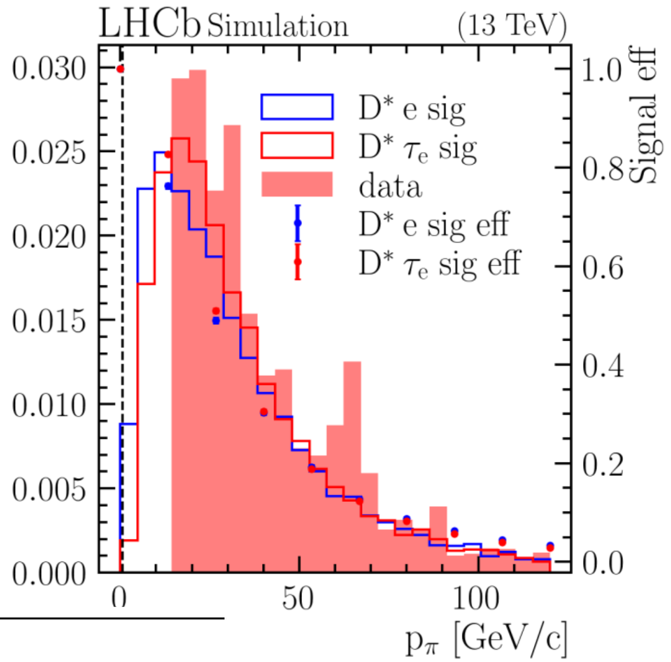
These filtered B^- ($\to D^0e^-$) candidates and all of the extra $\pi's$ are stored in this stage

Re-tuning of Hlt2 line

In October, we re-tuned the trigger lines (!3972) to

Fix the extreme tight p cuts on D^0 children (bug)

Reduce the bandwidth ~80%



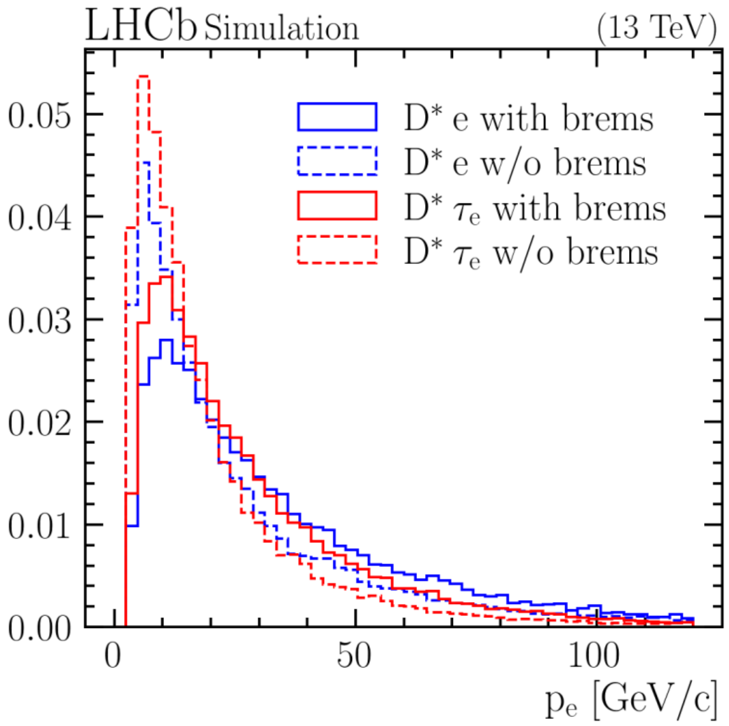
B^+ selections	
Mass (MeV/c^2)	$1500 < M < 10000$
χ^2_{vertex}/ndf	< 9
χ^2_{FD}	> 20
DIRA	> 0.9995
DOCA (mm)	< 0.3
$z^{D^0}_{end\,vtx} - z^{B^{*+}}_{end\,vtx}$ (mm)	> -5
D^0 selections	
Mass (MeV/c^2)	$1784.84 < M < 1944.84$
p (GeV/c)	> 15
χ^2_{DOCA}	< 5
χ^2_{FD}	> 25
χ^2_{vertex}/ndf	< 5
DIRA	> 0.9998

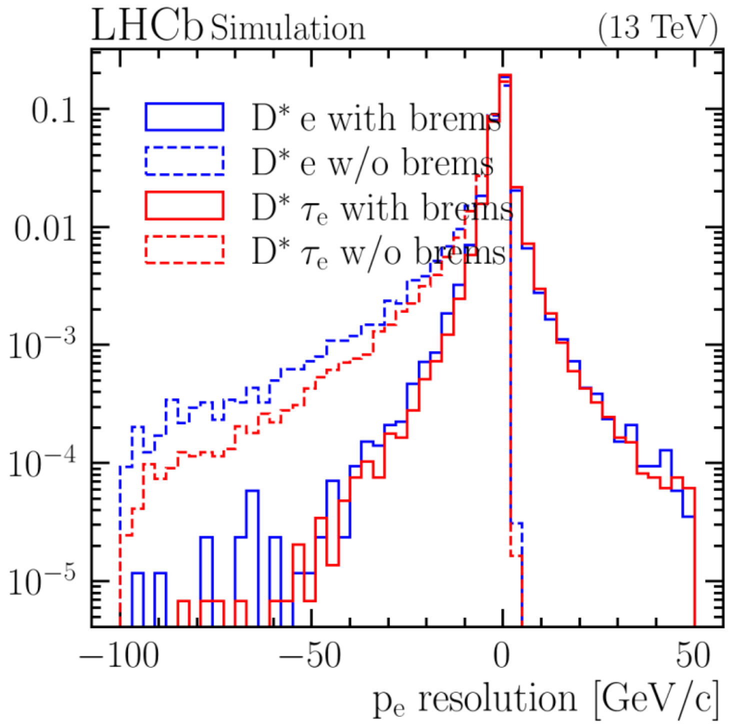
$K^\pm, \pi^\pm$ selections	
p (GeV/c)	> 5
p_T (MeV/c)	> 300
PID K for $K(\pi)$	> 4 (< 2)
$min.\ \chi^2_{IP}$	> 45
$GhostProb$	< 0.35
$e^\pm$ selections	
p (GeV/c)	> 3
p_T (MeV/c)	> 300
Pseudorapidity	$2.2 < \eta < 4.2$
PID e	> 2
$min.\ \chi^2_{IP}$	> 15

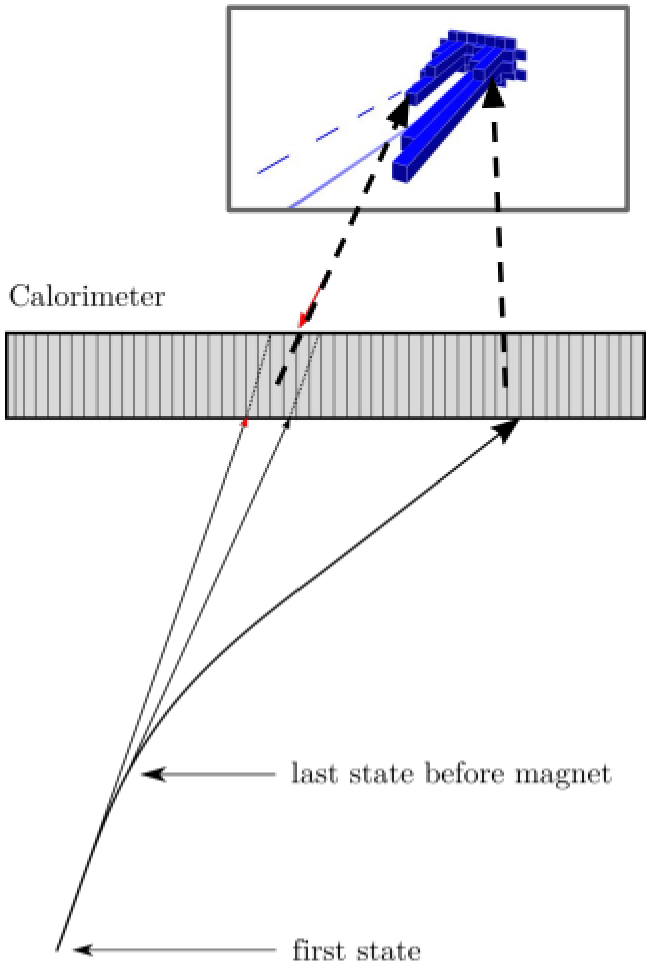
Decay	Signal efficiency [%] Before	After
$B \to D^* e$	31	25
$B \to D^* \tau$	33	32

Bremstralung correction

Due to the presence of electron in final states, we need to consider the correction to kinematics from the recovered photons emitted due to bremsstrahlung





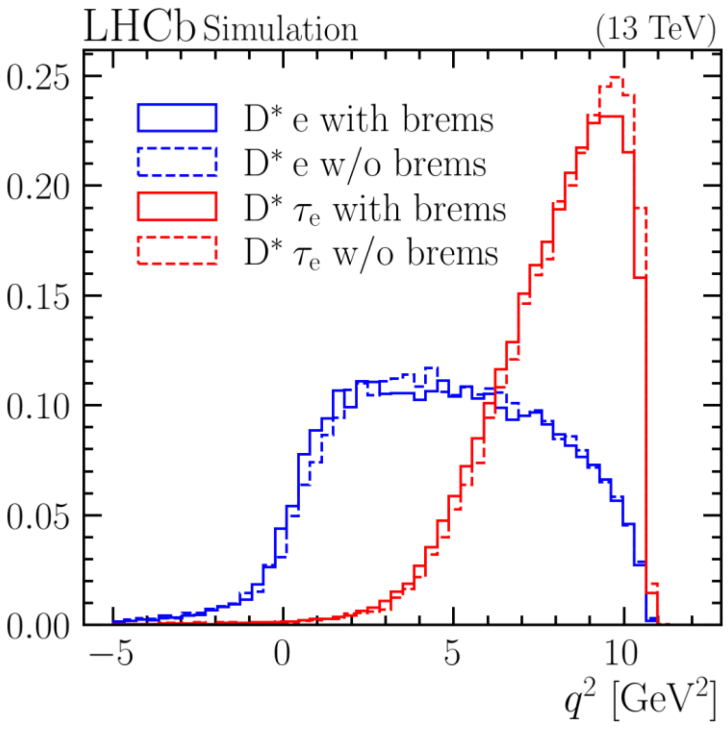


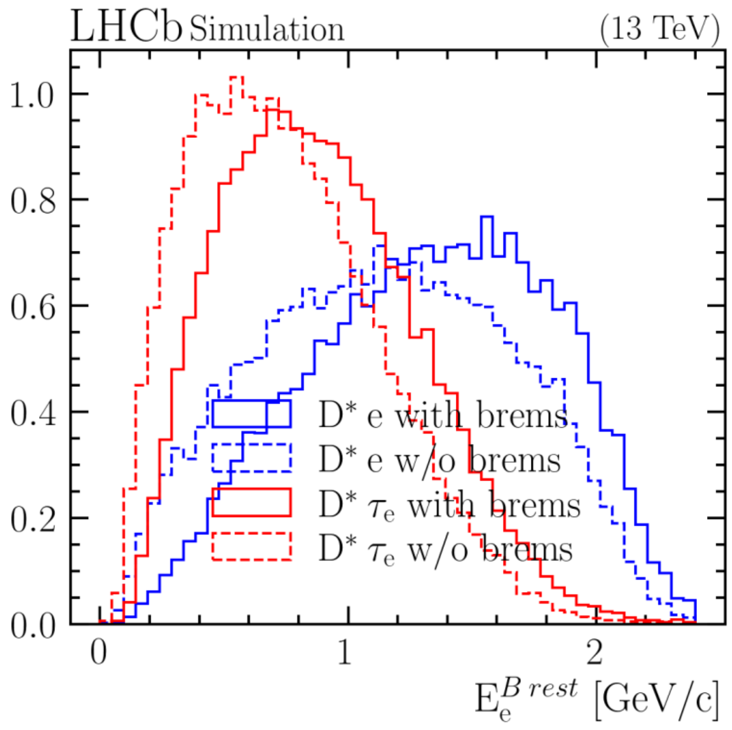
From Maarten's talk

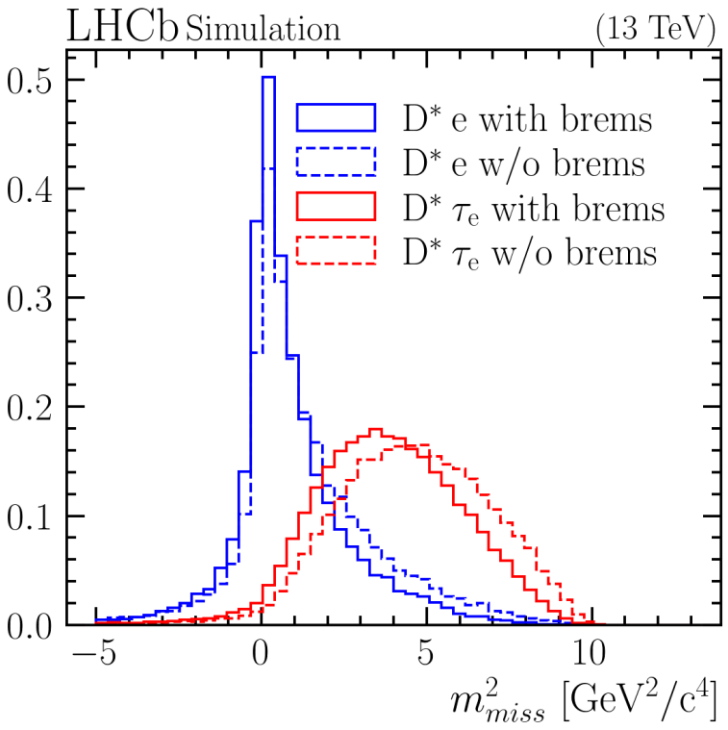
Resolution : $p^{Rec} - p^{Truth}$

Brems correction in fitting variables

Following the Run 2 $R(D^*)_{e/\mu}$ analysis, due to the difference in the shapes with and w/o brems correction. We will separate the data into two categories (with and w/o brems).
These datasets will then be fitted simultaneously.



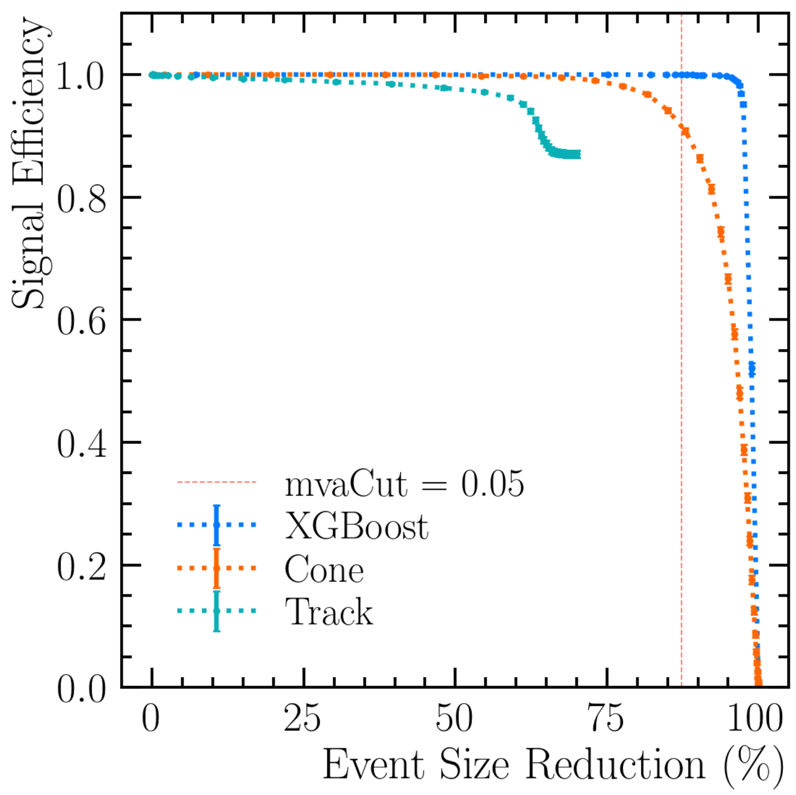




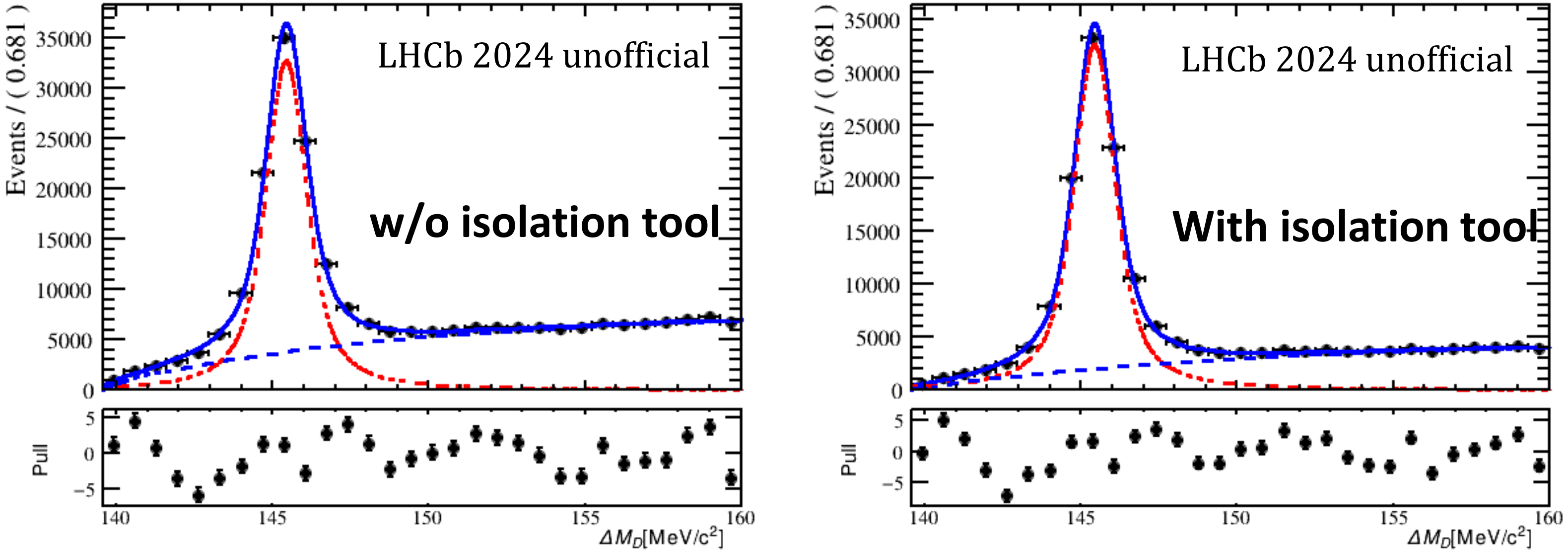
Charge isolation mva

We train a model using **XGBoost** to identify tracks that likely originate from the B hadron decay. Details can be found in [Talk at A&S week]

Training features :
Geometric or kinematic properties of the extra track either by itself or relative to the signal candidate
Samples :
MC of 19 semileptonic decay channels with μ or e in the final state



ROC curve of isolation mva and the legacy methods



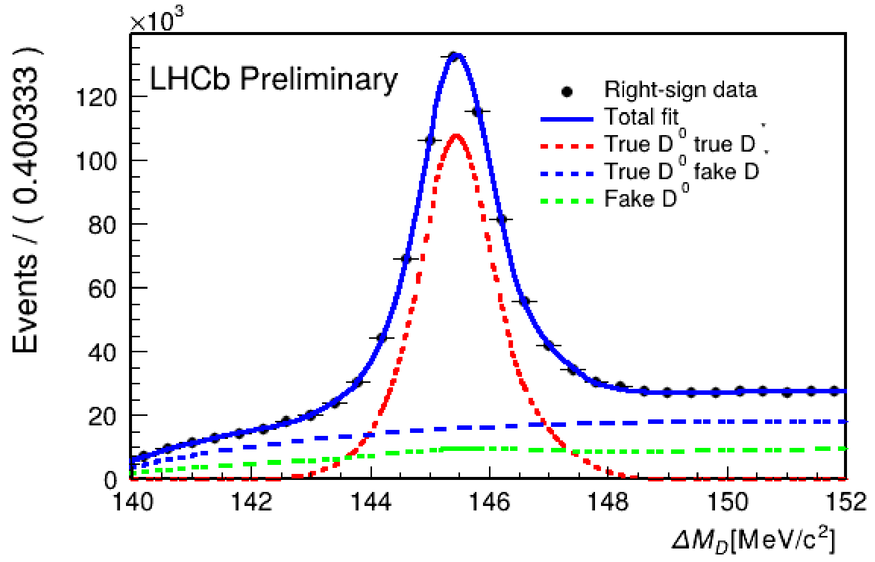
Fit plots for $M_{D^*} - M_{D^0}$ distribution in data

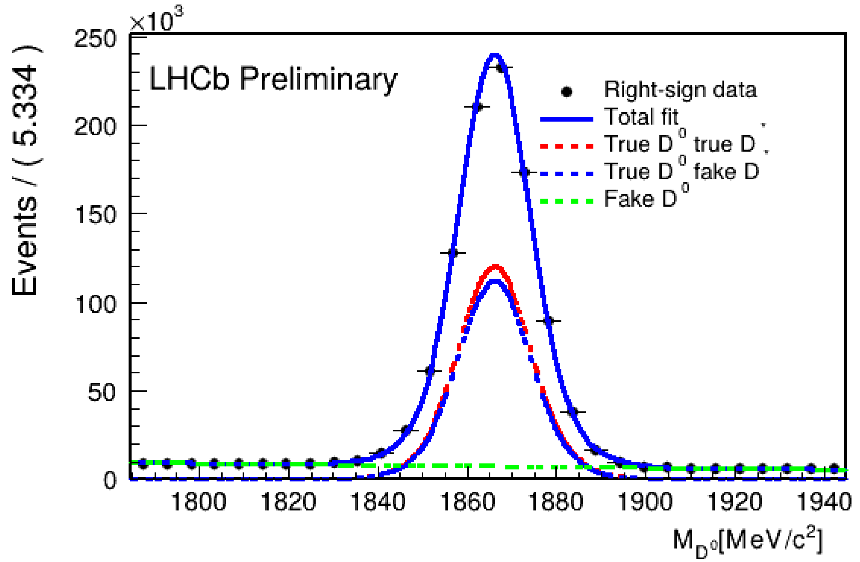
Remove ~ 40% background and only lose ~ 0.4% signal !

Fitting strategy and results

Parameters of

- M_{D^0} distributions are floated
- ΔM_{D^*} distributions of wrong π_{extra} background are Gaussian constraint
- ΔM_{D^*} distributions of fake D^0 and signal are fixed if the extracted parameters from control data have small uncertainties, otherwise use Gaussian constraints





Vertex isolation efficiencies

We test the performance of
vertex isolation (detailed in backup)
and offline selections (see below) across the different signal and background samples.

The vertex isolation and offline selections works well in most of the background samples !

Offline selections

$140 < M_{D^*} - M_{D^0} < 152\ (GeV/c^2)$

Electron ID > 2.0

$2.2 < \eta_e < 4.2$

	Vertex isolation	Offline selection
$B^+ \to D^* e v$	89.5	80.2
$B^+ \to D^* \tau v$	94.0	79.0
$B^0 \to D^{**} e v$	45.8	12.1
$*B^0 \to D^{**} \tau v$	74.7	48.0
$B^+ \to D^{**} e v$	64.3	23.6
$B^+ \to D^{**} \tau v$	65.7	28.0
$B^+ \to D^0(X_c \to evX')\, X$	53.2	13.9
$B^+ \to D^{**}(D_s \to \tau v)X$	73.7	28.5
$*B^0 \to D^{**}(D_s \to \tau v)X$	80.4	52.3
$B^0 \to D^0(X_c \to evX')\, X$	53.8	15.7
$*B^0 \to D^{**-}_{0,1,2} e v$	69.9	43.3
$B^+ \to D^{**0}_{0,1,2} e v$	61.0	26.0
$*B^0_s \to D^{**}_s e v$	87.7	45.5

Efficiency [%]

Vertex isolation

			BDT	
Pairs 1	B_1^-	π_1^+	0.99836	Passed
	B_1^-	π_2^+	0.96941	
	B_1^-	π_3^+	0.90353	
	B_1^-	π_4^+	0.85655	
	B_1^-	π_5^+	0.82173	
	B_1^-	π_6^+	0.41149	
Pairs 2	B_2^-	π_1^+	0.99954	Remove entire pairs
	B_2^-	π_2^+	0.99992	
	B_2^-	π_3^+	0.95614	
	B_2^-	π_4^+	0.82733	
	B_2^-	π_5^+	0.73321	
	B_2^-	π_6^+	0.59188	

second highest mva < 0.9999

			BDT	
Pairs 1	B_1^-	π_1^+	0.998	Passed
	B_1^-	π_2^+	0.969	
	B_1^-	π_3^+	0.903	
	B_1^-	π_4^+	0.856	
	B_1^-	π_5^+	0.821	
	B_1^-	π_6^+	0.411	
Pairs 2	B_2^-	π_1^+	0.999	Remove entire pairs
	B_2^-	π_2^+	0.996	
	B_2^-	π_3^+	0.994	
	B_2^-	π_4^+	0.827	
	B_2^-	π_5^+	0.733	
	B_2^-	π_6^+	0.591	

Sum of top 3 MVA < 2.9985

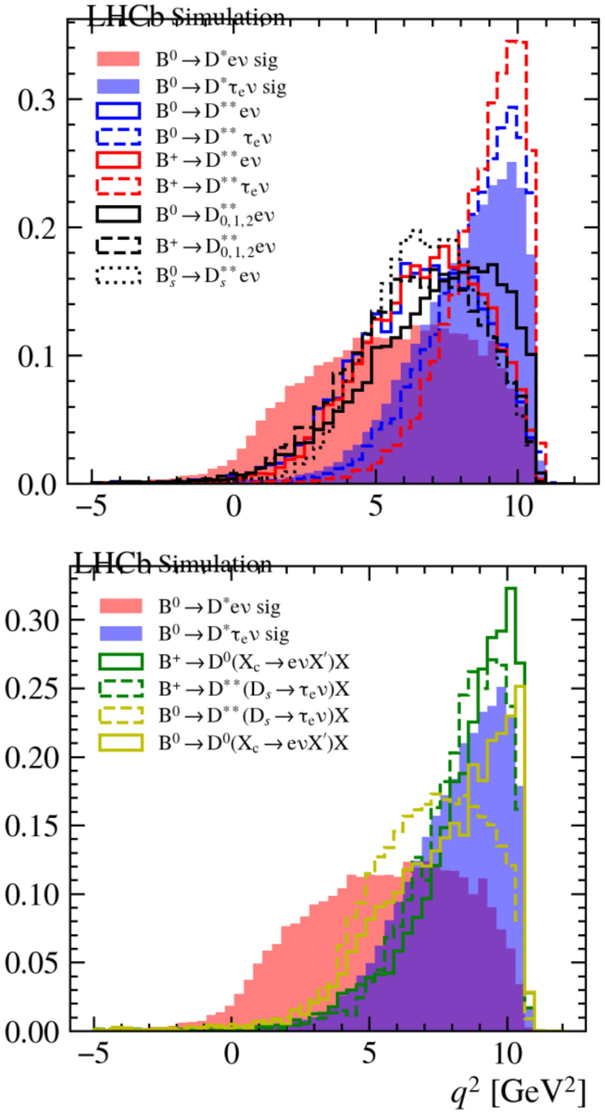
D^{**} sample

	Cocktail Notes
$B^+ \to D^{**} e\nu$	$D^*(2640)^0, D(2S)^0$ $D_1(2420)^0, D_2^*(2460)^0$
$B^+ \to D^{**} \tau\nu$	$D_1(H)^0, D_1(2420)^0$ $D_2^*(2460)^0$
$B^0 \to D^{**} e\nu$	$D^*(2640)^-, D(2S)^-$ $D_2^*(2460)^-, D_1(2420)^-$
$B^0 \to D^{**} \tau\nu$	$D_0^{*-}, D_1(H)^-$ $D_2^*(2460)^-, D_1(2420)^-$

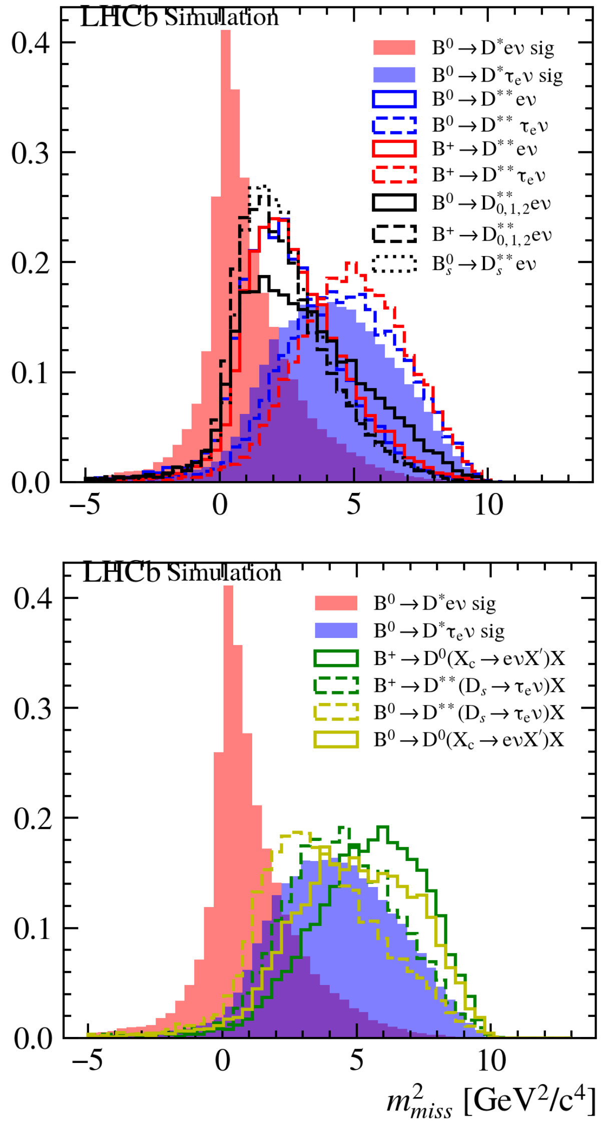
*With 2024 condition

Fitting variables in background sample

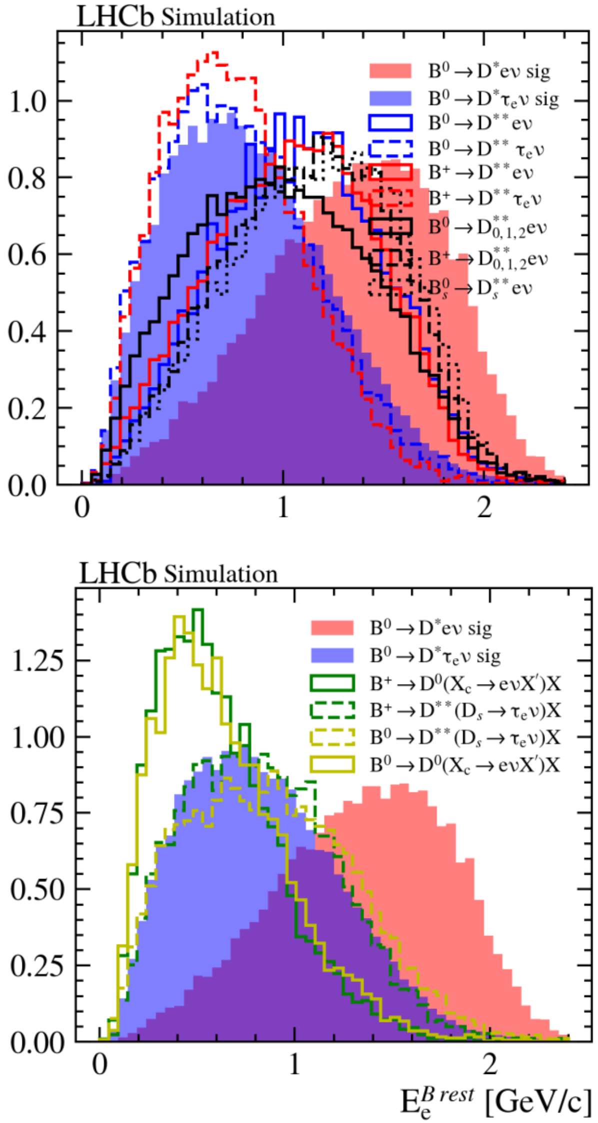
$q^2 = (p_{B^0} - p_{D^*})^2$



$M^2_{miss} = (p_{B^0} - p_{D^*} - p_e)^2$



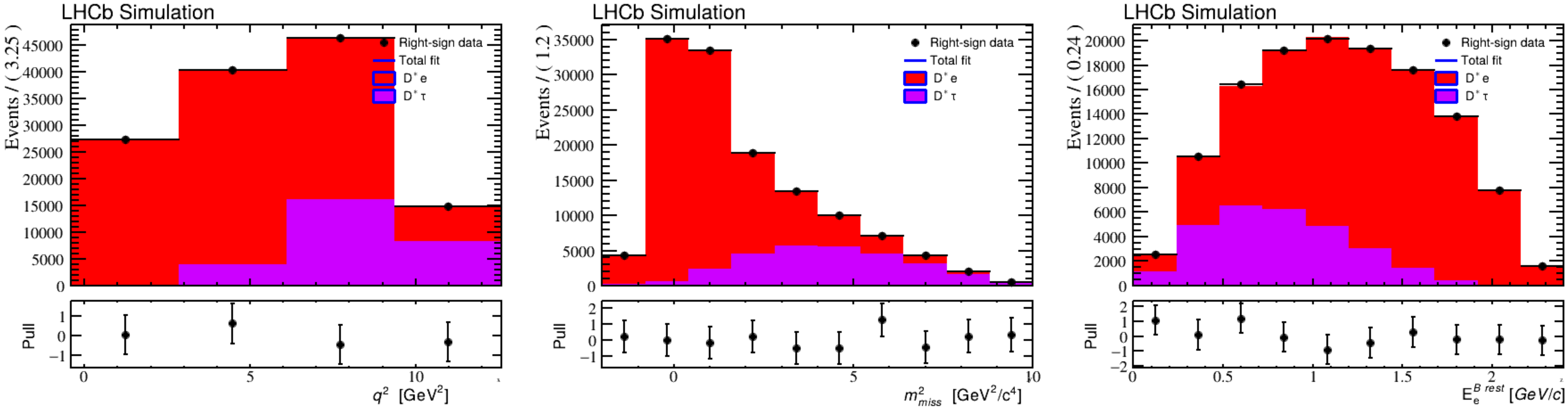
Lepton energy in the B rest frame



RooHistFactory test - result

With the templates from simulated sample, we got

Floating Parameter	Input value	Fit value	+/-	Error	GblCorr.
# of $B^0 \to D^{*-} e\nu$	1.0000e+05	1.0019e+05	+/-	6.13e+02	<none>
$R(D^*)_{\tau/e}$	2.9000e-01	2.8762e-01	+/-	4.52e-03	<none>



$\bar{B} \to D^* H_c(\to l\nu X)X$,