



SORBONNE
UNIVERSITÉ



SORBONNE UNIVERSITÉ

École doctorale des Sciences de la Terre et de l'Environnement et
Physique de l'Univers, Paris

Excerpts from the LHCb cookbook

—

RECEPTS for testing lepton universality and reconstructing primary vertices

Florian Reiß

Submitted in fulfilment of the requirements for the degree of

DOCTEUR DE SORBONNE UNIVERSITÉ

Supervised by Dr. Vladimir Gligorov

Defended on the 18th December 2020 in front of the committee

Mr. Pascal	Vincent	Jury president
Ms. Marie-Helene	Schune	Referee
Mr. David	Rousseau	Referee
Mr. Guy	Wilkinson	Examiner
Mr. Renaud	Le Gac	Examiner
Ms. Caterina	Doglioni	Examiner
Ms. Svjetlana	Fajfer	Examiner
Mr. Vladimir	Gligorov	Supervisor



Abstract

Lepton universality is an accidental symmetry of the Standard Model of particles physics, which expects that the charged leptons have the same characteristics with the exception of their masses. Experimental tests of lepton universality have shown hints of deviations from the Standard Model. This thesis presents a test of lepton universality using semileptonic b -hadron decays to final states involving a muon or an electron using data recorded with the LHCb detector in the years 2016, 2017 and 2018. The test is performed by measuring the branching fraction of the $\bar{B}^0 \rightarrow D^{*+} e^- \bar{\nu}_e$ decay relative to the branching fraction of the $\bar{B}^0 \rightarrow D^{*+} \mu^- \bar{\nu}_\mu$ decay, with the result

$$\begin{aligned} R(D^{*+})_{light} &= \frac{\mathcal{B}(\bar{B}^0 \rightarrow D^{*+} e^- \bar{\nu}_e)}{\mathcal{B}(\bar{B}^0 \rightarrow D^{*+} \mu^- \bar{\nu}_\mu)} \\ &= X \times (2.937 \pm 0.014(stat.) \pm 0.063(syst.)), \end{aligned}$$

where X is a blinding factor. This is the first time this ratio is measured at a hadron-hadron collider. It is shown that the systematic uncertainties related to the different behaviour of electrons and muons in the detector can be controlled to a level that a similar sensitivity to measurements at electron-positron colliders can be achieved.

The LHCb Upgrade for Run 3 of the Large Hadron Collider will rely on a completely software-based trigger, whose first stage needs to process events at a rate of up to 30 MHz. An implementation of the complete first trigger stage on GPUs was developed by the LHCb collaboration. During this thesis, the algorithm to reconstruct the proton-proton interaction points, necessary to identify displaced signatures, is implemented on GPUs. Changes to the original CPU implementation to achieve a fast and efficient algorithm are presented. The GPU algorithm is shown to fulfil the physics performance and throughput requirements of the Run 3 trigger.

Acknowledgements

The next couple of lines I want to dedicate to all the people who supported me for three years and helped me with bringing this thesis to fruition. Many of whom I can not thank enough.

My gratitude goes to my supervisor, Vladimir Gligorov, who guided me through this process, while leaving me a lot of freedom to grow as a person and as a scientist. I also have to thank all the amazing people I had the pleasure to have met in the lab, various workshops and conferences and LHCb collaboration meetings. The list is too long to mention everyone sufficiently.

I want to highlight my fellow "RECEPTniks" Dorothea vom Bruch, Renato Quagliani and Da Yu Tou, from whom I learned a lot. Dorothea in particular I need to thank for the productive collaboration on various projects and the many useful comments improving this document. Further I extend my gratitude to my fellow PhD students in the lab making lunches and coffee breaks very pleasurable.

Finally, I would like to give thanks to my family and friends providing the emotional support, without which this endeavour surely would not have succeeded. Shout-out to my parents, my sisters and my niece.

In memory of those who did not make it.

FR was funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 724777 "RECEPT".

Contents

Introduction	5
I Lepton flavour universality	7
I.1 The Standard Model	7
I.1.1 Gauge symmetries	9
I.1.2 The electroweak interaction	11
I.1.3 Tests of the Standard Model	13
I.1.4 Beyond the Standard Model	15
I.2 Experimental tests of lepton flavour universality	18
I.2.1 Electroweak penguin decays	19
I.2.2 Semileptonic B-hadron decays	21
I.3 Theoretical treatment	27
I.3.1 Hadronic form factors	27
I.3.2 Standard Model prediction	31
I.4 Explaining lepton universality violation	33
II The LHCb experiment	38
II.1 The Large Hadron Collider	38
II.2 The LHCb detector	40
II.3 Charged particle tracking performance	51
II.4 Particle identification	56
II.5 Data acquisition and trigger	58
II.6 Offline processing & simulation	63
III Testing lepton-flavour universality in semileptonic decays using electrons and muons	64
III.1 Analysis strategy	65
III.1.1 Data samples	67
III.1.2 Background sources	67
III.1.3 Simulated samples	68

III.1.4	Kinematic reconstruction	70
III.1.5	Initial studies	70
III.1.6	Signal reconstruction and selection	74
III.1.7	The $B^0 \rightarrow J/\psi K^{*0}$ control channel ▲	81
III.1.8	Combinatorial background	87
III.2	Corrections to the simulation	92
III.2.1	Particle identification efficiency ▲	93
III.2.2	Track reconstruction	94
III.2.3	L0 trigger ▲	95
III.2.4	Event multiplicity	99
III.2.5	Kinematic corrections	100
III.2.6	HLT corrections for signal modes ▲	101
III.2.7	HLT corrections for control modes ▲	104
III.2.8	Comparison of data and simulation distributions of control mode decays	107
III.2.9	$R(J/\psi)$ flatness cross-check	114
III.2.10	Form factor weighting	118
III.2.11	Comparison of data and simulated distributions of signal mode decays	118
III.2.12	Summary	119
III.3	Extraction of the relative signal yields	122
III.3.1	Description of the template fit	122
III.3.2	Templates from simulation	125
III.3.3	Data-driven templates	126
III.3.4	Fit results	126
III.4	Determination of the ratio $R(D^{*+})_{light}$	129
III.4.1	Selection efficiencies	129
III.4.2	$R(D^{*+})_{light}$ flatness cross-check	129
III.4.3	Combination of results	130
III.5	Estimation of systematic uncertainties	133
III.5.1	Tracking efficiency correction	134
III.5.2	Particle identification efficiency	134
III.5.3	Size of simulated samples	135
III.5.4	Form factor model	137
III.5.5	Estimation of selection efficiency	138
IV	Primary vertex reconstruction for the LHCb Upgrade	139
IV.1	The LHCb Upgrade	141

IV.1.1	Run 3 trigger	142
IV.2	The Allen project	145
IV.2.1	Modern processing architectures	145
IV.3	Primary vertex reconstruction	147
IV.3.1	The Run 3 algorithm ▲	148
IV.3.2	Implementation on GPUs	156
IV.3.3	Performance evaluation	159
IV.3.4	Timing performance	161
IV.4	Compilation on CPU	165
V	Conclusion	169
Appendix A	Meaning of selection requirements	172
Appendix B	Control channel selection	173
Appendix C	Matrix elements of misidentification matrix used for lepton misID background	177
Appendix D	Background studies	187
Appendix E	TISTOS efficiencies on data and simulation ▲	189
E.1	L0 control mode	189
E.2	HLT control mode	190
E.3	HLT signal mode	191
Appendix F	Template projections	192
Appendix G	Template fit projections in q^2 bins	214
Appendix H	Template fit projections without MC shape varia- tions	218
References		225

Introduction

“There are no mistakes. Only new paths to explore.”

Gregory David Roberts, *Shantaram*

This thesis summarises a study of lepton universality in semileptonic $\bar{B}^0$ hadron decays to final states including either a muon or an electron with data recorded by the LHCb detector in the years 2016, 2017 and 2018. The test of lepton universality is performed by measuring the ratio $R(D^{*+})_{light}$ defined as the ratio of the branching fraction $\bar{B}^0 \rightarrow D^{*+} e^- \bar{\nu}_e$ relative to $\bar{B}^0 \rightarrow D^{*+} \mu^- \bar{\nu}_\mu$,

$$R(D^{*+})_{light} = \frac{BR(\bar{B}^0 \rightarrow D^{*+} e^- \bar{\nu}_e)}{BR(\bar{B}^0 \rightarrow D^{*+} \mu^- \bar{\nu}_\mu)}, \quad (1)$$

precisely predicted to be unity by the Standard Model (SM). In all the following charge conjugation is implied meaning that this ratio also includes the decays of the B^0 hadron. Another convention is the usage of natural units, where the speed of light c and the reduced Planck constant $\hbar$ are set to unity $c = \hbar = 1$. As a consequence invariant masses are expressed in the same units as energies, usually electronvolt (eV).

This work is the first time this ratio has been measured at a hadron collider, where different experimental challenges arise compared to electron-positron colliders, where it has already been measured with good precision confirming the SM prediction. In proton-proton collisions b hadrons are produced from $b\bar{b}$ -pairs, whose initial energy can not be determined. Additionally, the LHCb detector only covers the forward-facing region, so that not all decay products of the collision can be detected. Neutrinos produced in semileptonic decays can not be reconstructed in the detector and thus carry away an unknown amount of momentum and energy, which can not easily be reconstructed since the total and missing 4-momentum is unknown.

Furthermore, charged particles, in particular electrons, emit Bremsstrahlung photons when passing through the detector material degrading their momentum

resolution and reconstruction efficiency. In that context, this work should be considered as a proof-of-principle that single electron final states can be used to measure lepton universality at hadron colliders, thus completing the picture of lepton universality in semileptonic b -hadron decays.

The first and main part of this thesis will be dedicated to the measurement of the ratio $R(D^{*+})_{light}$. After a brief description of the SM of particle physics and an introduction to the relevant theoretical background, a review of recent tests of lepton universality is given in Chapter I. The LHCb experiment, which recorded the data used in this analysis, is described in more detail in Chapter II. The main body of the analysis is described in Chapter III. The strategy to select the two signal modes and determine the relative yield with a template fit is presented, as well as the calibration of the simulation using an independent control channel.

The second part of the thesis will change focus from the present and by now retired LHCb detector to the LHCb Upgrade detector set to start taking data in 2022. The trigger system of the Upgrade detector will be completely software-based, putting an enormous requirement on the speed at which it has to (partially) reconstruct events. For this purpose, the first stage of the trigger will be implemented completely on graphics processing units (GPUs), a novelty for a high-energy physics experiment. This work describes the personal contribution to this effort consisting of adapting and tuning the primary vertex reconstruction to be executed on GPUs in the context of the *Allen* project [1], presented in Chapter IV.

The work is summarised and concluded in Chapter V. Figures created as part of this work are marked by a ● in the caption, while figures taken from other sources are marked by a ■ in the caption with a reference citation. The combination ■● is used and a reference is given if the figures were produced during the course of the thesis and appear in this form in public documents. In rare cases, this symbolism is used in the caption of tables. The Chapters III and IV contain the original work by the author marked in this way. Some sections are mainly based on the work of collaborators, where no major contribution by the author is claimed, and are marked by a ▲ in the section title. Figures marked with a ▲ are based on that work and were re-produced and in some cases visually adapted for this thesis. Figures marked with a ● in these sections still signify a contribution by the author.

Chapter I

Lepton flavour universality

“The world breaks everyone and afterward many are strong at the broken places.”

Ernest Hemingway, *A Farewell to Arms*

In recent years, various experimental results hint at a deviation from the predictions of the SM of particle physics in the sector of charged leptons. This has been met with cautious excitement, as concrete evidence of the violation of the universality of lepton flavour would unambiguously confirm the existence of particles and forces beyond the SM (BSM). The purpose of this chapter is to briefly introduce the SM and the concept of lepton flavour universality. It summarises recent tests thereof, as well as the required theoretical concepts. It is concluded by reviewing additions to the SM based on effective field theory approaches, which can accommodate the observed tensions in a generic way.

I.1 The Standard Model

The SM is a quantum field theory encapsulating the current knowledge of the fundamental particles and their interactions. Fundamental here means that these particles can not further be broken down into smaller parts within the theory. A common schematic representation of these particles and the force carriers which mediate these interactions is shown in Figure 1.

The SM is built from three generations of (anti-)quarks, three generations of (anti-)leptons as well as the gauge vector bosons and the scalar Higgs boson. The quarks further subdivide into up-type quarks (u, c, t) with positive charge $q = +\frac{2}{3}e$ and down-type quarks (d, s, b) with negative charge $q = -\frac{1}{3}e$, where the charge is given as fraction of the absolute value of the elemental electron charge e . The leptons on the other hand subdivide into either charged (e^-, μ^-, τ^-) or neutral

Standard Model of Elementary Particles

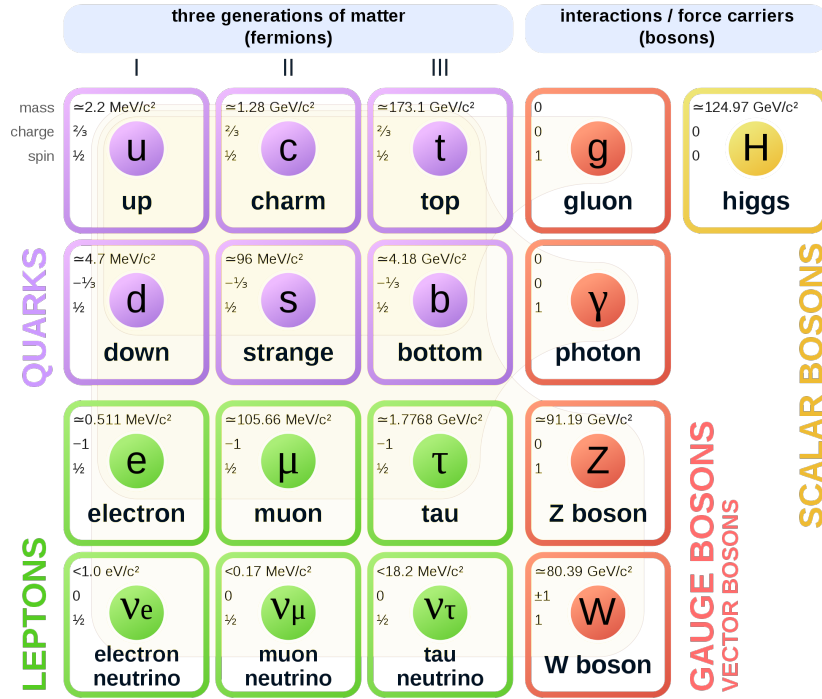


Figure 1: ■ Diagrammatic overview of the Standard Model of particle physics showing the three generations of leptons and quarks, as well as the gauge and Higgs bosons [2].

$(\nu_e, \nu_\mu, \nu_\tau)$. For each quark and lepton there exists an anti-particle with opposite quantum numbers. The fundamental charges carried by both the quarks and leptons corresponding to the gauge symmetries of the SM are the weak isospin T_3 and hypercharge Y , where the weak isospin and hypercharge are related to the electric charge Q by

$$Q = T_3 + \frac{1}{2}Y. \quad (2)$$

In addition, quarks carry a colour charge — (anti-)red, (anti-)green or (anti-)blue — of the strong interaction.

Interactions between the quarks and leptons are mediated by the gauge bosons. The gauge bosons are the photon γ responsible for the electromagnetic interaction between charged particles, the Z and $W^\pm$ bosons involved in the weak interaction and the gluons g implied in the strong interaction. While the gluon and photon are massless, the Z and $W^\pm$ are massive. Another difference is that gluons also carry colour and are self-interacting, leading to phenomena like *colour confinement*. Unlike QED, where the force weakens if two charged particles are separated, in QCD it increases. Qualitatively, this can be understood by the fact that the gluon self-interaction field lines of QCD are bundled together into a 'flux-tube', where

the field line density and thus energy density is constant. This means that if one would try to separate two quarks, the energy in the gluon field would increase with the distance. At some point, this energy is large enough to produce a quark anti-quark pair, where each forms a new bound state with one of the initial quarks. As a result, only bound objects which are colour-neutral can exist.

With the discovery of the Higgs boson in 2012 [3] [4], whose existence is a consequence of the Higgs mechanism giving fundamental particles their masses, the SM can be considered to be a complete and internally consistent theory of nature at microscopic scales. However, it is not a 'theory of everything'. For example, it does not include gravitational forces or contains a dark matter candidate. For this and other reasons, searches for BSM physics are well motivated, as explained below.

I.1.1 Gauge symmetries

The SM can be succinctly described by the local gauge symmetries

$$\text{SU}(3) \times \text{SU}(2)_L \times \text{U}(1)_Y, \quad (3)$$

where $\text{SU}(3)$ describes QCD and $\text{SU}(2)_L \times \text{U}(1)_Y$ the electroweak interaction. Each symmetry group is associated to charges conserved by the respective interaction: colour charge (anti-)red, (anti-)green, (anti-)blue) for QCD, and weak isospin and hypercharge Y for the electroweak interaction. Built into the theory is the distinction between left-handed doublets and right-handed singlets of the $\text{SU}(2)_L$ transformation, where left- and right-handed refers to the left- and right-handed chiral projection of a fermion, created by using the chiral projection operators

$$P_L = \frac{1}{2}(1 - \gamma_5), P_R = \frac{1}{2}(1 + \gamma_5), \quad (4)$$

where $\gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ is the product of the Dirac γ^μ matrices. This results in the total fermion content of $\begin{pmatrix} u^i \\ d^i \end{pmatrix}_L, \begin{pmatrix} \nu \\ \ell \end{pmatrix}_L, (u^i)_R, (d^i)_R, (\ell)_R$, where the index i indicates the colour charge of the quarks. The fermion charges relevant to the electroweak sector are summarised in Table 1.

Applying the gauge symmetries to the Lagrangian $\mathcal{L}_{free}^\psi$ of a free, massless fermion field ψ

$$\mathcal{L}_{free}^\psi = \bar{\psi}(i\gamma^\mu\partial_\mu)\psi, \quad (5)$$

where the fields transform as

$$\begin{aligned}\text{SU}(3) : \psi(x) &\rightarrow \psi(x)' = \exp^{ig_s \alpha(x) \hat{U}} \psi(x) \\ \text{SU}(2)_L : \psi(x) &\rightarrow \psi(x)' = \exp^{ig \alpha(x) \hat{T}} \psi(x) \\ \text{U}(1)_Y : \psi(x) &\rightarrow \psi(x)' = \exp^{ig' \frac{Y}{2} \alpha(x)} \psi(x),\end{aligned}\tag{6}$$

local gauge invariance is only achieved by also introducing the covariant derivatives

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + i \frac{g'}{2} Y B_\mu + i \frac{g}{2} \sigma_j W_\mu^j + i \frac{g_s}{2} \lambda_\alpha G_\mu^\alpha,\tag{7}$$

where now the additional vector boson fields B_μ , W_μ^j and G_μ^α are introduced and g' , g and g_s are the respective coupling constants. The generators of the groups $\text{SU}(2)_L$ and $\text{SU}(3)$, $\hat{T}$ and $\hat{U}$, can be expressed in terms of the Pauli matrices σ and Gell-Mann matrices λ by $T_j = \frac{\sigma_j}{2}$ and $U_\alpha = \frac{\lambda_\alpha}{2}$. These fields have to transform as

$$\begin{aligned}G_\mu^k &\rightarrow G_\mu'^k = G_\mu^k - \partial_\mu \alpha^k - g_s f_{ijk} \alpha_i G_\mu^j \\ W_\mu^k &\rightarrow W_\mu'^k = W_\mu^k - \partial_\mu \alpha^k - g \epsilon_{ijk} \alpha_i W_\mu^j \\ B_\mu &\rightarrow B'_\mu = B_\mu - \partial_\mu \alpha,\end{aligned}\tag{8}$$

where f_{ijk} and ϵ_{ijk} are the structure constants of the respective symmetry groups. By introducing those fields, the local gauge symmetry of the Lagrangian is restored. Now the Lagrangian not only contains free fermions, but also interactions to vector boson fields. These include the four gauge bosons of the electroweak sector, B_μ and $W_\mu^{1,2,3}$. The observed physical photon, Z and $W^\pm$ fields are defined by

$$\begin{aligned}W_\mu^\pm &= \frac{1}{\sqrt{2}} (W_\mu^1 \mp i W_\mu^2) \\ A_\mu &= B_\mu \cos \theta_W + W_\mu^3 \sin \theta_W \\ Z_\mu &= -B_\mu \sin \theta_W + W_\mu^3 \cos \theta_W,\end{aligned}\tag{9}$$

with the weak mixing angle θ_W and the photon field A_μ . The strong interaction is mediated by the 8 gluon fields $G_\mu^{1..8}$.

The gauge symmetries would be explicitly broken by trying to add mass terms for the gauge bosons and fermions. Instead, particle masses are included in the SM through the Higgs mechanism, which introduces an additional scalar field, the Higgs field, with a non-zero vacuum expectation value. By this, the $\text{SU}(2)_L \times \text{U}(1)_Y$ symmetry is spontaneously broken to $\text{U}(1)_{\text{QED}}$.

Table 1: Charges of Standard Model fermions.

	Electric charge	Weak isospin	Hypercharge
$(\nu)_L$	0	$+\frac{1}{2}$	-1
$(\ell)_L$	-1	$-\frac{1}{2}$	-1
$(\ell)_R$	-1	0	-2
$(u)_L$	$+\frac{2}{3}$	$+\frac{1}{2}$	$+\frac{1}{3}$
$(d)_L$	$-\frac{1}{3}$	$-\frac{1}{2}$	$+\frac{1}{3}$
$(u)_R$	$+\frac{2}{3}$	0	$+\frac{4}{3}$
$(d)_R$	$-\frac{1}{3}$	0	$-\frac{2}{3}$

During this spontaneous symmetry breaking the Z and $W^\pm$ gauge bosons acquire their masses, while the fermion mass terms can be added by Yukawa-interactions with the non-zero vacuum expectation value of the Higgs field, which respects the gauge symmetries. Since this mass term uses left- and right-handed fermions and right-handed neutrinos do not exist in the SM, neutrinos remain massless in the Higgs mechanism. From the Higgs mechanism, it follows that the Higgs boson couples to Z and $W^\pm$ gauge bosons and the fermions with coupling strengths proportional to their masses.

I.1.2 The electroweak interaction

One particularity of the weak interaction is that parity symmetry is not conserved, where the parity transformation amounts to switching the signs of space-related variables $\mathbf{x} \rightarrow -\mathbf{x}$. This is written into the theory of the weak interaction by the V-A (vector-axial vector) structure of the weak coupling. Left- and right-handed (chiral) projections of fermions can be created by using the chiral projection operators introduced above.

The fermion current j^μ associated to a fermion- W vertex is given by

$$j_{weak}^\mu = \bar{f}'(-i\frac{g_W}{\sqrt{2}}\gamma^\mu\frac{1}{2}(1-\gamma^5))f. \quad (10)$$

From this, it follows that only left-handed particles and right-handed anti-particles participate in the weak interaction mediated by the charged W as the currents for other combinations would be zero due to the appearance of the projection operator.

Fermi theory of weak interaction

Historically, the weak interaction was formulated to explain the β -decay of neutrons. It was imagined as a point-like interaction without the exchange of a prop-

agator. Since then it became clear that the weak interaction is mediated by the exchange of W and Z bosons, but in the limit where the energy of the process is small compared to the gauge boson masses, Fermi's original description as point-like interaction is valid. This principle of using a point-like interaction in place of an unknown propagator can also be generalised to BSM scenarios, where the expected new particles have much larger masses compared to the scale currently reached by particle colliders. In the case of the β -decay, the matrix element of the 4-point interaction is described by

$$\mathcal{M} = \frac{1}{\sqrt{2}} G_F g_{\mu\nu} [\bar{e} \gamma^\mu (1 - \gamma^5) \nu_e] [\bar{u} \gamma^\mu (1 - \gamma^5) d], \quad (11)$$

where the effect of the heavy boson exchange is absorbed into the Fermi constant G_F , measured to be $G_F = 1.1663787(6) \times 10^{-5} \text{ GeV}^{-2}$ [5].

At larger energies, this approximation breaks down and the full W propagator has to be considered. Using the unitary gauge it can be written as

$$\frac{-i \left(g_{\mu\nu} - \frac{q_\mu q_\nu}{M_W^2} \right)}{q^2 - M_W^2}, \quad (12)$$

where M_W is the mass of the W boson and $g_{\mu\nu}$ the metric tensor of special relativity. Therefore once the involved energies are large enough, the point-like interaction can be resolved in the sense that now short-range effects have to be considered. In the limit $q^2 \ll M_W^2$ the propagator can be written as

$$i \frac{g_{\mu\nu}}{M_W^2}. \quad (13)$$

Comparing the effective and full theory, one can obtain the relation of the Fermi constant and the coupling strength g_w of the weak interaction

$$\frac{G_F}{\sqrt{2}} = \frac{g_w^2}{8M_W^2}. \quad (14)$$

CKM mechanism

The weak eigenstates of quarks participating in the weak interaction and the mass eigenstates are not the same but are related by a mixing matrix V_{CKM} (Cabibbo–Kobayashi–Maskawa matrix)

$$\begin{pmatrix} u' \\ d' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \quad (15)$$

This allows quarks in charged weak currents mediated by $W^\pm$ exchange to change flavour to a different generation. The parameters of the CKM matrix are not predicted by the SM but have to be determined experimentally. The CKM matrix also includes an imaginary part, which introduces charge-parity violation.

The CKM matrix can be described using three angles θ_{12} , θ_{23} , θ_{13} and a phase δ_{13}

$$V_{CKM} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{bmatrix} \begin{bmatrix} c_{13} & 0 & s_{13}e^{-i\delta_{13}} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta_{13}} & 0 & c_{13} \end{bmatrix} \begin{bmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (16)$$

$$= \begin{bmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{13}} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta_{13}} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta_{13}} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta_{13}} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta_{13}} & c_{23}c_{13} \end{bmatrix},$$

where $c_{ij} = \cos(\theta_{ij})$ and $s_{ij} = \sin(\theta_{ij})$. Another parametrisation was devised by Wolfenstein, writing the CKM matrix in terms of

$$\begin{aligned} \lambda &= s_{12} \\ A\lambda^2 &= s_{23} \\ A\lambda^3(\rho - i\eta) &= s_{13}e^{-i\delta}, \end{aligned} \quad (17)$$

where the matrix is now approximated up to third order in λ as

$$V_{CKM} = \begin{bmatrix} 1 - \frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{bmatrix} + O(\lambda^4). \quad (18)$$

Since the parameter λ is small, it can be seen in this parametrisation that the diagonal elements are almost unity while the off-diagonal elements are small. This means that a quark transition changing the generation is suppressed compared to a transition within a generation. Historically, this is known as Cabibbo suppression.

The CKM matrix has to be unitary in the SM. This leads to a set of equations, each of which can be represented in the complex plane. The condition

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0 \quad (19)$$

can be represented as a triangle in the complex plane, commonly called the unitary triangle.

I.1.3 Tests of the Standard Model

Generally speaking, there are two ways to discover new particles beyond the SM: *direct* and *indirect* observations. In a direct measurement the decay products of a

new particle produced in a particle collider are detected. It would then be observed by a significant peaking structure in the invariant mass spectrum of the decay products. An example of such is the direct observation of the Higgs boson in various decay channels first reported in July 2012 by the ATLAS and CMS collaborations. Direct measurements are limited by the centre-of-mass energy of the collider as well as the sensitivity reachable considering all backgrounds.

In indirect measurements, the contribution of BSM particles to Feynman diagrams as virtual particles, which can change observables like branching ratios and angular distributions, is constrained by precisely measuring these quantities and comparing them to the SM expectation. A disagreement between the expected and measured value would then indicate the presence of effects not taken into account in the theory calculation, possibly caused by BSM particles. The achievable sensitivity depends on the mass and the coupling strength of the new particle. As a historical example of such indirect searches, it can be mentioned that the existence of the Higgs boson and constraints on its mass could be derived from measurements of the t quark and W boson masses years before it was finally directly observed. By precisely measuring the parameters of the electroweak sector a prediction for the W mass can be obtained. Loop corrections to this prediction introduce a dependency on the t quark and Higgs boson masses. Measuring the t quark and W boson masses then constrains the mass of the Higgs boson.

The achievements of the SM are manifold with experiment and theory prediction agreeing to great precision with measurements performed over a large energy range. Especially in the QED sector, it allows for extremely precise predictions taking many higher-order effects into account in perturbative calculations. For example, the anomalous magnetic moment of the muon is predicted to be [6]

$$a_{\mu}^{SM} = 116\,591\,810(43) \times 10^{-11} \quad (20)$$

and measured to be [7]

$$a_{\mu}^{meas} = 116\,592\,080(54)(33) \times 10^{-11}. \quad (21)$$

The big success of the SM is its ability to deliver such extremely precise predictions, which can be tested experimentally to a similar precision, which in turn, as in the case of the muon anomalous magnetic moment, can reveal tensions between measurement and prediction, which can point in the direction of BSM theories.

The unitarity conditions of the CKM mechanism are another expectation of the SM. By measuring and over-constraining the angles and side lengths of the unitarity triangle introduced earlier the CKM mechanism in the SM can be tested

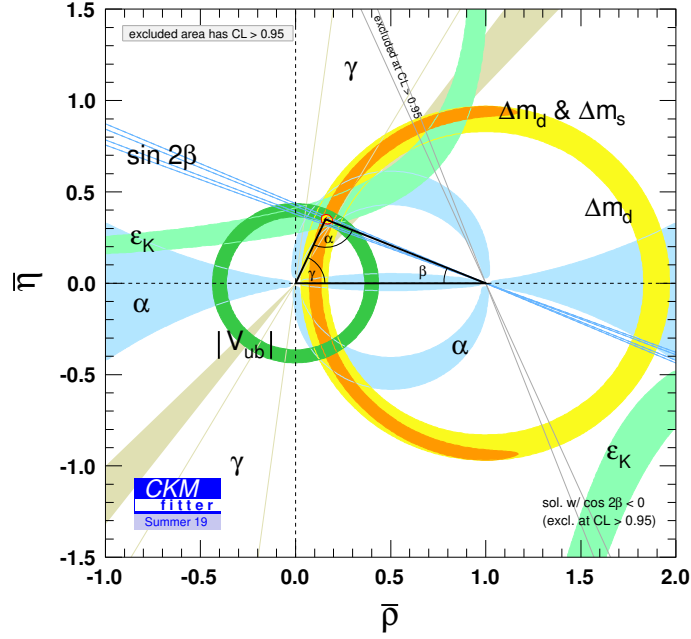


Figure 2: ■ Measurements of the sides and angles of the unitarity triangle [8].

to great precision, as shown in Figure 2. So far the measurements of the CKM parameters show no deviations from the unitarity conditions.

The success story of the SM is completed by the discovery of a boson consistent with the SM Higgs boson, whose existence is predicted from the Higgs mechanism introducing particle masses to the SM. Its discovery allows for additional tests of the SM by measuring its production cross sections and branching fractions in various channels, as shown in Figures 3 and 4. Again the measurements are in agreement with the SM predictions.

1.1.4 Beyond the Standard Model

While the SM has been scrutinised to excruciating detail mastering almost every challenge brought against it experimentally, it is considered to be incomplete. This assessment originates from more fundamental considerations – its formulation is not ‘elegant’ or ‘beautiful’ enough to be a complete theory of everything – and the difficulties to unify it with the gravitational force or from its lack of explanation of certain phenomena, be it the asymmetry between matter and anti-matter in the universe, the nature of dark matter or the observed hierarchy in the masses of its constituents. Adding non-zero neutrino masses, which follow from the observation of neutrino oscillations [10], also require additions to the basic formulation of the SM laid out in this chapter. The SM can then be regarded as a low energy

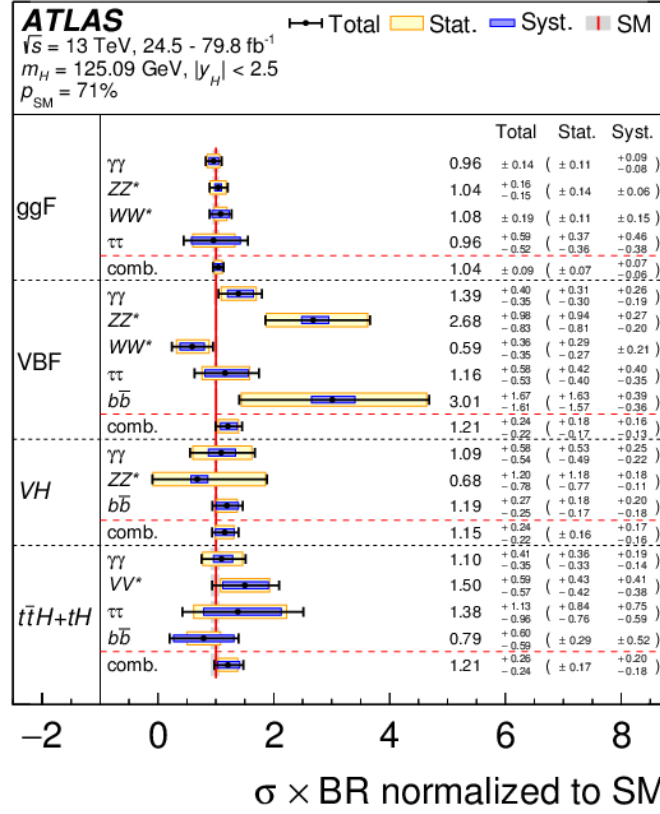


Figure 3: ■ Cross sections times branching fraction for various Higgs production and decay channels normalised to the SM expectation [9].

approximation of an unknown underlying theory.

There are numerous indications for the existence of dark matter, which does not interact by strong or electroweak interaction. For example, it is postulated to explain the measurements of rotation velocities of galaxies [11]. By fitting the cosmic microwave background anisotropy in the ΛCDM model, dark matter is determined to make up about 26% of the energy-matter content of the universe [12]. However, the SM does not contain a candidate for a dark matter particle. An even bigger mystery is the nature of dark energy, which is thought to accelerate the expansion of the universe, and which in this model is the largest part of the matter-energy content of the universe with about 68%. For both dark matter and energy, the SM can not provide a description and thus BSM models are required.

Another problem is the observed difference between the amount of matter and anti-matter in the universe. It was proposed that three conditions must be met to explain the matter anti-matter difference to be generated during baryogenesis [13]:

1. baryon number conservation must be violated

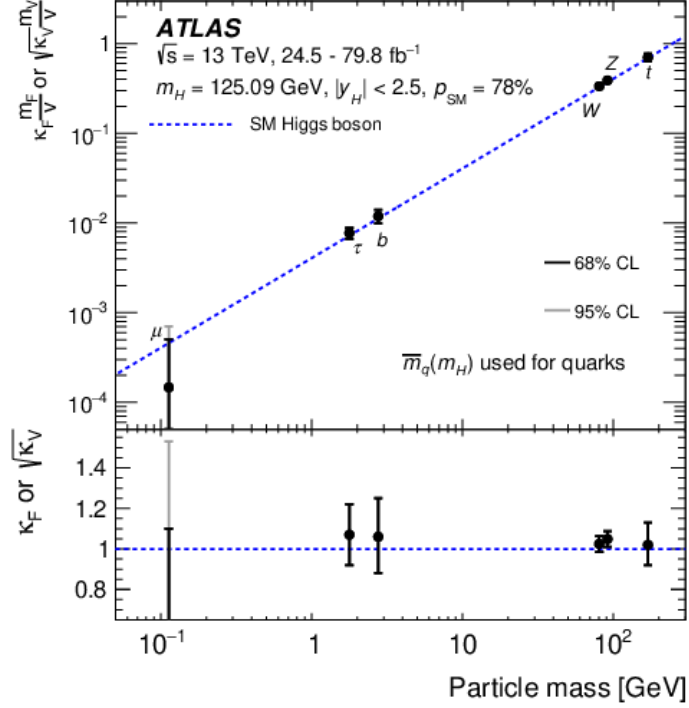


Figure 4: ■ Reduced vector boson and fermion Higgs coupling strength as function of the boson and fermion masses. The SM expectation is shown as blue dotted line [9].

2. C- and CP-symmetry must be broken

3. interactions must have taken place out of thermal equilibrium

In the SM, no processes violating baryon number conservation are allowed, but such decays, for example the decay of a free proton, are possible in BSM models such as grand unifying theories (GUT). While C- and CP-symmetry are violated in the SM in the weak interaction, the size of this effect measured so far is thought to be too small to generate the observed matter anti-matter asymmetry.

Taking all this together, it is certainly uncontroversial to say that the SM can not and will not be the last word spoken. Although more and more precise tests of the SM have been performed, reaching higher and higher energy scales, its remarkable resilience has been barely scratched. However in recent years, some exciting hints of possible deviations from the SM predictions have been observed in several measurements performed at different experiments, which indicate that lepton flavour universality, as implicitly assumed in the SM, might be broken. These studies are discussed and an overview of the experimental results are given in the next section.

I.2 Experimental tests of lepton flavour universality

The SM implicitly assumes that the gauge couplings of the charged leptons to the gauge bosons of the electroweak interaction, the γ , $W^\pm$ and Z , are the same. Since there is no fundamental symmetry requiring this, it is often called an 'accidental' symmetry. In that sense, the leptons of different flavours are universal with their only difference being their masses. Lepton flavour universality in the sense of the coupling strength to the Z and W bosons can be directly tested in the decays $Z \rightarrow \ell^+ \ell^-$ and $W^- \rightarrow \ell^- \bar{\nu}_\ell$, for example at LEP expressed as ratios of partial decay widths of the bosons [14] [15]

$$\begin{aligned} \frac{\Gamma_{Z \rightarrow \mu^+ \mu^-}}{\Gamma_{Z \rightarrow e^+ e^-}} &= 1.0009 \pm 0.0028, \\ \frac{\Gamma_{Z \rightarrow \tau^+ \tau^-}}{\Gamma_{Z \rightarrow \mu^+ \mu^-}} &= 1.0019 \pm 0.0032, \\ \frac{\Gamma_{W^- \rightarrow \tau^- \bar{\nu}_\tau}}{\Gamma_{W^- \rightarrow e^- \bar{\nu}_e}} &= 1.063 \pm 0.027, \\ \frac{\Gamma_{W^- \rightarrow \tau^- \bar{\nu}_\tau}}{\Gamma_{W^- \rightarrow \mu^- \bar{\nu}_\mu}} &= 1.070 \pm 0.026. \end{aligned} \tag{22}$$

The observed tensions in the case of W decays have been resolved by measurements performed by the LHC experiments, with the story being closed by the latest ATLAS result [16]

$$\frac{\Gamma_{W^- \rightarrow \tau^- \bar{\nu}_\tau}}{\Gamma_{W^- \rightarrow \mu^- \bar{\nu}_\mu}} = 0.992 \pm 0.013, \tag{23}$$

in combination with a previous measurement [17] of

$$\frac{\Gamma_{W^- \rightarrow e^- \bar{\nu}_e}}{\Gamma_{W^- \rightarrow \mu^- \bar{\nu}_\mu}} = 0.997 \pm 0.010. \tag{24}$$

From this experimental evidence, lepton universality in the boson couplings is confirmed to a high precision. Observing a process where lepton universality is broken would be evidence of a BSM interaction.

Thus, when measuring the relative branching ratio of two decays which differ only in the flavour of the involved lepton, one would expect from the SM that this ratio only deviates from unity due to the different available phase space in these decays. Recently, such tests of lepton flavour universality have been performed at e^+e^- - and pp colliders, resulting in observed deviations from the SM

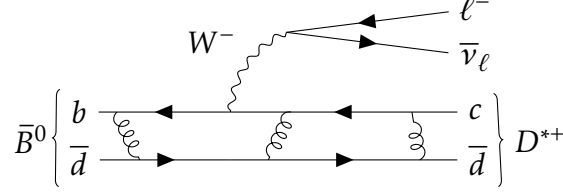


Figure 5: Feynman diagram of the decay $\bar{B}^0 \rightarrow D^{*+} \ell \bar{\nu}_\ell$.

which are in some cases above the typical 3σ limit used to claim evidence for a new phenomenon. Performing these measurements at an e^+e^- or pp collision environment has several implications. Generally speaking, e^+e^- collisions produce less background and the well known centre-of-mass energy of the collision, together with full angular coverage of the detector, makes the reconstruction of undetected particles like neutrinos easier. On the other hand, pp colliders typically have higher signal yields and reach higher centre-of-mass energies, so that such measurements can be also performed with heavier b -hadrons like B_c^+ or Λ_b^0 . In particular, a lot of interest has focused on decays of hadrons containing b -quarks with the transitions $b \rightarrow c \ell \nu$ or $b \rightarrow s \ell^+ \ell^-$. While the former is a tree-level transition with typically a relative large branching fraction, the later is forbidden at tree-level in the SM. But still this transition can occur mediated via loop diagrams, which results in an extremely rare process. As example the Feynman diagrams for the decays $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu$ and $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ are shown in Figures 5 and 6. Their respective branching fractions are different by multiple magnitudes with $\mathcal{B}(B^0 \rightarrow D^{*-} \mu^+ \nu_\mu) = 5.05\%$ and $\mathcal{B}(B^0 \rightarrow K^{*0} \mu^+ \mu^-) = 9.4 \times 10^{-7}$ [18].

From a theoretical point of view, such ratios of branching fractions are interesting as they can be precisely predicted in the SM as uncertainties related to describing the hadronic transition in the decay cancel to first order. From the experimental side ratios of branching fractions can generally be measured to a better precision than absolute values, as systematic uncertainties can cancel. In the following sections, the measurements performed with $b \rightarrow c \ell \nu$ and $b \rightarrow s \ell^+ \ell^-$ transitions are summarised and more detail is given how the cancellation of theoretical and experimental uncertainties occurs.

1.2.1 Electroweak penguin decays

Electroweak penguin decays are suppressed strongly in the SM due to the appearance of the loop in the Feynman diagram. They consist of a flavour changing neutral quark current, forbidden at tree-level in the SM. New particles could enter these loops as virtual particles and measurably change observables like the branch-

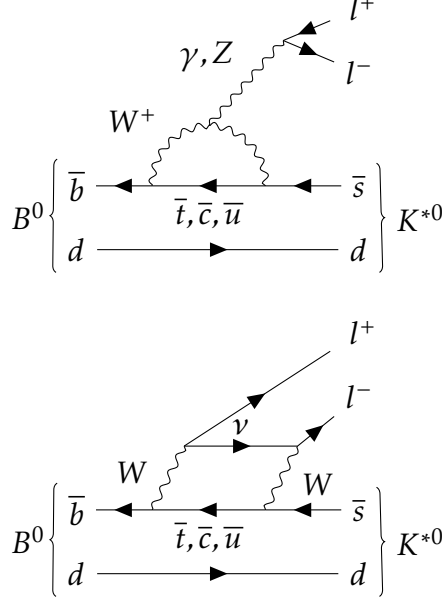


Figure 6: Examples of Feynman loop diagrams contributing to the process $B^0 \rightarrow K^{*0} \ell^+ \ell^-$.

ing fraction of the decay or angular distributions of the decay products. Since these decays have small branching ratios experiments have only recently collected enough data to make measurements with enough precision to make meaningful comparisons with the SM. As an example for such an electro-weak penguin decay, where lepton universality has been tested, the Feynman diagram of the process $B^0 \rightarrow K^{*0} \ell^+ \ell^-$ is shown in Figure 6. Here the lepton-antilepton pair consists of either muons or electrons. To reduce systematic uncertainties the measurement at LHCb is performed as double-ratio relative to the control channel $B^0 \rightarrow J/\psi K^{*0}$

$$R_{K^{*0}} = \frac{\mathcal{B}(B^0 \rightarrow K^{*0} \mu^+ \mu^-)}{\mathcal{B}(B \rightarrow K^{*0} J/\psi (\rightarrow \mu^+ \mu^-))} \frac{\mathcal{B}(B \rightarrow K^{*0} J/\psi (\rightarrow e^+ e^-))}{\mathcal{B}(B^0 \rightarrow K^{*0} e^+ e^-)}, \quad (25)$$

using the observation that lepton universality is fulfilled in J/ψ decays to a lepton pair [19]. Systematic effects, for example related to the difference of reconstruction performance between muons and electrons, are expected to cancel by measuring this double-ratio.

The measurement is typically performed in different regions of q^2 , which is the square of the invariant mass of the lepton-antilepton system, as the SM prediction typically depends on this variable. A comparison of this ratio measured at LHCb with the SM prediction is shown in Figure 7. A similar measurement was performed using $B^+ \rightarrow K^+ \ell^+ \ell^-$ decays, where a 2.5σ deviation from the SM has been observed in the range $1.1 < q^2 < 6.0 \text{ GeV}^2$ [20]. More recently such a ratio was measured using the b -baryon decay $\Lambda_b^0 \rightarrow p K^- \ell^+ \ell^-$, found to be compatible

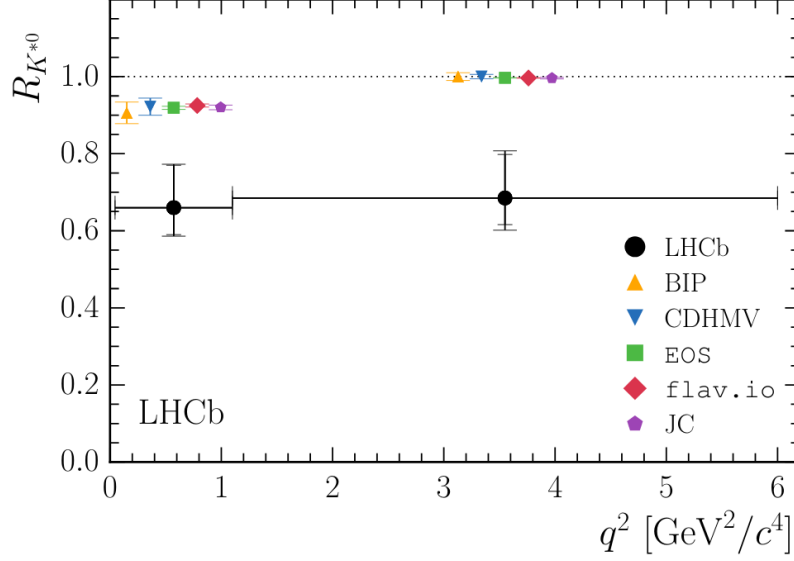


Figure 7: ■ Comparison of the measured values of $R(K^{*0})$ with the SM prediction in two different ranges of the squared invariant mass of the lepton-lepton system [22].

with the SM within the experimental uncertainties [21].

Additional tensions with the SM are seen in angular analyses, as shown in Figure 8. The P'_5 variable used here is constructed in a way to reduce the uncertainties on the theoretical predictions. The measurement is performed in bins of q^2 and a tension with the SM at a 2.9σ level can be observed in one of the bins. Another tension is seen in the measurement of the differential branching fraction $\mathcal{B}(B_s^0 \rightarrow \phi \mu^+ \mu^-)$ as function of q^2 [23] in shown in Figure 9. The tension becomes more evident when considering the $1.0 < q^2 < 6 \text{ GeV}^2$ region, where the differential branching fraction is measured to be $(2.58_{-0.31}^{+0.33} \pm 0.08 \pm 0.19) \times 10^{-8} \text{ GeV}^{-2}$, which is more than 3σ below the SM prediction of $(4.81 \pm 0.56) \times 10^{-8} \text{ GeV}^{-2}$.

To summarise, deviations from the SM have been observed at several places in electroweak penguin decays of b -hadrons. Although no single measurement has yet reached the 5σ threshold to claim the observation of a new effect, taking all the observations together in a global analysis shows that allowing for BSM contributions can improve the description of the data with respect to the SM, as discussed in Section I.4.

I.2.2 Semileptonic B-hadron decays

For semileptonic $b \rightarrow c \ell \nu$ decays, mediated by a charged weak current at the tree-level, the SM branching fraction is large and new physics effects might be hard to notice. Still these decays are interesting if one assumes that BSM effects primarily affect the third lepton generation. This could happen if the coupling strength of

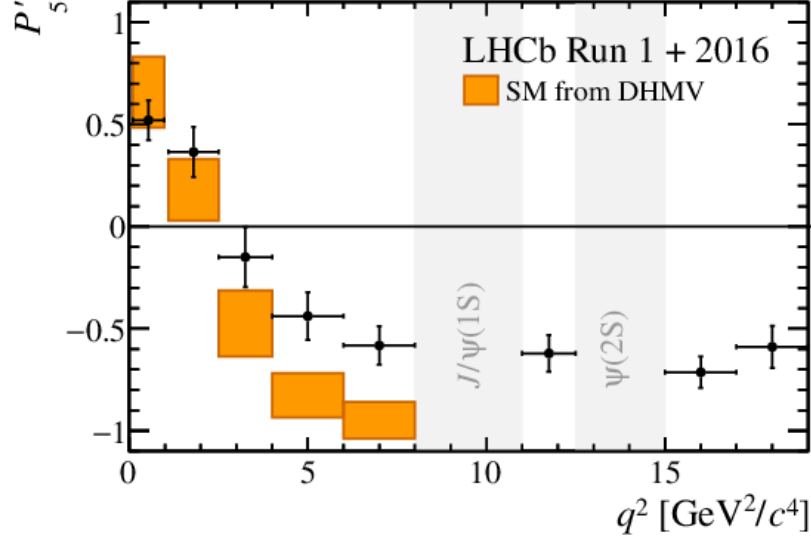


Figure 8: ■ Measurement of the angular variable P'_5 as function of q^2 [24].

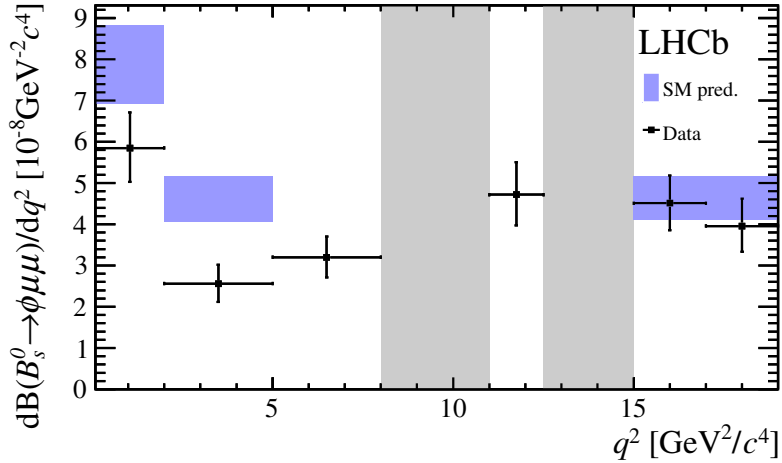


Figure 9: ■ Measured $B_s^0 \rightarrow \phi \mu^+ \mu^-$ differential decay rate as function of q^2 . Predictions by the SM, if available, are shown as blue bands [23].

the BSM particles depends on the mass of the involved lepton. Then the BSM contribution to the decay to the heaviest lepton, the τ , can become significant enough to modify the observed branching fraction while leaving the branching fraction of decays to lighter leptons almost unchanged with respect to the SM expectation. Typically relative branching fractions defined as

$$R(H_b) = \frac{\mathcal{B}(H_B \rightarrow H_c \tau \nu_\tau)}{\mathcal{B}(H_B \rightarrow H_c \ell \nu_\ell)} \quad (26)$$

are measured, where, depending on the experimental set-up, the lepton in the denominator is either an electron, muon or a combination of both. Measurements us-

Table 2: Experimental results of $R(H_c)$ measurements performed at LHCb.

H_b	H_c	τ decay mode	$R(H_c)$	Reference
B^0	D^{*-}	hadronic	$0.291 \pm 0.019(stat) \pm 0.026(syst) \pm 0.013(ext)$	[25]
B^0	D^{*-}	muonic	$0.336 \pm 0.027(stat) \pm 0.030(syst)$	[26]
B_c^+	J/ψ	muonic	$0.71 \pm 0.17(stat) \pm 0.18(syst)$	[27]

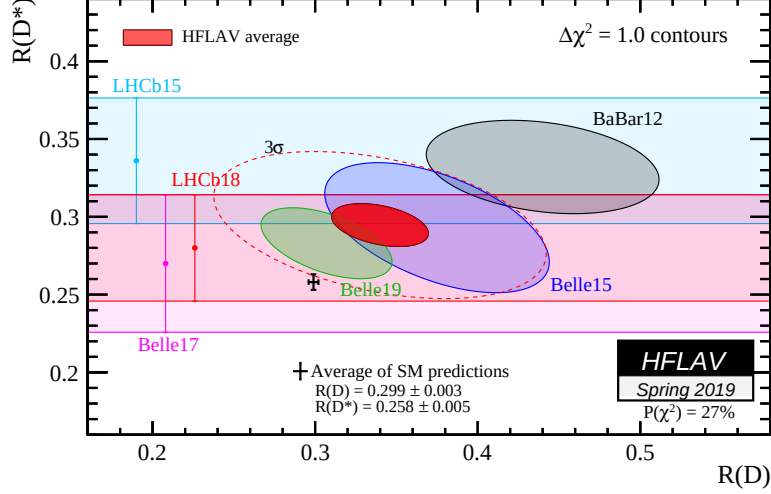


Figure 10: ■ Summary of measurements of $R(D^*)$ and $R(D)$ compiled by the Heavy Flavor Averaging Group [28]. The average of the experimental results is shown as red ellipsis with the dotted line indicating the 3σ region. The average of SM predictions is shown as black point.

ing the b hadrons B^0 , B^+ or B_c^+ decaying to $D^{(*)+}$, $D^{(*)0}$ and J/ψ have already been performed, while measurements using b -baryons decays, for example $\Lambda_b^0 \rightarrow \Lambda_c^+$, are foreseen. These measurements have been performed with different decay channels of the τ , either decaying to a muon or electron or to a different number of charged and neutral pions, as summarised in Table 2 for measurements performed at LHCb.

A recent summary of measurements of lepton universality in $b \rightarrow c\ell^-\bar{\nu}_\ell$ transitions is shown in Figure 10. Here, $R(D)$ signifies the charm hadrons D^+ and D^0 , while $R(D^*)$ signifies the charms hadrons D^{*+} and D^{*0} . Depending on the measurement, not all charm hadrons are included when determining $R(D^*)$ or $R(D)$, but they can be related by isospin considerations, where the spectator quark is exchanged, by $R(D^*) = R(D^{*0}) = R(D^{*+})$ and $R(D) = R(D^0) = R(D^+)$. Since the excited charm hadrons decay to their ground state emitting a photon or a pion, which can be missed, the decays used for measuring $R(D^*)$ can enter as a background contribution when measuring $R(D)$ ('feed-down' background) leading to a correlation when doing a combined measurement of both, which shows as an

ellipsis in Figure 10.

For the measurements performed at LHCb, reconstructing the τ either as $\tau \rightarrow \mu \nu_\mu \nu_\tau$ or as $\tau \rightarrow \pi^+ \pi^- \pi^+ \nu_\tau$ leads to a different analysis strategy. In the first case the ratio

$$R(D^*) = \frac{\mathcal{B}(B^0 \rightarrow D^{*-} \tau^+ \nu_\tau)}{\mathcal{B}(B^0 \rightarrow D^{*-} \mu^+ \nu_\mu)} \quad (27)$$

is measured directly by determining the relative yield of $B \rightarrow D^* \tau \nu_\tau$ to $B \rightarrow D^* \mu \nu_\mu$. The D^{*-} is reconstructed in the decay $D^* \rightarrow \bar{D}^0 \pi^-$ and the $\bar{D}^0$ in the decay $\bar{D}^0 \rightarrow K^+ \pi^-$. As the τ decays to a μ and additional neutrinos, the reconstructed final state particles are the same in both cases. The relative yield can then be determined from the $(\mu^+ K^+ \pi^- \pi^-)$ sample.

When reconstructing the τ in the hadronic decay mode $\tau \rightarrow \pi^+ \pi^- \pi^+ \nu_\tau$ the measured quantity is

$$R_{had}(D^*) = \frac{\mathcal{B}(B^0 \rightarrow D^{*-} \tau^+ \nu_\tau)}{\mathcal{B}(B^0 \rightarrow D^{*-} \pi^+ \pi^- \pi^+)} \quad (28)$$

The relative branching fraction is then determined by

$$R(D^*) = R_{had}(D^*) \times \frac{\mathcal{B}(B^0 \rightarrow D^{*-} \pi^+ \pi^- \pi^+)}{\mathcal{B}(B^0 \rightarrow D^{*-} \mu^+ \nu_\mu)}, \quad (29)$$

using the already measured values of $\mathcal{B}(B^0 \rightarrow D^{*-} \pi^+ \pi^- \pi^+)$ and $\mathcal{B}(B^0 \rightarrow D^{*-} \mu^+ \nu_\mu)$ as external inputs.

Light leptons

The universality of the light electron and muon can be tested at low energies using decays of pseudoscalar mesons, whose invariant mass is too low to decay into a tau lepton. The relative rates of the $X^- \rightarrow \ell^- \bar{\nu}_\ell$ decay have been measured for different mesons X and can be predicted by the SM, where the $X^- \rightarrow e^- \bar{\nu}_e$ decay is typically helicity suppressed. Kaon decays have been used by the NA62 collaboration [29] to measure

$$\frac{\Gamma_{K^- \rightarrow e^- \bar{\nu}_e}}{\Gamma_{K^- \rightarrow \mu^- \bar{\nu}_\mu}} = (2.488 \pm 0.007 \pm 0.007) \times 10^{-5}, \quad (30)$$

where the first uncertainty is statistical and the second systematic. This is found to be in agreement with the SM expectation of [30]

$$\left(\frac{\Gamma_{K^- \rightarrow e^- \bar{\nu}_e}}{\Gamma_{K^- \rightarrow \mu^- \bar{\nu}_\mu}} \right)^{SM} = (2.477 \pm 0.001) \times 10^{-5}. \quad (31)$$

For pion decays the corresponding ratio has been measured by the *PIENU* Collaboration to be [31]

$$\frac{\Gamma(\pi^+ \rightarrow e^+ \nu_e + \pi^+ \rightarrow e^+ \nu_e \gamma)}{\Gamma(\pi^+ \rightarrow \mu^+ \nu_\mu + \pi^+ \rightarrow \mu^+ \nu_\mu \gamma)} = (1.2344 \pm 0.0023 \pm 0.0019) \times 10^{-4}, \quad (32)$$

where the first uncertainty is statistical and the second systematic. This is consistent with the SM prediction of [30]

$$\left(\frac{\Gamma(\pi^+ \rightarrow e^+ \nu_e + \pi^+ \rightarrow e^+ \nu_e \gamma)}{\Gamma(\pi^+ \rightarrow \mu^+ \nu_\mu + \pi^+ \rightarrow \mu^+ \nu_\mu \gamma)} \right)^{SM} = (1.2352 \pm 0.0001) \times 10^{-4}. \quad (33)$$

The universality of the couplings g_μ and g_e of the light leptons to the W boson have also been tested using the $\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$ and $\tau^- \rightarrow e^- \bar{\nu}_e \nu_\tau$ decays [32]

$$\left(\frac{g_\mu}{g_e} \right) = 1.0036 \pm 0.0020, \quad (34)$$

which again is found to be consistent with the SM expectation.

Due to the agreement of experiment and SM prediction, it is usually assumed that BSM effects are negligible for semileptonic decays involving muons or electrons. This assumption has further been substantiated by the B-factories for b -meson decays, but single electron final states have not been included yet in measurements at hadron colliders. Defining the ratio $R(D^{*+})_{light}$ in a similar way to $R(D^*)$ as

$$R(D^{*+})_{light} = \frac{\mathcal{B}(\bar{B}^0 \rightarrow D^{*+} e^- \bar{\nu}_e)}{\mathcal{B}(\bar{B}^0 \rightarrow D^{*+} \mu^- \bar{\nu}_\mu)}, \quad (35)$$

it was measured by Belle to be $1.01 \pm 0.01(stat.) \pm 0.03(syst.)$ [33], where the first uncertainty is statistical and the second systematic, consistent with the SM expectation of one, since the muon and electron masses are small compared to the typical momentum transfers in this interaction so that their phase space factors are very similar.

As usual for relative branching fraction measurements, many systematic uncertainties are expected to cancel to first order and only effects due to the different behaviour of electrons and muons in the detector remain. In the case of the Belle measurement, the dominating uncertainty is due to the electron and muon identification. For a measurement at a hadron collider, additional uncertainties are related to the partial reconstruction of the decay. Due to Bremsstrahlung, affecting electrons more than muons, the reconstruction efficiencies can be different.

When selecting events with electrons or muons care must be taken not to introduce differences between the two samples, for example when triggering on the electron or muon.

I.3 Theoretical treatment

An effective field theory approach is typically used to make precise predictions of the observables of interest and to interpret potential deviations in a generic way. This is motivated by the fact that the energy scales involved in the weak decay of a hadron are much smaller than the mass of the W boson and in particular much smaller than the mass scale of potential BSM particles. In the following, the focus will be on semileptonic decays. An effective Hamiltonian $\mathcal{H}_{eff}$ for the $b \rightarrow c\ell^-\bar{\nu}_\ell$ can be written as

$$\mathcal{H}_{eff}(b \rightarrow c\ell^-\bar{\nu}_\ell) = \frac{4G_F}{\sqrt{2}}V_{cb} \sum_i \mathcal{C}_i \mathcal{O}_i, \quad (36)$$

where $\mathcal{C}_i$ are Wilson coefficients, encapsulating the effects at larger energies and short-distances, and the operators $\mathcal{O}_i$ describing the low energy, long-distance part. In the case of the SM, the relevant operator is the product of the two axial-vector currents:

$$\mathcal{O}_{VA} = (\bar{c}\gamma_\mu P_L b)(\bar{\ell}\gamma^\mu P_L \nu_\ell), \quad (37)$$

while the normalisation of (36) was chosen such that the corresponding coefficient is unity. In the SM, the coefficient and operator is the same for all three generations of leptons, while BSM models could add new operators or modify the coefficient for a particular flavour. Since the decay can be split into a purely leptonic part and a hadronic part, it is possible to further factorise the problem. Because hadrons are not point-like particles but composite objects consisting of quarks held together by the strong force, this part requires careful treatment as described in the next section.

I.3.1 Hadronic form factors

The transition $b \rightarrow c\ell^-\bar{\nu}_\ell$ can be described using the formalism of effective Hamiltonians. In this way, it can be separated into a leptonic and hadronic part. While the calculation of the leptonic part is easier, as the involved particles are point-like and the QED mediator does not self-interact, the hadronic part requires QCD interactions by (soft) gluon exchange to be taken into account. In this regime, QCD is non-perturbative as the strong coupling constant α_s is large and more involved approaches are necessary for a precise calculation. For example, in the lattice QCD approach space-time is discretised as a four-dimensional 'lattice' to perform numerical calculations. The precision of this approach is limited by the spacing of the lattice and the computational expense involved in reducing this spacing.

Alternatively, some approximation has to be made or symmetry arguments have to be employed to simplify the calculations, which is the approach for the calculations described below. The matrix elements involving hadrons, like

$$\langle H_c | \bar{c} \gamma_\mu (1 - \gamma_5) b | H_b \rangle \quad (38)$$

will be described using form factors encapsulating the effects of gluon exchange and the presence of virtual quark-antiquark pairs, which is difficult to calculate as this lies in the non-perturbative regime of QCD. Since these form factors are not predicted by this approach, they have to be measured for a certain decay. In some cases, it is possible to calculate form factors using lattice QCD.

A very fruitful approach to obtain a parametrisation for the form factors has been Heavy Quark Effective Theory (HQET) [34], exploiting the fact that the b and c -quarks are much heavier than the involved spectator quarks. The partition of quarks in 'heavy' and 'light' is typically done according to the scale $m_u, m_d, m_s \ll \Lambda_{\text{QCD}} \ll m_c, m_b, m_t$, where $\Lambda_{\text{QCD}} \approx 200$ MeV is the hadronisation scale of QCD, where the strong coupling constant starts to become large so that the perturbative approach breaks down. The basic assumption of HQET in the case of a meson consisting of a heavy and a light quark is that the heavy quark has the same velocity as the meson and the light quark only sees it as a static source of colour. Spin interactions between the heavy quark and light degrees of freedom (the gluon exchanged between the two quarks) are suppressed by the heavy quark mass. This means that the heavy quark can be replaced by another heavy quark of different flavour (in practice exchanging the b - by a c -quark) and its spin can be flipped without any difference to the light degrees of freedom (spin-flavour symmetry) if the velocity remains unchanged. If the velocity is changed, the light degrees of freedom need to 'follow' the heavy quark, resulting in suppression by a so-called form factor.

For the concrete case of a $B \rightarrow D^*$ transition, four form factors are necessary. By splitting the hadronic current into vector and axial-vector parts, these form factors are defined by [35]

$$\begin{aligned} \langle D^*(p_{D^*}, \epsilon_\alpha) | \bar{c} \gamma_\mu b | B(p_B) \rangle &= \frac{2iV(q^2)}{m_B + m_{D^*}} \epsilon_{\mu\nu\alpha\beta} \epsilon^{*\nu} p_B^\alpha p_{D^*}^\beta, \\ \langle D^*(p_{D^*}, \epsilon_\alpha) | \bar{c} \gamma_\mu \gamma_5 b | B(p_B) \rangle &= 2m_{D^*} A_0(q^2) \frac{\epsilon^* \cdot q}{q^2} q_\mu + (m_B + m_{D^*}) A_1(q^2) \left(\epsilon_\mu^* - \frac{\epsilon^* \cdot q}{q^2} q_\mu \right) \\ &\quad - A_2(q^2) \frac{\epsilon^* \cdot q}{m_B + m_{D^*}} \left((p_B + p_{D^*})_\mu - \frac{m_B^2 - m_{D^*}^2}{q^2} q_\mu \right), \end{aligned} \quad (39)$$

where ϵ is the polarisation vector of the D^* and the form factors $V(q^2)$, $A_0(q^2)$, $A_1(q^2)$ and $A_2(q^2)$ depend on the square of the four-momentum q^2 carried away by the vector boson. The factor $A_0(q^2)$ only contributes to the term suppressed by helicity and is therefore only relevant for decays to the heavy τ lepton.

In the heavy quark approximation, where the spin-flavour symmetry holds, the four form factors can be expressed by a single function, called the Isgur-Wise function $h_{A_1}(w)$ [36] [37], which describes the transition probability from one rest-frame to the other by a Lorentz boost by encapsulating the overlap of the light degrees of freedom of two heavy quarks moving with v_B and v_{D^*} . The four-velocity transfer $w = v_B \cdot v_{D^*}$ is related to q^2 by

$$q^2 = m_B^2 + m_{D^*}^2 - 2m_B m_{D^*} w. \quad (40)$$

As the heavy quark mass is finite, the spin-flavour symmetry is not exact and short-distance effects have to be taken into account, complicating the picture. The form factors are then written as

$$\begin{aligned} A_0(w) &= \frac{R_0(w)}{r_{D^*}} h_{A_1}(w), \\ A_1(w) &= \frac{w+1}{2} r_{D^*} h_{A_1}(w), \\ A_2(w) &= \frac{R_2(w)}{r_{D^*}} h_{A_1}(w), \\ V(w) &= \frac{R_1(w)}{r_{D^*}} h_{A_1}(w), \end{aligned} \quad (41)$$

using the form factor ratios $R_0(w)$, $R_1(w)$ and $R_2(w)$ to express the difference between the form factors and defining the factor $r_{D^*} = \frac{2\sqrt{m_B m_{D^*}}}{(m_B + m_{D^*})}$.

These functions can further be parametrised around the point of zero recoil ($w = 1$) by the so-called CLN parametrisation [38] [35]:

$$\begin{aligned} h_{A_1}(w) &= h_{A_1}(1) \left[1 - 8\rho^2 z + (53\rho^2 - 15)z^2 - (231\rho^2 - 91)z^3 \right], \\ R_1(w) &= R_1(1) - 0.12(w-1) + 0.05(w-1)^2, \\ R_2(w) &= R_2(1) + 0.11(w-1) - 0.06(w-1)^2, \\ R_0(w) &= R_0(1) - 0.11(w-1) + 0.01(w-1)^2, \end{aligned} \quad (42)$$

using $z = \frac{\sqrt{w+1}-\sqrt{2}}{\sqrt{w+1}+\sqrt{2}}$ and introducing the free parameters $h_{A_1}(1)$, $R_0(1)$, $R_1(1)$, $R_2(1)$ and ρ^2 , which can be measured in the $B^0 \rightarrow D^* \ell \nu$ decay, for example re-

cently by Belle [33] to be

$$\begin{aligned}\rho^2 &= 1.106 \pm 0.031 \pm 0.007, \\ R_1(1) &= 1.229 \pm 0.028 \pm 0.009, \\ R_2(1) &= 0.852 \pm 0.021 \pm 0.006.\end{aligned}\tag{43}$$

The parameter $R_0(1)$ was not measured since the analysis only used transitions to electrons and muons, where the helicity-suppressed amplitude, the only one $R_0(1)$ contributes to, is negligible. Since the normalisation factor $h_{A_1}(w)$ appears in all factors it will cancel when considering the relative branching fraction and is not important when predicting the ratio of interest. For calculating ratios involving a τ , it is necessary to also know $R_0(1)$, which can be approximated from leading order perturbative estimations using the experimental results for the other factors to be $R_0(1) = 1.14$ [35], where an uncertainty estimate of 10% is applied to account for higher order corrections.

An alternative approach is the BGL parametrisation [39], where the form factors are parametrised by a generic expansion in z following the form

$$f(z) = \frac{1}{P(z)\phi(z)} \sum_{n=0}^{\infty} a_n^f z^n.\tag{44}$$

so that the four form factors are parametrised by

$$\begin{aligned}f(z) &= \frac{1}{P_{1+}(z)\phi_f(z)} \sum_{n=0}^{\infty} a_n^f z^n, \\ g(z) &= \frac{1}{P_{1-}(z)\phi_g(z)} \sum_{n=0}^{\infty} a_n^g z^n, \\ \mathcal{F}_1(z) &= \frac{1}{P_{1+}(z)\phi_{\mathcal{F}_1}(z)} \sum_{n=0}^{\infty} a_n^{\mathcal{F}_1} z^n, \\ \mathcal{F}_2(z) &= \frac{1}{P_2(z)\phi_{\mathcal{F}_2}(z)} \sum_{n=0}^{\infty} a_n^{\mathcal{F}_2} z^n,\end{aligned}\tag{45}$$

with the Blaschke factor $P(z)$ and the weighting factor $\phi(z)$. The Blaschke factor depends on the number of B_c states below the BD^* mass threshold. When measuring the parameters of this series, the expansion in z^n has to be truncated at some point. Usually including terms up to $n = 2$ or $n = 3$ is sufficient to describe the

data well. These parameters have been measured to be [33]

$$\begin{aligned}
\tilde{a}_0^f \times 10^3 &= -0.506 \pm 0.004 \pm 0.008 \\
\tilde{a}_1^f \times 10^3 &= -0.65 \pm 0.17 \pm 0.09 \\
\tilde{a}_1^{\mathcal{F}_1} \times 10^3 &= -0.270 \pm 0.064 \pm 0.023, \\
\tilde{a}_2^{\mathcal{F}_1} \times 10^3 &= +3.27 \pm 1.25 \pm 0.45 \\
\tilde{a}_0^g \times 10^3 &= -0.929 \pm 0.018 \pm 0.013,
\end{aligned} \tag{46}$$

where the parameters are scaled by $|V_{cb}|$ and an electroweak correction η_{EW}

$$\tilde{a}_i^b = |V_{cb}| \eta_{EW} a_i^b. \tag{47}$$

As for the CLN parametrisation, one of these form factors, $\mathcal{F}_2(z)$, is only sensitive to the helicity-suppressed amplitude and was not measured here and needs to be predicted, for example described in [40].

I.3.2 Standard Model prediction

It is instructive to briefly repeat the calculation of the expected value for the $R(D^*)$ ratio in the SM, which uses these form factor measurements. From this, it can be understood how new physics could enter, modifying the decay rate to tauons, while leaving the rate to electrons and muons almost untouched as present experimental results suggest. From the effective Hamiltonian describing the $b \rightarrow c \tau \nu_\tau$ transition

$$\mathcal{H}_{eff} = \frac{4G_F}{\sqrt{2}} (\bar{c} \gamma_\mu P_L b) (\bar{\tau} P_L \nu_\tau) + h.c. \tag{48}$$

the double-differential decay rate of $B^0 \rightarrow D^{*-} \ell^+ \nu$ can be written [35] as

$$\begin{aligned}
\frac{d^2 \Gamma_\tau}{dq^2 d \cos \theta} &= \frac{G_F^2 |V_{cb}|^2 |\mathbf{p}| q^2}{256 \pi^3 m_B^2} \left(1 - \frac{m_\tau^2}{q^2} \right)^2 \times \\
&\left[(1 - \cos \theta)^2 |H_{++}|^2 + (1 + \cos \theta)^2 |H_{--}|^2 + 2 \sin^2 \theta |H_{00}|^2 + \right. \\
&\left. \frac{m_\tau^2}{q^2} \left(\sin^2 \theta (|H_{++}|^2 + |H_{--}|^2) + 2 |H_{0t} - H_{00} \cos \theta|^2 \right) \right],
\end{aligned} \tag{49}$$

Table 3: Recent SM predictions for the value of $R(D^*)$.

Reference	Predicted $R(D^*)$
[41]	0.257 ± 0.003
[42]	0.260 ± 0.008
[40]	0.257 ± 0.005

where θ is the angle between the D^* and τ three-momenta in the $\tau - \bar{\nu}_\tau$ rest frame and using the helicity amplitudes

$$\begin{aligned}
H_{\pm\pm}(q^2) &= (m_B + m_{D^*})A_1(q^2) \mp \frac{2m_B}{m_B + m_{D^*}} |\mathbf{p}| V(q^2), \\
H_{00}(q^2) &= \frac{1}{2m_{D^*}\sqrt{q^2}} \left[(m_B^2 - m_{D^*}^2 - q^2)(m_B + m_{D^*})A_1(q^2) - \frac{4m_B^2 |\mathbf{p}|^2}{m_B + m_{D^*}} A_2(q^2) \right] \\
H_{0t}(q^2) &= \frac{2m_B |\mathbf{p}|}{\sqrt{q^2}} A_0(q^2),
\end{aligned} \tag{50}$$

which depend on the the form factors $V(q^2)$, $A_0(q^2)$, $A_1(q^2)$ and $A_2(q^2)$ introduced earlier.

Integrating (49) over θ , one can obtain

$$\frac{d\Gamma_\tau}{dq^2} = \frac{G_F^2 |V_{cb}|^2 |\mathbf{p}| q^2}{96\pi^3 m_B^2} \left(1 - \frac{m_\tau^2}{q^2} \right)^2 \left[(|H_{++}|^2 + |H_{--}|^2 + |H_{00}|^2) \left(1 + \frac{m_\tau^2}{2q^2} \right) + \frac{3}{2} \frac{m_\tau^2}{q^2} |H_{0t}|^2 \right], \tag{51}$$

which is now used to obtain the ratio of differential decay rate between a tau and a light lepton in the final state:

$$R_{D^*}^*(q^2) = \frac{d\Gamma_\tau/dq^2}{d\Gamma_\ell/dq^2} = \left(1 - \frac{m_\tau^2}{q^2} \right)^2 \left[\left(1 + \frac{m_\tau^2}{2q^2} \right) + \frac{3}{2} \frac{m_\tau^2}{q^2} \frac{|H_{0t}|^2}{|H_{++}|^2 + |H_{--}|^2 + |H_{00}|^2} \right], \tag{52}$$

where the mass of the light lepton is neglected. By integrating over q^2 and using the measured form factors, one can obtain SM predictions for $R(D^*)$ summarised in Table 3, with the average

$$R(D^*)^{avg} = 0.258 \pm 0.005. \tag{53}$$

By substituting the τ by a light lepton in (52) the ratio between two light leptons, i.e. an electron and muon, becomes unity. Since the term including $|H_{0t}|^2$ only becomes relevant for heavy lepton masses, BSM effects modifying this helicity amplitude would change the relative branching ratio between taus and the light leptons, but not the ratio between an electron and a muon.

I.4 Explaining lepton universality violation

It would be a desirable feature of a BSM theory to be able to consistently explain the observed deviations from the SM for penguin decays as well as those measured in semileptonic decays. At the same time, other constraints, like the lifetimes of B -hadrons and the measured upper limits on lepton flavour violating decays have to be respected, as BSM contributions could have an impact on them. There are various attempts to extend or augment the SM to be able to account for the observed deviations. As an example, the existence of leptoquarks are hypothesised, which are particles coupling to quarks and leptons. As a consequence lepton flavour is not conserved anymore. In this way, tests for lepton universality and observed tensions there serve as a motivation of the search for lepton-flavour violating decays [43]. Another possible BSM scenario is the introduction of additional heavy bosons. Figure 11 shows examples of Feynman diagrams where BSM particles enter.

However, it is quite difficult to devise a specific model, which explains all tensions while leaving other experimentally confirmed expectations intact. As mentioned earlier, instead of regarding a specific BSM model the experimental results can be analysed using the effective Hamiltonian approach in a generic way. Global fits, which include different observations, can be performed and it can be tested if introducing additional Wilson coefficients or modifying the SM ones can yield a better description of the data. In the following, the results of such fits in the $b \rightarrow s\ell^+\ell^-$ and $b \rightarrow c\ell\nu$ cases are discussed.

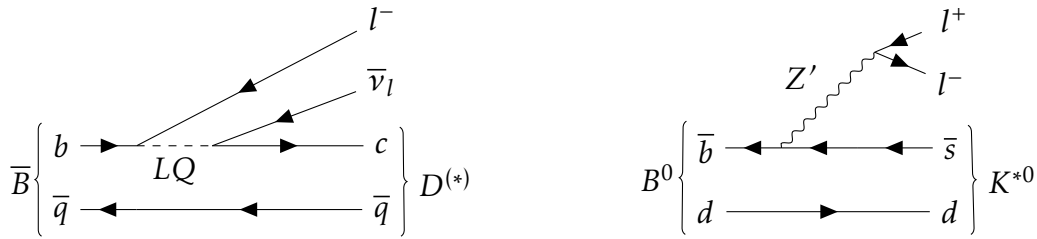


Figure 11: Example of a leptoquark contribution to the $b \rightarrow c\ell\nu$ transition (left) and a BSM neutral gauge boson Z' to the $b \rightarrow s\ell^+\ell^-$ transition (right).

Penguin decays

The experimental results summarised earlier can be used in global fits to Wilson coefficients including terms representing non-SM contributions to obtain a total significance of the potential BSM effects. Allowing the contribution of some combination of BSM Wilson coefficients can improve the description of the data significantly by modifying the $b \rightarrow s\mu^+\mu^-$ currents. For example, the analysis

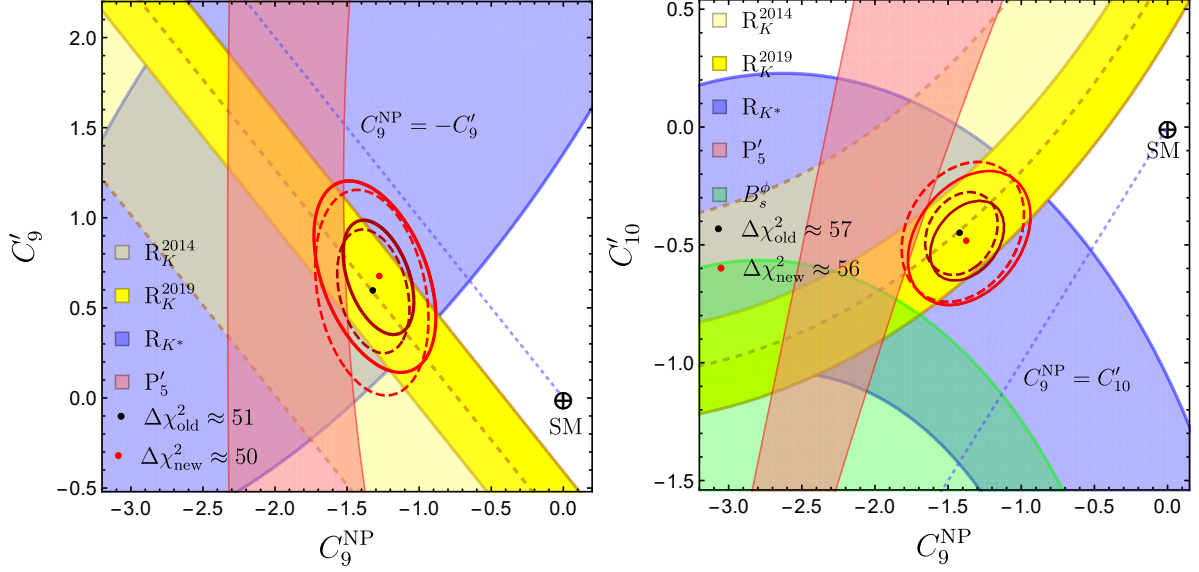


Figure 12: ■ Comparison of the fitted Wilson coefficients with the constraints obtained from experiment for the scenarios (C_9^{NP}, C'_9) and $(C_9^{\text{NP}}, C'_{10})$ [44].

presented in [44] uses as input the observed $R(K)$, $R(K^*)$, P'_5 and $B_s^0 \rightarrow \phi \mu^+ \mu^-$ branching ratio measurements discussed earlier. Without going into detail here, the following BSM Wilson coefficient-operator combinations are considered:

$$\begin{aligned}
 &C_9^{\text{NP}}(\bar{s}\gamma^\mu P_L b)(\bar{\mu}\gamma_\mu \mu), \\
 &C_{10}^{\text{NP}}(\bar{s}\gamma^\mu P_L b)(\bar{\mu}\gamma_\mu \gamma_5 \mu), \\
 &C'_9(\bar{s}\gamma^\mu P_R b)(\bar{\mu}\gamma_\mu \mu), \\
 &C'_{10}(\bar{s}\gamma^\mu P_R b)(\bar{\mu}\gamma_\mu \gamma_5 \mu),
 \end{aligned} \tag{54}$$

where the C_9 and C_{10} terms also appear in the SM Hamiltonian. These terms are used in the fit in pairs of two, testing six scenarios in total. It is seen that adding the combinations $(C_9^{\text{NP}}, C'_9) = (-1.28, +0.68)$ or $(C_9^{\text{NP}}, C'_{10}) = (-1.38, -0.48)$ to the Hamiltonian improves the description the most compared to the SM. These fitted values are compared with the measurement constraints in Figure 12.

Semileptonic decays

One way to introduce new physics contributions in the $b \rightarrow c \tau \nu_\tau$ transition within the effective approach described earlier is by introducing new operators which modify the helicity-suppressed amplitude. In this way, the heavier the lepton, the more the influence of new physics contributions are apparent, matching the experimental observations. It is clear that with the current experimental results tensions with the SM only reveal themselves by looking at $R(D)$ and $R(D^*)$

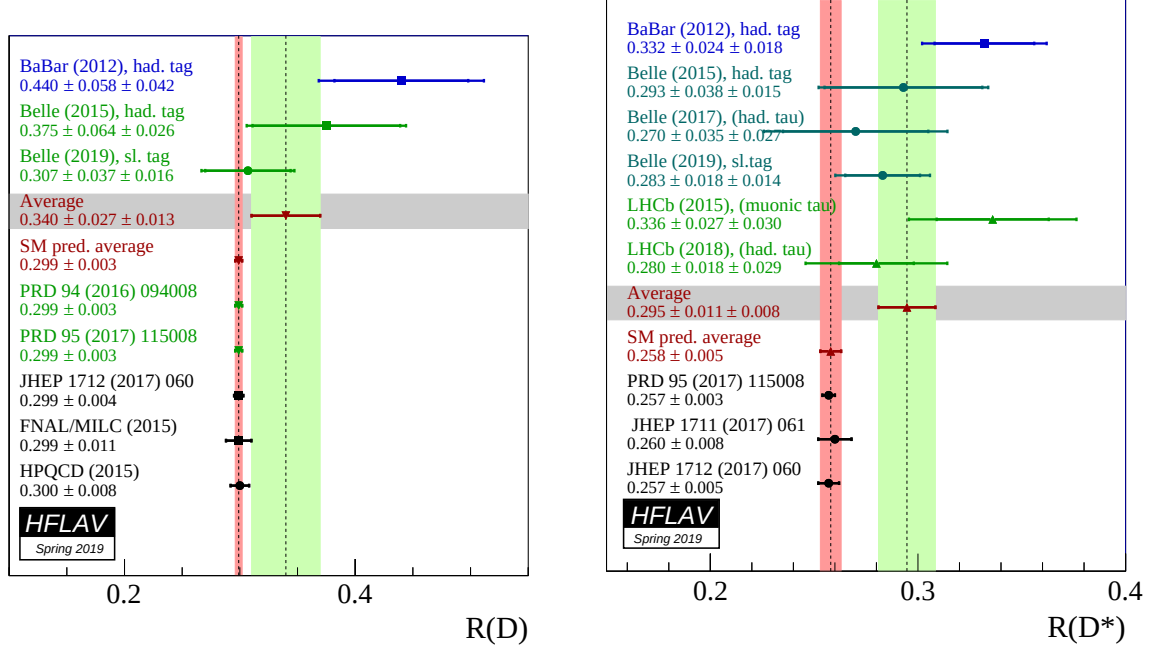


Figure 13: ■ Summary of measurements and SM predictions of $R(D^*)$ and $R(D)$ compiled by the Heavy flavour averaging group [28].

together, taking correlations into account if a combined measurement was done. Looking only at the single ratios, shown in Figure 13, no large tension with the SM is observed. Maybe more worryingly, one could get the impression that the more recent the measurement, the closer one gets to the SM prediction, especially in the case of $R(D)$. The latest combined measurement from 2019 by Belle, which has the smallest uncertainties of the measurement included in the averaged value, is very consistent with the SM prediction. In fact, in the shown update of the HFLAV averages in the spring of 2019, the tension was reduced from about 4σ to about 3σ . One could interpret this in several ways. One could argue that lepton flavour universality holds in the SM and the 2012 BaBar measurement, which is one of the main drivers of the tension underestimated its uncertainties, overlooked a systematic effect shifting the result or was just an unlucky fluctuation. From the point of BSM, this would be very disappointing.

On the other hand, one could argue that a tension remains, which should motivate further measurements, and together with the observations in electro-weak penguin decays, this chapter is far from closed. In the end, it is evident that more measurements are required to finally confirm if the SM expectation is violated. Especially exciting here are the ongoing measurements using LHCb data and the data collected by the currently running Belle II experiment, which should bring more clarity and resolve this issue.

In the effective field theory approach, global fits to the Wilson coefficients are made to find the values, which best describe the data. In many cases, only one or two coefficients are allowed to vary. For the study of $b \rightarrow c\ell\nu_\ell$ transitions, it is commonly assumed that new physics only contributes to the τ decay mode, as the experimental evidence does not show deviations in the muon or electron modes. From measurements of $R^{\mu/e}(D)$ and $R^{\mu/e}(D^*)$, consistent with the SM expectations, it seems unlikely that there are BSM contributions to the $b \rightarrow c\mu\bar{\nu}_\tau$ transition, so that it is well motivated to try to explain the tensions by only looking at the τ currents.

To describe the data and fit for BSM effects, the effective Hamiltonian defined in (36) can be expanded in a generic way as [45]

$$H_{eff} = \frac{4G_F}{\sqrt{2}} V_{cb} \left[O_{V_L} + \frac{\sqrt{2}}{4G_F V_{cb}} \frac{1}{\Lambda^2} \left\{ \sum_i (C_i O_i + C'_i O'_i + C''_i O''_i) \right\} \right], \quad (55)$$

where O_{V_L} is the SM operator and Λ is the BSM scale. The possible operators are summarised in Table 4. The O'_i and O''_i operators correspond to quark-lepton currents, which could be generated in leptoquark models.

Table 4: ■ All possible four-fermion operators that can contribute to $b \rightarrow c\tau\bar{\nu}$ transition assuming only left-handed neutrinos exist. Taken from [46].

	Operator
O_{V_L}	$(\bar{c}\gamma_\mu P_L b)(\bar{\tau}\gamma^\mu P_L \nu)$
O_{V_R}	$(\bar{c}\gamma_\mu P_R b)(\bar{\tau}\gamma^\mu P_L \nu)$
O_{S_R}	$(\bar{c}P_R b)(\bar{\tau}P_L \nu)$
O_{S_L}	$(\bar{c}P_L b)(\bar{\tau}P_L \nu)$
O_T	$(\bar{c}\sigma^{\mu\nu} P_L b)(\bar{\tau}\sigma_{\mu\nu} P_L \nu)$
O'_{V_L}	$(\bar{\tau}\gamma_\mu P_L b)(\bar{c}\gamma^\mu P_L \nu)$
O'_{V_R}	$(\bar{\tau}\gamma_\mu P_R b)(\bar{c}\gamma^\mu P_L \nu)$
O'_{S_R}	$(\bar{\tau}P_R b)(\bar{c}P_L \nu)$
O'_{S_L}	$(\bar{\tau}P_L b)(\bar{c}P_L \nu)$
O'_T	$(\bar{\tau}\sigma^{\mu\nu} P_L b)(\bar{c}\sigma_{\mu\nu} P_L \nu)$
O''_{V_L}	$(\bar{\tau}\gamma_\mu P_L c^c)(\bar{b}^c\gamma^\mu P_L \nu)$
O''_{V_R}	$(\bar{\tau}\gamma_\mu P_R c^c)(\bar{b}^c\gamma^\mu P_L \nu)$
O''_{S_R}	$(\bar{\tau}P_R c^c)(\bar{b}^c P_L \nu)$
O''_{S_L}	$(\bar{\tau}P_L c^c)(\bar{b}^c P_L \nu)$
O''_T	$(\bar{\tau}\sigma^{\mu\nu} P_L c^c)(\bar{b}^c\sigma_{\mu\nu} P_L \nu)$

As an example, the new physics contributions are fitted in [46] using the CLN form factor parametrisation and the BSM scale $\Lambda = 1$ TeV. As input the then-available measurements of $R(D^*)$, $R(D)$ and $R(J/\psi)$ and of the τ polarisation

are used, also taking into account a constraint on the upper limit of $BR(B_c^+ \rightarrow \tau \bar{\nu}_\tau) < 0.1$ [47]. One particularly appealing result is that the relatively simple shift in the contribution of the SM operator $O_{V\tau} = (\bar{c}\gamma_\mu P_L b)(\bar{\tau}\gamma_\mu P_L \nu)$ by a coefficient of value $C_{V_L} = 0.149 \pm 0.032$ describes the measured values well. In the study presented this goes however with the caveat that the predicted value of $R(J/\psi)$ using this fit result is smaller than the measured one. But since the experimental uncertainties are fairly large, this tension is not significant and can only be resolved by more precise measurements. In addition, three alternative solutions were identified. One tensor solution with $C_T = 0.516 \pm 0.015$, one with $C_{S_L}'' = -0.526 \pm 0.102$ and one with both $C_{V_L} = -1.286$ and $C_{V_R} = 1.512$. It is pointed out that the tensor contribution can be distinguished from the other solution by more precise measurements of $R(J/\psi)$ and the τ polarisation. To differentiate between the other viable results angular analysis and measuring the forward-backward asymmetry is necessary [48].

In an update to this work [49], which now includes the 2019 Belle measurement of $R(D^{(*)})$ and a measurement of the D^* polarisation fraction $f_L^{D^*}$, only the C_{V_L} solution remains with a value of $C_{V_L} = 0.10 \pm 0.02$. It should be noted that here only combinations of 'similar' operators were allowed, meaning that the scalar, vector and tensor operators are not mixed when allowing two coefficients to vary. In the effective field theory picture, the combination of 'non-similar' operators corresponds to the contribution of more than one BSM particle with different spins to the transition. If this mixing is allowed, additional combinations of two coefficients can describe the data. It is argued that a precise measurement of the branching fraction $\mathcal{B}(B_c^+ \rightarrow \tau \bar{\nu})$ would allow distinguishing between these solutions.

Chapter II

The LHCb experiment

“»Es ist ein eigentümlicher Apparat«,
sagte der Offizier zu dem
Forschungsreisenden und überblickte mit
einem gewissermaßen bewundernden
Blick den ihm doch wohl bekannten
Apparat.”

Franz Kafka, *In der Strafkolonie*

The LHCb detector is built, maintained, operated and the recorded data is analysed by an international collaboration of over 1000 people. It is situated at the Large Hadron Collider (LHC) operated by the European Organization for Nuclear Research (CERN). In the following, an overview of the LHC and the detector components of LHCb are given.

Since the data sample used in the presented analysis was taken exclusively during Run 2 in the years 2016, 2017 and 2018, the description of the detector and trigger focuses on that period, with changes and improvements compared to Run 1 (2011-2012) briefly described where necessary.

II.1 The Large Hadron Collider

The LHC is a circular proton-proton collider. It has been running in the years 2009-2013 (Run 1) and 2015-2018 (Run 2). After starting with low-energy proton beams at the end of 2009, the beam energy was ramped up so that the first proton-proton collisions with a centre-of-mass energy $\sqrt{s}$ of 7 TeV could be produced in March 2010. The LHC continued operating at this $\sqrt{s}$ in 2011, while it was increased for the 2012 period to 8 TeV. The first months of 2013 were used for proton-lead and lead-lead collisions, which were followed by the first long shut-down. When restarting the LHC in 2015 a centre-of-mass energy of 13 TeV could

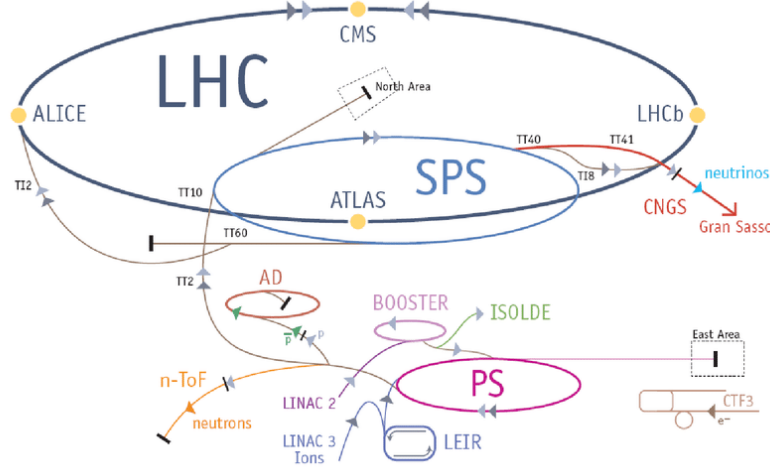


Figure 14: ■ Accelerator chain feeding the LHC [50].

be reached, which were used for the remainder of Run 2 until the LHC was shut down for maintenance at the end of 2018.

The protons pass through several pre-accelerators and are injected into the LHC ring with an energy of 450 GeV, where they are accelerated to their final collision energy. A scheme of the different parts involved in the accelerator chain is shown in Figure 14. The protons are accumulated in bunches and brought to collision at several places ('interaction points') where the different detectors are installed. In total seven experiments record and analyse the data generated by these collisions. The 'general-purpose' detectors ATLAS and CMS, responsible for observing the Higgs boson in 2012, ALICE, optimised for studying heavy-ion collisions, and LHCb, and three 'smaller' experiments (TOTEM, LHCf and MoEDAL). In addition to the proton-proton collision program, heavy-ion beams can also be accelerated and brought to collision. The proton bunches are spaced in time with a distance of 25 ns resulting in a bunch-crossing rate of 40 MHz. Since some of the bunches must be empty the average collision rate is about 30 MHz. To increase the interaction rate the beams are squeezed at the interaction points. In contrast to other experiments like ATLAS and CMS the luminosity at LHCb is kept constant at a lower value by shifting the beam focus to counteract the depletion of the circulating protons. In this way, the average number of inelastic pp interactions per collision is kept around $\mu = 1.1$ during Run 2. The total pp collision rate also includes elastic scattering, which is of no interest in this analysis. The instantaneous luminosity reached at LHCb during Run 2 was around $4 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$.

II.2 The LHCb detector

The LHCb detector is a forward-facing spectrometer, primarily envisioned for the precise study of CP-violation, but by now with a physics program wide enough to rightly deserve the description as a 'general purpose detector in the forward direction'. Its design is optimised for detecting the decay products of charm and bottom hadrons produced in a pp collision with a boost along the beam axis. As described in more detail below, its basic components are a tracking system up- and downstream of a magnet, whose bending of charged particle trajectories allows to measure their momenta, a tracker close to the interaction region to reconstruct the pp collision points, the RICH system and calorimeters for particle identification and the muon stations for muon reconstruction and identification.

A scheme of the detector and the definition of the coordinate system used is shown in Figure 15. The positive z -axis is aligned with the beam axis in direction of the detector with the origin being placed at the centre of the interaction region. The remaining horizontal and the vertical directions are described by the x - and y -axis, respectively, to form a right-handed coordinate system. The commonly used transverse plane is spanned by x and y . Additionally, the coordinates r , ϕ and θ are utilised. The angles ϕ and θ are the angles in the transverse plane and with respect to the z -axis, while the radius r is the radius in the x - y -plane. Another commonly used variable is the pseudorapidity η defined by the angle θ by $\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]$. Expressed in this way, the LHCb detector has a angular coverage of $2 < \eta < 5$.

The forward-facing layout was chosen since the main mechanism of $b\bar{b}$ production at the collision energies at LHCb is gluon-gluon fusion, where the $b\bar{b}$ -pair is boosted along the beamline since the momenta of the two interacting gluons from each proton are asymmetric, as illustrated in Figure 16. The lifetime of these boosted b hadrons is large enough that they decay at a discernible distance at the order of a centimetre from the original proton-proton interaction. For this reason, the first level of the software trigger is based on displaced topologies, as described below.

A proton-proton bunch crossing producing particles traversing the detector is called an *event*. The interaction point of a collision is called a *primary vertex* (PV), displaced vertices of particles with large enough lifetime are called secondary vertices (SV). In Run 2 events are produced at a rate of up to 30 MHz with about one PV on average per event. In the following, a brief description of the detector components and chosen hardware is given.

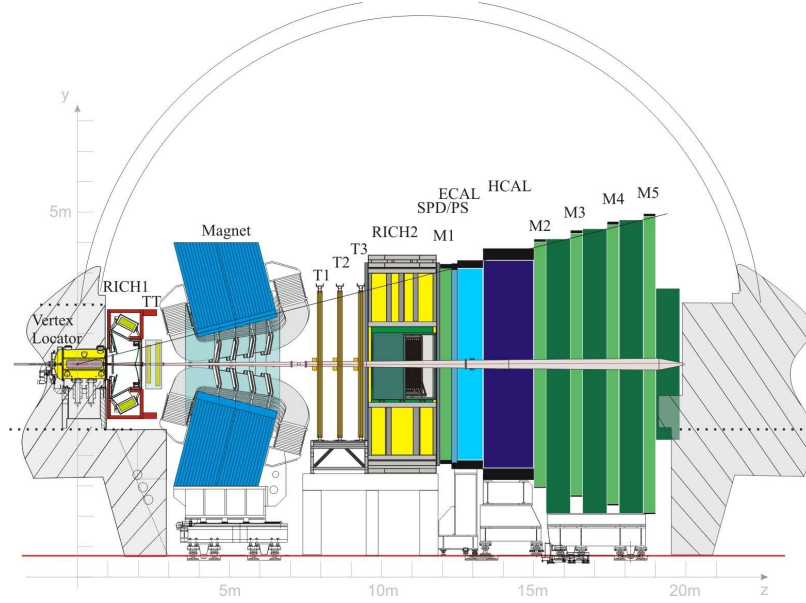


Figure 15: ■ Schematic side view of the LHCb detector. The coordinate system is defined by the horizontal z -axis from left to right and the vertical y -axis from bottom to top. Particles are produced on the left side and traverse the detector elements [51].

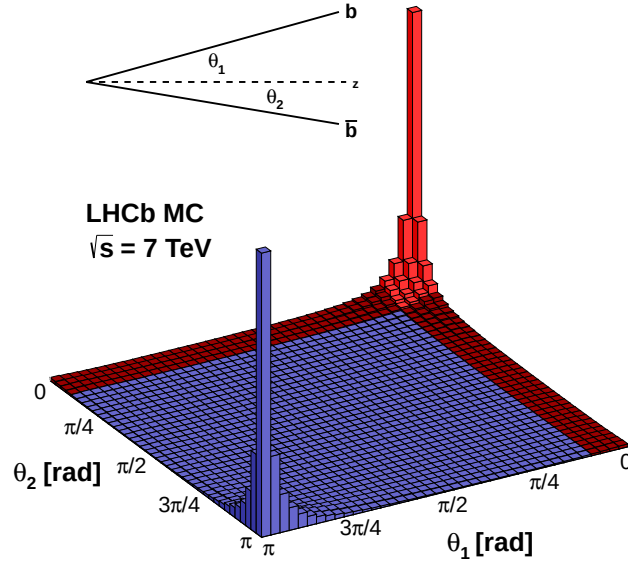


Figure 16: ■ Angular distribution of the $b\bar{b}$ pair produced in pp-collisions at LHCb [52].

Vertex Locator

The Vertex Locator (VELO) consists of 21 silicon modules arranged along the beam axis, where each module is made up of an R-sensor and a ϕ -sensor to provide a measurement of the corresponding radial coordinates. The VELO is designed to

reconstruct the position of the proton-proton collision with the highest possible resolution. For this purpose, the modules are placed at a distance of about 8 mm from the beamline. This reduces the distance a charged particle travels to its first measured point and therefore improves the resolution of the reconstructed trajectory. The VELO modules can be opened to protect the sensors from being damaged before the beams are stabilised and closed again once safe beam conditions are reached. Because of this, the VELO modules can be slightly displaced between different fills, which has to be corrected by an alignment procedure.

The modules are contained within a box with a secondary vacuum separated from the beam vacuum by a thin corrugated aluminium sheet (RF-foil) shown in Figure 17. The RF-foil introduces additional material traversed by the particles, up to about a 10% increase in radiation length for particles travelling along the beam axis passing through many layers of the foil.

Particles produced in the interaction region traverse the VELO in approximately straight lines since the remaining magnetic field is very weak. Charged particles deposit charge when passing through the detector planes ('hits'), which are used to reconstruct the particle trajectories. When traversing the material, the charged particles undergo multiple scattering depending on their momentum, the thickness and properties of the material. The design and layout of the VELO take several considerations into account. The innermost planes are placed close to the interaction region to precisely reconstruct the PVs. The number of detector planes needs to be sufficient for charged particles to leave enough hits that they can be reconstructed. At the same time, the overall traversed material has to be kept small to reduce the amount of multiple scattering. A scheme of the VELO layout is shown in Figure 18.

Upstream tracker

The next tracker upstream of the magnet is the Tracker Turicensis (TT). It consists of four stations, each made up of four layers of silicon microstrips in an $(x-u-v-x)$ arrangement, meaning that the two inner layers are rotated by an angle of $\pm 5^\circ$ with respect to the vertical x -layers ($0^\circ, +5^\circ, -5^\circ, 0^\circ$). A residual magnetic field in the TT allows for a rough momentum estimation before extrapolating the tracks to the downstream tracking layers. As their naming implies, the x -layers provide a measurement of the x -position of a hit, which then can be used in conjunction with the information from the tilted layers to get a y -position. The layout of the TT with the tilted inner layers is shown in Figure 19. This so-called stereo layout was chosen to be able to ensure mechanical stability and to be able to place read-out

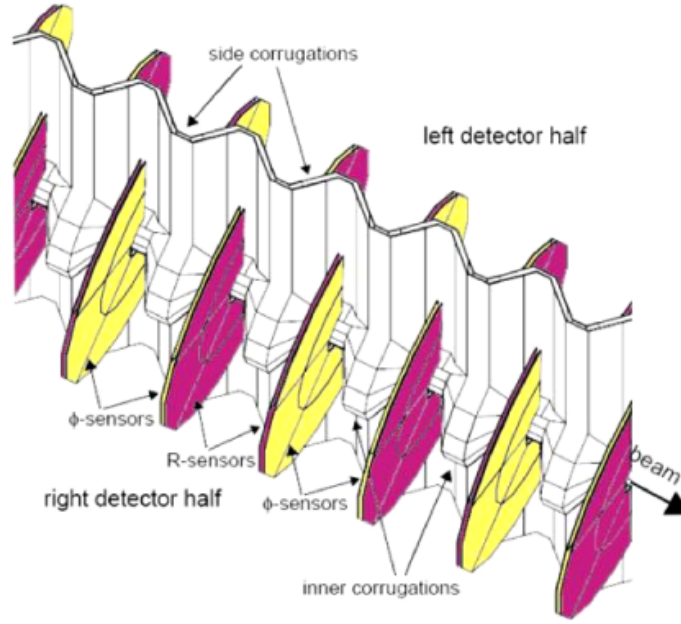


Figure 17: ■ Scheme of the RF-foil separating the two VELO halves [51].

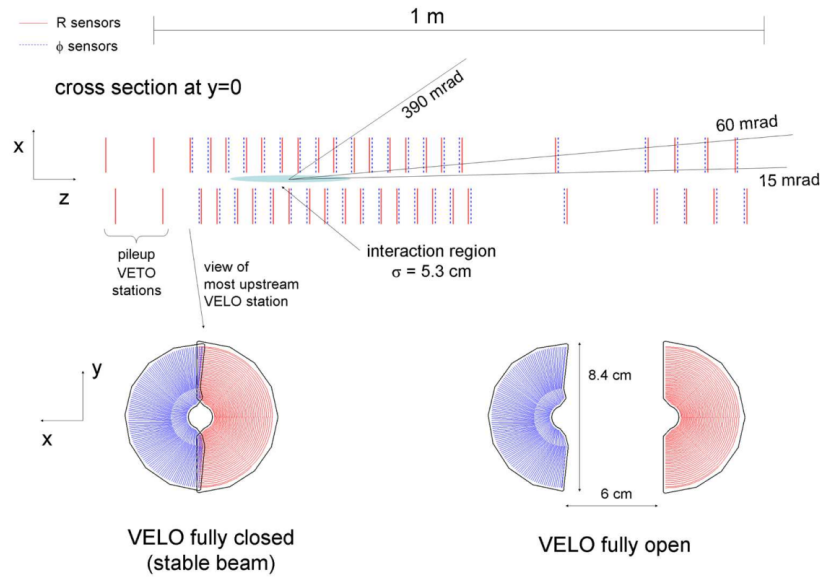


Figure 18: ■ Schematic side-view of the LHCb VELO detector (top) and the VELO in open and closed position (bottom) [51].

electronics and cooling and support structures at the top and bottom outside of the detector acceptance. This design also ensures that the most precise measurement is made in the x -direction, the bending plane of the residual magnetic field, while the y -coordinate is still measured with sufficient precision.

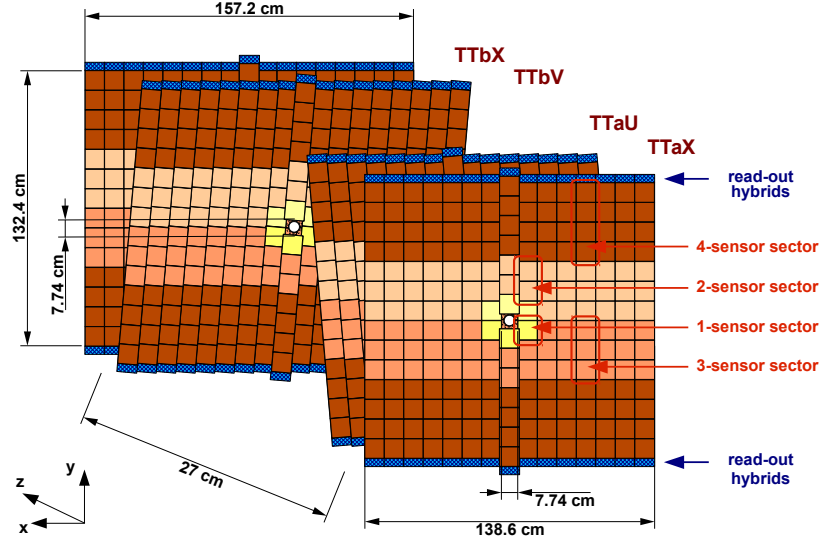


Figure 19: ■ Schematic view of the TT layout [53].

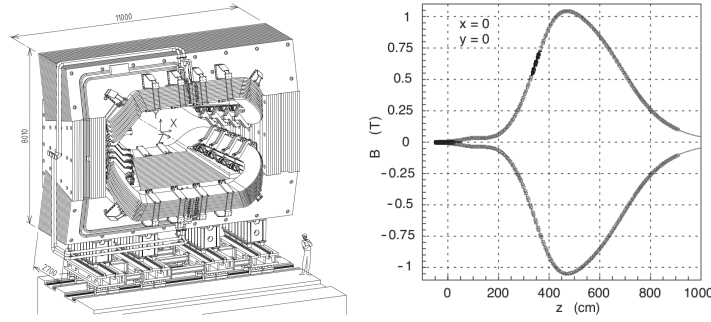


Figure 20: ■ Sketch of the LHCb magnet (left) and magnetic field strength along the beamline (right) [51].

Magnet

A warm dipole magnet with an integrated magnetic field of 4Tm is used to deflect charged particles to measure their momenta. The main component B_y of the magnetic field is along the vertical axis, while the components B_x and B_z are almost negligible, so that charged particles are mainly bent in the horizontal plane. The RICH detectors described below have to be shielded from the magnetic field to function properly. A residual magnetic field remains in the region of the TT, which is used to get a rough estimate of the particle momenta. A scheme of the magnet and the magnetic field strength along the beam-axis are shown in Figure 20. The polarity of the magnet is regularly switched within a data-taking period (*MagDown* and *MagUp*) to be able to check for detector asymmetries.

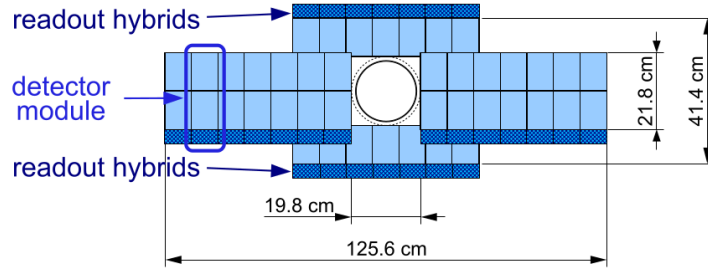


Figure 21: ■ Layout of the inner part of the downstream tracker [51].

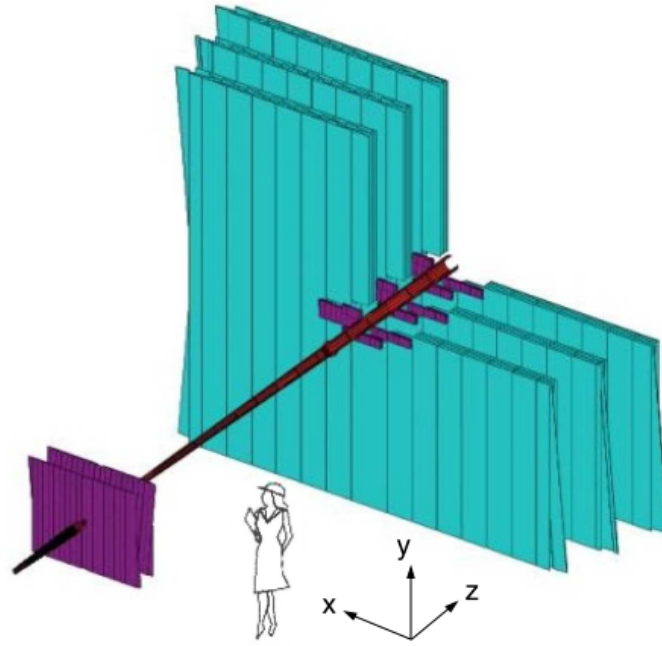


Figure 22: ■ Layout of the outer part of the downstream tracker [51].

Downstream tracker

The tracking system downstream of the magnet consists of three stations T1-T3, each made of the Inner Tracker (IT) and Outer Tracker (OT). The IT, covering the high occupancy inner region, uses silicon microstrips, while the OT uses straw tubes. Similar to the TT, a $(x-u-v-x)$ layout is chosen for the layers within each station. Again the tilting angle is chosen such that the most precise measurement is made in x -direction, as charged particles are mostly bent in that direction, while the y -coordinate can be measured with a precision sufficient for track reconstruction. The layout of the IT and OT are shown in Figure 21 and 22, respectively. The downstream tracker allows for the reconstruction of tracks after they have been bent by the magnet to be able to determine their momenta.

Calorimeters

The principle of calorimeters, generally speaking, is to measure the energy of particles by stopping them in an absorbing material and measure the deposit of energy in the resulting particle shower. Furthermore, by using differently built calorimeters and the observation that showers produced by electrons and photons look distinct to showers caused by hadrons, the calorimeter system also provides information about the species of showering particle. For this purpose, the calorimeter system of LHCb is composed of an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL). Hadrons deposit some energy in the ECAL and then produce showers in the HCAL, while photon and electron showers are mostly contained in the ECAL.

The ECAL is preceded by a preshower detector (PS) to be able to separate backgrounds due to charged pions and a scintillating-pad detector (SPD) to distinguish electrons from photons. The ECAL follows a shashlik design, where layers of lead and sampling scintillators, collecting the produced photons, are alternated. The energy resolution of the ECAL is designed to be $\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 1\%$, where the energy is given in units of GeV.

The absorber material used in the HCAL is iron with scintillating tiles being the active material running parallel to the beam axis. The energy resolution achieved is $\frac{\sigma_E}{E} = \frac{(69 \pm 5)\%}{\sqrt{E}} \oplus (9 \pm 2)\%$ (energy in units of GeV). The segmentation of the calorimeters is shown in Figure 23. The innermost section of the calorimeters has a finer segmentation as the occupancy is higher in this region. The inner section of the ECAL is made up of cells of size $40.4 \text{ mm} \times 40.4 \text{ mm}$, the middle section of size $60.6 \text{ mm} \times 60.6 \text{ mm}$ and the outer section of size $121.2 \text{ mm} \times 121.2 \text{ mm}$. The inner region of the HCAL has cells of size $131.3 \text{ mm} \times 131.3 \text{ mm}$ and the outer region of size $262.6 \text{ mm} \times 262.6 \text{ mm}$.

Each calorimeter cell is equipped with a photomultiplier (PMT), which detects photons collected by wavelength-shifting fibres from the scintillating elements. The photon yield corresponds to the energy E deposited in the cell. The ECAL and HCAL responses are adjusted so that each cell provides the same dynamic range in transverse energy E_T , which is defined as

$$E_T = E \sin \theta, \quad (56)$$

where θ is the angle between the beamline and the line connecting the centre of the cell and the interaction point.

The ECAL is calibrated using $\pi^0 \rightarrow \gamma\gamma$ candidates. Photon candidates are reconstructed as neutral clusters consisting of 3×3 cells. After enough π^0 can-

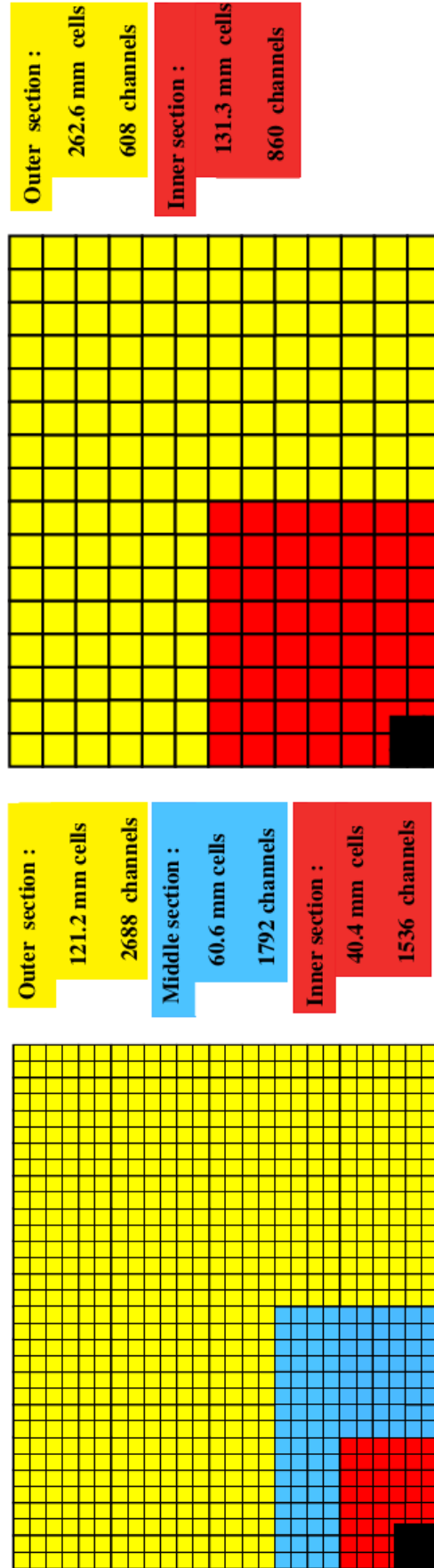


Figure 23: ■ Segmentation of the ECAL (bottom) and HCAL (top) [51].

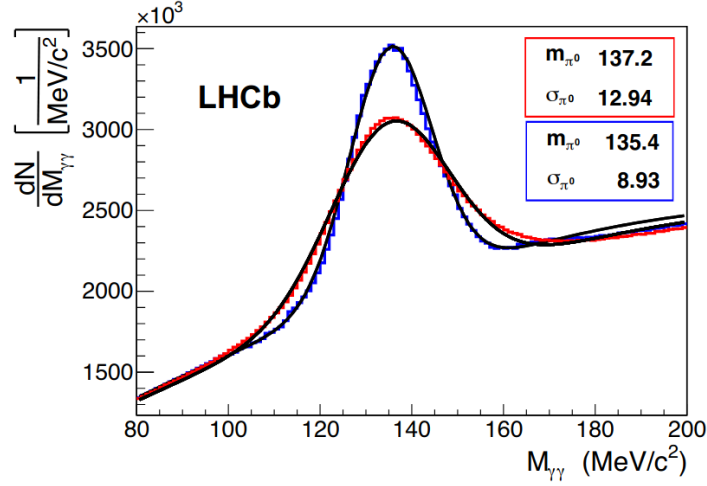


Figure 24: ■ Invariant mass spectrum for $\pi^0 \rightarrow \gamma\gamma$ candidates before (red) and after (blue) ECAL calibration [54].

didates are collected, the $\gamma\gamma$ invariant-mass spectrum is fitted to determine the reconstructed π^0 mass. The energy deposit in each cell is then scaled so that the difference between reconstructed and known π^0 mass is reduced. This process is repeated until a stable calibration has been achieved. The effect of this calibration on the invariant mass of π^0 candidates during Run 1 is shown in Figure 24. To account for the ageing effect on the photo-multipliers collecting the photon signals the calorimeter system is equipped with a LED-system¹, which emit light flashes of known intensity. The gain of the PMTs can then be adjusted after each fill to match a reference value.

During Run 2 an additional online calibration of the ECAL was implemented, which, by monitoring the relative occupancy in the ECAL cells, kept its response at a stable level by adjusting the high voltages if necessary. In particular, this ensured that the hardware trigger using the ECAL (described in Section II.5) had a near-constant output rate. Before this method was implemented it was observed that this rate would drop over time due to ageing effects.

RICH system

Particle identification at LHCb makes use of two Ring-Imaging Cherenkov (RICH) detectors (RICH1 and RICH2). Particles travelling in a medium at a speed faster than the speed of light in that medium emit Cherenkov radiation at an angle θ_C depending on the refraction index n of the medium and the velocity v of the

¹light emitting diodes

particle

$$\cos \theta_C = \frac{1}{nv}, \quad (57)$$

where v is given as a fraction of the speed of light. If the momentum of a particle has been determined by the tracking system, measuring the Cherenkov angle allows for the identification of different particle species. The main purpose of the RICH system is to distinguish pions, kaons and protons. The RICH1 detector uses as medium fluorobutane (C_4F_{10}) gas. In Run 1 it also contained an aerogel, which was removed for Run 2, since its ability to provide PID information below the Cherenkov threshold of the gas was reduced by the large photon occupancy. Also, it blocked some of the photons produced in the gas radiator from reaching the photon-detectors. The momentum range where the RICH1 can provide particle identification to distinguish between kaons and pions is about $1 - 60$ GeV. For RICH2 CF_4 gas is used, providing this separation in the momentum range of about $10 - 100$ GeV. The produced Cherenkov light is collected by mirrors and detected using hybrid photon detectors. The Cherenkov angle produced by particles with a certain momentum in the material at LHCb is shown in Figure 25. The angular acceptance of RICH1 is from ± 25 mrad to ± 300 mrad in the horizontal and to ± 250 mrad in the vertical direction. RICH2 has a reduced acceptance from ± 15 mrad to ± 120 mrad in the horizontal and to ± 100 mrad in the vertical direction. Schemes of RICH1 and RICH2 are shown in Figure 26.

Muon stations

The five muon stations M1-M5 dedicated to the reconstruction of muons are placed upstream of the calorimeters in the case of M1 and downstream in the case of M2-M5. The muon system is composed of multi-wire proportional chambers (except for the inner region of M1 where triple-GEM² are used). The detection chambers of M2-M5 are interleaved with absorbers made of 80 cm thick iron to reduce the number of non-muon particles traversing all the stations. Muons with energy above 6 GeV typically are able to traverse all muon stations. A schematic view of the muon stations is shown in Figure 27.

²gas electron multiplier

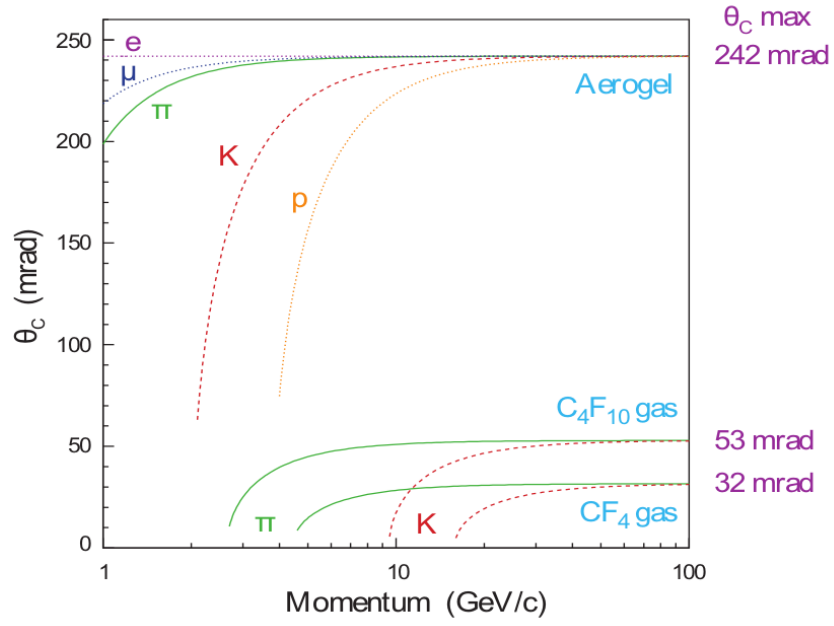


Figure 25: ■ Cherenkov angle as function of particle momentum for the material used in the RICH system [51].

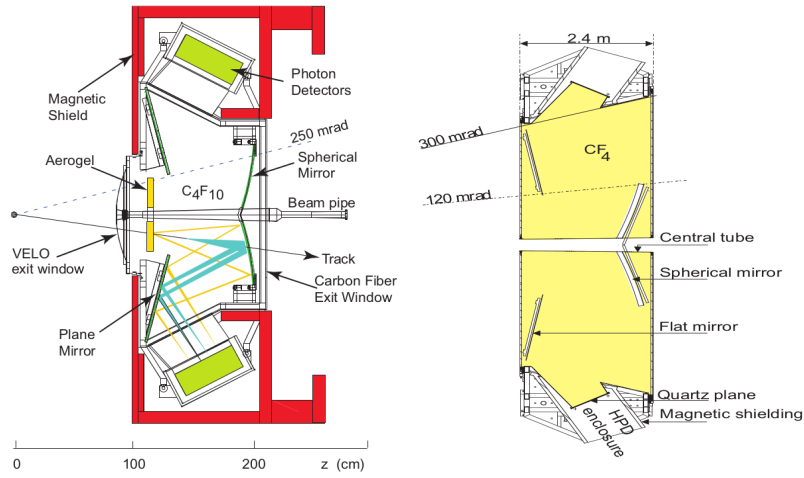


Figure 26: ■ Schematic side view of the RICH1 (left) and top view of the RICH2 (right) detectors [51].

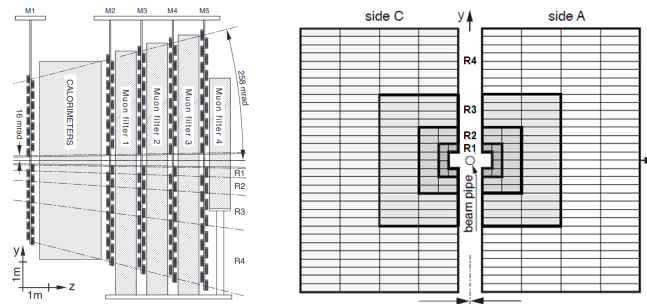


Figure 27: ■ Schematic view of the side (left) and front (right) of the muon stations [55].

II.3 Charged particle tracking performance

An overview of different track types reconstructed at LHCb is shown in Figure 28. Most relevant for this work are the 'long' tracks used in the analysis, which leave hits in the VELO, TT and T1-3 tracking stations and typically have the best momentum resolution and 'Velo' tracks, which are used to reconstruct primary vertices. After a long track has been reconstructed, it is fitted with a Kalman filter [56] [57] to obtain the best estimate of the particle momentum and its uncertainty.

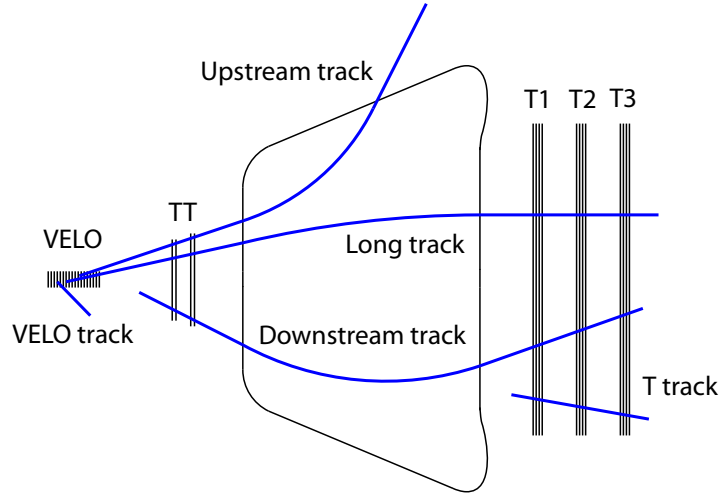


Figure 28: ■ Track types used by LHCb [58].

The purpose of the vertex locator is the reconstruction of the (x, y, z) -coordinates of the proton-proton-interaction points, as well as the associated uncertainties, and to provide the starting point to extrapolate the track through the following tracking systems downstream. The resolution of reconstructed PVs is shown in Figure 29.

The so-called impact parameter (IP), defined as the minimum distance of a track to a PV, can be determined with a resolution of $(15 + [29 \text{ GeV}]/p_T) \mu\text{m}$ dependent on the transverse momentum p_T measured in units of GeV, as shown in Figure 30. The impact parameter is typically used to identify displaced decays of b hadrons. The momentum resolution of long tracks depends on the track momentum and ranges from about 0.5% at low momentum ($p < 20 \text{ GeV}$) to about 1.1% at large momenta ($p > 200 \text{ GeV}$), as shown in Figure 31. The tracking efficiency determined for muon tracks is shown in Figure 32. The lower efficiency in Run 2 compared to Run 1 is caused by larger spillover effects in the OT due to a reduced

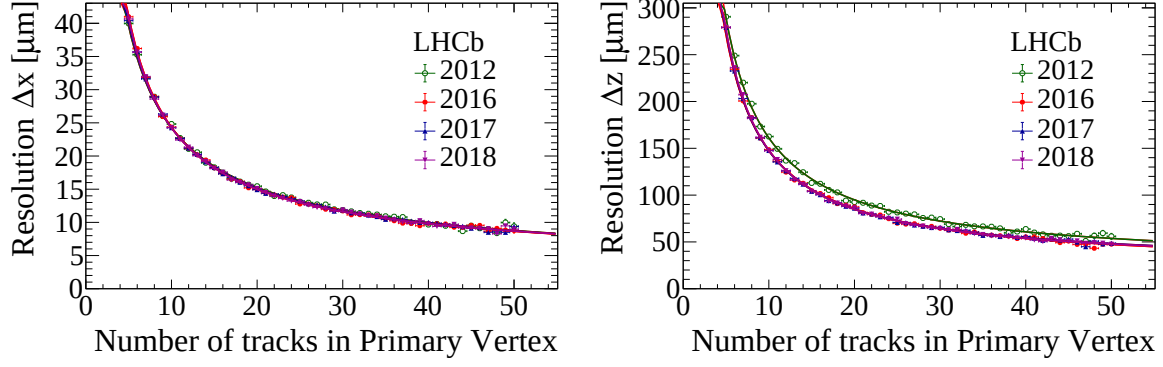


Figure 29: ■ Primary vertex resolution as function of number of track in the PV in the x - (left) and z - (right) coordinates [59].

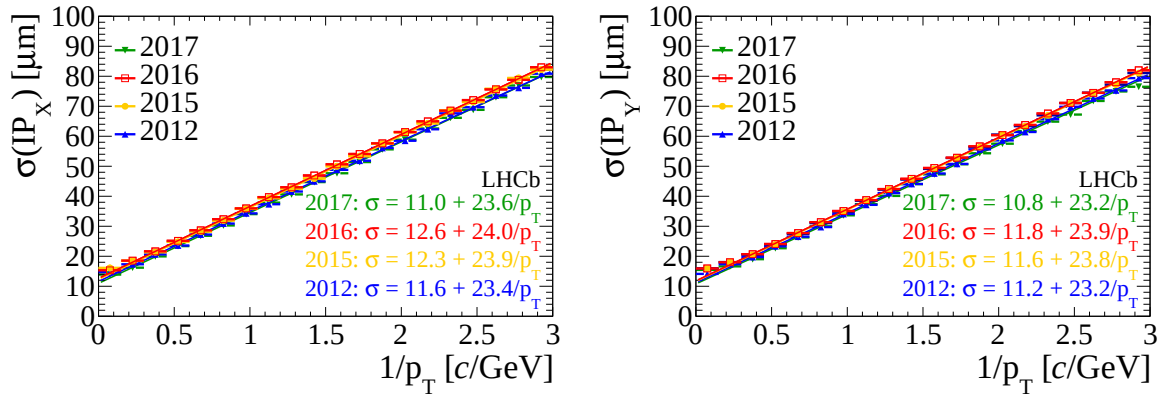


Figure 30: ■ Impact parameter significance in x (left) and y (right) as function of the inverse transverse momentum for the 2012, 2015, 2016 and 2017 data-taking periods [59].

proton-bunch spacing.

To reduce tracks consisting of random combinations of hits ('fake' tracks), which do not belong to the same particle, a neural network is employed, tuned on simulation to reduce 60% of fake tracks and keep 99% of 'real' tracks. The performance of this classifier is validated on data, shown in Figure 33, where $D \rightarrow K^- \pi^+$ decays are used.

Electron reconstruction

Compared to other charged particles traversing the LHCb detector, electrons are a special case, as they produce the most Bremsstrahlung due to their low mass. This has consequences on the reconstructed momentum of electrons and the efficiency of reconstructing them. If an electron radiates Bremsstrahlung before passing through the magnet, its momentum will be estimated to be lower than it originally was. This results in long tails in invariant mass distributions of particles decaying to electrons extending to lower masses. If the Bremsstrahlung is emitted

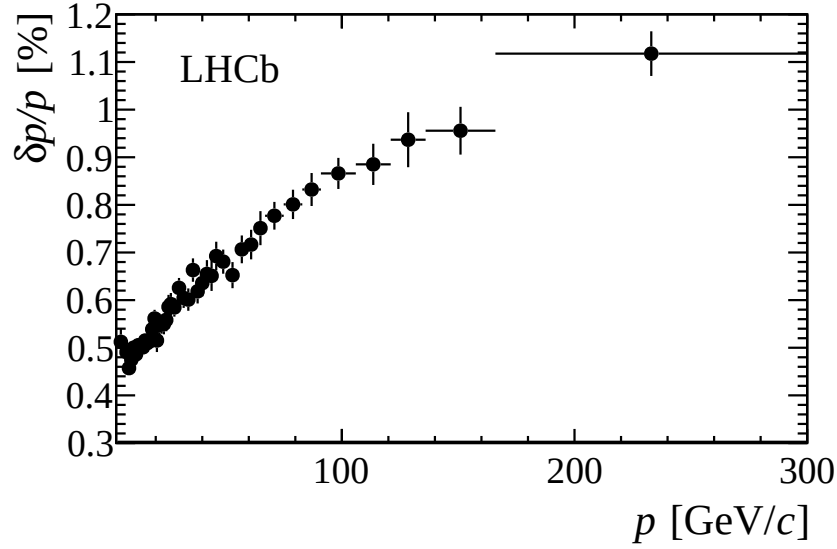


Figure 31: ■ Relative momentum resolution as function of momentum for long tracks [60].

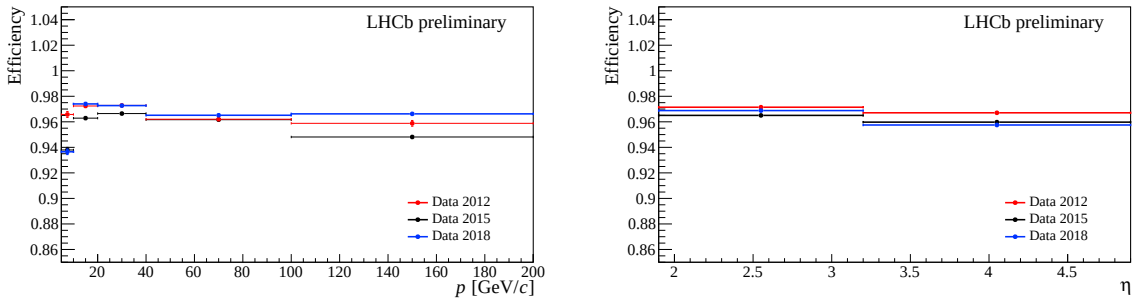


Figure 32: ■ Muon tracking efficiencies determined on data obtained from a $J/\psi \rightarrow \mu^+ \mu^-$ sample as function of the momentum (left) and pseudorapidity (right) of the muon for the 2012, 2015 and 2018 data-taking periods [61].

after passing through the magnet, the emitted photon and electron will likely be hit the same region in the electronic calorimeter, so that the energy carried away by the photon is restored. If the photon is radiated before the magnet, the electron and photon will hit different parts of the calorimeter. There are procedures in place to recover the energy carried away by these photons. They are however not perfect resulting in the aforementioned radiative tails. Since the electron-photon separation is based on the ratio E/p of the particle, the identification performance is not disturbed by Bremsstrahlung.

For electrons with low momentum, it can happen that the electron will be swept out of the downstream tracker acceptance by the magnet after radiating Bremsstrahlung. The reconstruction efficiency for electron long tracks is therefore expected to be smaller compared to other particle species. This demonstrated

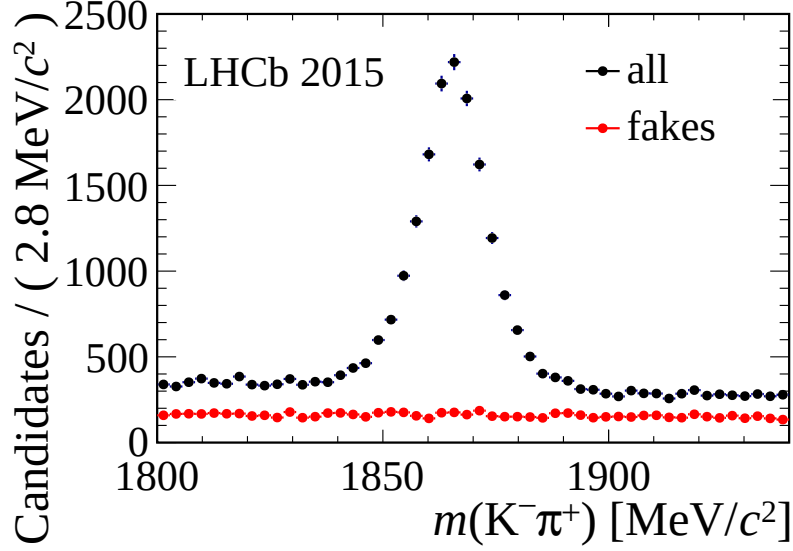


Figure 33: ■ Performance of the classifier rejecting fake tracks [59].

on simulated data in Figure 34, where it can be seen that the efficiency to reconstruct a muon or an electron in the VELO is similar over a large kinematic range, but the efficiency to reconstruct an electron as long track is smaller compared to the muon case, especially at low transverse momenta. For precision measurements, it is required to perform a data-driven measurement of the electron reconstruction efficiency together with an estimation of associated systematic uncertainties, as was already done for muons. While it is not straight-forward to get a background-subtracted data sample to measure this reconstruction efficiency due to Bremsstrahlung, it was shown that it is feasible [62]. Special care has to be taken to account for regions where the amount of material traversed is larger, such as the RF-foil region, which has to be taken into account when making the measurement. For this reason, it is performed in regions of p_T , η of the electron separately for electrons travelling in parallel or not to the RF-foil, as shown in Figure 35.

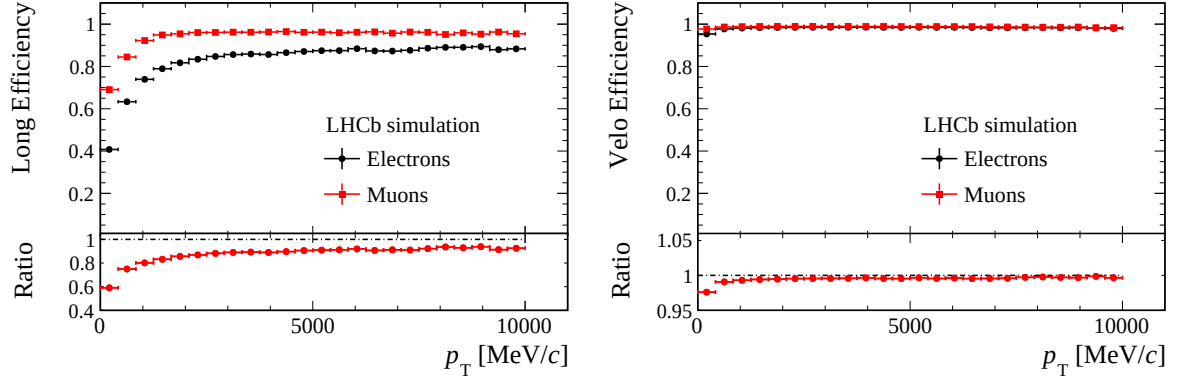


Figure 34: ■ Simulated long track (left) and VELO-track reconstruction efficiency (right) for electrons and muons [62].

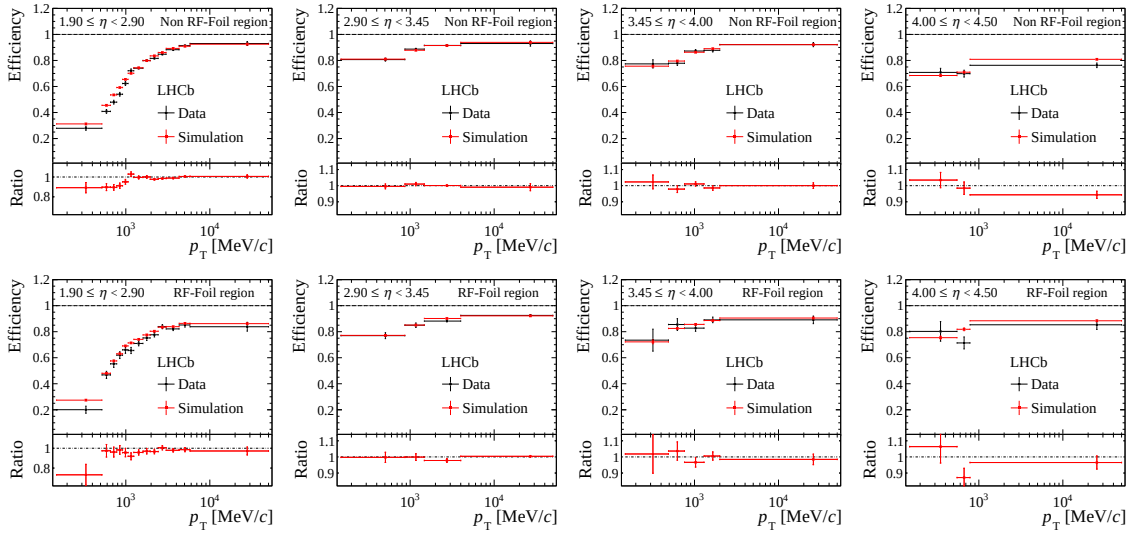


Figure 35: ■ Electron reconstruction efficiency obtained on data and simulation as function of the transverse momentum and pseudo-rapidity of the electron. The top rows show electrons not travelling parallel to the RF-foil, while the bottom row shows electrons that do [62].

II.4 Particle identification

Particle identification (PID) information is used at various places in the selection and analysis of data. In the case of the RICH, particle identification is achieved by comparing the measured light ring with the expected ring of a particle of a specific mass. Assigning a particular mass hypothesis to a particle, its velocity can be estimated using its measured momentum and the expected Cherenkov angle can be compared with the measured one. In this way, the probabilities of different mass hypotheses can be compared and the mass assignment most consistent with the produced angle can be found.

Energy deposits in different parts of the calorimeter system indicate if the particle was a hadron, photon or electron, as photons and electrons typically deposit all their energy in the ECAL, while hadrons deposit some energy in the ECAL and the majority in the HCAL. The presence or absence of a track consistent with an energy deposit in the calorimeters can be used to distinguish between charged and neutral particles. Muons can be identified by the presence of hits in the muon stations consistent with the corresponding particle track.

It is possible to define differential logarithmic likelihoods for the different systems used for particle identification that a particle is a particular species compared to the pion hypothesis. In addition, information of all subdetectors used for particle information can be combined to a single differential logarithmic likelihood. The efficiency of correctly identifying a kaon and the probability of misidentifying a pion as kaon for different selections on the differential likelihood are shown in Figure 36. The performance of the particle identification depends on the momentum and angle of the particle, as well as the overall occupancy in the detector. Particles must have a certain momentum to produce Cherenkov light and at large momentum it becomes more difficult to distinguish the rings produced by different types of particles.

Electrons and muons behave very differently in the detector. Electrons can be identified by having energy deposits in the SPD, PS and ECAL. While muons leave some energy in the calorimeter system, they are minimally ionizing and are likely to reach the muon stations. By matching hits in the muon system to tracks muons can be distinguished from other particle species. This is exploited by the *IsMuon* criterion [64], which, based on the momentum of the track, classifies tracks as muons depending on the number of muon stations with hits consistent with the track trajectory. The performance of the electron and muon identification is shown in Figure 37.

The input from different parts of the detector can be combined using multi-

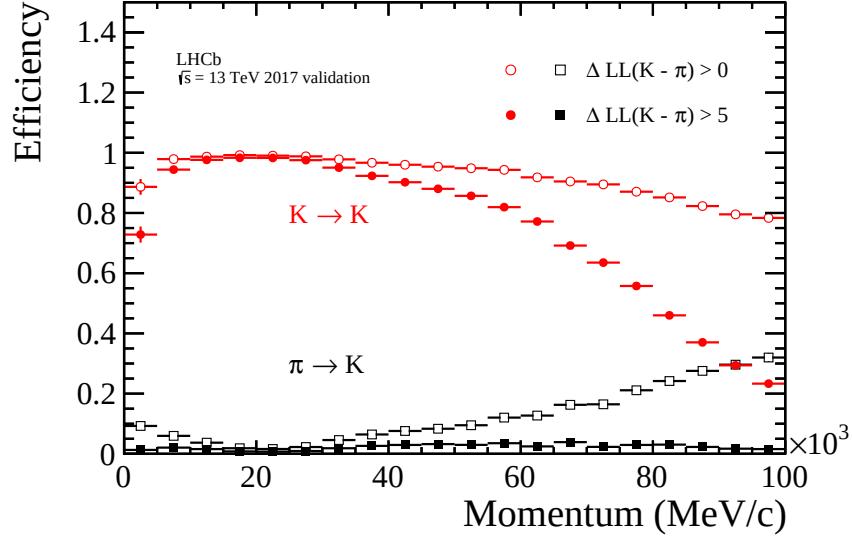


Figure 36: ■ Kaon identification and pion misidentification performance obtained on the 2017 calibration data sample as function of the momentum of the particle [63].

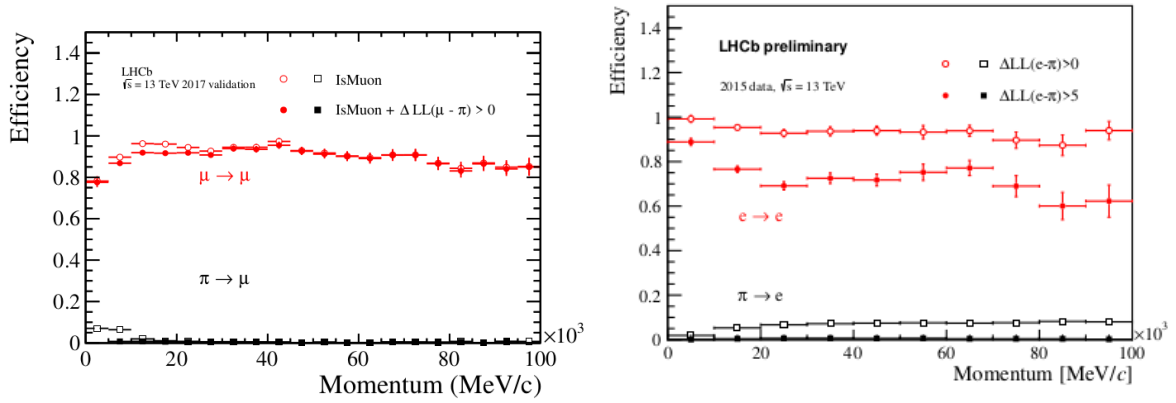


Figure 37: ■ Muon identification (left) and electron identification (right) performance as function of the momentum of the respective particle for two PID selections. Also shown is the π mis-identification rate [65].

variate classifiers or neural networks. One such centrally trained and optimised neural network provides the so-called *ProbNN* variables combining various inputs from the tracking, RICH, calorimeter and muon systems into a single number, which is akin to the probability of the particle being of a particular species.

II.5 Data acquisition and trigger

Since it is impossible to record all the data produced by the LHCb detector at the event rate of up to 30 MHz due to constraints on computing power and disk space, a trigger system implemented at the hardware and software level is used to select and only save interesting events. First, the hardware-level trigger (L0) selects events based on energy deposits in the electronic or hadronic calorimeters and track "stubs" in the muon stations to reduce the rate by about a factor 30 to 1 MHz. If the event meets any of the L0 requirements, it is passed to a software-level trigger consisting of two stages (HLT1 and HLT2).

HLT1 performs a partial event reconstruction using information from the tracking system (VELO, TT, T1-T3 and muon system) to select displaced tracks or displaced two-track vertices as well as (di-) muons. By varying the selection criteria applied in HLT1 its output rate can be tuned to a typical value around 110 kHz. The events selected by HLT1 are saved to a disk buffer until they are processed by HLT2. At this input rate, it is possible to use the full detector information including particle-identification from the RICH systems and calorimeters. The event rate is further reduced to about 12.5 kHz (in the Run 2 case) and written to final storage. During Run 2, it was made possible to do an online calibration and alignment of the detector, such that HLT2 could perform a selection with offline-like quality. This means that physics analysis could already be performed using these events without requiring an additional re-reconstruction step after data-taking. This quality is exploited in the *Turbo* stream [66].

A trigger system typically consists of a set of classification algorithms (trigger 'lines') deciding if the event at hand will be kept or discarded. Generally speaking, those lines are independent but can be combined, and the final decision is the logical 'OR' of all lines meaning that the positive decision of a single line is sufficient to save the event. It is often useful to have access to the decisions of all lines even if the particular trigger selection is not used in an analysis to calibrate the efficiency of the simulated trigger response in a data-driven way using the TISTOS method [67], as will be used in Section III.2.

From Run 1 to Run 2

To expand the physics program in the areas of the electroweak sector and soft QCD, the software-level trigger of LHCb was evolved during Run 1 and beyond. During Run 1 the transverse momentum threshold, below which tracks are not reconstructed, in the first stage of the software-level trigger was tighter due to throughput constraints. This reduced the efficiency in particular of selecting decays of c

hadrons. Also this meant that the data had to be re-reconstructed offline to be able to reach the final precision, resulting in differences between the online and offline selection, which were hard to account for. For these reasons, the high-level trigger was re-designed and improved for Run 2. Due to the timing constraints in Run 1 in the HLT1 step only long tracks of non-muon particles with a minimum transverse momentum of $p_T > 1.2$ GeV could be reconstructed. Low transverse momentum muons could be recovered by directly matching VELO-tracks with hits in the muon stations down to a threshold $p > 6$ GeV and $p_T > 0.5$ GeV [68]. The line selecting displaced tracks ('TRACKALLL0') required $p > 10$ GeV and $p_T > 1.7$ GeV, while the line selecting single displaced muons ('TRACKMUON') used $p > 8$ GeV and $p_T > 1$ GeV.

Due to the optimisation for Run 2, the reconstruction threshold applied in HLT1 could be lowered to $p_T > 500$ MeV for all tracks, and for reconstructing muons down to $p_T > 80$ MeV. A new line for selecting two displaced tracks forming a vertex was introduced, which increased the efficiency for b and c hadrons with low transverse momentum.

In Run 2 events could be reconstructed in full offline-like quality at the HLT2 level. This was in particular made possible by the real-time alignment and calibration performed run-by-run on the events selected by HLT1 stored to a buffer before they are reconstructed by HLT2. All this brought an improvement in the selection efficiencies of the trigger system, especially for decays from c - or s -hadrons.

L0

The hardware-level trigger L0 is made up of three components: the pile-up system upstream of the main detector, the calorimeter trigger and the muon trigger. The calorimeter trigger is based on particles leaving large transverse energy E_T deposits in the calorimeter. Electrons, photons and π^0 converting to two photons leave the main part of their energy in the electronic calorimeter, while hadrons typically produce clusters starting in the ECAL and spreading out into the HCAL. The energy deposits of 2×2 adjacent cells are clustered for the trigger decision and photons are distinguished from electrons by additionally using information from the SPD and PS. The calorimeter trigger decision is thereby divided into the 'L0ELECTRON', 'L0PHOTON' and 'L0HADRON' decisions.

Muons are reconstructed in the L0 by building crossings of the strip and pads of the muon system. The transverse momentum of the muons is estimated by assuming that they originate from the interaction region with a resolution of about $\sim 20\%$. The two muons with the largest transverse momentum are used in the

trigger decision. Either the largest p_T ('L0MUON' decision) or the product of both ('L0DIMUON' decision) are required to surpass a certain threshold for the respective line reaching a positive decision.

HLT1

The first stage of the software-level trigger HLT1 uses information from the tracking systems to do a partial event reconstruction. First the tracks of charged particles in the VELO are reconstructed and fitted with a simplified Kalman filter. Since no momentum estimate is known at the stage of the VELO reconstruction, an average of 300 MeV is assumed for each particle to estimate the track uncertainties due to multiple scattering. The obtained track states are used to reconstruct the primary vertices. By finding matching hits in the TT the charge of the particle associated with a VELO track can be estimated. The tracks are propagated through the magnet field to the T stations and combined with the hits there to a full (long) track, using a search window for finding T station hits assuming a transverse momentum above 500 MeV, taking into account the sign of the charge. Finally, they are fitted with a Kalman filter to get a more precise estimation of their momentum and to reject 'fake' tracks. By associating long tracks with hits in the muons stations it is possible to identify muons.

Using the reconstructed long tracks with muon identification information and primary vertices, the following decisions are mainly used for analysis: A trigger on a single track with high p_T displaced from all PVs ('HLT1TRACKMVA'), another line selecting two-track vertices ('HLT1TWOTRACKMVA'), also with large p_T and displacement from any PV. These lines are optimised for selecting decaying c and b hadrons, whose decay vertex is typically displaced from the PV. The efficiency of selecting various b - and c -hadron decays of the single-track line, two-track line and a combination of both is shown Figures 38 and 39.

The muon-ID information is used to select displaced muons with large p_T ('HLT1TRACKMUON'), non-displaced muons with even higher p_T , dimuon pairs within a certain invariant mass range ('HLT1DIMUONHIGHMASS') and displaced dimuon vertices. These lines are chosen to be able to identify and select (di-)muons from c and b hadrons as well as from W and Z boson decays.

HLT2

The incoming event rate in HLT2 is low enough that the full detector information can be used to make the final decision if an event is saved or discarded. This means that now particle identification information from the calorimeters and RICH sys-

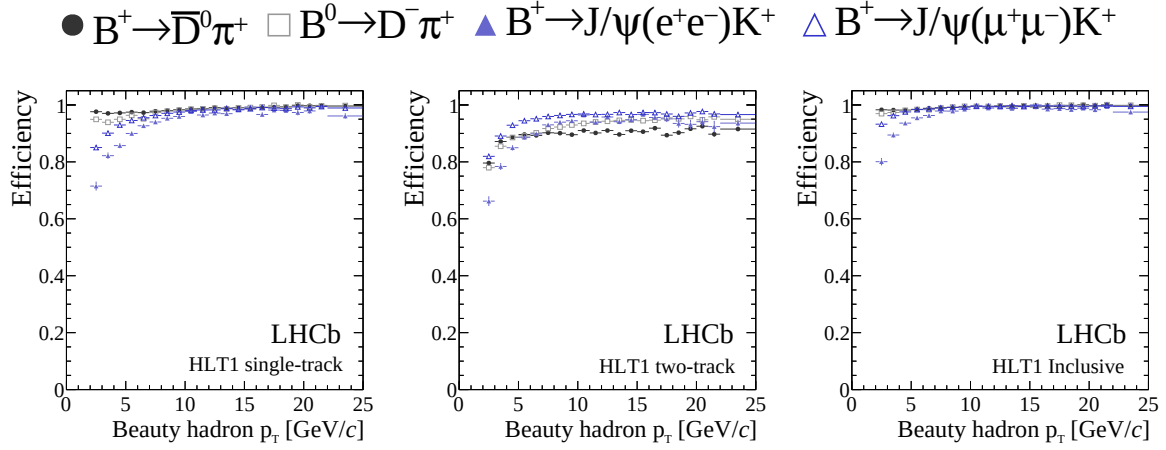


Figure 38: ■ Efficiency of the HLT1 single-track line (left), two-track line (middle) and combination of both (right) as function of the b hadron p_T [59].

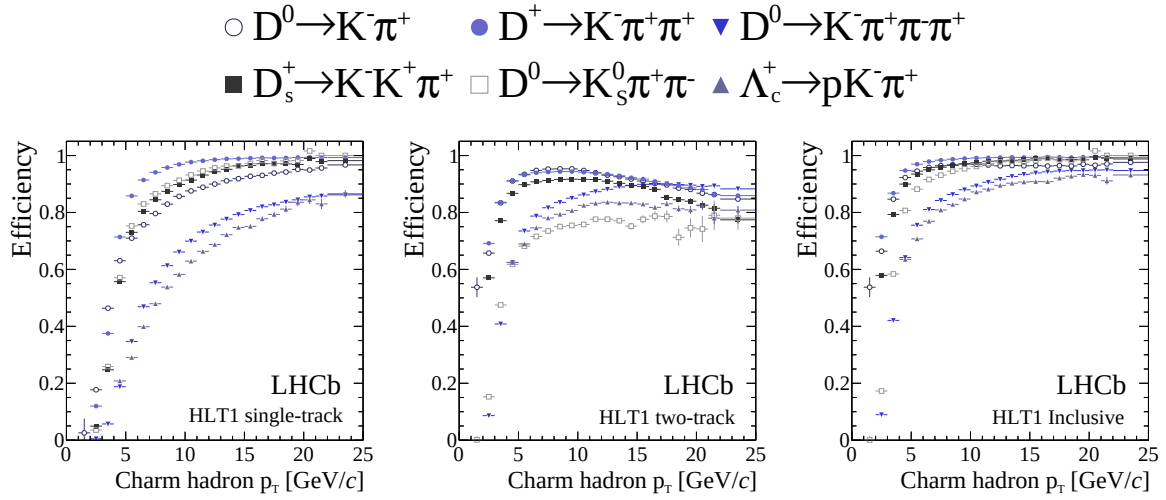


Figure 39: ■ Efficiency of the HLT1 single-track line (left), two-track line (middle) and combination of both (right) as function of the c -hadron p_T [59].

tems is used and more complicated selection sequences can be employed. Of central importance is the topological trigger on two-, three- and four-track vertices to inclusively select b -hadron decays ('HLT2TOPO{2,3,4}BODY'). These vertices are required to have significant p_T and displacement from the PVs. A multi-variate classifier is employed [69] to select the track combinations most likely to originate from b hadron decays. Since for decays into three or more final states, the transverse momentum of a final state particle might be below the 500 MeV threshold used in the HLT1 sequence when propagating tracks to the downstream tracking stations, an additional step is taken to recover tracks with low transverse momentum. After the same track reconstruction as in HLT1 is repeated, the VELO tracks are extrapolated through the magnetic field to the T-stations with a search win-

dow assuming a minimum p_T of 80 MeV without using hits in the TT stations. To further increase the long track reconstruction efficiency an independent algorithm is executed, which first finds tracks segment in the T-stations, which are then matched to VELO tracks to construct long tracks.

After all long tracks are identified they are fitted with a Kalman filter. A neural network is used to reject fake tracks, tuned to keep 99% of 'real' tracks and to reject 60% of fake tracks.

II.6 Offline processing & simulation

After the data has been selected by the HLT and is stored, it is made available for analysis via the so-called 'stripping' step. In the stripping, various 'lines' are defined, applying selections on the particles and combining them to reconstruct certain decays. Similar stripping lines are grouped into streams so that one does not need to regard the whole data sample when looking for events of interest. The user has then access to the candidate particles reconstructed and saved by the stripping and can define which information like momenta or PID variables they want to have accessible in the next analysis steps. This analysis is largely based on the ROOT software package [70]. Since in Run 2 the best reconstruction performance was already achieved in the online processing, re-running the reconstruction software offline is generally not necessary.

A commonly used tool during the offline processing is decay tree fitting (DTF) [71], where the decay is (re-)fitted using all external particles and four-momentum conservation constraints at the decay vertices. It is possible to impose additional constraints on a DTF such as setting the mass of an intermediate particle to its known value. This process can improve the momentum and mass resolution of the particles at the top of the decay chain.

Simulation using Monte-Carlo (MC) methods is essential for performing any analysis. It is used to describe the shapes of signal and background contributions to the observed data, as well as to estimate selection efficiencies. For this purpose pp collisions are first generated using PYTHIA [72] with a configuration adapted for LHCb [73]. Particle decays are described using EVTGEN [74] while final-state radiation is simulated using PHOTOS [75]. The interaction of the particles with the detector and its response are modelled using GEANT4 [76] [77].

Since the simulation can not model the measured data perfectly, it is corrected using data-driven methods. For example, particle identification information is usually not simulated well enough and so the efficiency of PID selections is estimated using dedicated calibration samples. Correction weights to the simulated sample can be obtained by comparing distributions in the background-subtracted data and the simulated samples. The kinematic distributions of simulated semileptonic decays depend on the choice of form factor parametrisation. While the decays are generated with a specific model and parameter values, the simulated data can be weighted to a different choice of form factor model using the HAMMER package [78].

Chapter III

Testing lepton-flavour universality in semileptonic decays using electrons and muons

“All animals are equal, but some animals are more equal than others.”

George Orwell, *Animal Farm*

As described in the previous section, tests of lepton flavour universality begin to paint a path towards BSM physics. In the case of hadron collider experiments, one missing piece is the test of lepton universality in semileptonic b -hadron decays using electrons. This work describes the first measurement of the ratio $R(D^{*+})_{light}$ with pp -collisions at LHCb, which is the relative branching fraction of the $\bar{B}^0 \rightarrow D^{*+} e^- \bar{\nu}_e$ decay relative to the $\bar{B}^0 \rightarrow D^{*+} \mu^- \bar{\nu}_\mu$ decay

$$R(D^{*+})_{light} = \frac{\mathcal{B}(\bar{B}^0 \rightarrow D^{*+} e^- \bar{\nu}_e)}{\mathcal{B}(\bar{B}^0 \rightarrow D^{*+} \mu^- \bar{\nu}_\mu)}. \quad (58)$$

Measuring relative branching fractions is advantageous, as previously explained. The $\bar{B}^0$ -production fraction and the branching fraction of the D^{*+} meson decay chain do not need to be known. Some systematic uncertainties are expected to cancel to first order in the ratio, for example, systematic uncertainties related to the reconstruction and identification of the hadrons, as they appear in both decay chains. However there are systematic uncertainties that can not cancel in the ratio as muons and electrons behave differently when traversing the detector. Such uncertainties are related to the PID and tracking performance of electrons and muons. The main aim of this work is to demonstrate that these uncertainties can be controlled to a high precision and that it is possible to perform precise measurements with single electron final states in pp collisions.

The reason why $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ decays are chosen for this work is to profit from the experience of a previous measurement of $R(D^*)$ by the LHCb collaboration [79], where the tauon was reconstructed in the muonic decay mode. The procedure chosen here to determine the signal yields follows the same strategy. Of all the possible decays to choose from to test lepton universality in the $b \rightarrow c e \nu_e$ transition, $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ decays are one of the simpler choices when compared to, for example, the similar $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$ decay. While the additional pion present in the $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ final state results in a lower expected signal yield, as it can not be reconstructed with 100% efficiency, the background when using $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$ decays would be larger. This is due to $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ and similar decays entering the $B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$ decay sample as background when the additional pion is not reconstructed. Other possible decays like $\bar{B}^0 \rightarrow \pi^+ \ell^- \bar{\nu}_\ell$ or $\bar{B}_s^0 \rightarrow K^+ \ell^- \bar{\nu}_\ell$ are more difficult to use for such an analysis as the signal yields are smaller due to Cabibbo-suppression and the expected combinatorial background is larger as kaons and pions are more abundant than D^* mesons.

This chapter first outlines the analysis strategy and steps taken to measure this ratio in Section III.1. The data and simulated samples are described and the control channel $B^0 \rightarrow J/\psi K^{*0}$ is introduced. The simulated samples used need to be weighted to account for differences between data and simulation as described in Section III.2. The control channel is employed for this purpose. The extraction of the relative signal yield with a template fit is described in Section III.3. The relative efficiency is determined and combined with the relative yield to $R(D^{*+})_{light}$ in Section III.4. Sources of systematic uncertainties are discussed and their impact is estimated in Section III.5. To improve the readability of the text some plots of intermediate results are shown only for the 2017 period if there is no significant difference between the years. Figures which do not indicate the period in their caption are produced from the 2017 samples.

III.1 Analysis strategy

The relative branching fraction can be determined by measuring the yield N of each signal mode in the used data sample and the efficiency ϵ of reconstructing and selecting the decays

$$\begin{aligned} R(D^{*+})_{light} &= \frac{\mathcal{B}(\bar{B}^0 \rightarrow D^{*+} e^- \bar{\nu}_e)}{\mathcal{B}(\bar{B}^0 \rightarrow D^{*+} \mu^- \bar{\nu}_\mu)} \\ &= \frac{N(\bar{B}^0 \rightarrow D^{*+} e^- \bar{\nu}_e)}{N(\bar{B}^0 \rightarrow D^{*+} \mu^- \bar{\nu}_\mu)} \times \frac{\epsilon(\bar{B}^0 \rightarrow D^{*+} \mu^- \bar{\nu}_\mu)}{\epsilon(\bar{B}^0 \rightarrow D^{*+} e^- \bar{\nu}_e)}. \end{aligned} \quad (59)$$

In the following the $\bar{B}^0 \rightarrow D^{*+} e^- \bar{\nu}_e$ and $\bar{B}^0 \rightarrow D^{*+} \mu^- \bar{\nu}_\mu$ decays are referred to as 'electron signal mode' and 'muon signal mode', respectively. The D^{*+} is reconstructed in the $D^{*+} \rightarrow D^0 \pi^+$ decay and the D^0 in the $D^0 \rightarrow K^- \pi^+$ decay. Defining the 'raw' ratio R_{raw} and the relative efficiency ϵ_{rel}

$$R_{raw} = \frac{N(\bar{B}^0 \rightarrow D^{*+} e^- \bar{\nu}_e)}{N(\bar{B}^0 \rightarrow D^{*+} \mu^- \bar{\nu}_\mu)}, \quad \epsilon_{rel} = \frac{\epsilon(\bar{B}^0 \rightarrow D^{*+} \mu^- \bar{\nu}_\mu)}{\epsilon(\bar{B}^0 \rightarrow D^{*+} e^- \bar{\nu}_e)}, \quad (60)$$

it can be written as

$$R(D^*)_{e\mu} = R_{raw} \times \epsilon_{rel}. \quad (61)$$

The ratio R_{raw} is determined using a simultaneous template fit to the muon and electron data samples. Since the neutrino is not reconstructed, the $\bar{B}^0$ invariant-mass distribution is very broad with a considerable tail to lower values. Therefore separating the signal contribution from background is more involved. As first step the $\bar{B}^0$ momentum is approximated from the partially reconstructed values. Using this approximation kinematic variables are calculated in the rest frame of the $\bar{B}^0$, which allows to distinguish between the signal decays and background sources. The shapes used in the template fit to estimate R_{raw} are obtained using simulated samples and data-driven methods. The efficiency ratio is calculated using simulated signal mode decays, which is corrected for differences to the data.

The offline selection requirements are chosen to select a data sample with high signal purity and to focus on a kinematic region where the muon and electron channels have similar kinematic distributions and reconstruction performances to minimise systematic uncertainties. In particular, a requirement on the pseudorapidity of the charged lepton is made to reduce the systematic uncertainties associated with the reconstruction efficiency of the lepton. In general, as the uncertainty of the $R(D^{*+})_{light}$ measurement is expected to be systematic-dominant, a strategy of 'sacrificing' statistical power to reduce the expected systematic uncertainty is employed in the selection.

The determination of $R(D^{*+})_{light}$ is performed separately for each year and then combined to the final result. The analysis is performed blinded to avoid biasing the result. This is achieved by multiplying the fitted R_{raw} with a random factor. To avoid being able to infer R_{raw} by comparing the signal contributions in the fit projections shown in Section III.3 with the background contributions, all freely floating yields are blinded by an additional common random factor. This factor is shared since the relative size of the background yields with respect to each other and the signal is of interest when developing the analysis.

III.1.1 Data samples

The analysis is based on proton-proton collision data recorded in the years 2016-2018 with the LHCb detector. Data samples collected during Run 1 (2011-2012) and in 2015 are not included since the sensitivity of the measurement is not expected to be statistically limited. Adding these samples would increase the analysis complexity without offering many benefits as in these periods the calibration of the calorimeter system was not as well understood. The occupancy calibration described in Section II.2 was not used yet and ageing effects of the ECAL were not accounted for.

The signal modes are selected starting from the centrally defined stripping selections¹, which select combinations of $D^0\mu^-$ and D^0e^- particles. At first, the $D^0\ell$ are 'separated' again and the D^0 candidates are combined with an additional π^+ candidate, referred to as 'soft pion', to form a D^{*+} candidate. To form the $\bar{B}^0$ candidate the D^{*+} candidate is in turn combined with the μ^- or e^- from the original combination. This is done in this way since the D^{*+} decays via the strong interaction and is thus short-lived so that it makes almost no practical difference if using D^{*+} or D^0 candidates when making the initial combination with a lepton. The advantage is that the stripping selection is kept more flexible and the $D^0\ell$ combinations could also be used for other analyses. Both stripping selections also contain ($D^0\ell^+$) combinations built with a wrong-sign lepton, which is used to model the combinatorial background. To study the background due to different particles misidentified as muons or electrons ('fake' muons or electrons), candidates selected by a dedicated stripping line² are used, which is the same as the line selecting $D^0\mu^-$ candidates but with the muon PID requirement inverted.

The 2016, 2017 and 2018 data samples are treated separately since the corrections to the simulation might be different between the years and the trigger efficiency can vary because of different trigger thresholds.

III.1.2 Background sources

Possible background sources are decays with additional charged or neutral particles in the final state compared to the signal modes. Since the signal mode is only partially reconstructed, these decays can look similar to the signal modes when looking at quantities like the $\bar{B}^0$ invariant-mass distribution.

Decays of a b hadron to two charm mesons can mimic the signal mode if the first charm meson is a D^{*+} or decays into one and the second charm meson de-

¹STRIPPINGB2DMuNuX_D0 and STRIPPINGB2DMuNuX_D0_ELECTRON

²STRIPPINGB2DMuNuX_D0_FAKEMUON

cays semileptonically to a muon or electron final state. For example the decay $B^0 \rightarrow D^{*+} D^-$ can occur, where the D^- meson then decays as $D^- \rightarrow K^0 \mu^- \bar{\nu}_\mu$. Similar decays are also possible for B^+ hadrons resulting in the same reconstructed particles as the signal modes. The case where the second charm meson first decays to a τ , which in turn decays to an electron or muon and neutrinos, is also considered. This contribution is referred to in the following as 'double-charm' background.

Other background sources are semileptonic b -hadron decays to higher excited D^{**} states, with decay chains that contain a D^{*+} and additional charged and neutral particles. For example the decay $B^0 \rightarrow D_1'^- e^+ \nu_e$ needs to be considered as potential background source, as the $D_1'^-$ meson decays as $D_1'^- \rightarrow D^{*-} \pi^0$. Similar decays of B^+ and B_s^0 hadrons also can contribute to the total background. These decays are labelled as 'excited D^{**} ' background. Most of these backgrounds are expected to affect the electron and muon signal modes in the same way as the muon and electron masses are small compared to the b -hadron masses and the available phase space in the decays is similar. This is a main difference to the $R(D^*)$ measurement with tauons and muons where the backgrounds contribute differently due to the large mass of the tauon and one of the main challenges was the separation of the tauon mode from similar-looking backgrounds.

The $\bar{B}^0 \rightarrow D^{*+} \tau^- \bar{\nu}_\tau$ decay, with the τ^- decaying as $\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$ and $\tau^- \rightarrow e^- \bar{\nu}_e \nu_\tau$, is also considered a background for this analysis.

III.1.3 Simulated samples

Various decays are simulated to model the signal and background contributions. The simulated samples include the muon and electron signal modes, as well as various background decays to excited charmed meson and double-charm intermediate states. The simulation is known to not model the data perfectly well. The procedure to correct for these discrepancies is described in Section III.2.

Similar background decays are combined in 'cocktail' samples. The excited charm mesons considered in the D^{**} background samples are summarised in Table 5. In addition to that, decays to even heavier excited states, which are not well measured, are included in a separate sample ('higher' excited background), which are summarised in Table 6. In total 11 different samples are simulated for each the muon and electron signal mode covering different background sources and the signal mode decay. The samples are summarised in Table 7.

Table 5: Excited D^{**} states contained in the $B^+ \rightarrow D^{**}\ell\nu$ and $B^0 \rightarrow D^{**}\ell\nu$ simulated cocktail samples.

D^{**}	Mass (MeV)	Γ (MeV)
D_1^0	2421.4	27.1
$D_{1\frac{1}{2}}^0$	2445	250.27
D_2^{*0}	2462.6	49
D_1^-	2423	20.006
$D_{1\frac{1}{2}}^-$	2445	250.27 2
D_2^{*-}	2464.3	37

Table 6: Higher excited D^{**} states contained in the $B^+ \rightarrow D^{**}\ell\nu$ and $B^0 \rightarrow D^{**}\ell\nu$ simulated cocktail samples.

D^{**}	Mass (MeV)	Γ (MeV)	cocktail fraction (%)
$D^{*0}(2S)$	2640	-	5
$D^0(2S)$	2580	-	45
$D^0(2750)$	2462.6	49	45
$D^0(3000)$	2421.4	27.1	5
$D^{*-}(2S)$	2640	-	5
$D^-(2S)$	2580	-	45
$D^-(2750)$	2464.3	37	45
$D^-(3000)$	2423	20.01	5

Table 7: Simulated samples used in the analysis. Cocktail samples do not include a single decay, but a mixture of various similar decay chains.

initial hadron	sample type
$\bar{B}^0$	$\bar{B}^0 \rightarrow D^{*+}\ell^-\bar{\nu}_\ell$ (signal decay)
	$\bar{B}^0 \rightarrow D^{*+}\tau^-\bar{\nu}_\tau$
	double-charm background cocktail
	double-charm via τ decay cocktail
	excited D^{**} background cocktail
B^+	higher excited D^{**} background cocktail
	double-charm background cocktail
	double-charm via τ decay cocktail
	excited D^{**} background cocktail
B_s^0	higher excited D^{**} background cocktail

III.1.4 Kinematic reconstruction

The z-component of the $\bar{B}^0$ momentum $(p_{\bar{B}^0})_z$ is approximated using the partially reconstructed momentum vector $p_{\bar{B}^0, reco}$, the known $\bar{B}^0$ mass $m_{\bar{B}^0, PDG}$ and the reconstructed $\bar{B}^0$ mass $m_{\bar{B}^0, reco}$ by

$$(p_{\bar{B}^0})_z \approx \frac{m_{\bar{B}^0, PDG}}{m_{\bar{B}^0, reco}} \times (p_{\bar{B}^0, reco})_z. \quad (62)$$

The other components of the $\bar{B}^0$ momentum are calculated using this estimated $(p_{\bar{B}^0})_z$ and the $\bar{B}^0$ direction of flight obtained using the reconstructed positions of the PV and the $\bar{B}^0$ decay vertex. With this estimated momentum the $\bar{B}^0$ rest frame is approximated and the following quantities are calculated using this frame: the square of the missing mass $m_{missing}^2$ carried away by the additional not-reconstructed particle(s), the energy of the muon or electron E_l^* and the square of the invariant mass q^2 of the lepton-neutrino system. The values of $m_{missing}^2$ and q^2 are calculated as

$$\begin{aligned} m_{missing}^2 &= (p_{\bar{B}^0} - p_{D^*} - p_{lepton})^2, \\ q^2 &= (p_{\bar{B}^0} - p_{D^*})^2, \end{aligned} \quad (63)$$

using the measured four-momenta p_{D^*} and p_{lepton} of the D^{*+} and the lepton and the approximated four-momentum $p_{\bar{B}^0}$ of the $\bar{B}^0$ candidate. As will be shown, these three variables are suitable to distinguish the signal decays from the various background sources. The resolution of the approximated variables is dominated by the partially reconstructed $\bar{B}^0$ momentum and to a lesser extent by the resolution of the $\bar{B}^0$ decay vertex position. The other used variables – the PV position and the $\bar{B}^0$ daughter momenta – are reconstructed with higher precision, as the PV resolution is determined by the other tracks it produces and the D^{*+} and charged lepton momenta are fully reconstructed.

III.1.5 Initial studies

Semileptonic decays to an electron-neutrino pair have not been used for analysis in LHCb before. Therefore initial studies are performed to establish the feasibility of this analysis and to devise a strategy, where systematic uncertainties are expected to cancel in the ratio measurement or at least to be minimised. First, kinematic quantities of the muon and electron decay channels are compared after a selection requirement on the number of hits in the SPD ($n_{SPDHits} < 450$) on the data events selected by the stripping procedure. This requirement on the multiplicity

Table 8: Relative yield of inclusive trigger categories, normalised to sum of all three categories.

	μ			e		
	2016	2017	2018	2016	2017	2018
L0H	11%	12%	12%	20%	22%	23%
L0L	62%	59%	59%	39%	30%	29%
L0I	27%	29%	29%	41%	47%	48%

is chosen since it is also applied in most of the L0 lines used in this analysis. For the cancellation to occur the kinematic distributions should be similar between the electron and muon modes.

The three variables used in the fit, as well as the momentum and pseudorapidity of the lepton and the L0 trigger decision, are of interest. In the latter, the three trigger categories **L0H**, **L0L** and **L0I** are defined. The category **L0H** signifies that a hadron which is part of the signal decay deposited enough energy in the calorimeters to trigger the *L0Hadron* line. Similarly, for **L0L** the lepton of the signal decay triggered the relevant line (*L0Muon* or *L0Electron*). Finally, **L0I** corresponds to the *L0Global_TIS* decision, where any of the L0 lines is fired by particles not belonging to the signal decay. The trigger categories are not exclusive, such that an event could be included in multiple categories. The relative number of events observed in each category is summarised in Table 8. For the muon signal mode, a large portion of the candidates is included in the **L0L** category, while for the electron mode the fraction in the **L0I** category is equal or larger than the one in the **L0L** category. The difference between the muon and electron modes can be explained by the threshold in the L0 trigger being higher for the *L0Electron* line than for the *L0Muon* line. This results in an increased fraction of candidates selected by the **L0H** and **L0I** categories compared to the **L0L** category in the case of the electron signal mode. It is expected that the statistical sensitivity is driven by the electron mode yield and rejecting proportionally more muons than electrons should not heavily impact the final statistical uncertainty.

Trigger categories

Looking at the total and transverse momentum of the lepton in those trigger categories, shown in Figures 40 and 41, it is observed that for **L0H** and **L0I** the distributions are fairly similar, but different for **L0L**. This can be explained by the hardware used and thresholds required when comparing the muon and electron hardware triggers. While the former is based on the muon system and lower thresholds can be chosen, the latter uses the ECAL, where higher thresholds are necessary to suppress backgrounds from QCD processes and non-electron charged particles.

This results in the distribution for electrons being shifted to larger values. From this, the **L0L** category seems not to be suitable for this analysis, as the shifted kinematic spectrum could result in systematic uncertainties, which do not cancel in the ratio. The **L0H** category in principle could be used, but can introduce other complications. Since the simulated efficiency of the event selection at the L0-level does not reflect the actual one well enough, it needs to be corrected in a data-driven way using the TISTOS method [67], explained in more detail in Section III.2.3. In the HCAL showers from different particles can overlap, making a determination if a decision was caused by a signal or unrelated particles more difficult. Since the L0Hadron decision also takes into account the energy deposited by an electron in the ECAL, the events with an electron final state could be affected differently than ones with a muon. To avoid the systematic uncertainties associated with this correction, only the **L0I** category is used in the analysis, where a signal-independent trigger selection should yield in cancelling uncertainties when correcting for the trigger efficiency. The small difference between the momentum distribution for muons and electrons can be explained by the electron emitting Bremsstrahlung, which is not recovered.

Lepton pseudorapidity range

Now only considering the **LOTIS** category, a fiducial region, where the fit variables agree reasonably well has to be identified. In particular, the pseudorapidity η of the lepton is of interest, since the amount of material traversed and so the amount of Bremsstrahlung emitted depends on it. The fit variables m_{miss}^2 and E_l^* are shown for different regions of q^2 and η in Figures 42 and 43, respectively. For larger values of the pseudo-rapidity, the disagreement between the distributions of the muon and electron final states increases. This motivates rejecting this region by imposing $\eta_{lepton} < 3.5$.

Another motivation for selecting a specific η region is the lepton reconstruction efficiency. The analysis is performed in the pseudorapidity region $2.5 < \eta_{lepton} < 3.5$ of the lepton, since the simulated VELO reconstruction efficiency is very similar in this region for electrons and muons and the simulated overall reconstruction efficiency is stable, as shown in Figure 44.

Charm reconstruction

The analysis relies on tagging the signal mode decay with the D^{*+} to D^0 decay, where a pure sample is selected already at the stripping stage. This has been used for the muon mode in the past and also works well with the electron mode. To

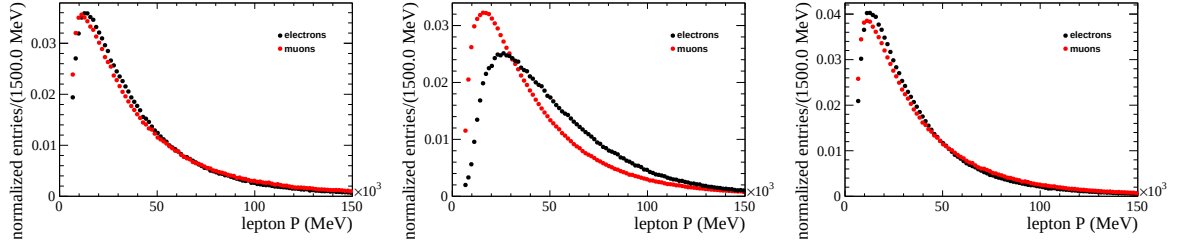


Figure 40: ● Muon (red) and electron (black) momentum in the trigger categories L0H (left), L0L (middle) and L0I (right).

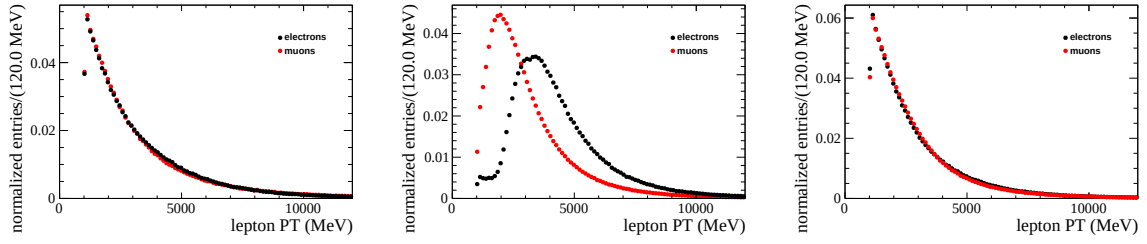


Figure 41: ● Muon (red) and electron (black) transverse momentum in the trigger categories L0H (left), L0L (middle) and L0I (right).

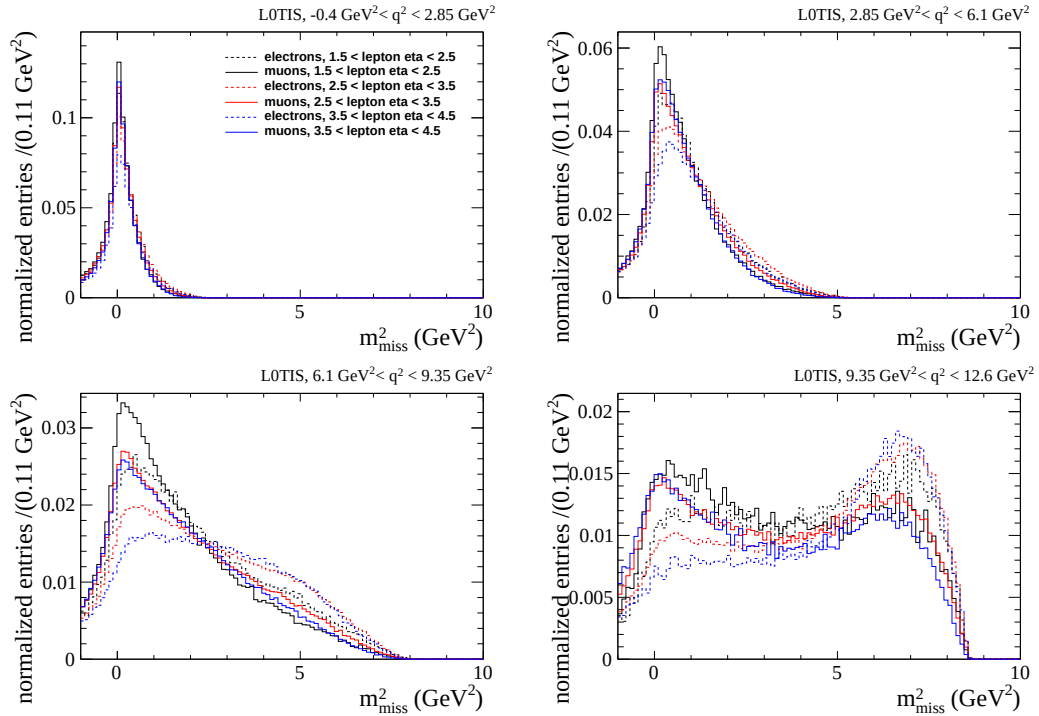


Figure 42: ● M_{miss}^2 for the muon (solid lines) and electron (dashed lines) final states in four q^2 bins in the L0I trigger category. The different colours indicate different lepton pseudorapidity ranges.

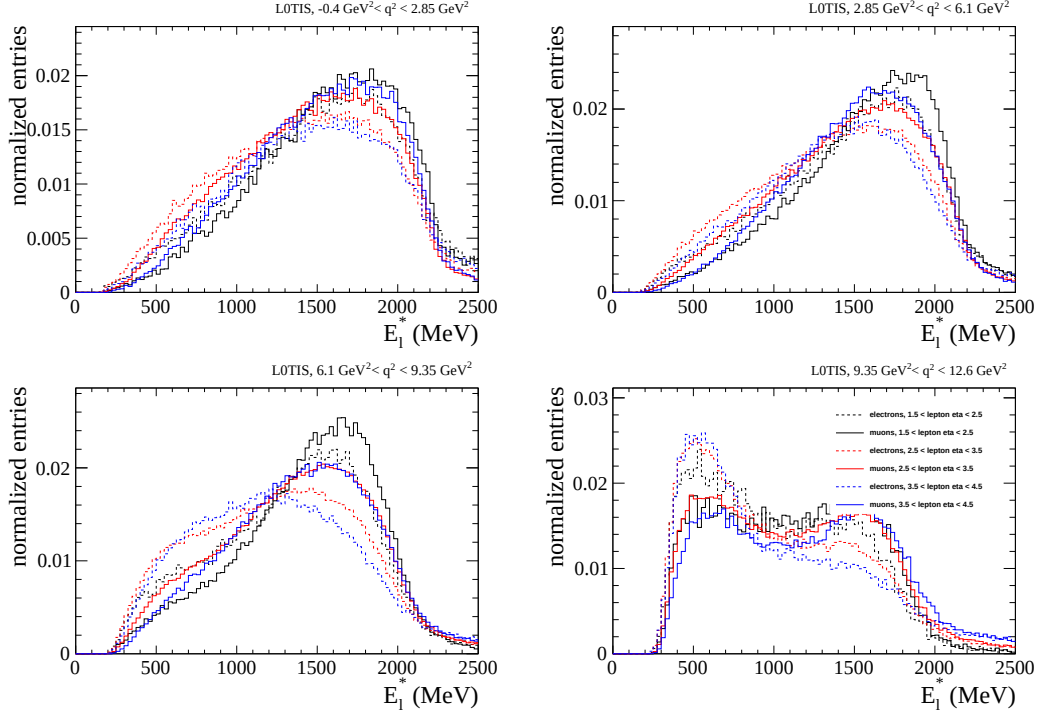


Figure 43: E_l^* for the muon (solid lines) and electron (dashed lines) final states in four q^2 bins in the L0I trigger category. The different colours indicate different lepton pseudorapidity ranges.

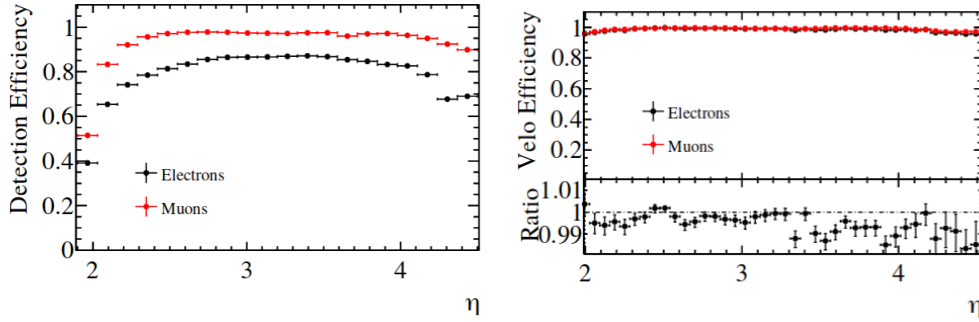


Figure 44: ■ Overall (left) and Velo (right) reconstruction efficiency determined with a simulated $B^+ \rightarrow J/\psi K^+$ sample [80].

demonstrate this, the invariant mass of the D^0 candidate and the $D^{*+}-D^0$ mass difference are shown in Figure 45.

III.1.6 Signal reconstruction and selection

In general, the selection requirement on the electron and muon signal modes are kept as similar as possible to reduce systematic uncertainties. Signal candidates are first built from events selected by the central stripping step and then a final 'offline' selection is performed to reduce the amount of background. The meaning of the different variables used in the selection is explained in Appendix A.

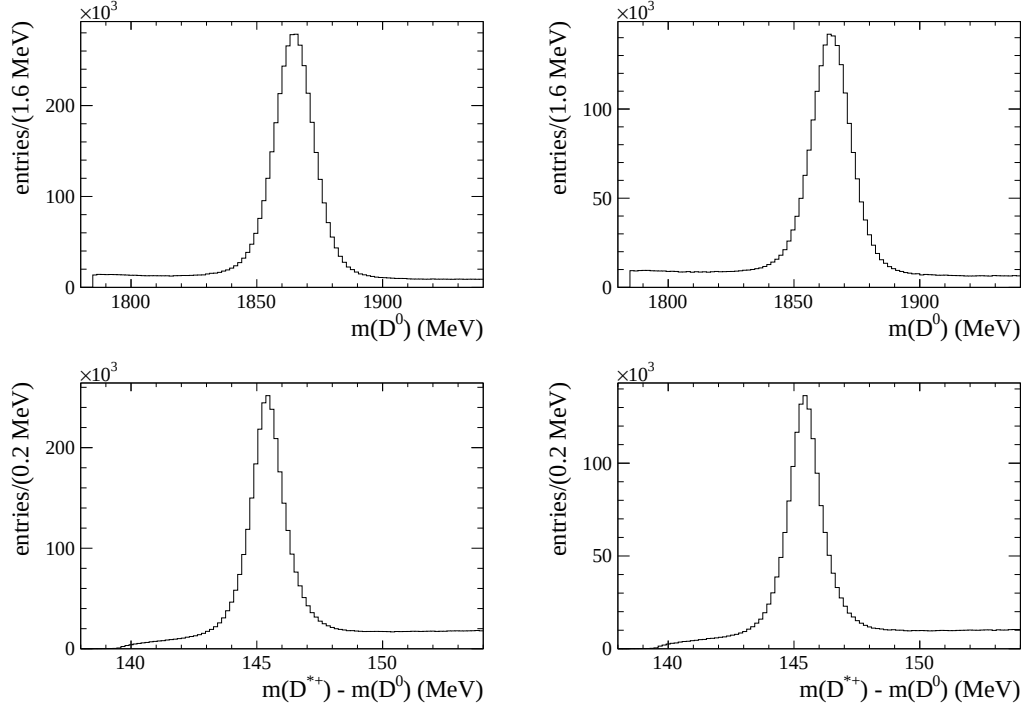


Figure 45: ● Invariant mass of the D^0 (top) and $D^{*+} - D^0$ mass difference (bottom) for the muon (left) and electron (right) signal modes.

Stripping selection

In the stripping, D^0 candidates are centrally reconstructed from $K^-\pi^+$ combinations and combined with μ^- or e^- candidates, with the selection requirements summarised in Table 9 applied. The K^- , π^+ and lepton tracks are required to be of good quality and displaced with respect to their associated PV. PID requirements are included to reduce the number of misidentified particles selected. The tracks are required to be inconsistent with being fake tracks. The D^0 candidate is required to be within 80 MeV of the known D^0 mass, have a good quality vertex and the distance of its decay vertex to the associated PV has to be significant. Finally, the $(D^0\ell^-)$ combination should have an invariant mass between 2200 MeV and 8000 MeV, result in a good quality vertex and the combined momentum vector should point in the same direction as the vector between the associated PV and the vertex. The stripping selection is the same for the muon and electron final states except for the PID requirement on the charged lepton. The stripping selection does not include any requirement on the HLT decision.

Additional selection requirements are imposed when combining the particles selected by the stripping to the final candidates and are summarised in Table 10. The vertices of the B^0 and D^{*+} candidates are required to be of good quality. The difference between the invariant masses of the D^{*+} and D^0 candidates is restricted

Table 9: Selection applied by the stripping step for the signal modes.

	common selection	
B^0	$m > 2200.0$ MeV $m < 8000.0$ MeV $\chi^2_{DOCA} < 10.$ $\chi^2_{vtx}/\text{DOF} < 9.0$ $DIRA > 0.999$	
$\bar{D}^0$	$\chi^2_{DOCA} < 20.$ $ m - \langle m_{D^0, PDG} \rangle < 80.0$ MeV $\chi^2_{vtx}/\text{DOF} < 6.0$ $\chi^2_{VD} > 25.0$ $DIRA > 0.99$ $DIRA > 0.99$	
K^+	$p > 2000.0$ MeV $p_T > 250.0$ MeV $\chi^2_{tr}/\text{DOF} < 3.0$ min. $\chi^2_{IP} > 4.0$ $PIDK > -2.0$ GHOSTPROB < 0.35	
π^-	$p > 2000.0$ MeV $p_T > 250.0$ MeV $\chi^2_{tr}/\text{DOF} < 3.0$ min. $\chi^2_{IP} > 4.0$ $PIDK < 10.0$ GHOSTPROB < 0.35	
e^+/μ^+	$p > 6000.0$ MeV $p_T > 1000.0$ MeV $\chi^2_{tr}/\text{DOF} < 3.0$ min. $\chi^2_{IP} > 9.0$ FAKEPROB < 0.35	
e^+/μ^+ PID	μ $PIDmu > 0.0$ isMuon	e $PIDe > 3.0$

to suppress the combinatorial background contribution from adding random soft pions to the D^0 candidates. The soft pion track is required to be inconsistent with being a fake track. These selection requirements are chosen to reject some combinatorial background while being efficient on the signal decays to reduce time spent on processing the data. The final selection requirements to further reduce the amount of background are described below.

Table 10: Selection applied by the event building on top of the stripping selection.

B^0	$m > 1500.0 \text{ MeV}$ $m < 10000.0 \text{ MeV}$ $\chi^2_{DOCA} < 10.$ $\chi^2_{vtx}/\text{DOF} < 12.0$ $DIRA > 0.999$
D^{*+}	$m_{D^*} - m_{D^0} < 160. \text{ MeV}$ $\chi^2_{vtx}/\text{DOF} < 9.0$
soft π^+	FAKEPROB < 0.25

Trigger selection

At the L0 level, only events triggered independently of the signal decay (TIS) are considered for further analysis. To reduce the possible difference between muons and electrons in the HLT selection, trigger-on-signal (TOS) decisions on the D^0 and its decay products are used. For HLT1 the single track and two-track combination TOS decisions and for HLT2 the two-body topological TOS decision on the two daughters of the D^0 are chosen. The trigger selection is summarised in Table 11.

Table 11: Summary of the trigger requirements for the signal modes.

L0	L0ELECTRONDECISION_TIS L0PHOTONDECISION_TIS L0MUONDECISION_TIS L0DiMUONDECISION_TIS
HLT1	HLT1TRACKMVADECISION_TOS HLT1TWOTRACKMVADECISION_TOS on D^0
HLT2	HLT2TOPO2BODYDECISION_TOS on D^0

Offline selection

The selection applied after the events are reconstructed by the stripping and event building are summarised in Table 12. A tight window around the known value of the $D^{*+} - D^0$ mass difference is chosen to minimise combinatorial background contributions.

To reduce backgrounds from decays with additional charged tracks, an MVA-based isolation criterion is applied, which considers all tracks in an event, which are not part of the signal decay. If an unrelated track is found to be consistent with the $\bar{B}^0$ vertex, the event should be classified by the MVA as background rather than signal and be rejected. For this purpose, a boosted decision tree (BDT) classifier is trained using the TMVA package [81]. Isolated tracks from the $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu$ decay and associated tracks from the $B^0 \rightarrow D^{*+} \mu^- \nu_\mu$ decay are used as signal and

Table 12: Selection applied by the offline selection.

	common selection
B^0	$m < 5480.$ MeV $2.5 < \eta < 3.5$
D^{*+}	$ m_{D^*} - m_{D^0} < 6.0$ MeV
D^0	$\chi_{IP,PV}^2 > 9.$ $1835.$ MeV $< m < 1895.$ MeV
e^+/μ^+	$p > 3000.0$ MeV $2.5 < \eta < 3.5$
occupancy	$n_{SPDHits} < 450$
isolation MVA	maximum MVA output < 0.1

Table 13: Input variables in the isolation MVA training.

B	$\log(PT)$
additional particle	$\log(PT)$ $\Delta\phi$ $\Delta\theta$ $\alpha_{opening}$ $\arccos(\alpha_{opening})$ $\cos(\alpha_{opening})$
	$\log(IP_{new})$ $\log(IP_{old})$ $\log(\chi_{IP,new}^2)$ $\log(\chi_{IP,old}^2)$
	$\log(\chi_{FD,new}^2 - \chi_{FD,old}^2)$ $\log(\chi_{FD,new}^2)$ $\log(\min(\chi_{IP}^2))$
vertex	$\log(\chi_{vtx,new}^2)$ $\log(\chi_{vtx,old}^2)$ $\log(\chi_{vtx,new}^2 - \chi_{vtx,old}^2)$

background input samples. 'Isolated' signifies here additional tracks in the events, which are not part of the $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu$ decay, whereas associated tracks are the additional tracks from the $B^0 \rightarrow D^{*+} \mu^+ \nu_\mu$ decay. In both cases, all the long, VELO and upstream tracks in the event are used to not miss any additional particles. The input variables used for the classifier are summarised in Table 13. They are based on geometric quantities of the additional tracks with respect to the $\bar{B}^0$ candidate, as well as quantities related to the vertex fit with and without the additional track included. The BDT classifier is configured such that its output is large if a track is likely to originate from the $\bar{B}^0$ vertex. For each event, the maximum value of the BDT output of all tracks is used in the selection.

Different values for the hyperparameters have been tried out when training the BDT to find the configuration which maximises the area under the ROC (receiver operating characteristic) curve, which is the background rejection rate as function of the signal efficiency obtained on the BDT input samples for different requirements on the BDT output. To optimise the BDT selection requirement the signal significance is used as the figure of merit (FOM) defined as

$$FOM = \frac{S}{\sqrt{S+B}}, \quad (64)$$

where S and B are the expected signal and background number of events. The FOM is evaluated at different requirements of the maximum BDT output to esti-

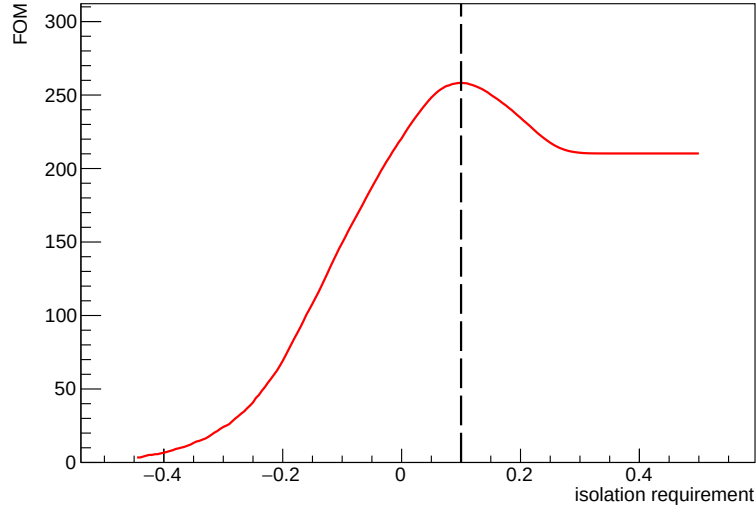


Figure 46: ● Signal significance as function of the isolation requirement. The dashed black line indicates the value where the signal significance is at its maximum.

mate the position of the maximum, as shown in Figure 46. The expected number of events is estimated from the simulated samples, where the background includes the excited and double-charm samples with yields scaled to account for the relative branching fraction of signal and the backgrounds and the different number of events generated in the simulation. Also the expected relative amount of combinatorial background is included when determining B , which is estimated from a fit to the signal sample without any BDT requirement. The rejection rate of the isolation requirement for the $B \rightarrow D^{**}$ samples as function of the efficiency on the signal mode, estimated from the simulated cocktail samples, is shown in Figure 47. Since decays of D^{**} from B^0 mostly produce additional π^0 particles, the rejection power on this sample is smaller compared to the B^+ sample, where mostly additional charged pions are produced. The performances on the double-charm background cocktail samples are also shown. The larger rejection performance compared to the D^{**} cocktail samples can be explained by the double-charm cocktail samples including more decay paths with additional charged particles.

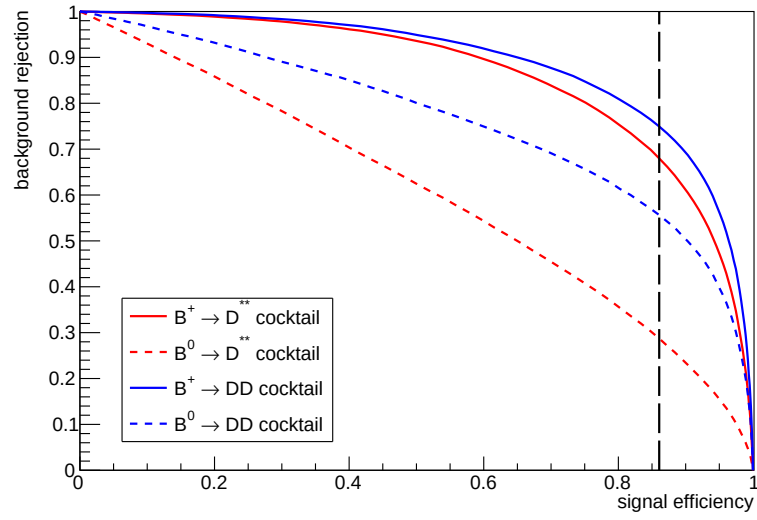


Figure 47: ● Background rejection as function of the signal efficiency for the isolation selection. The dashed black line indicates the signal efficiency at the BDT requirement where the FOM is at its maximum.

III.1.7 The $B^0 \rightarrow J/\psi K^{*0}$ control channel ▲

The channel $B^0 \rightarrow J/\psi K^{*0}$ is used extensively to verify the analysis procedure and to calculate corrections to the simulation where applicable. For the reconstruction of the J/ψ meson, both the $J/\psi \rightarrow e^+e^-$ and $J/\psi \rightarrow \mu^+\mu^-$ channels are used, and named as dimuon and dielectron control mode, respectively. The K^{*0} particle is reconstructed in the $K^+\pi^-$ decay.

The dimuon control mode, $B^0 \rightarrow J/\psi K^{*0}$ with $J/\psi \rightarrow \mu^+\mu^-$, which is chosen because of its large data and simulation sample sizes and signal purity, is reconstructed by the stripping procedure³ with the selection requirements summarised in Table 26 in Appendix B. The stripping selection requires B^0 , J/ψ and K^{*0} candidates with good quality vertices. The final state particles are required to be displaced with respect to the associated PV. For the dielectron control mode, a stripping line⁴ similar to the dimuon mode is used, with the main difference being the lepton PID selection. The stripping selection requirements are summarised in Table 27, also found in Appendix B. The dielectron mode is less clean compared to the dimuon mode due to Bremsstrahlung, but still can be used to cross-check the simulation corrections obtained on the dimuon mode.

The invariant-mass distributions of the reconstructed J/ψ , K^{*0} and B^0 candidates after the stripping selection and a requirement on the number of hits in the SPD ($nSPDHits < 450$) are shown in Figures 48 and 49 for the dimuon and dielectron modes, respectively. Also shown is the B^0 invariant mass distribution after the decay chain is re-fitted with a DTF where the J/ψ mass is constrained to its known value, which improves the resolution of the B^0 mass peak, especially for the dielectron mode.

As alternative decay mode to obtain the correction weights the $B^+ \rightarrow K^+ J/\psi$ mode could be used. The advantage would be its larger signal yields. The reason why the $B^0 \rightarrow J/\psi K^{*0}$ decay was chosen instead is that the large simulated samples necessary were already available for the $B^0 \rightarrow J/\psi K^{*0}$ decay for all years at the beginning of the analysis, while $B^+ \rightarrow K^+ J/\psi$ simulation samples still had to be produced. The effect of this choice on the sensitivity is estimated in Section III.5.3.

Selection

To increase the purity of the control channel and reject specific backgrounds additional selection requirements are applied after the stripping selection. Kinematic, topological and PID variables are used, and the detailed selections are listed in

³Bu2LLK_mmLine

⁴Bu2LLK_eeLine

Table 28 for the dimuon control mode, and Table 29 for the dielectron control mode, both found in Appendix B. The trigger requirements are the same for both dimuon and dielectron modes, which are listed in Table 14. The simulation uses the same thresholds for the L0 trigger lines within each year, while they were varied during data-taking. To account for this only the fraction of the control mode data where the thresholds align with the simulation is used.

Additional selection requirements are imposed to reduce contributions from specific background sources. The $\Lambda_b^0 \rightarrow J/\psi p K^-$ decay can be reconstructed as a control mode candidate if the proton is misidentified as a pion. The amount of this background contribution is estimated in the dimuon control mode. The invariant mass distribution when assigning the pion the proton mass is shown in Figure 50 as blue curve. A peaking structure at the known value of the Λ_b^0 mass can be seen after the stripping selection but vanishes after the following selection step. Still,

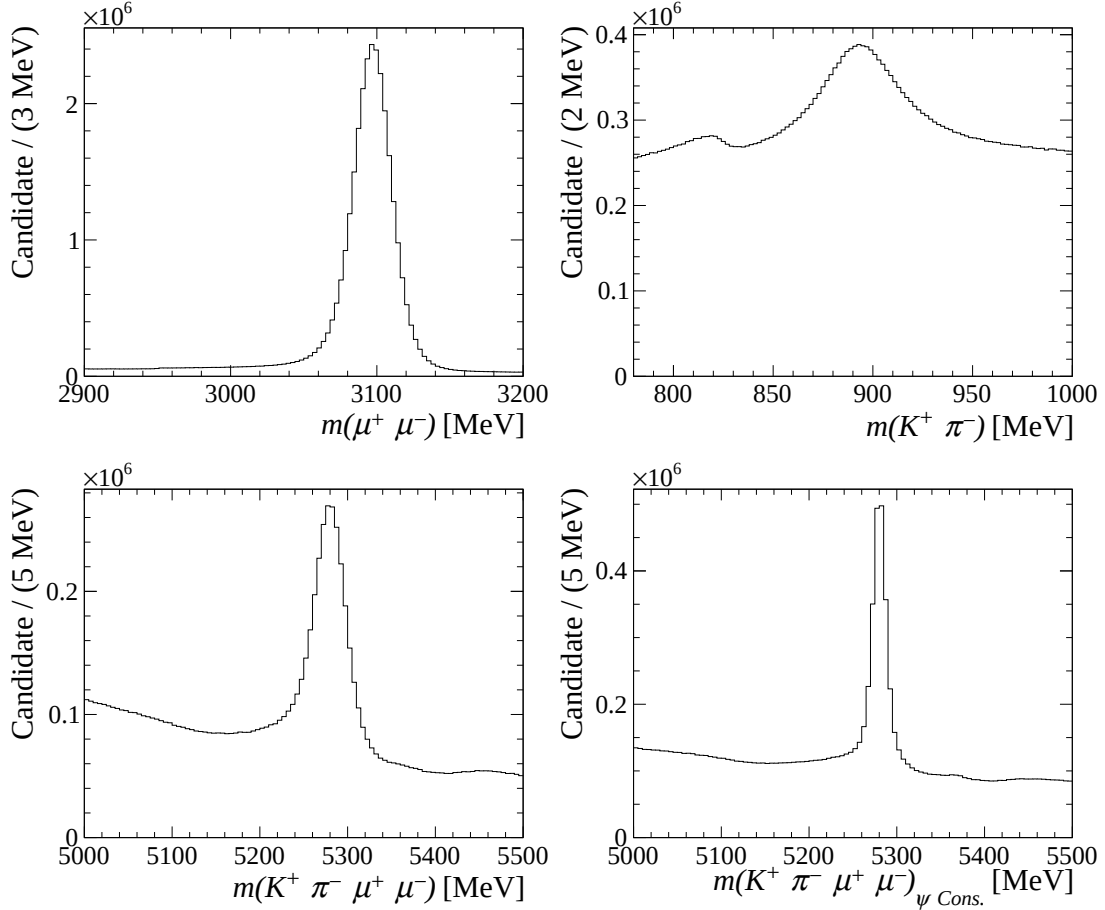


Figure 48: ▲ Invariant-mass spectra of the J/ψ (top left), K^{*0} (top right) and B^0 (bottom) candidates of the dimuon control mode. The bottom left figure is the B^0 candidates' mass spectrum without a $m(J/\psi)$ constraint, and the bottom right one is for the B^0 candidates' mass spectrum with the $m(J/\psi)$ constraint. Data samples of 2016, 2017 and 2018 data-taking years are combined when making these figures.

to decrease this background contribution a veto is applied when the pion is likely a misidentified proton by rejecting events with $m(J/\psi p K^-) \in (5570, 5670) \text{ MeV}/c^2$ and $probNNp(\pi \text{ misID to } p) > 0.2$.

Another potential background contribution is the decay $B^+ \rightarrow J/\psi K^+$ in combination with an additional random pion. To estimate the size of this background the invariant-mass distribution of the $J/\psi K^+$ system, leaving out the pion, is shown in Figure 51. A peaking structure can be seen at the B^+ invariant-mass indicating the presence of these decays. They are vetoed by rejecting events with $m(J/\psi K^+) > 5230 \text{ MeV}/c^2$.

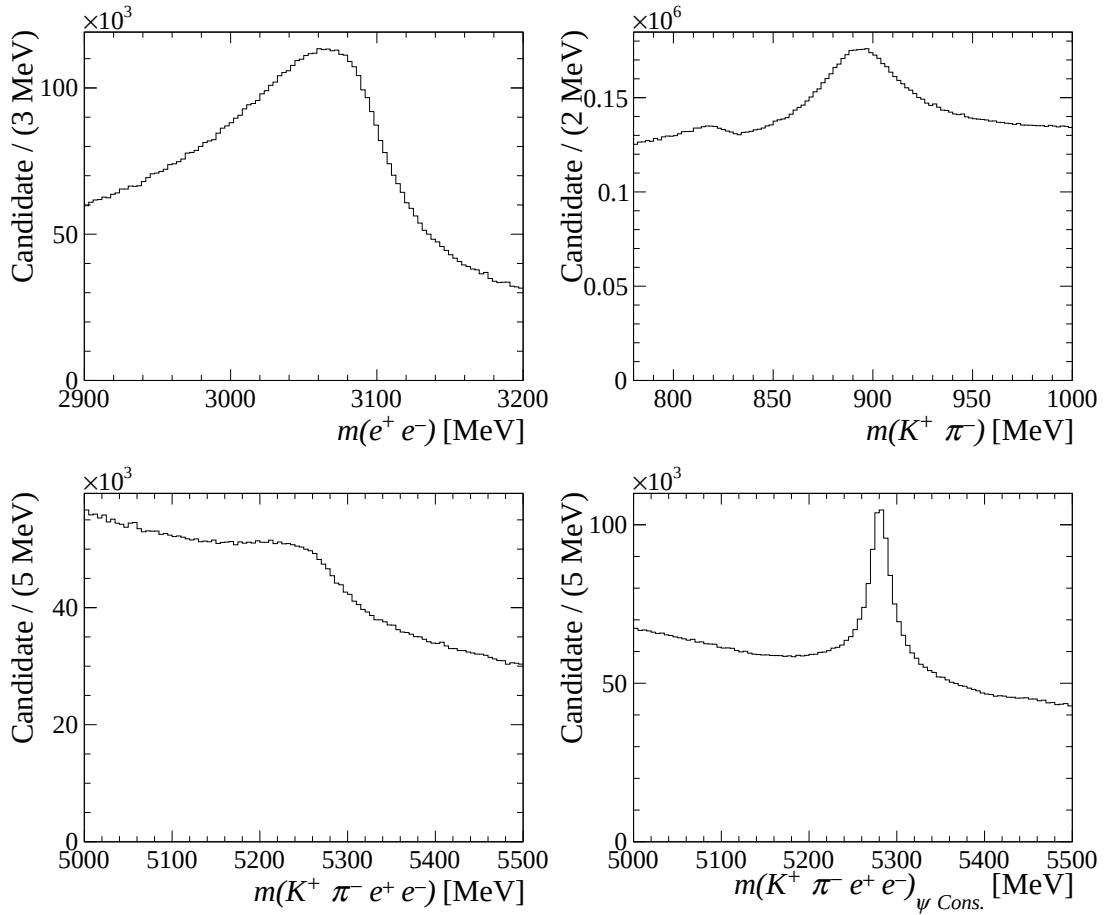


Figure 49: ▲ Invariant-mass spectra of the J/ψ (top left), K^{*0} (top right) and B^0 (bottom) candidates of the dielectron control mode. The bottom left figure is the B^0 candidates' mass spectrum without a $m(J/\psi)$ constraint, and the bottom right one is for the B^0 candidates' mass spectrum with the $m(J/\psi)$ constraint. Data samples of 2016, 2017 and 2018 data-taking years are combined when making these figures.

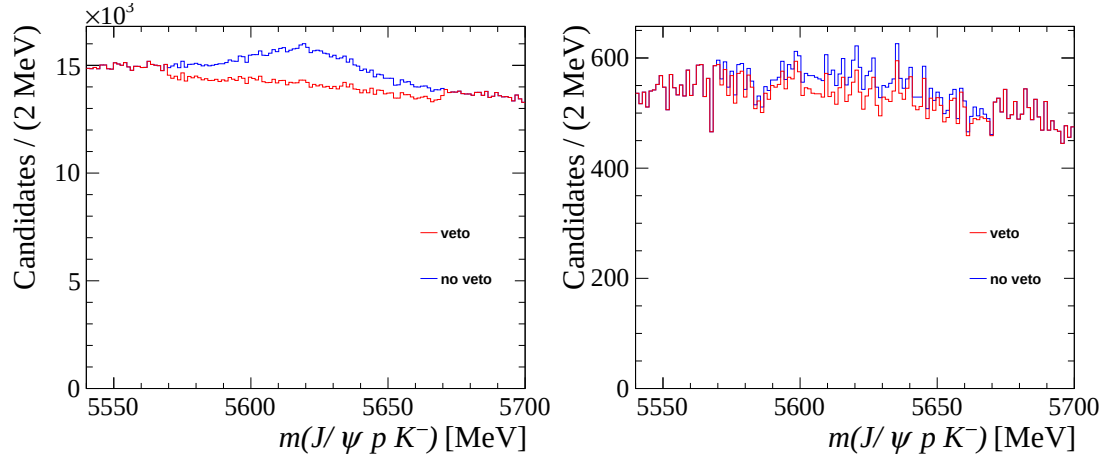


Figure 50: ▲ Invariant-mass spectrum of the $J/\psi \pi^+ K^-$ system when assigning the proton mass to the pion. The left distribution is obtained after the stripping selection and the right one after the following selection step. The blue curve is the distribution before the veto to the $\Lambda_b^0 \rightarrow J/\psi p K^-$ misID background and $B^+ \rightarrow J/\psi K^+$ partially reconstruction background, and the red one shows the distribution after the veto.

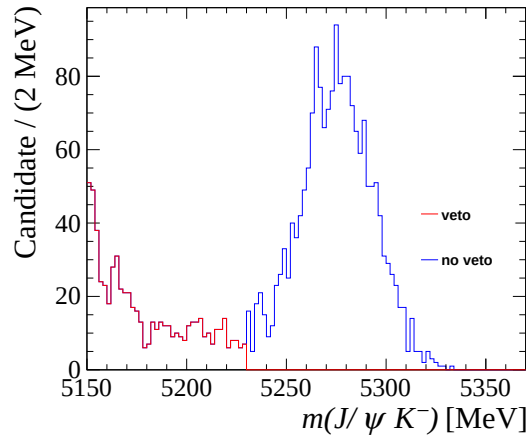


Figure 51: ▲ Invariant-mass spectrum of the $J/\psi K^-$ system. The blue curve is the distribution before the veto to the $\Lambda_b^0 \rightarrow J/\psi p K^-$ misID background and $B^+ \rightarrow J/\psi K^+$ partially reconstruction background, and the red one is the distribution after the veto.

To subtract the contribution from the remaining combinatorial background, the *sPlot* technique [82] is used. This method allows the calculation of per-event weights based on the distribution of a discriminating variable. For each category – in this case signal and combinatorial background – a weight is obtained. By construction, the sum of the respective weights is equal to the yield of the contribution and by using these weights the distributions of other variables can be extracted. One limitation to this approach is that the discriminating variable must be uncorrelated to the variable in question.

To obtain these so-called 's-weights' the invariant mass distribution of the B^0 candidates is used as the discriminating variable and fitted for both the dimuon and dielectron modes. The correlation of the B^0 invariant mass distribution with other variables has been checked and found to be small so that the *sPlot* method is applicable. The signal component is described by the sum of two single-sided Crystal ball functions [83] with shared mean and the combinatorial background by an exponential function. The parameters describing the tails of the Crystal Ball functions are fixed to the values obtained from a fit of the signal component to the distribution in simulation. In addition, the contribution from the $B_s^0 \rightarrow J/\psi K^{*0}$ decay is considered in the fit model. The shape of the B_s^0 peak is assumed to be the same as that of the signal component of the $B_s^0 \rightarrow J/\psi K^{*0}$ decay. The distance between the B^0 and B_s^0 peak is fixed to the difference of known masses of the B_s^0 and B^0 particles [84]. For the dimuon control mode, the yield of the $B_s^0 \rightarrow J/\psi K^{*0}$ component is left as a free parameter, while for the dielectron control mode, the yield ratio between the $B_s^0 \rightarrow J/\psi K^{*0}$ and $B_s^0 \rightarrow J/\psi K^{*0}$ components is fixed to that obtained from the fit of the dimuon mode. The fitted distributions are shown in Figure 52 for the dimuon control mode, and Figure 53 for the dielectron control mode.

Table 14: Trigger selection of the control channel $B^0 \rightarrow J/\psi K^{*0}$, $J/\psi \rightarrow l^+ l^-$, where l stands for a muon or electron particle. The DiMuon and TrackMuon lines are added to enlarge the statistic of the dimuon control mode, and they are kept also for the dielectron control mode to have a consistent trigger requirement between the two control channels.

L0	Photon, Electron, Muon, DiMuon Decision TIS	
HLT1	Hlt1DiMuonHighMassDecision_TOS	Hlt1TrackMuonDecision_TOS
	Hlt1TrackMVADecision_TOS	Hlt1TwoTrackMVADecision_TOS
HLT2	Hlt2DiMuonDetachedJPsiDecision_TOS	Hlt2Topo2BodyDecision_TOS
	Hlt2Topo3BodyDecision_TOS	Hlt2Topo4BodyDecision_TOS

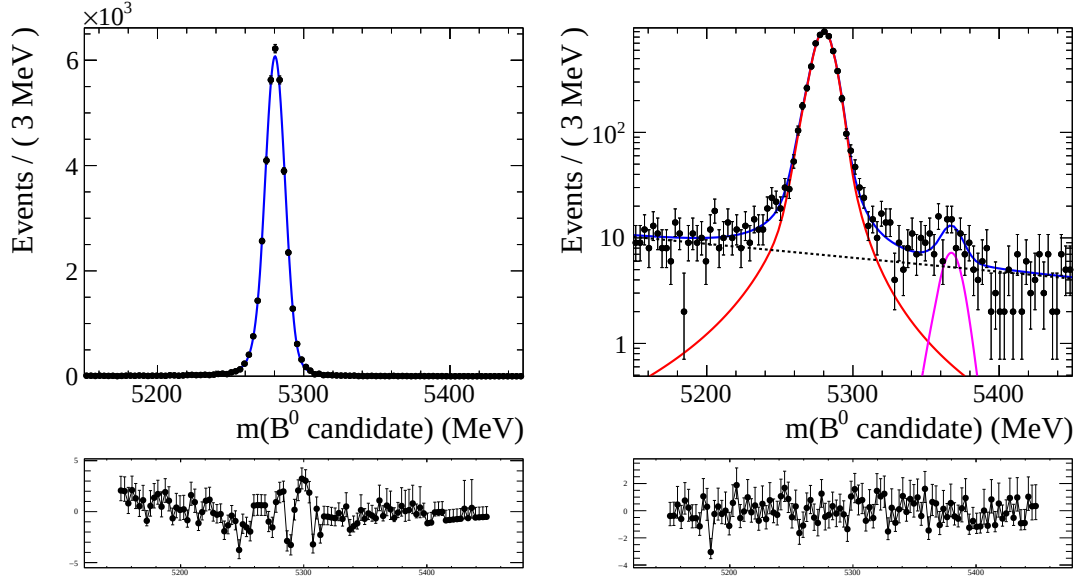


Figure 52: $\blacktriangle$ Fit of the signal component to simulated sample (left) and fit of total p.d.f to data (right). The invariant mass of the B^0 candidate is obtained after a $m(J/\psi)$ constraint. The distribution is obtained using the $B^0 \rightarrow J/\psi K^{*0}, J/\psi \rightarrow \mu^+ \mu^-$ sample collected in 2017 after all the event selections. The data distribution is displayed as the black dots. The total PDF is shown in the blue solid line. The red solid line, magenta solid line and black dashed line show the contribution from the signal $B^0 \rightarrow J/\psi K^{*0}$ decays, the $B_s^0 \rightarrow J/\psi K^{*0}$ decays and the combinatorial background, respectively.

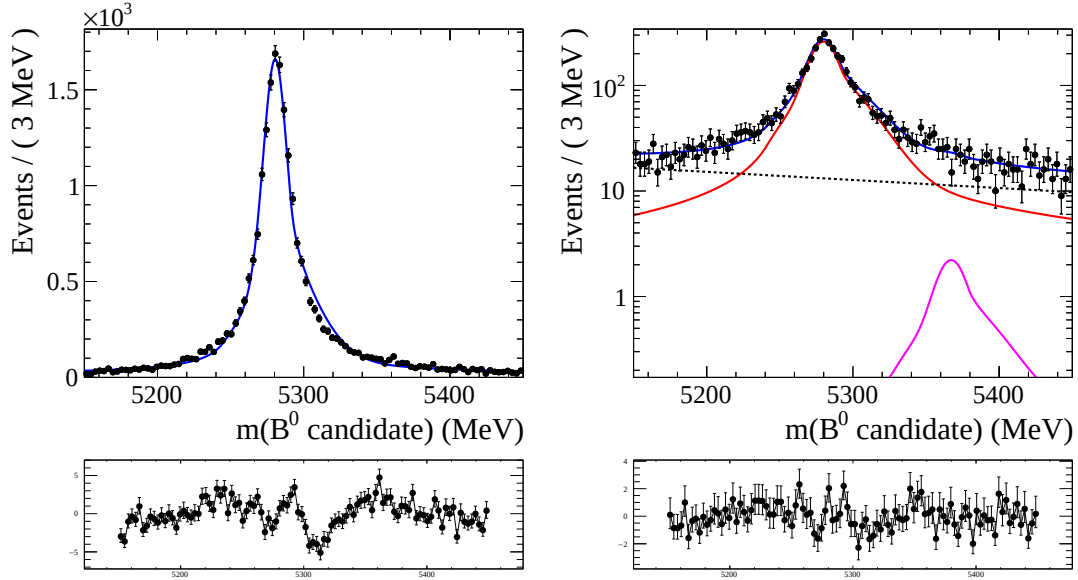


Figure 53: $\blacktriangle$ Fit of the signal component to simulated sample (left) and fit of total p.d.f to data (right). The invariant mass of the B^0 candidate is obtained after a $m(J/\psi)$ constraint. The distribution is obtained using the $B^0 \rightarrow J/\psi K^{*0}, J/\psi \rightarrow e^+ e^-$ sample collected in 2017 after all the event selections. The data distribution is displayed as the black dots. The total PDF is shown in the blue solid line. The red solid line and black dashed line show the contribution from the signal $B^0 \rightarrow J/\psi K^{*0}$ decays and the combinatorial background, respectively.

III.1.8 Combinatorial background

Background caused by random combinations of particles needs to be considered to describe the signal mode data well. Three different sources of combinatorial background are included. A real D^{*+} combined with a fake muon or electron (track misidentified as lepton), a 'fake' D^* (random D^0 soft π combination) with a real lepton and the random combination of a real D^{*+} with a real lepton. In all three cases, the shape of the corresponding background template is determined using a dedicated data sample.

Real D^* with misidentified lepton

Three different types of lepton misidentification are considered: $h \rightarrow e$, $\mu \rightarrow e$ and $(h,e) \rightarrow \mu$, where $h = K, \pi, p$.

$h \rightarrow e$ and $(h, e) \rightarrow \mu$ misidentification

To estimate the shape of the $h \rightarrow e$ and $(h, e) \rightarrow \mu$ background, the data sample produced by a dedicated stripping line⁵ is used. It consists of the same selections as the line used for the muon signal channel selection described earlier, with the only difference being that the PID requirement on the muon is inverted to ($PID_{mu} < 0.0$). The same offline selection without PID requirements is applied.

Four sub-samples are selected from the FakeMuon sample, each containing mainly pions, kaons, protons or electrons. The true number of each particle species for the PID selections in bins of p , η and $nTracks$ is estimated by

$$\begin{pmatrix} N_{\pi,true} \\ N_{K,true} \\ N_{p,true} \\ N_{e,true} \end{pmatrix} = H^{-1} \begin{pmatrix} N_{\pi,sel} \\ N_{K,sel} \\ N_{p,sel} \\ N_{e,sel} \end{pmatrix}, \quad (65)$$

where H is the matrix

$$\begin{pmatrix} \epsilon_{\pi \rightarrow \pi} & \epsilon_{K \rightarrow \pi} & \epsilon_{p \rightarrow \pi} & \epsilon_{e \rightarrow \pi} \\ \epsilon_{\pi \rightarrow K} & \epsilon_{K \rightarrow K} & \epsilon_{p \rightarrow K} & \epsilon_{e \rightarrow K} \\ \epsilon_{\pi \rightarrow p} & \epsilon_{K \rightarrow p} & \epsilon_{p \rightarrow p} & \epsilon_{e \rightarrow p} \\ \epsilon_{\pi \rightarrow e} & \epsilon_{K \rightarrow e} & \epsilon_{p \rightarrow e} & \epsilon_{e \rightarrow e} \end{pmatrix}. \quad (66)$$

The selection criteria for pions, kaons, protons and electrons within the FakeMuon sample are listed in Table 15. The PID selection efficiency $\epsilon_{i \rightarrow k}$ for a particle

⁵B2DMuNuX_D0_FakeMuon

Table 15: Criteria used to select pions, kaons, protons and electrons from the FakeMuon sample. When processing the PIDCalib performance histograms for PID (mis)ID efficiencies, $DLL_{\mu} < 0$ is required for all four particle species in addition to the cuts listed in the table, since this cut is applied within the stripping of the FakeMuon line.

Pions	$MC15TuneV1_ProbNN_{\pi} > 0.2 \ \&\& \ DLL_p < 1 \ \&\& \ DLL_K < 1$
Kaons	$MC15TuneV1_ProbNN_K > 0.2 \ \&\& \ DLL_p - DLL_K < 0 \ \&\& \ DLL_K > 0$
Protons	$MC15TuneV1_ProbNN_p > 0.2 \ \&\& \ DLL_p - DLL_K > 4 \ \&\& \ DLL_p > 0$
Electrons	$MC15TuneV1_ProbNN_e > 0.2 \ \&\& \ DLL_e > 0 \ \&\& \ DLL_p < 1 \ \&\& \ DLL_K < 1$

Table 16: Binning scheme used to create PIDCalib performance histograms for PID (mis)identification efficiencies.

p	[3, 6, 9.3, 10, 15.6, 20, 40, 60, 100] GeV
η	[2.5, 3.5]
$nTracks$	[0, 150, 225, 500]

of species i to pass the selection requirement of the species k is estimated with the *PIDCalib* package [85], which employs centrally selected calibration samples for this purpose. Further requirements are imposed on the calibration samples to avoid biases when estimating the $h \rightarrow e$ and $h \rightarrow \mu$ misidentification rate and to ensure that acceptance effects are separated from the misidentification probability. The $(p, \eta, nTracks)$ binning scheme is listed in Table 16. The obtained misidentification matrices H are shown in Appendix C.

The PID requirements to select the different particle species in the FakeMuon sample are not exclusive and each event can pass or fail each requirement. Using the estimated true number of each particle species in each bin the probability that an event in the FakeMuon sample contained a certain particle species can be estimated taking into account which of the PID selections are passed by the event. For this purpose the modified matrix H' is calculated

$$H'_{ij} = \begin{cases} \epsilon_{q_i \rightarrow q_j}, & \text{if the track fulfils the PID criteria of } q_j \\ 1 - \epsilon_{q_i \rightarrow q_j}, & \text{if the track does not fulfil the PID criteria of } q_j \end{cases} \quad (67)$$

where $q_i, q_j \in K, \pi, p, e$. The probability p_k for a track to be of a certain particle species is then calculated using this matrix by

$$N_{k,exp} = N_{k,true} \times \prod_i H'_{ik} \quad (68)$$

$$N_{tot} = \sum_k N_{k,exp} \quad (69)$$

$$p_k = N_{k,exp}/N_{tot}. \quad (70)$$

Finally the shape of the $(h, e) \rightarrow \mu$ background is obtained by weighting the Fake-Muon sample with weights calculated by

$$w = \sum_k p_k \times \epsilon_{k \rightarrow \mu}, \quad (71)$$

where $k = K, \pi, p, e$. The weights for $h \rightarrow e$ are obtained similarly by

$$w = \sum_k p_k \times \epsilon_{k \rightarrow e}, \quad (72)$$

where $k = K, \pi, p$. The probabilities $\epsilon_{k \rightarrow \mu}$ and $\epsilon_{k \rightarrow e}$ that a particle of species k is misidentified as a muon or an electron are again determined from the PIDCalib calibration samples using the same binning and taking the PID requirements imposed on the muon or electron when selecting the signal samples into account.

$\mu \rightarrow e$ misidentification ▲

The amount of $\mu \rightarrow e$ misidentification is estimated from the PIDCalib performance histogram of muons with the electron selection requirement of $DLL_e > 3$ applied, in bins of p , p_T and $nTracks$. The average $\mu \rightarrow e$ misidentification efficiency for the kinematic distribution of muons in data is less than 0.05% for all data samples. Therefore, this component is negligible and its shape is not accounted for in the template fit.

Combinatorial D^* with real lepton

The template shape used in the fit of the combinatorial background due to fake D^* is estimated using a data sample where a D^0 is combined with a wrong-sign π^- . This sample is reconstructed and selected in the same way as the signal modes when constructing the template used in the fit, while the requirement on the $D^{*+}D^0$ mass difference is removed for the study below. To verify that this sample accurately describes the kinematic distribution in the signal samples, various quantities are compared between the signal and fake D^* sample in different bins in the upper sideband of the $\Delta(M_{D^{*+}} - M_{D^0})$ distribution, defined by $150 \text{ MeV} < \Delta(M_{D^{*+}} - M_{D^0}) < 160 \text{ MeV}$, as shown in Figures 118 and 119 in Appendix D. The distributions agree reasonably well and do not depend on the value of $\Delta(M_{D^{*+}} - M_{D^0})$, giving confidence that the sample should be able to model the background below the signal peak in the $\Delta(M_{D^{*+}} - M_{D^0})$ distribution.

To estimate the expected yield of this background, a fit to the mass difference $\Delta(M_{D^{*+}} - M_{D^0})$ spectrum of the signal samples is used, shown in Figure 54. The

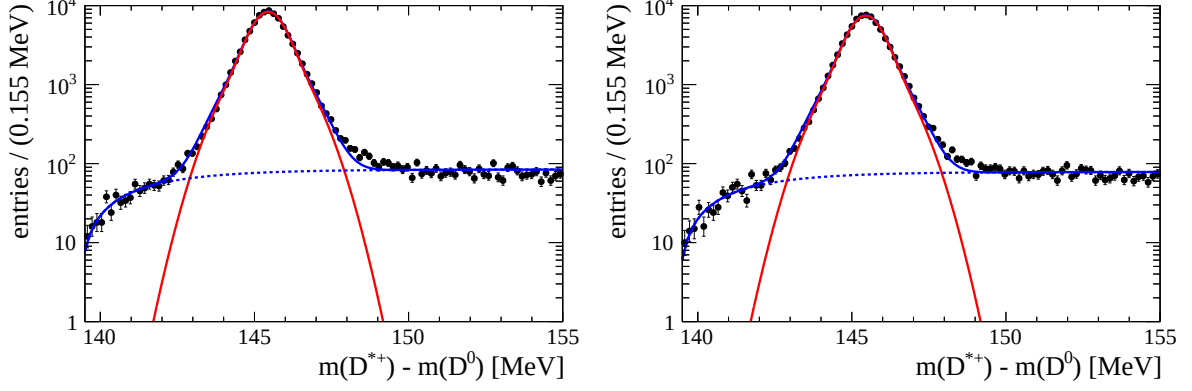


Figure 54: $\bullet$ $\Delta(M_{D^{*+}} - M_{D^0})$ distribution (black points) overlaid by the fitted total function (blue) for the muon (left) and electron (right) final states. The signal distribution is shown as red solid line, the background as blue dotted line.

signal shape is described by the sum of a CrystalBall and Gaussian function, while the phenomenological **DstD0BG** function, implemented in the *RooFit* package [86], is used to describe the distribution of the background. It can be seen that the background distribution is well described. The expected yield is then taken as the integral of the fitted background shape in the nominal selection region.

Real D^* with real lepton

The shape of the background due to random combinations of real D^{*+} with a real μ^-/e^- is obtained by adding wrong-sign leptons to a D^0 in the stripping and then building the $\bar{B}^0$ candidate as before. To estimate the expected yield of this contribution, the number of events of the wrong-sign (WS) and right-sign (RS) sample in the upper $\bar{B}^0$ mass region are compared as shown in Figure 55. The ratio of RS to WS number of events is determined in bins of the $\bar{B}^0$ mass. A dependence of this ratio on the $\bar{B}^0$ mass is observed and described by a linear function. The expected yield of this combinatorial background is then determined by integrating the fitted linear function over the $\bar{B}^0$ mass region included in the signal selection, which gives an overall scale of RS to WS number of events in the signal region. The expected number of WS events in the signal region is then multiplied by this scale to get the expected number of background events in the RS sample.

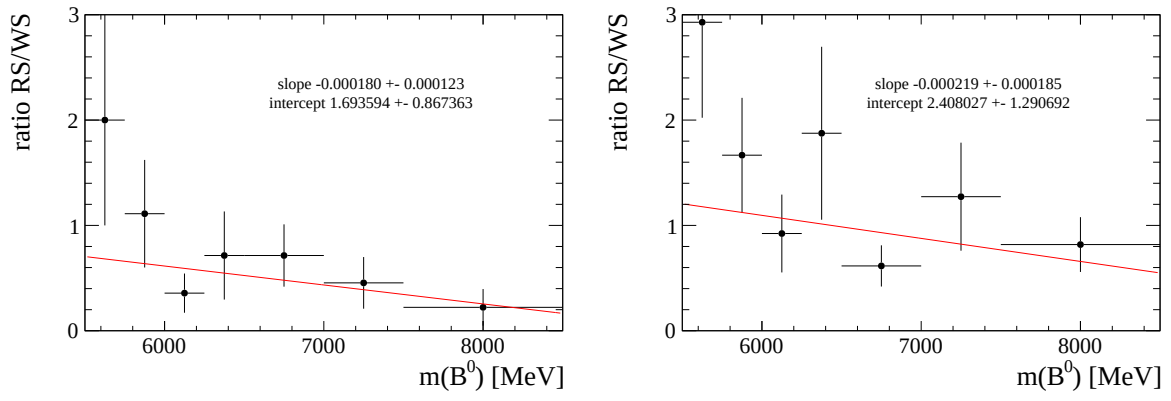


Figure 55: ● Fit of the ratio of number of events between the right- and wrong-sign sample in bins of the $\bar{B}^0$ mass for the muon (left) and electron (right) signal modes for 2017 data.

III.2 Corrections to the simulation

The simulation is known not to describe the data perfectly well. This can impact the shapes used in the template fit obtained from simulation and the determination of the efficiency ratio. There are many well-known discrepancies, which must be corrected to allow for a precise measurement. It is known that the kinematic distribution of B^0 particles generated in simulation differs from the one observed in data. This is due to the difficulties of accurately simulating the pp collision and the strong processes taking place there. Variables related to particle identification (PID), especially those from the RICH detectors, are not well described by the simulation as these depend on external conditions, which are hard to emulate in the simulation. For example, the temperature and pressure of the RICH gases vary during data-taking but are set to a fixed value in simulation. The multiplicity of particles traversing the detector is in general underestimated by simulation. The efficiency of the track reconstruction requires a correction to achieve consistency with data. Also the response of the trigger system is not perfectly modelled.

Such discrepancies can be studied by comparing the normalised distribution of a particular variable for a particular decay mode in background-subtracted data to the one obtained when simulating that decay. Agreement between data and simulation is then achieved by defining a correction weight as the ratio of data and simulation in bins of the variable in question. The binning scheme needs to be chosen in a way that each bin has enough statistics to reduce the statistical uncertainty, but the bin widths can not be too broad, as variation within a bin would increase the systematic uncertainty. By correcting the simulated sample using these weights on an event-by-event basis the discrepancies can be reduced. In practice, this means down- or up-weighting events lying in regions where the simulation produced too many or too few decays. In a similar way, the simulated tracking and trigger efficiencies can be corrected. The challenge is to obtain the background-subtracted data sample to be able to do this comparison. This is achieved by using calibration samples with high purity and obtaining the signal component using the *sPlot* method. When comparing trigger efficiencies they are obtained on simulation and data using the TISTOS method.

The strategy followed for obtaining the correction weights for the simulated signal decay samples is to rely on the control channel as much as possible, as it is more difficult to separate the signal contribution from the background in case of the signal modes. For this purpose, the simulated sample of the control channel is corrected in an iterative way, where each correction step is applied before the next one is calculated. The corrections are determined and applied in the following

order:

1. Particle identification efficiency
2. Tracking efficiency
3. L0 trigger efficiency
4. Event multiplicity
5. B^0 kinematics
6. HLT trigger efficiency

This order is chosen to be able to obtain correction weights for generator-level effects, which affect the B^0 kinematics and event multiplicity, without being distorted by the trigger effect. This is necessary as the generator-level events are also weighted with the B^0 kinematics and multiplicity corrections when determining the efficiency from simulated samples. With the exception of the HLT trigger efficiency corrections, all the correction weights obtained on the control modes are applied on the signal modes. The PID and tracking efficiency correction steps rely on centrally produced calibration samples and can be directly applied on the signal and control modes. Since the multiplicity and kinematic corrections aim to correct for generator-level effects, it should be possible to apply them to the signal modes, as they should be independent of the decay channel. In a similar vein, the L0 efficiency corrections obtained from the control mode are applied to the signal modes, as in both cases the same TIS decisions are used, which are not expected to depend on the decay. Only the HLT1 and HLT2 trigger efficiency corrections are determined on the signal channel itself.

III.2.1 Particle identification efficiency

The PID variables in simulation are calibrated using high statistics, well-understood calibration samples [85] [87]. For this analysis, the PID variables used in the selection are replaced in the simulation samples by variables sampled from a distribution obtained from the calibration samples. This is done by estimating a four-dimensional distribution of the PID variable, p_T , η and number of tracks in the event using a kernel-density method [88]. The replacement can then be done for each event by looking at the p_T and η of the corresponding particle and the number of tracks in the event and randomly sampling a new value for the PID variable from the distribution estimated by the kernel-density method ('PID re-sampling'). These kernel distributions are provided by a centralised package. The

calibration sample used for electrons consists of $B^+ \rightarrow K^+ J/\psi (\rightarrow e^+ e^-)$ decays, $J/\psi \rightarrow \mu^+ \mu^-$ decays for muons and $D^{*+} \rightarrow D^0 (\rightarrow K^- \pi^+) \pi^+$ decays for kaons and pions.

III.2.2 Track reconstruction

The efficiency of the track reconstruction is known to be different between data and simulation. This has been observed by measuring the efficiency using $J/\psi \rightarrow \mu^+ \mu^-$ decays in a tag-and-probe approach [89]. The probe muon is reconstructed using hits in the TT and muon stations to obtain the efficiency of it being reconstructed as long track with hits in the VELO and T-stations. By measuring this efficiency on data a correction weight to the simulated samples can be calculated by dividing it by the efficiency determined on simulated $J/\psi \rightarrow \mu^+ \mu^-$ decays. This is done by the LHCb collaboration in a centralised manner. The correction weights depend on the p and η of the particle in question and therefore the weights are obtained as function of those variables. While this correction weight is obtained using muons, it is typically also used for hadron tracks. In this case, hadronic interactions, which modify the tracking efficiency, are not taken into account in the correction weight but are considered as a source of systematic uncertainty. This additional systematic uncertainty is expected to cancel in the ratio as the same hadrons appear in the signal mode final states.

Electrons have to be treated as special case [62] since the emission of Bremsstrahlung impacts the reconstruction efficiency. Electrons which emit Bremsstrahlung before the magnet might lose so much momentum that they are 'swept out' of the downstream tracker acceptance by the magnet. This effect might not be properly reflected in simulation and therefore the reconstruction efficiency correction is determined in a dedicated approach. The corrections are binned in ϕ , η and true p_T of the electron, as the probability of Bremsstrahlung emission depends on the amount of material traversed. In some regions of ϕ the electron will cross through the RF-foil multiple times increasing the material budget.

This method requires the electron to be reconstructed as VELO track and thus potential differences between muon and electron VELO reconstruction efficiencies are a source of systematic uncertainty. But the fiducial region of the analysis was chosen such that the VELO reconstruction efficiencies of electrons and muons are very similar and nearly constant, so the associated systematic uncertainty is expected to be small.

The soft pion track from the D^* decay is not assigned a track reconstruction correction as its kinematic distribution is for a large part outside of the coverage

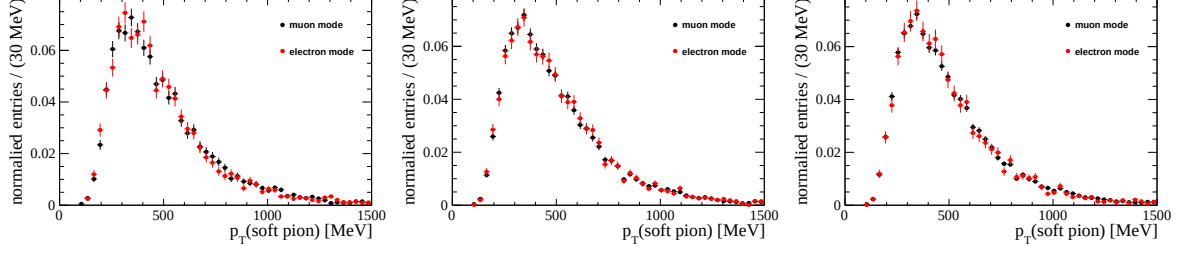


Figure 56: ● Transverse momentum of the soft pion in the muon (black) and electron (red) signal modes for the years 2016 (left), 2017 (middle) and 2018 (right).

of the control sample used to determine these corrections. Since the kinematic distribution of the soft pion agrees well between the muon and electron signal modes, as shown in Figure 56, the effect of omitting this correction is expected to cancel in the ratio.

The tracking efficiency corrections are applied to the control modes samples before calculating the ratio between background-subtracted data and simulated distributions for the L0 efficiency correction step described next.

III.2.3 L0 trigger ▲

Since the selection strategy relies on L0 being triggered independently of the signal, the $L0$ efficiency should not strongly depend on the channel and cancel to first order in the efficiency ratio. Still the simulated samples can be weighted to better reflect the L0 efficiency determined on data.

The control mode decays $B^0 \rightarrow J/\psi K^{*0}$, with $J/\psi \rightarrow \mu^+ \mu^-$ and $J/\psi \rightarrow e^+ e^-$, are used to obtain the L0 efficiency on data and simulation employing the TISTOS method. This method is based on selecting a sub-sample of data with a certain trigger decision, which contains also events selected by the trigger decision of interest. The trigger efficiency is then obtained on this sub-sample by additionally requiring the trigger decision in question. This means that the trigger efficiency is determined on a sub-sample of data with the assumption that this efficiency is representative of the trigger efficiency on the whole sample. Here the sub-sample is selected by requiring a TOS decision and the efficiency of the TIS requirement is determined by

$$\epsilon_{L0}^{TIS} = \frac{N(L0TIS \text{ AND } L0TOS)}{N(L0TOS)}, \quad (73)$$

where $N(L0TOS)$ is the number of events in the sample selected by the TOS decision and $N(L0TIS \text{ AND } L0TOS)$ the number of events when additionally requiring the TIS decision.

The efficiency of triggering an event independent of the decay is to a large part influenced by the kinematics of the other b hadron of the initial $b\bar{b}$ pair. To take this into account the L0 efficiency is determined as function of the p_T of the B^0 hadron in the event, as this is correlated with the other b hadron. The L0 corrections for different data-taking years are obtained individually using the samples in the corresponding year. To obtain the number of signal decays in data the *sPlot* technique is used as discussed in Section III.1.7. The yield of the simulated sample is obtained by taking into account the tracking reconstruction efficiency correction weights. The fit projections to obtain the s-weights are shown in Figure 57 for 2017.

The obtained values of $\epsilon_{L0,MC}^{tis}$ and $\epsilon_{L0,data}^{tis}$ in different $p_T(B^0)$ regions are shown in Figure 120 in Appendix E, and their ratios are shown in Figure 58. As the L0 trigger requirements are TIS on the B^0 decay products, the correction of the L0 efficiency for candidates in a certain B^0 kinematic region should be consistent for different channels. To improve the consistency between the dimuon and dielectron control modes the B^0 momentum is estimated by the DTF with the J/ψ constrained to its known value, which reduces the effect of Bremsstrahlung on the B^0 momentum resolution. To compare the correction factors obtained on the dimuon and dielectron modes, the ratio of both are shown in Figure 59. The correction factors show some discrepancies, especially at small and large p_T of the B^0 . This could be due to statistical fluctuations as the uncertainties on the correction factor are the largest in these regions or due to residual differences between the simulation of electrons and muons. While it is worrisome that for all three years the ratio of correction weights shows a downwards trend when going to lower values of p_T of the B^0 , the electron correction factor does not consistently decrease and the muon correction factor does not consistently increase between the years in this region. This might further indicate that the downwards trend and disagreement is caused by fluctuations instead of a systematic effect. Still, to verify the assumption that the correction factor is independent of the specific decay channel, as is expected for the LOTIS efficiency, the correction factors obtained from the dimuon mode are taken as nominal choice and the dielectron correction factors are used for cross-checking. This choice is motivated by the dimuon mode generally being cleaner. When performing the cross-check by using the dielectron correction factors instead, a negligible difference in the final result is observed.

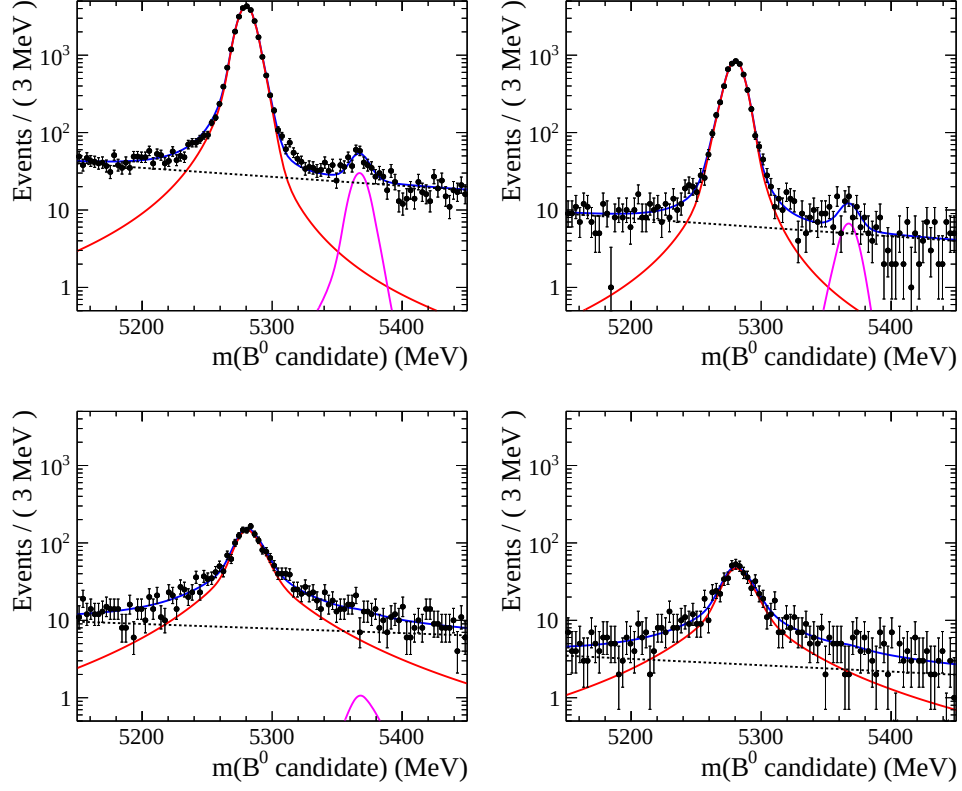


Figure 57: $\blacktriangle$ Projections of the fit to the B^0 invariant-mass spectrum in the (*L0TOS*) (left) and (*L0TIS AND L0TOS*) (right) categories. The top row shows the dimuon and the bottom row the dielectron control modes for 2017.

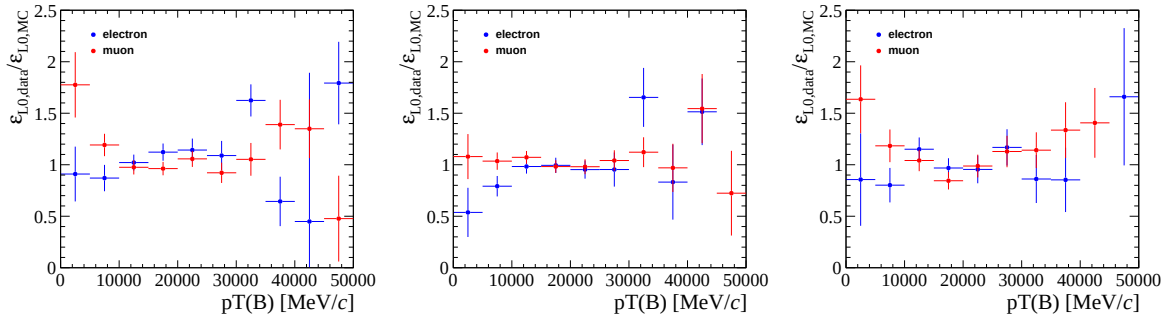


Figure 58: $\blacktriangle$ Ratios between $\epsilon_{L0,data}^{tistos}$ and $\epsilon_{L0,MC}^{tistos}$ as a function of $p_T(B^0)$ for the dimuon (red) and dielectron (blue) control modes for the years 2016 (left), 2017 (middle) and 2018 (right).

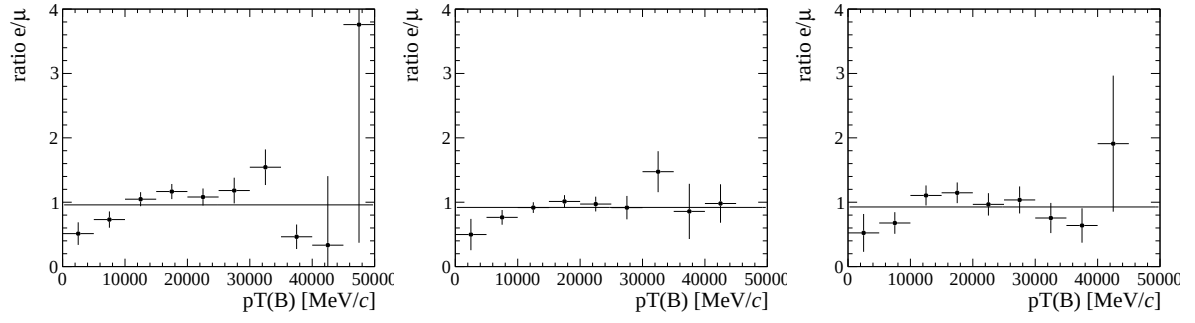


Figure 59: • Ratio of the L0 correction weight obtained on the dielectron control mode over the correction weight obtained on the dimuon control channel as a function of $p_T(B^0)$ for the years 2016 (left), 2017 (middle) and 2018 (right).

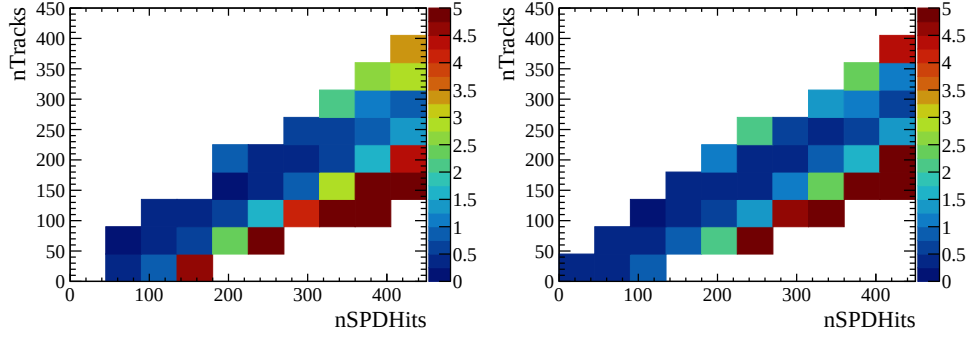


Figure 60: • Multiplicity corrections obtained from the 2017 background-subtracted data and simulated samples of the dielectron (left) and dimuon (right) control mode.

III.2.4 Event multiplicity

The simulation is corrected for differences in quantities related to the event multiplicity. This correction is determined as a function of the total number of reconstructed long tracks ($n\text{Tracks}$) and the number of hits in the SPD ($n\text{SPDHits}$). The correction weights are obtained as the ratio of the binned two-dimensional distribution of $n\text{Tracks}$ and $n\text{SPDHits}$ in data and in simulation. The *sPlot* procedure is used to obtain the background-subtracted distribution in data, while the weights from the previous correction steps are applied to the simulated samples. The chosen binning scheme of the multiplicity correction weights are

- $n\text{Tracks}$: [0, 45, 90, 135, 180, 225, 270, 315, 360, 405, 450]
- $n\text{SPDHits}$: [0, 45, 90, 135, 180, 225, 270, 315, 360, 405, 450]

The multiplicity correction weights obtained with this binning are shown in Figure 60 obtained for the 2017 sample. As the signal yield in the dielectron control channel is smaller than in the dimuon mode, only the weights obtained from the dimuon mode are used. Since the weighting is done in 2D with more bins in total than the previous correction step, the smaller yield of the electron control mode limits its usefulness as correction factors in more bins would have a larger uncertainty. The correction factors obtained on the dimuon mode are then applied on the dimuon and dielectron control modes, as well as the muon and electron signal modes. This is motivated by the event multiplicity being largely independent of which B^0 decay took place.

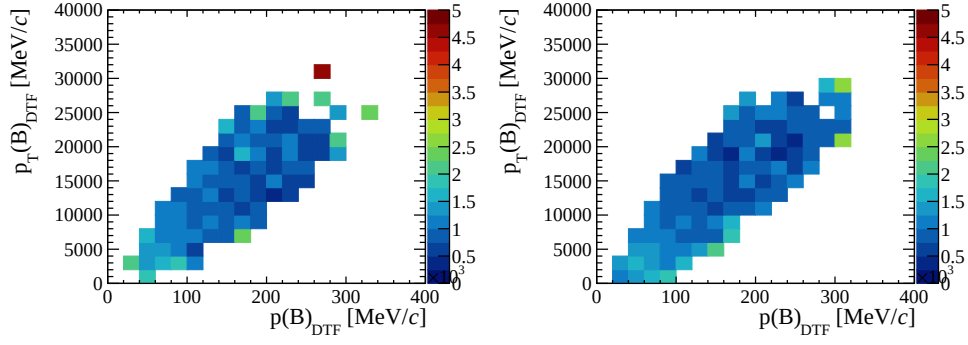


Figure 61: ● Kinematic weights in bins of $B^0 p$ and p_T , obtained from the 2017 background-subtracted data and the simulated signal samples of the dimuon (left) and dielectron (right) control modes. A J/ψ mass constrain is applied when calculating the $B^0 p$ and p_T .

III.2.5 Kinematic corrections

The simulation is known to not describe the kinematic distributions of B^0 -hadrons produced by the pp collision well. The correction weights to account for this are calculated in bins of the fully reconstructed p and p_T of the B^0 meson. The weights are obtained using only the dimuon control mode since it has a larger sample size and the effect corrected for is caused at the generation level of the B^0 and is independent of its decay. The binning scheme splits the p_T spectrum from 0 to 40 GeV/ c into 20 bins and the p spectrum from 0 to 400 GeV/ c into 20 bins, with a constant bin width in both cases.

As in previous steps, the correction weight is obtained as the ratio of the distribution in data and simulation. Again the s-weighted data is used and the tracking, L0 and multiplicity weights are taken into account in case of the simulation. The resulting weights for 2017 are shown in Figure 61.

In the dimuon control channel, the reconstructed B^0 momentum is a good measurement of the 'true' momentum with a small resolution, which is even more precise when using the DTF with the J/ψ and B^0 masses constrained to their known values. Since the correction weights obtained on the dimuon control channel are also used when correcting simulation of the dielectron control channel as well as the electron and muon signal channels, where the B^0 momentum resolution is much worse due to Bremsstrahlung in the former case and the missing neutrino in the latter, the weights are attached using the known values of the B^0 momentum generated in simulation.

III.2.6 HLT corrections for signal modes ▲

For the signal modes, the HLT selection requirements are based on decisions related to the D^0 candidate to reduce any potential bias on the $R(D^{*+})_{light}$ measurement, as it is the shared part between the electron and muon modes. For both the HLT1 and HLT2 selections only decisions which are TOS on the D^0 candidate are considered. As the HLT 2 topological trigger line implicitly depends on the HLT1 decision, an overall correction is obtained to account for differences between the HLT selection efficiency in data and simulation. Similar to the L0 efficiency, the HLT1 selection efficiency can be studied with the TISTOS method

$$\epsilon_{HLT}^{tos} = \frac{N(HLTTIS \text{ AND } HLTTOS)}{N(HLTTIS)}, \quad (74)$$

Here HLTTIS indicates the sample selected by a HLT decision independent of the D^0 candidate, and HLTTOS is the HLT requirement of the signal mode selection listed in Table 11. The correction weight is obtained as the ratio between the HLT efficiency determined on data and simulation as function of the p_T of the D^0 candidate. The binning scheme splits the p_T spectrum from 0 to 30 GeV into 20 bins with a constant bin width. The HLT selection efficiency is obtained on simulation taking into account the previous corrections steps. For data the background-subtracted $p_T(D^0)$ spectrum is determined using the *sPlot* technique. For this purpose, the distribution of the mass difference between the D^0 and D^* candidates is fitted with the sum of two CrystalBall functions with a shared mean to describe the signal and an exponential function to describe the background. The shape parameters of the CrystalBall functions are fixed to the values obtained from a fit to the signal shape found in simulation. The fit results for each period are shown in Figure 62 for 2017.

The HLT selection efficiency determined on data and simulation as function of the p_T of the D^0 candidate is shown in Figure 122 in Appendix E. The ratio between data and simulation is shown in Figure 63. To compare the correction factors obtained from the electron and muon signal modes their ratio is shown in Figure 64. Within the statistical uncertainties they are consistent. To reduce the uncertainty on the final correction factor in a certain $p_T(D^0)$ region it is obtained by averaging the correction factors obtained from the electron and muon signal modes, using $\frac{1}{\sigma^2}$ as the weights, where σ indicates the uncertainty of a certain correction factor.

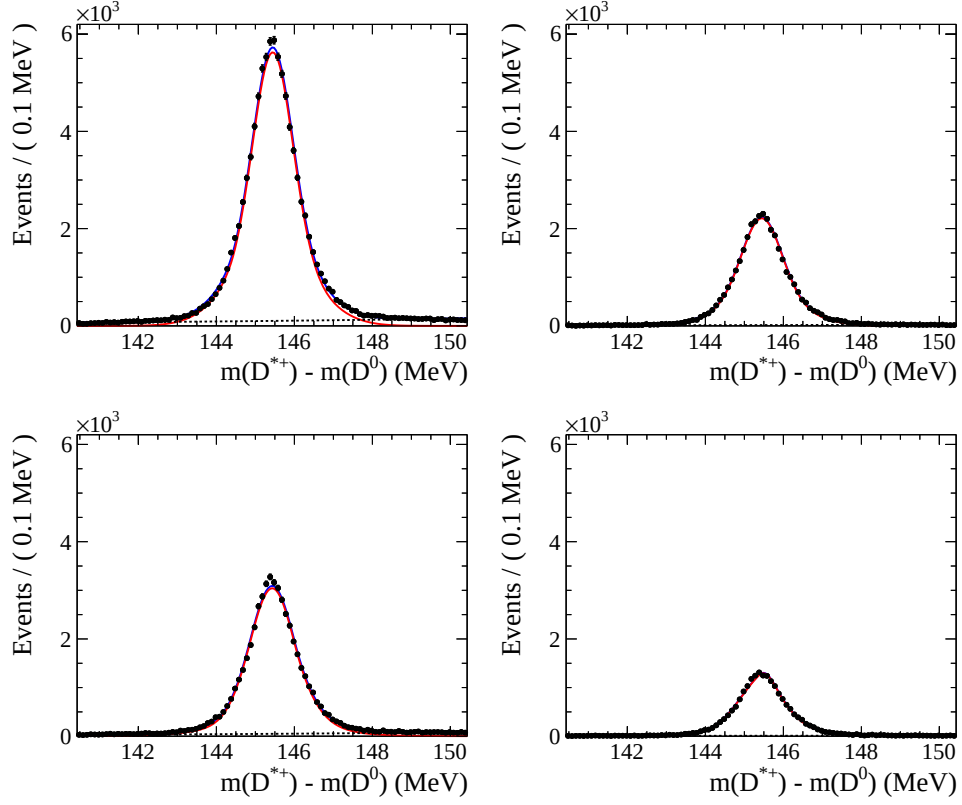


Figure 62: $\blacktriangle$ Projections of the fit to the $m(D^{*+}) - m(D^0)$ invariant-mass difference in the (*HLTTIS*) (left) and (*HLTTIS AND HLTT OSS*) (right) categories. The top row shows the muon and the bottom row the electron signal modes for 2017.

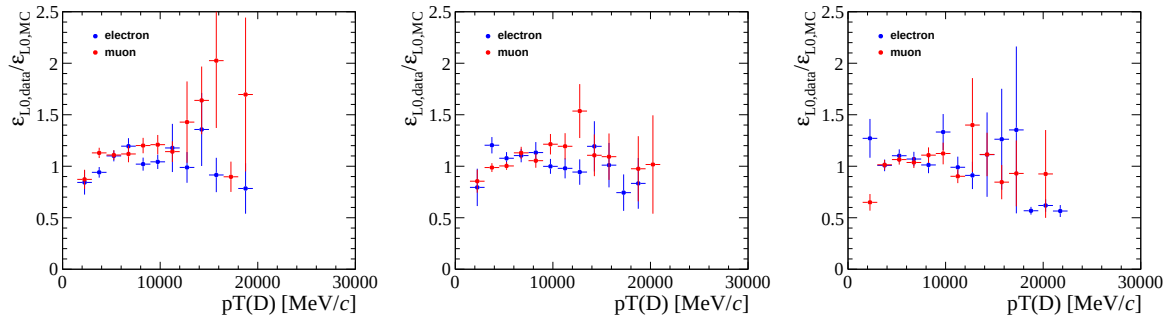


Figure 63: $\blacktriangle$ Ratios between $\epsilon_{HLT,sig,data}^{tistos}$ and $\epsilon_{HLT,sig,MC}^{tistos}$ as a function of $p_T(D^0)$ for the electron (blue) and muon (red) signal modes obtained in the years 2016 (left), 2017 (middle) and 2018 (right).

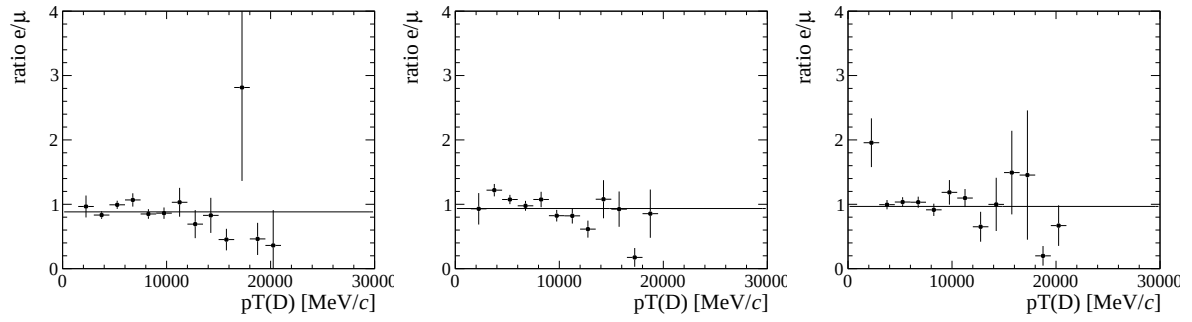


Figure 64: ● Ratios of the HLT correction weight obtained on the electron signal mode over the weight obtained on the muon signal mode of $p_T(D^0)$ for the years 2016 (left), 2017 (middle) and 2018 (right).

III.2.7 HLT corrections for control modes ▲

While it is not required for correcting the signal mode simulated samples, the HLT efficiency corrections are also obtained for the control mode. They are used in the $R(J/\psi)$ cross-check described in Section III.2.9. Similar to the weights obtained on the signal channel, the HLT1 and HLT2 efficiencies are corrected for in the same step. Again the TISTOS method is employed to estimate the HLT selection efficiency on data and simulation as

$$\epsilon_{HLT,con}^{tos} = \frac{N(HLTTIS \text{ AND } HLT\text{TOS})}{N(HLTTIS)}. \quad (75)$$

Here HLTTIS indicates that the positive HLT decision is not related to the B^0 decay, and HLT\text{TOS} is the HLT requirement for control mode listed in Table 14. The correction weight is determined as a function of the p_T of the B^0 meson. The binning scheme splits the p_T spectrum from 0 to 50 GeV into 10 bins with a constant bin width. The HLT selection efficiency is obtained on simulation taking into account the previous corrections steps. For data the background-subtracted $p_T(B^0)$ spectrum is again determined using the *sPlot* technique. The s-weights are obtained from fits to the B^0 mass spectrum, which are shown in Figure 65 for the 2017 data-taking year. The values of $\epsilon_{HLT,con,MC}^{tos}$ and $\epsilon_{HLT,con,data}^{tos}$ in different $p_T(B^0)$ regions are shown in Figure 121 in Appendix E. Their ratio are shown in Figure 66. The obtained correction weights are observed to be close to unity, which indicates that the HLT selection efficiency of the control channel is well modelled by the simulation. Their ratio is shown in Figure 67. The final correction factor in a certain $p_T(B^0)$ region is obtained by averaging the correction factors obtained from the electron and muon control modes, using $\frac{1}{\sigma^2}$ as the weights, where σ indicates the uncertainty of a certain correction factor. Again the averaging serves to reduce the uncertainty on the weights.

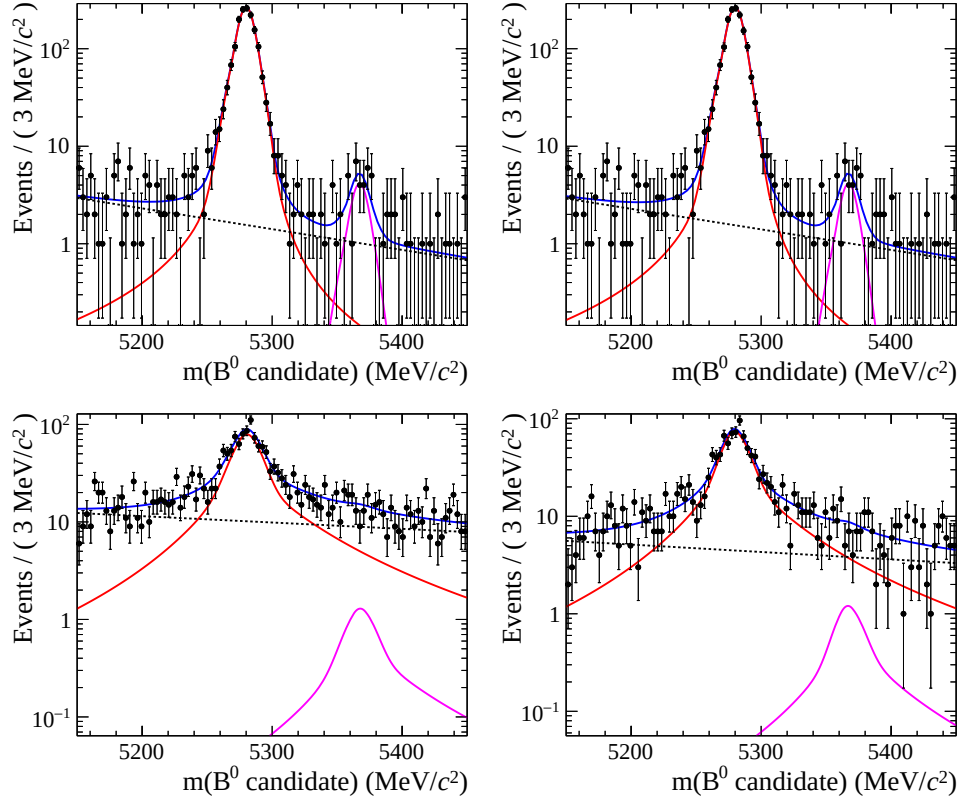


Figure 65: $\blacktriangle$ Projections of the fit to the B^0 invariant-mass spectrum in the (*HLTTIS*) (left) and (*HLTTIS AND HLTTOS*) (right) categories. The top row shows the dimuon and the bottom row the dielectron control mode for 2017.

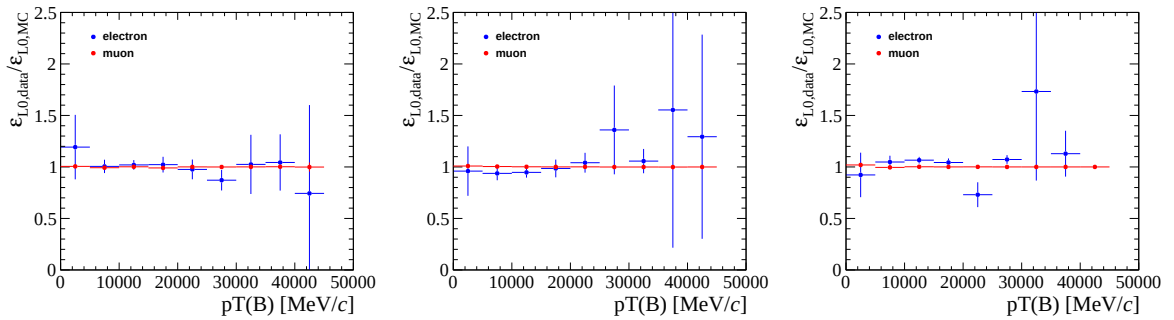


Figure 66: $\blacktriangle$ Ratios between $\epsilon_{HLT,con,data}^{tistos}$ and $\epsilon_{HLT,con,MC}^{tistos}$ obtained on the dielectron (blue) and dimuon (red) control modes as a function of $p_T(B^0)$ for the years 2016 (left), 2017 (middle) and 2018 (right).

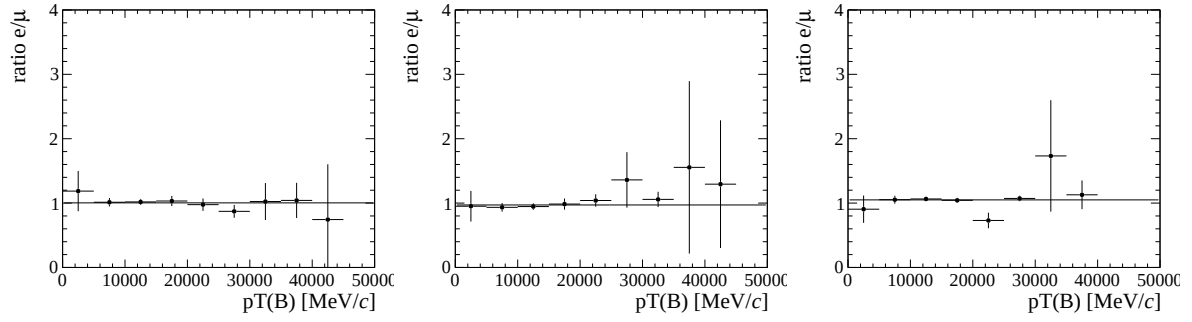


Figure 67: ● Ratios between $\epsilon_{HLT,con,data}^{tistos}$ and $\epsilon_{HLT,con,MC}^{tistos}$ obtained on the dielectron (blue) and dimuon (red) control modes as a function of $p_T(B^0)$ for the years 2016 (left), 2017 (middle) and 2018 (right).

III.2.8 Comparison of data and simulation distributions of control mode decays

To verify that the correction procedure to improve the agreement between data and simulation is yielding in the desired result the simulated control channel samples with and without the application of the weights is compared to s-weighted data distributions for different quantities in Figures 68- 73.

The application of the corrections improves the agreement between background-subtracted data, especially in the variables weighted for, like the momentum of the B^0 and multiplicity. However the agreement is still not perfect and in some bins the weighting procedure worsens the agreement. This indicates that there are other sources causing discrepancies between data and simulation, which are not addressed yet. If and how the remaining disagreement can be expected to be a source of systematic uncertainty on the measurement is studied using 'flatness' checks discussed next and in Section III.4 for the signal modes.

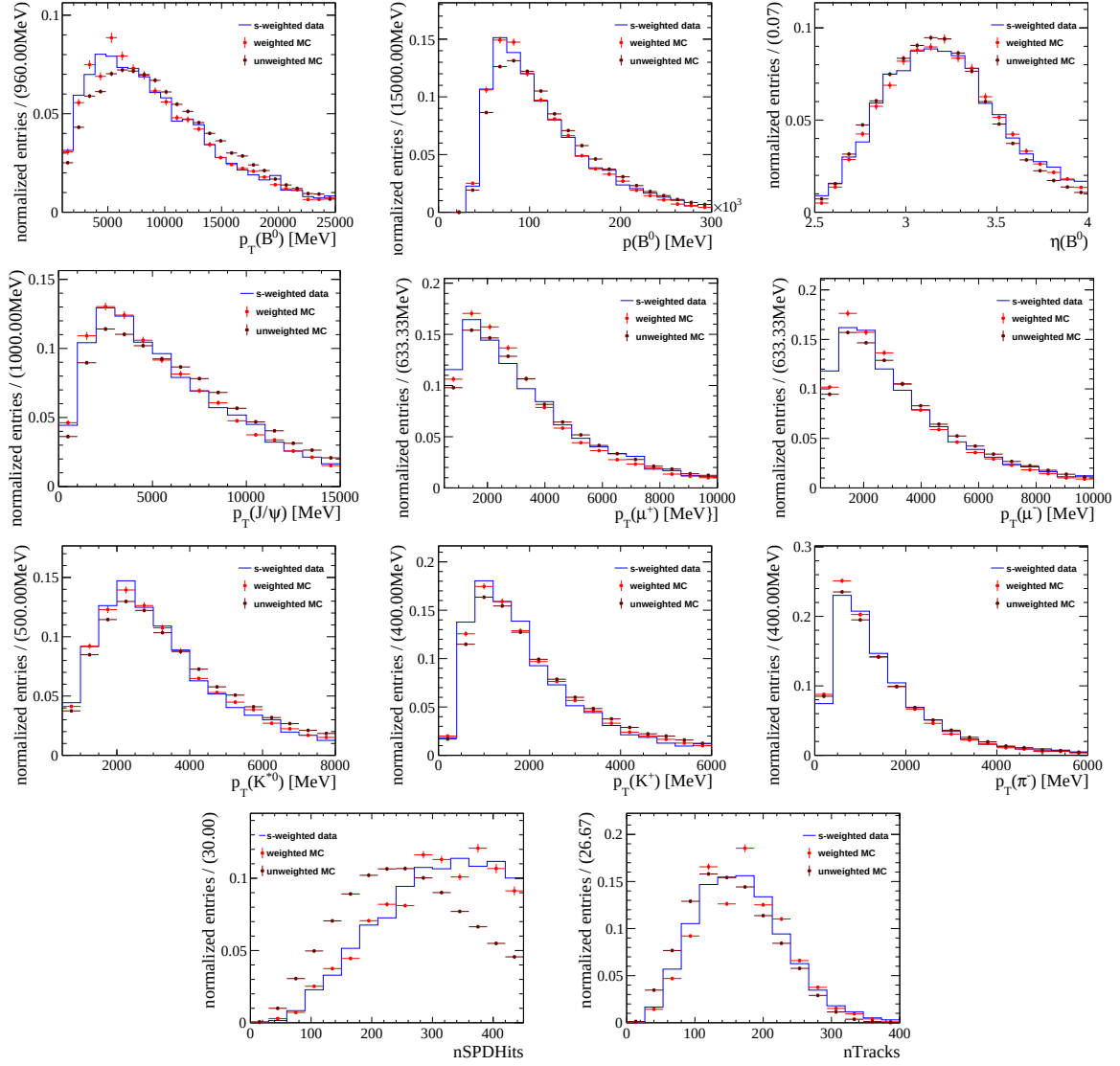


Figure 68: Comparison of background-subtracted data and simulated distributions in the control channel before and after weighting the simulation for the 2016 dimuon control mode sample.

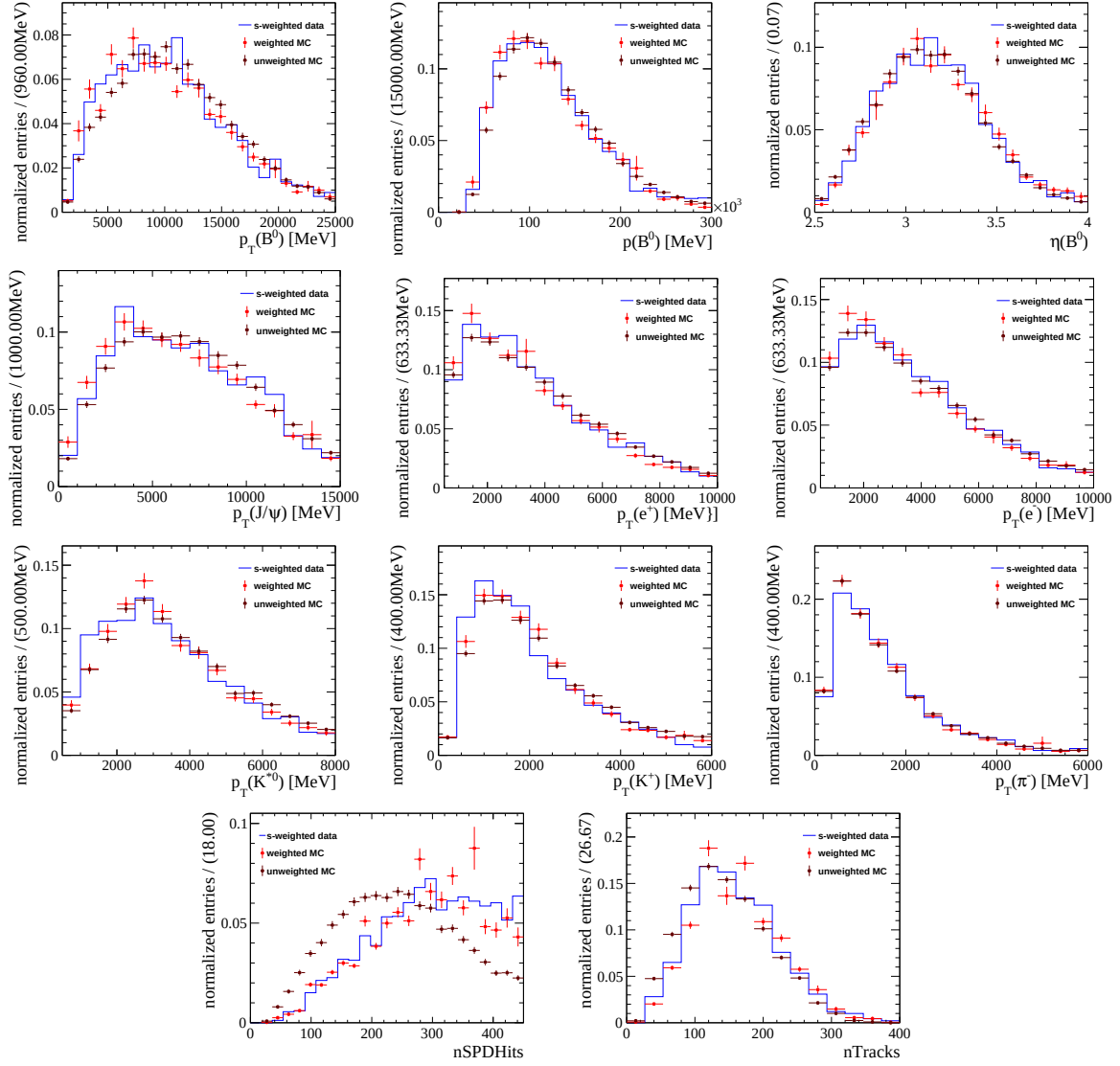


Figure 69: ● Comparison of background-subtracted data and simulated distributions in the control channel before and after weighting the simulation for the 2016 dielectron control mode sample.

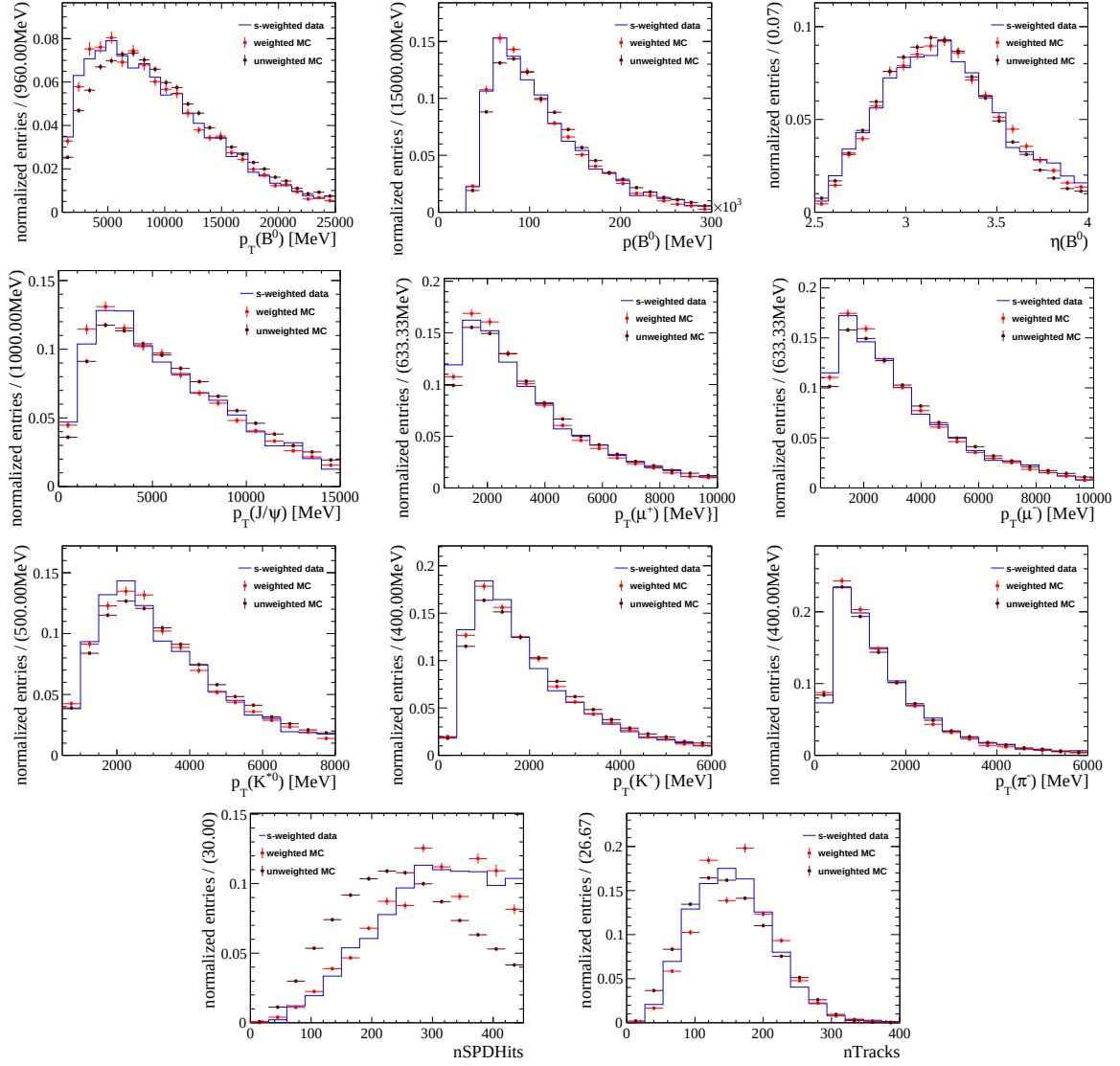


Figure 70: ● Comparison of background-subtracted data and simulated distributions in the control channel before and after weighting the simulation for the 2017 dimuon control mode sample.

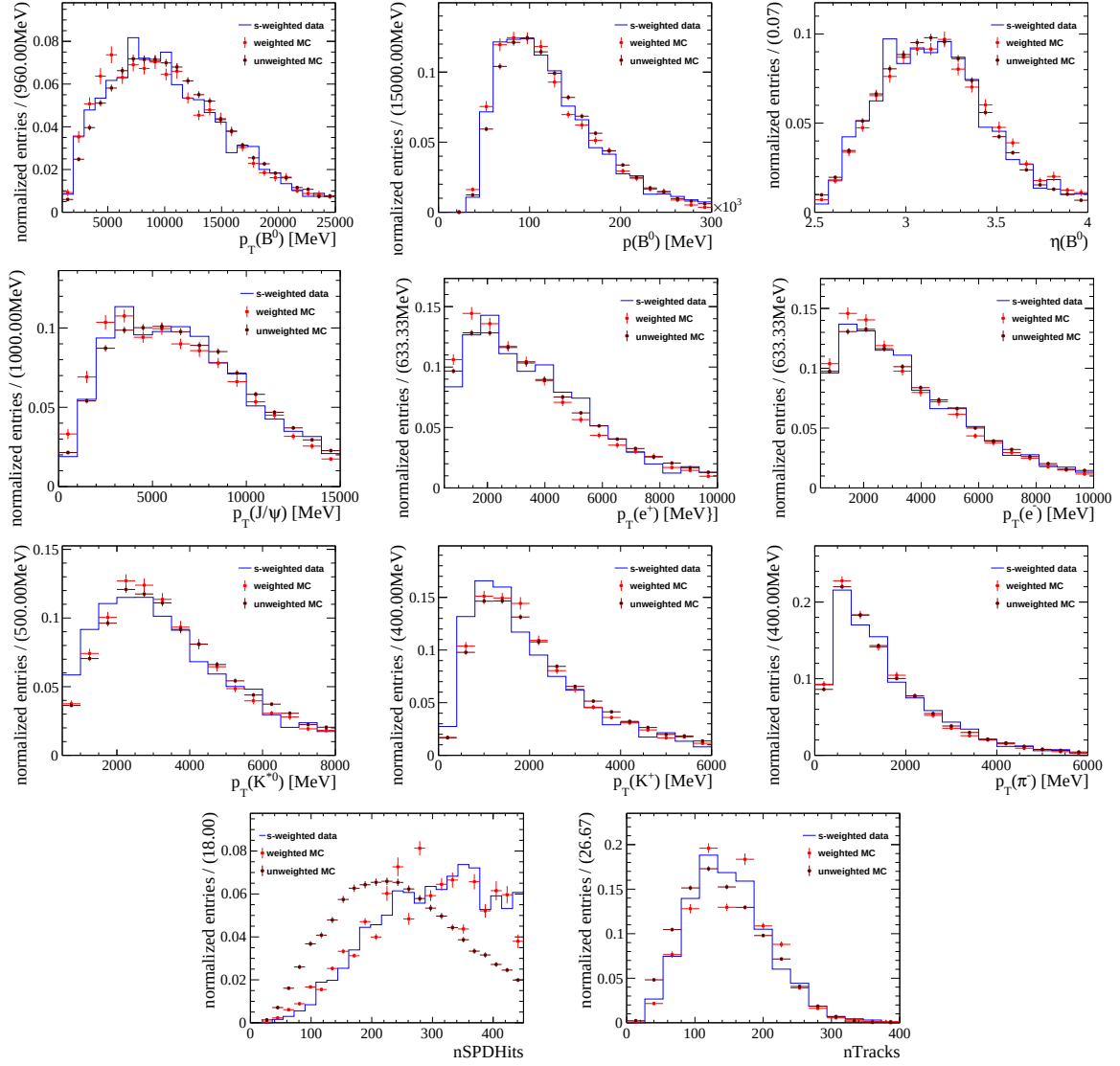


Figure 71: ● Comparison of background-subtracted data and simulated distributions in the control channel before and after weighting the simulation for the 2017 dielectron control mode sample.

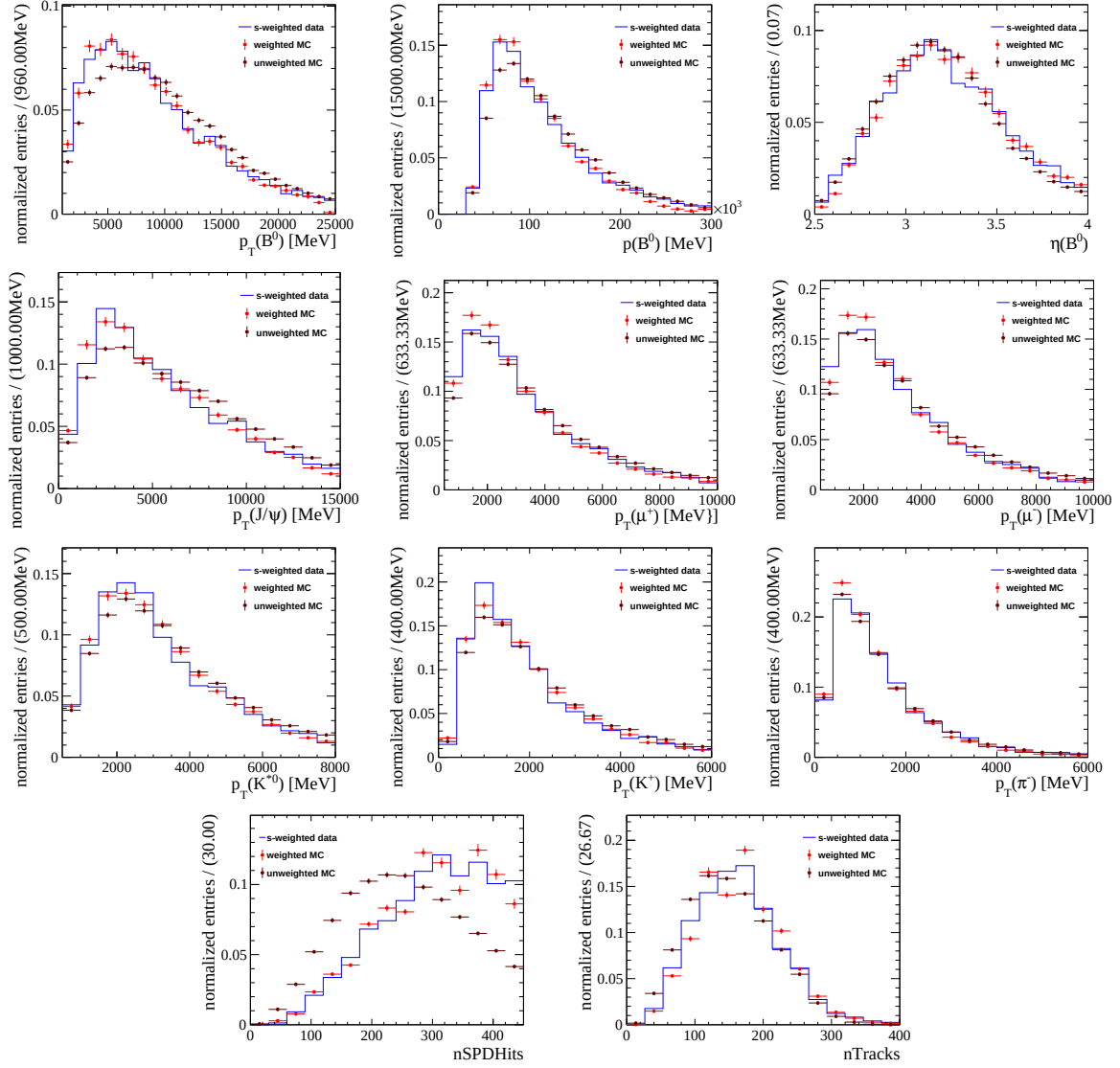


Figure 72: ● Comparison of background-subtracted data and simulated distributions in the control channel before and after weighting the simulation for the 2018 dimuon control mode sample.

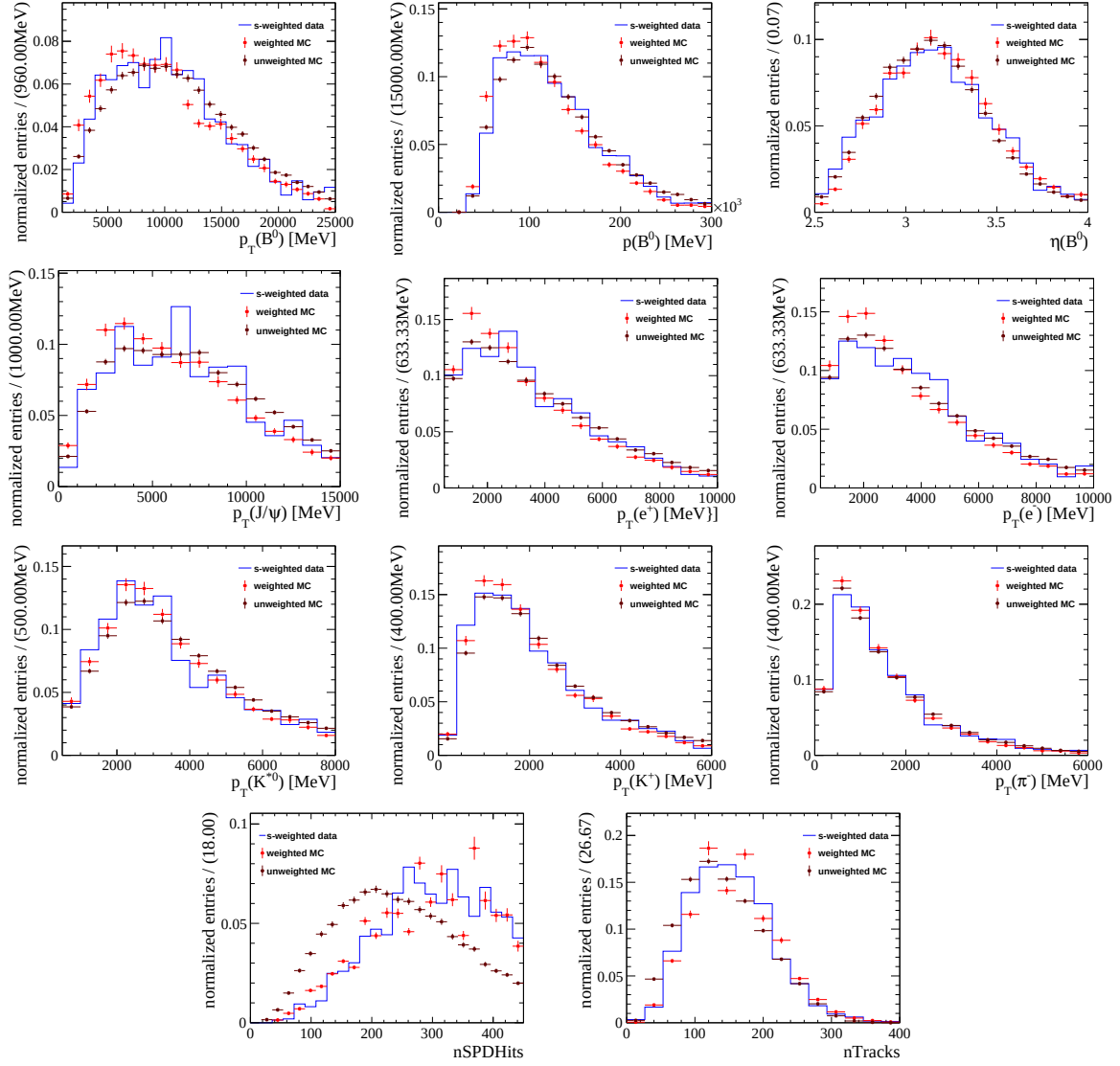


Figure 73: ● Comparison of background-subtracted data and simulated distributions in the control channel before and after weighting the simulation for the 2018 dielectron control mode sample.

III.2.9 $R(J/\psi)$ flatness cross-check

In order to validate the efficiency estimation and correction scheme of the simulated samples, the ratio $R(J/\psi)$, defined as

$$R(J/\psi) = \frac{N(B^0 \rightarrow J/\psi K^{*0}, J/\psi \rightarrow e^+ e^-)}{N(B^0 \rightarrow J/\psi K^{*0}, J/\psi \rightarrow \mu^+ \mu^-)} \times \frac{\epsilon(B^0 \rightarrow J/\psi K^{*0}, J/\psi \rightarrow \mu^+ \mu^-)}{\epsilon(B^0 \rightarrow J/\psi K^{*0}, J/\psi \rightarrow e^+ e^-)}, \quad (76)$$

is plotted as function of various quantities, shown in Figures 74, 75 and 76. Also shown is the result of a fit with a constant straight line to each distribution with the parameter $p0$ and the resulting fit probability. Since this ratio is expected to be unity, a flat distribution consistent with that value is suitable to demonstrate that the differences between the data and simulated distributions are well understood and corrected for in the control modes. The $R(J/\psi)$ are observed to be flat and are consistent with unity within the uncertainties. For the distribution in nSPDHits for 2017, it seems that a trend could be present. But this could also be explained by a fluctuation of the first bin, which has the largest uncertainty. Here the $R(J/\psi)$ distribution as function of the B^0 kinematic and multiplicity variables is studied as those variables are corrected for in the signal modes using the weights obtained from the control modes. Although the data-simulation agreement in the control modes was observed not to be perfect after weighting, it can be seen that this does not impact the $R(J/\psi)$ ratio in the relevant variables. This indicates that these remaining discrepancies will not introduce a large systematic uncertainty to the determination of $R(D^{*+})_{light}$.

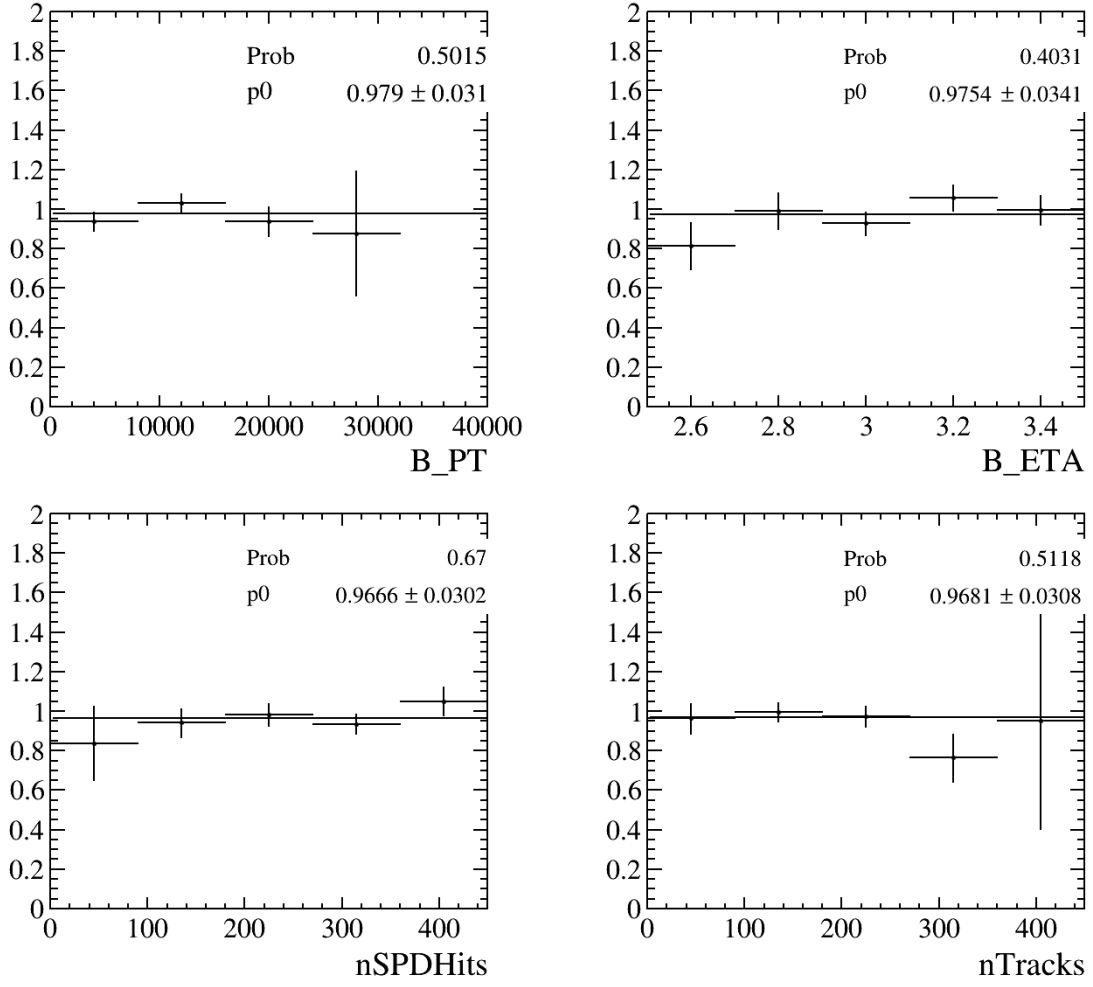


Figure 74: ● The ratio $R(J/\psi)$ as function of the B^0 transverse momentum and pseudorapidity and the number of hits in the SPD and number of tracks in the event for the 2016 sample.

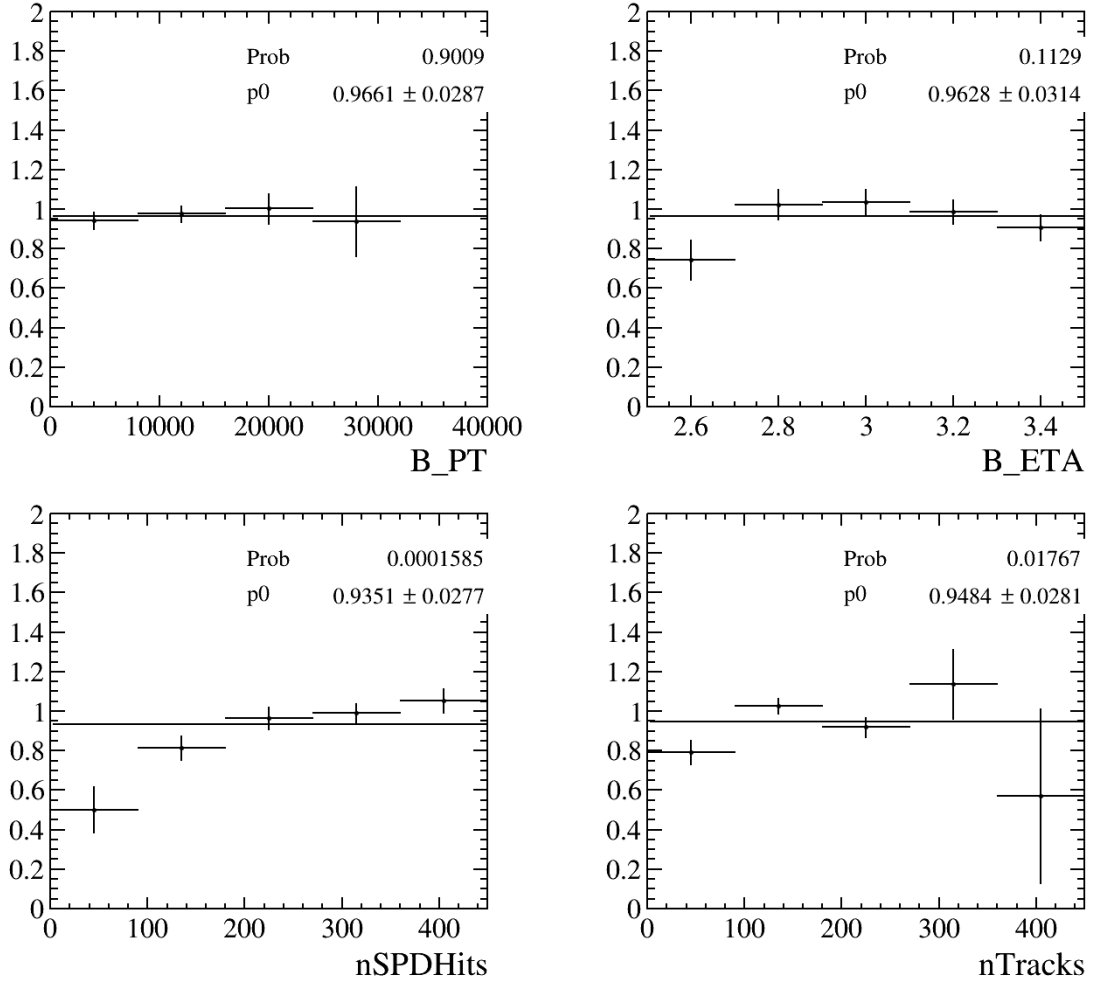


Figure 75: ● The ratio $R(J/\psi)$ as function of the B^0 transverse momentum and pseudorapidity and the number of hits in the SPD and number of tracks in the event for the 2017 sample.

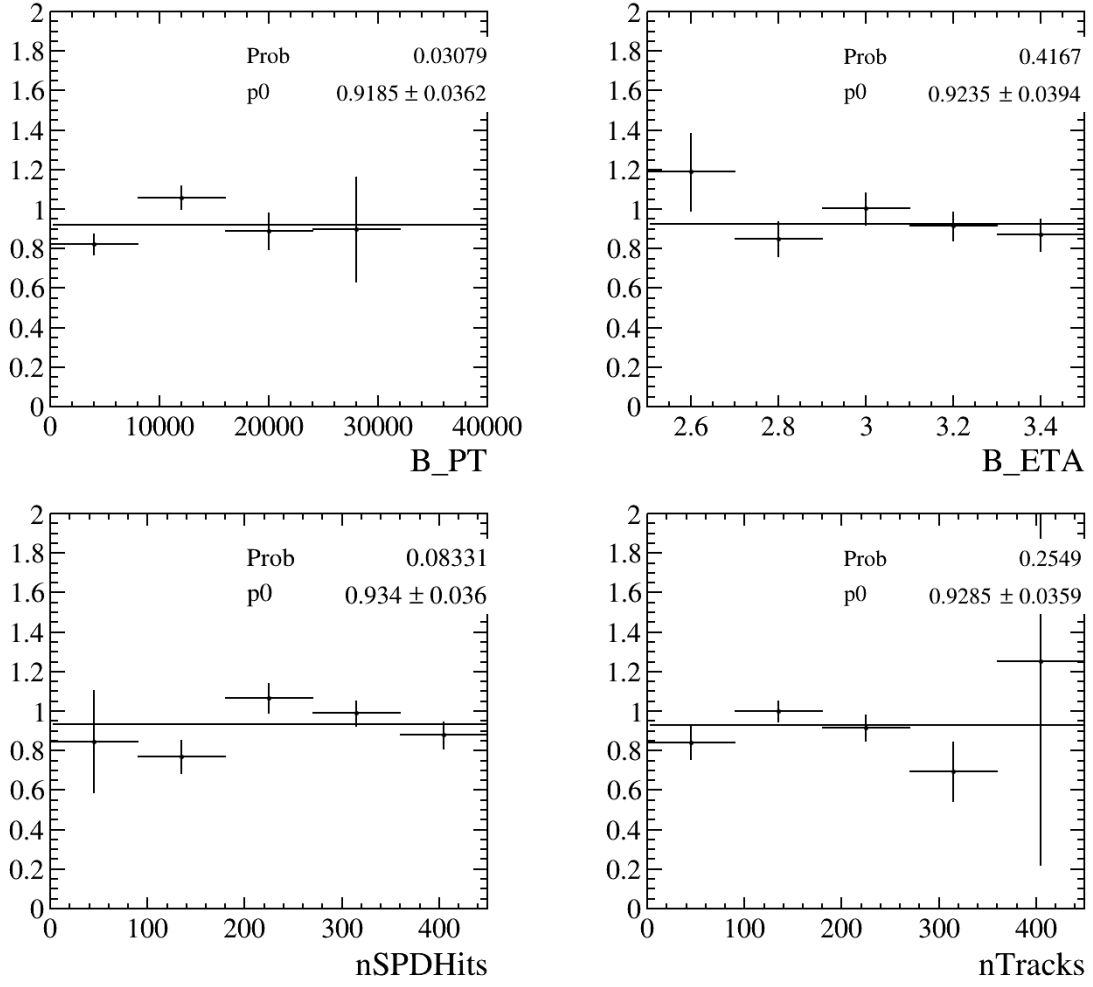


Figure 76: ● The ratio $R(J/\psi)$ as function of the B^0 transverse momentum and pseudorapidity and the number of hits in the SPD and number of tracks in the event for the 2018 sample.

III.2.10 Form factor weighting

For the semileptonic decays, a realistic description of the form factors in the simulated samples used to determine the selection efficiency is necessary since they can modify the decay kinematics of the process and so the efficiency of selecting that decay. The approximated q^2 distribution is one of the variables used to separate signal and background decays. The distribution of q^2 in simulated samples depends on the form factor model so that a reasonable form factor model is required to describe the data well. It is possible to use different form factor models when describing the data and to look for new physics by checking agreement of data with simulated samples using form factors models including BSM contributions.

Simulated samples generated with a specific form factor model and values for the form factors can be weighted to match a different form factor parametrisation. This form factor weighting is achieved using the HAMMER tool [78]. The signal modes, as well as the $B \rightarrow D^* \tau$ decay modes, are weighted to match the CLN and BGL parametrisations with the measured values from (43) and (46) described in Section I.3.1. The CLN parametrisation is used to extract the baseline result and the BGL parametrisation as cross-check to estimate the impact of using a particular form factor model, described in more detail in Section III.5.4.

To illustrate the effect of the form-factor weighting the distributions of the fit variables are shown in Figures 77 and 78 with the initial phase space distributions and the distributions weighted to match the CLN and BGL parametrisations. It is observed that the CLN and BGL parametrisations result in similar distributions. The E_l^* and q^2 distributions are seen to depend on the chosen form factor model the most, where large differences between using a phase space decay model and the CLN or BGL parametrisations are present.

The simulated higher excited D^{**} and double-charm background samples would also require a form factor weighting to get the most realistic description of the data. Especially for the double-charm decays, this is more involved due to the two hadron final state. At the present stage, no form factor weighting is performed for these samples.

III.2.11 Comparison of data and simulated distributions of signal mode decays

Using the respective yields of the signal and background contribution to the signal channels obtained from the template fit described below in Section III.3 and the respective shapes obtained from corrected simulation, it is possible to compare distributions in data and simulation for the signal modes. In practice, this means

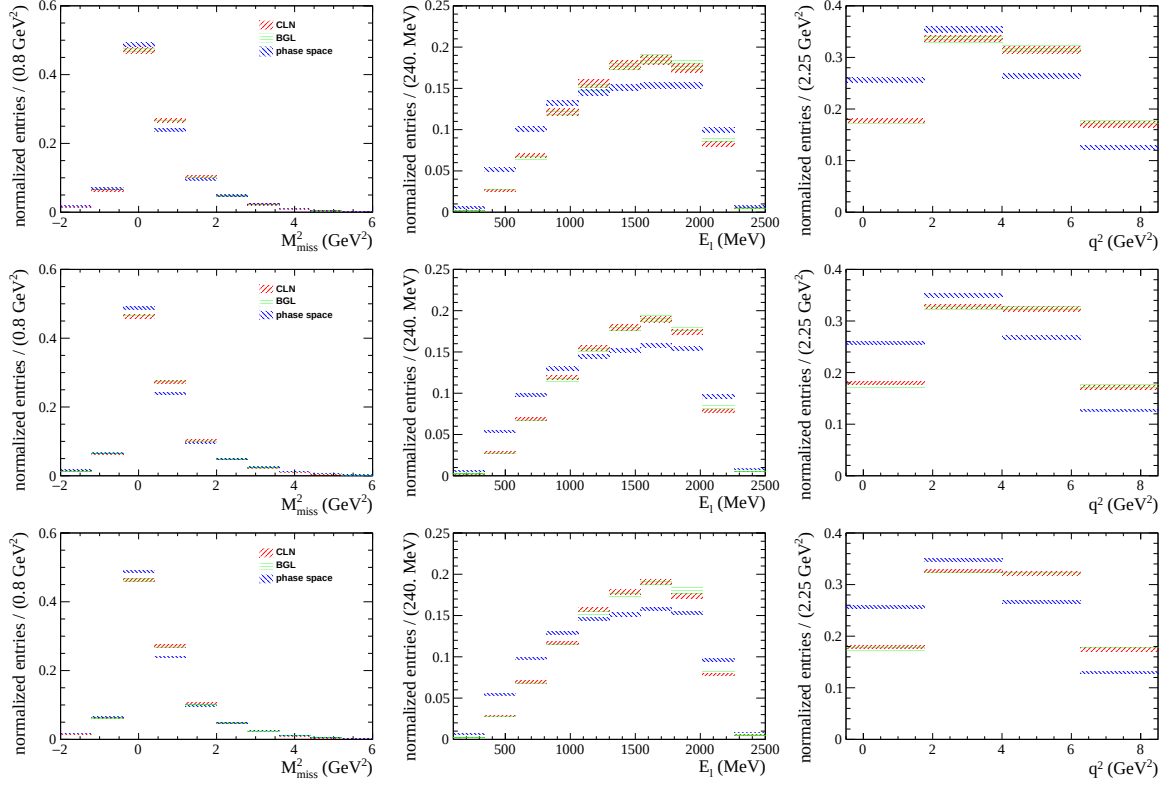


Figure 77: ● Fit variables in the simulated signal electron samples generated with a phase space model (blue) and weighted to the CLN (red) and BGL parametrisations for the $\bar{B}^0$ decay for the 2016 (top), 2017 (middle) and 2018 (bottom) periods.

creating the histograms of all samples contributing to the fit in the variable to be compared and adding them according to their respective yields as determined by the fit. The combined histogram can then be compared to the data. The comparison then shows the distribution of the signal candidates in data including background contributions with the distribution expected from the fitted yields. These comparisons are shown in Figure 79.

The weighting procedure improves the agreement in the multiplicity variables but worsens it in case of the $\bar{B}^0 p_T$. At the current stage of the analysis, this is something still under study. As the muon and electron modes are affected in the same way the effect of this increased discrepancy can be expected to cancel in the ratio. This is checked in Section III.4.2.

III.2.12 Summary

To correct the signal mode simulation samples, correction factors, addressing discrepancies due to the generated B^0 kinematics, the generated event multiplicity and the L0 trigger efficiency, are taken from the control modes. As the HLT selection differs between the signal and control modes the problem could arise that a

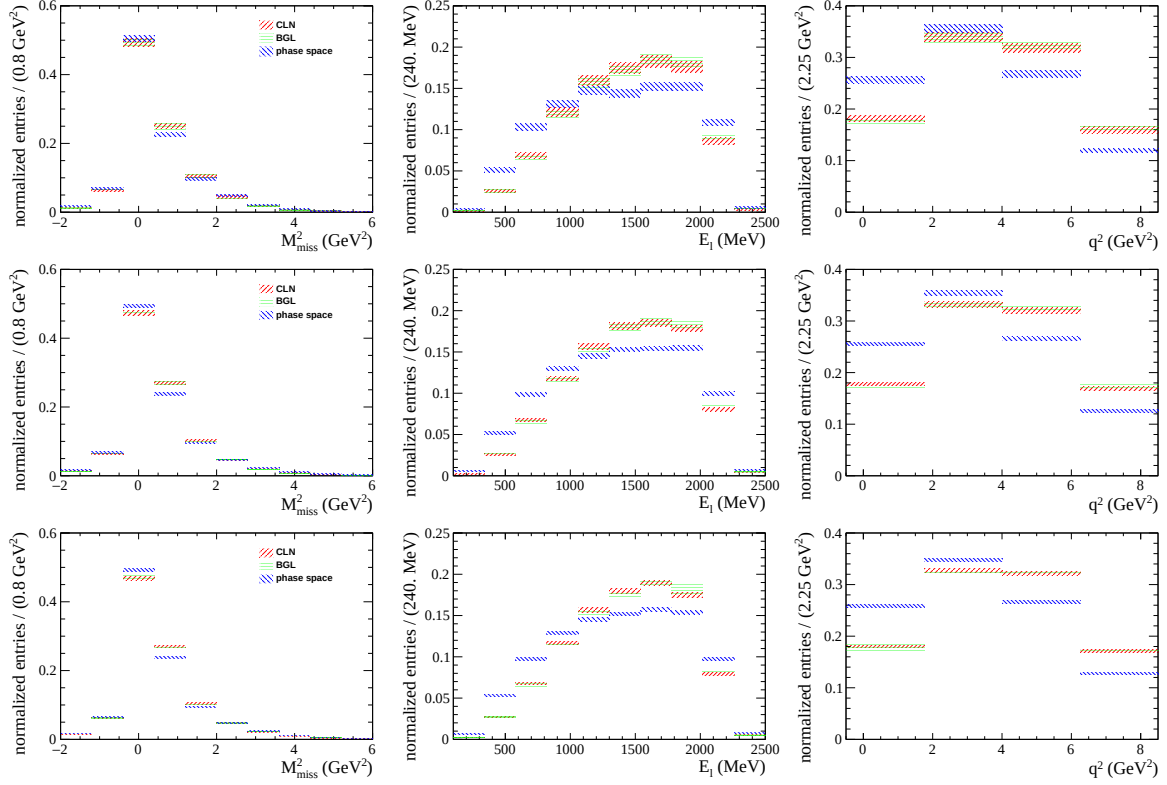


Figure 78: ● Fit variables in the simulated signal muon samples generated with a phase space model (blue) and weighted to the CLN parametrisation (red) for the $\bar{B}^0$ decay for the 2016 (top), 2017 (middle) and 2018 (bottom) periods.

mismodelling of the HLT efficiency in the control mode simulation would distort the other correction weights and make the application of the weights obtained from the control channel on the signal modes less justified. However it is seen that the HLT efficiency is well modelled in the control mode simulation and the associated correction weight could be expressed by a constant factor close to unity.

The HLT efficiency correction factor for the signal modes is obtained from the signal mode itself. The HLT selection for the signal modes was chosen to only use the D^0 decay products so that the HLT efficiency and its correction factor is independent of the lepton in the decay. It is seen that the HLT correction factors obtained on the electron and muon mode reasonably agree for a large region of the $D^0 p_T$. Therefore it can be expected that uncertainties due to the HLT efficiency correction will cancel in the ratio.

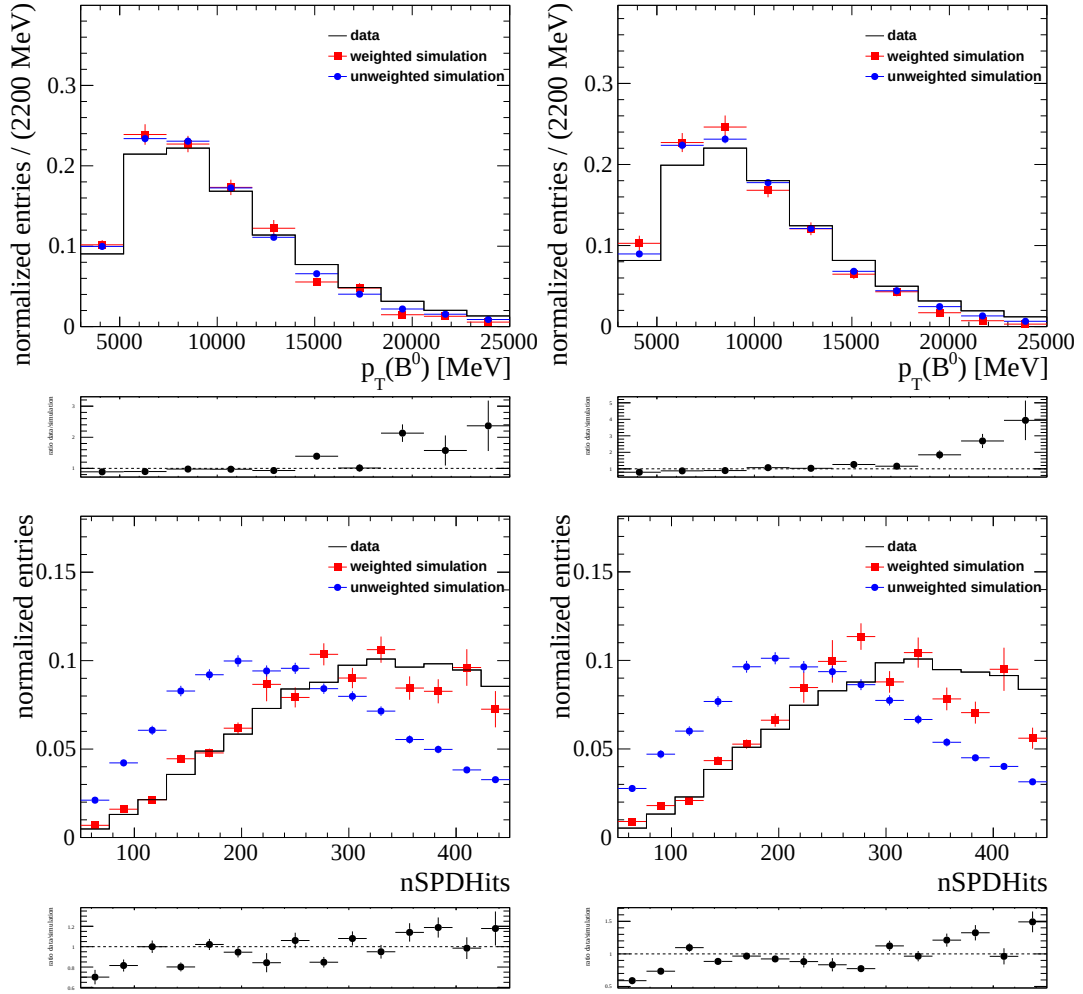


Figure 79: ● Comparisons of the data and expected distributions for the (left) muon and (right) electron signal modes. The expected distributions is obtained by adding the different components included in the template fit according to their fitted yields.

III.3 Extraction of the relative signal yields

Central to the analysis is the extraction of the relative yield R_{raw} of the electron and muon signal channels. To obtain the final relative branching fraction, R_{raw} then needs to be multiplied by the relative selection efficiencies. As alluded to previously, the missing neutrino and the resulting partial reconstruction of the $\bar{B}^0$ hadron makes the extraction of the signal yield more difficult than performing a 'simple' fit to the $\bar{B}^0$ invariant-mass distribution. Instead a template fit is performed where the probability density distributions are approximated using histograms. The distribution observed in data is described by considering the signal and several background template shapes. The shapes of the signal and most of the background decays are obtained from simulation taking into account the correction weights. The remaining background sources due to random combinations and misidentification of particles are described by control samples obtained in a data-driven way, as described earlier in Section III.1.8.

III.3.1 Description of the template fit

A three-dimensional template fit in q^2 , $m_{missing}^2$, E_{lepton}^* is performed using the data samples of each year of the signal modes to separate the signal components from the background contributions. The fit is performed simultaneously on the electron and muon samples, sharing the ratio R_{raw} . The choice of binning for the templates is summarised in Table 17 with 400 bins in total. The choice of the discriminating variables is motivated by a previous measurement of $R(D^*)$ by LHCb [79]. Compared to that analysis, the upper limit of the binning has been lowered in M_{miss}^2 and q^2 . This is motivated by the trigger selection chosen here, which greatly reduces the number of candidates in that region, which is expected to mostly contain events which are considered background for this analysis, demonstrated in Figure 80 showing the result of the template fit to the $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu$ sample collected during Run 1 at LHCb.

Table 17: Binning used for the templates.

	lower border	upper border	number of bins
$m_{missing}^2$	-2 GeV^2	6 GeV^2	10
E_{lepton}^*	100 MeV	2500 MeV	10
q^2	-0.5 GeV^2	8.5 GeV^2	4

The template fit uses the **HistFactory** [90] package. A simplified ('lite') version of the Beeston-Barlow method [91] is employed, allowing the sum of the tem-

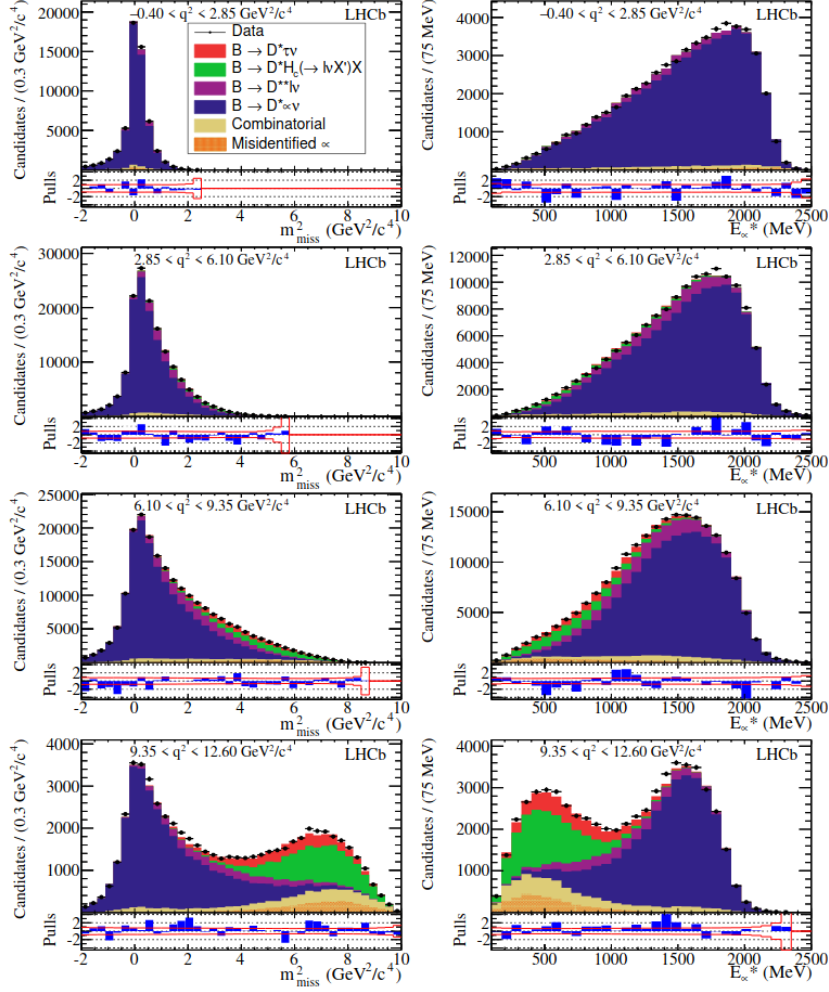


Figure 80: ■ Projections of the three-dimensional template fit of the Run 1 $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu$ sample [79].

plates obtained from simulated samples to vary bin-by-bin by the corresponding statistical uncertainties due to finite sample sizes. This 'lite' version makes the simplification of not varying each simulated template shape separately, which would introduce one parameter for each template per bin, but instead calculating the overall uncertainty in each bin as the weighted average of the simulated templates at the beginning and varying each bin by this value during the fit. This variation is implemented by introducing the parameters γ_i for each bin i , which are included in the likelihood as Poisson constraint terms. With this method, the uncertainty due to limited simulated sample size can be considered in the fit and its impact on the sensitivity can be estimated.

Since the uncertainty estimation of the Beeston-Barlow method is only accurate when the initial amount of each simulated component is close to the final result, the fit is performed in two steps. As a first step, a fit without the Beeston-Barlow method is performed to get a first estimation of the yields. In the second

Table 18: Parameters used in the template fit. A red colouring signifies that the parameter is constrained to a certain value.

muon	electron
N_{sig}^μ	-
R_{raw}	
$N(B^0 \rightarrow D^* D(\rightarrow \mu^+ \nu_\mu) X)$	$N(B^0 \rightarrow D^* D(\rightarrow e^+ \nu_e) X)$
$N(B^+ \rightarrow D^* D(\rightarrow \mu^+ \nu_\mu) X)$	$N(B^+ \rightarrow D^* D(\rightarrow e^+ \nu_e) X)$
$N(B \rightarrow D^{**} \mu \nu)$	$N(B \rightarrow D^{**} e \nu)$
α_{B^+/B^0}	
$\alpha_{BR(B^0 \rightarrow D_1^-)}$	
$\alpha_{BR(B^0 \rightarrow D_1'^-)}$	
$\alpha_{BR(B^0 \rightarrow D_2^{*-})}$	
$\alpha_{BR(B^+ \rightarrow D_1^0)}$	
$\alpha_{BR(B^+ \rightarrow D_1'^0)}$	
$\alpha_{BR(B^+ \rightarrow D_2^{*0})}$	
$N^\mu(B^0 \rightarrow D^{***})$	$N^e(B^0 \rightarrow D^{***})$
$N^\mu(B^+ \rightarrow D^{***})$	$N^e(B^+ \rightarrow D^{***})$
N_{misID}^μ	N_{misID}^e
$N_{comb,WS}^\mu$	$N_{comb,WS}^e$
$\alpha_{comb,WS}^\mu$	$\alpha_{comb,WS}^e$
N_{comb,D^*}^μ	N_{comb,D^*}^e
α_{comb,D^*}^μ	α_{comb,D^*}^e
$N^\mu(B^0 \rightarrow D^{*-} \tau^+ \nu_\tau)$	$N^e(B^0 \rightarrow D^{*-} \tau^+ \nu_\tau)$
$N^\mu(B^0 \rightarrow D^* D_s^+ (\rightarrow \tau^+ \nu_\tau) X)$	$N^e(B^0 \rightarrow D^* D_s^+ (\rightarrow \tau^+ \nu_\tau) X)$
$N^\mu(B^+ \rightarrow D^* D_s^+ (\rightarrow \tau^+ \nu_\tau) X)$	$N^e(B^+ \rightarrow D^* D_s^+ (\rightarrow \tau^+ \nu_\tau) X)$
$N(B_s^0 \rightarrow D_s^{**} \mu \nu_\mu)$	$N(B_s^0 \rightarrow D_s^{**} e \nu_e)$

step, these yields are taken as the starting point and the Beeston-Barlow method is enabled.

The parameters floating in the fit are summarised in Table 18. By expressing the electron signal yield as a function of the muon signal yield and the raw ratio, a simultaneous fit is performed. The α parameters correspond to Gaussian constraints imposed on various yield parameters.

The fitting procedure relies on the ability of the fitter to distinguish between the different contributions based on their template shapes. To illustrate this in the following the different background contributions considered in the fit are discussed and one-dimensional projections of the fit variable of the signal modes are compared with the various background shapes. The figures for all years can be found in Appendix F.

III.3.2 Templates from simulation

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

The $B^0 \rightarrow D^{*-}\tau^+\nu_\tau$ decay, where the tau decays to a muon or electron and a pair of neutrinos, is treated as background in this analysis. The form-factor model of the B^0 decay is weighted to the CLN parametrisation. A comparison of this background shape to the signal template is shown for both signal modes in Figures 123, 136 and 149 in Appendix F for 2016, 2017 and 2018, respectively.

Templates for $B \rightarrow D^{**}\ell^+\nu_\ell$

The contribution from $B \rightarrow D^{**}\ell\nu$ decays is described using the cocktail samples summarised in III.1.3. The different decays of both the B^0 and B^+ samples are added to the fit model with a shared total yield and their relative branching fractions, which are constrained to the known values [84].

Since for the $B^0 \rightarrow D^{**+}\ell^-\nu$ decay, the D^{**+} have only been measured using decays to $\bar{D}^{*0}\pi^-$, a factor of 0.5 is added, derived from the isospin symmetry when considering the decay to $\bar{D}^{*+}\pi^0$. An additional factor is added to account for the relative efficiency between the B^0 and B^+ modes, mostly caused by the isolation requirement having different rejection power for the two samples. Finally, a relative factor to account for differences between the abundance of B^0 and B^+ mesons is introduced constrained to the nominal value of 1 with an uncertainty of 15%. The nominal template shapes are shown for the electron mode in Figures 124, 137, 150 and for the muon mode in Figures 138 in Appendix F, separately for each year.

Templates for higher D^{**} resonances, which are less understood, are included in the fit with the templates shown in Figures 139 and 140. As their relative branching fractions are not well known, the composition of the different decay chains in the cocktail sample is kept to the fractions initially generated.

Templates for $B \rightarrow D^*DX$

The double-charm background, where the the additional D-meson decays to final states involving a muon or an electron, are included with the template shown for the electron mode in Figures 128, 141, 154 and for the muon mode in Figures 129, 142, 155 in Appendix F, separately for each year.

Also included are templates, where the D meson decays to a τ first, shown for the electron mode in Figures 130, 143, 156 and for the muon mode in Figures 131, 144, 157 in Appendix F, separately for each year.

Templates for $B_s^0 \rightarrow D_s^{**} \ell \nu_\ell$

The templates for backgrounds from semileptonic B_s^0 decays including D_s^{**} states, which can decay to D^{*+} , are shown for the electron mode in Figures 132, 145, 158 and for the muon mode in Figures 133, 146, 159 in Appendix F, separately for each year.

III.3.3 Data-driven templates

The combinatorial background shapes obtained from data, as described earlier, are shown for the electron mode in Figures 134, 147, 160 and for the muon mode in Figures 135, 148, 161 in Appendix F, separately for each year. The yields of the 'fake' D^{*} 'real' lepton and 'real' D^{*} 'real' lepton combinations are constrained in the fit to their respective expected value with an uncertainty of 15%. The fake lepton contribution is left freely floating due to an ongoing discussion within the collaboration about the limitations of the *PIDCALIB* approach, which could lead to a biased estimate of the yield of this contribution.

III.3.4 Fit results

One-dimensional projections in m_{miss}^2 , E_l^* and q^2 of the fit results are shown in Figures 81, 82 and 83 for the muon and electron signal modes, separately for the 2016, 2017 and 2018 data samples. The template shapes describe the data well in most of the kinematic regions and the pulls are small. At the edges of the m_{miss}^2 and E_l^* the description of the data gets worse, as seen from larger pulls. Especially the first bin in m_{miss}^2 is not well modelled. A possible explanation is that the combinatorial background templates have low statistics in this bin and so their contribution is not well accounted for. A possible improvement could be to allow the combinatorial background shapes also to vary according to their uncertainties with the Beeston-Barlow method as the simulated samples do. As the yield in those edge bins is much smaller compared to the overall yield this mismodelling should have a negligible effect on the signal mode yields and their ratio.

Projections in m_{miss}^2 and E_l^* in the four q^2 bins are shown in Appendix G in Figures 162, 164 and 166 for the muon final state, and in Figures 163, 165 and 167, separately for the 2016, 2017 and 2018 data samples.

To give an impression of how the templates are varied by the Beeston-Barlow method, the same projections, but with the shape variation set to unity are shown in Appendix H. It should be noted that the pulls shown in that case do not take into account the uncertainty on the simulated template shapes. From this, it can

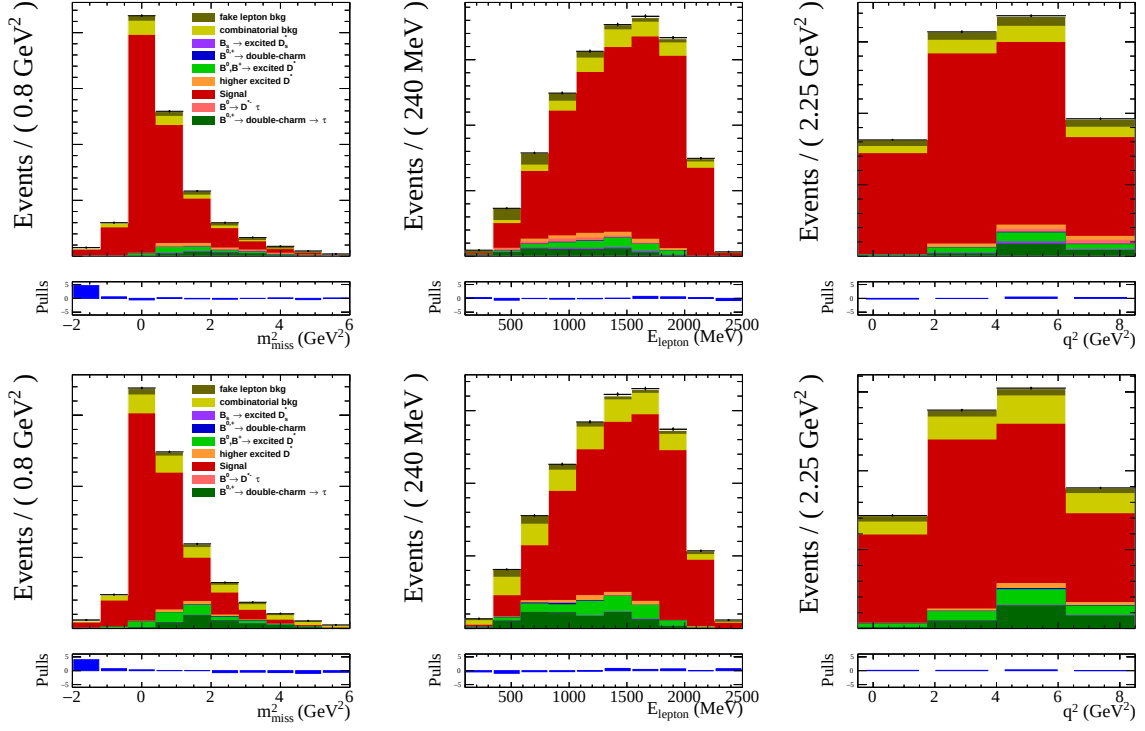


Figure 81: • One-dimensional projections in m_{miss}^2 (left), E_{μ}^* (middle) and q^2 (right) of the fit result of the model to data for the muon (top) and electron (bottom) signal modes for 2016.

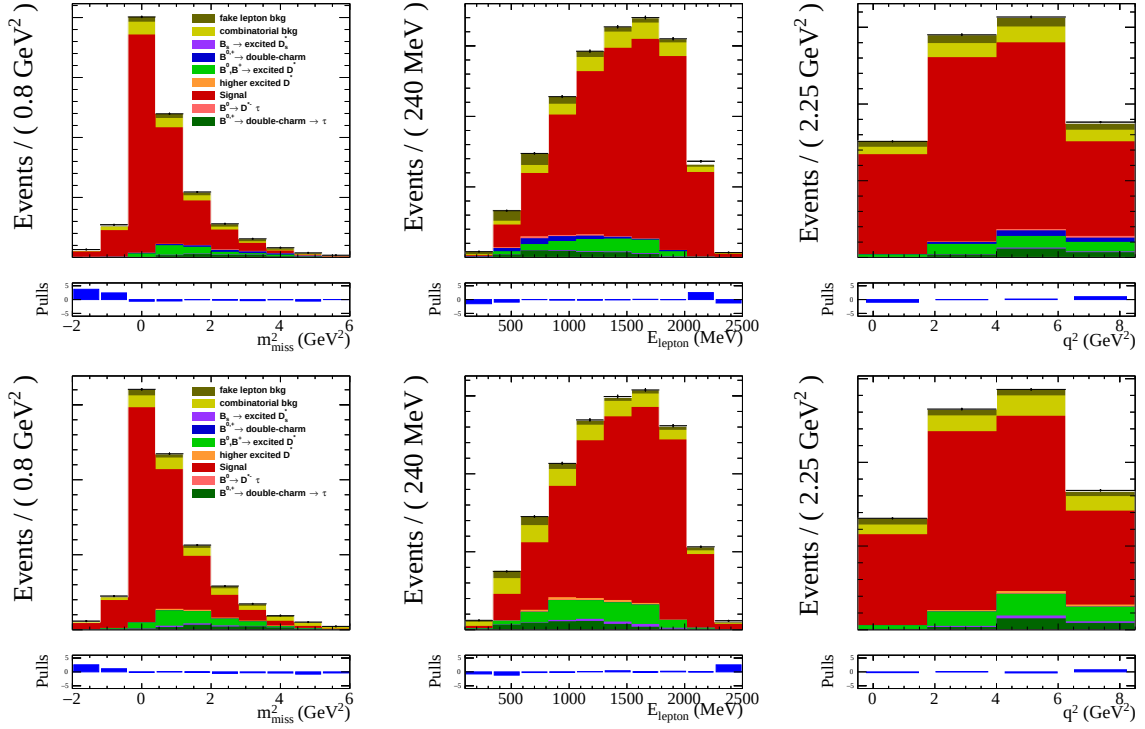


Figure 82: • One-dimensional projections in m_{miss}^2 (left), E_{μ}^* (middle) and q^2 (right) of the fit result of the model to data for the muon (top) and electron (bottom) signal modes for 2017.

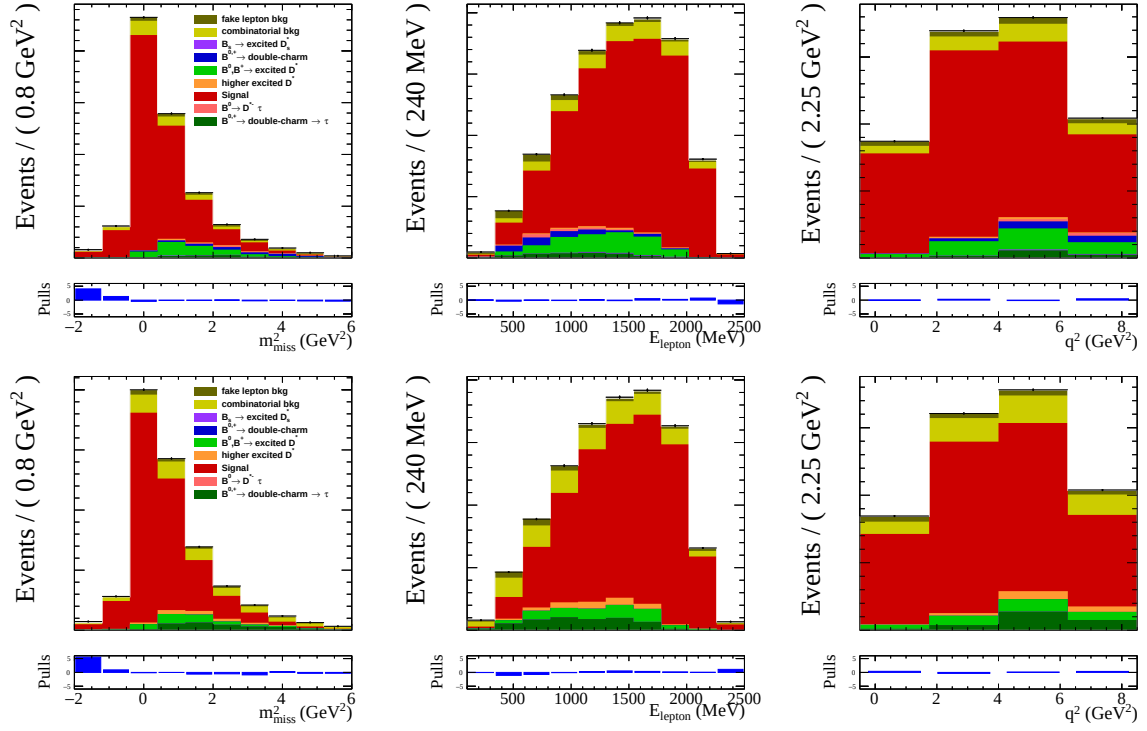


Figure 83: • One-dimensional projections in m_{miss}^2 (left), E_{μ}^* (middle) and q^2 (right) of the fit result of the model to data for the muon (top) and electron (bottom) signal modes for 2018.

be seen that the Beeston-Barlow method is necessary to obtain a good fit result as the simulated samples are too small to provide accurate template shapes.

III.4 Determination of the ratio $R(D^{*+})_{light}$

After obtaining the ratio of the yields R_{raw} of the electron and muon signal modes from the template fit, the only missing ingredient is the estimation of the relative reconstruction and selection efficiencies. The efficiency is obtained separately for the electron and muon signal channels.

III.4.1 Selection efficiencies

The total efficiency ϵ_{tot} to reconstruct and select an event estimated by simulation is given by the product

$$\epsilon_{tot} = \epsilon_{gen} \times \epsilon_{reco|gen} \times \epsilon_{strip|reco} \times \epsilon_{sel|strip} \times \epsilon_{trig|sel}, \quad (77)$$

where each efficiency is defined 'on top' of the preceding ones.

The generator-level efficiency ϵ_{gen} describes the efficiency of the selection applied at the generator-level of the simulation. This includes a selection, which requires the final state particles of the simulated decay to lie within the LHCb acceptance. Additional requirements on the momentum and pseudo-rapidity are applied to save time spent on simulating events, which would be later rejected due to the stripping and offline selection requirements, by already rejecting them at this stage. This efficiency can be directly obtained from the simulation step by generating a large enough sample. The reconstruction efficiency $\epsilon_{reco|gen}$ describes the efficiency of reconstructing all (visible) final state particle tracks of the simulated decay given that the generator-level selection is passed. The stripping efficiency $\epsilon_{strip|reco}$ is defined here as the efficiency of the stripping selection of the simulated decay if the tracks of the final state particles have been reconstructed. The selection efficiency $\epsilon_{sel|strip}$ is the efficiency of all further selections performed offline, with the exception of the trigger requirements, given that the event was selected by the stripping step. Finally the trigger selection efficiency $\epsilon_{trig|sel}$ is defined as the efficiency of events passing the L0 and HLT requirements given they passed all previous steps. The combined efficiency $\epsilon_{reco|gen} \times \epsilon_{strip|reco} \times \epsilon_{sel|strip} \times \epsilon_{trig|sel}$ is estimated from the corrected simulation samples. It is defined as the weighted number of events in the simulated sample after all selection steps and taking all the correction weights into account, divided by the weighted number of events passing the generator-level selection.

III.4.2 $R(D^{*+})_{light}$ flatness cross-check

Similarly to what was done for the control mode, the data is split in bins of various variables and the template fit is repeated in each. The ratio obtained in each bin

is then again corrected by the efficiency ratio, taking into account the additional selection efficiency for an event being in that bin. This serves as a cross-check that the fitting procedure and efficiency correction are well understood by checking the flatness of the resulting distribution, as shown in Figures 84, 85 and 86 as function of p_T and η of the $\bar{B}^0$ and nSPDHits, where the uncertainty is the statistical uncertainty and the systematic uncertainty due to the simulated sample size. No significant trends are observed which demonstrates that the remaining data-simulation discrepancies observed in Section III.2.11 cancel in the ratio.

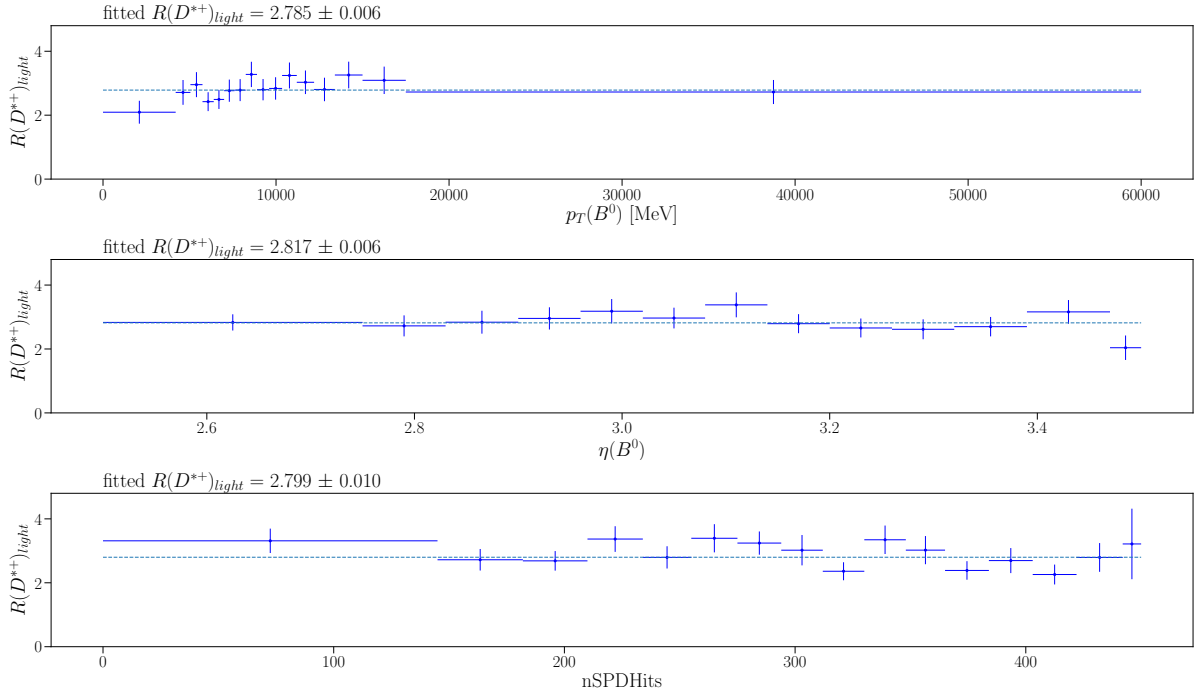


Figure 84: $\bullet R(D^{*+})_{light}$ as function of various quantities for the 2016 sample.

III.4.3 Combination of results

The ratio $R(D^{*+})_{light}$ is obtained separately for the three data samples, summarised in Figure 87. The values are combined as uncertainty-weighted average

$$R_{comb} = \frac{\sum_i \frac{R_i}{\sigma_i^2}}{\sum_j \frac{1}{\sigma_j^2}} \quad (78)$$

with the uncertainty

$$\sigma_{comb} = \sqrt{\left(\sum_i \frac{1}{\sigma_i^2} \right)^{-1}}, \quad (79)$$

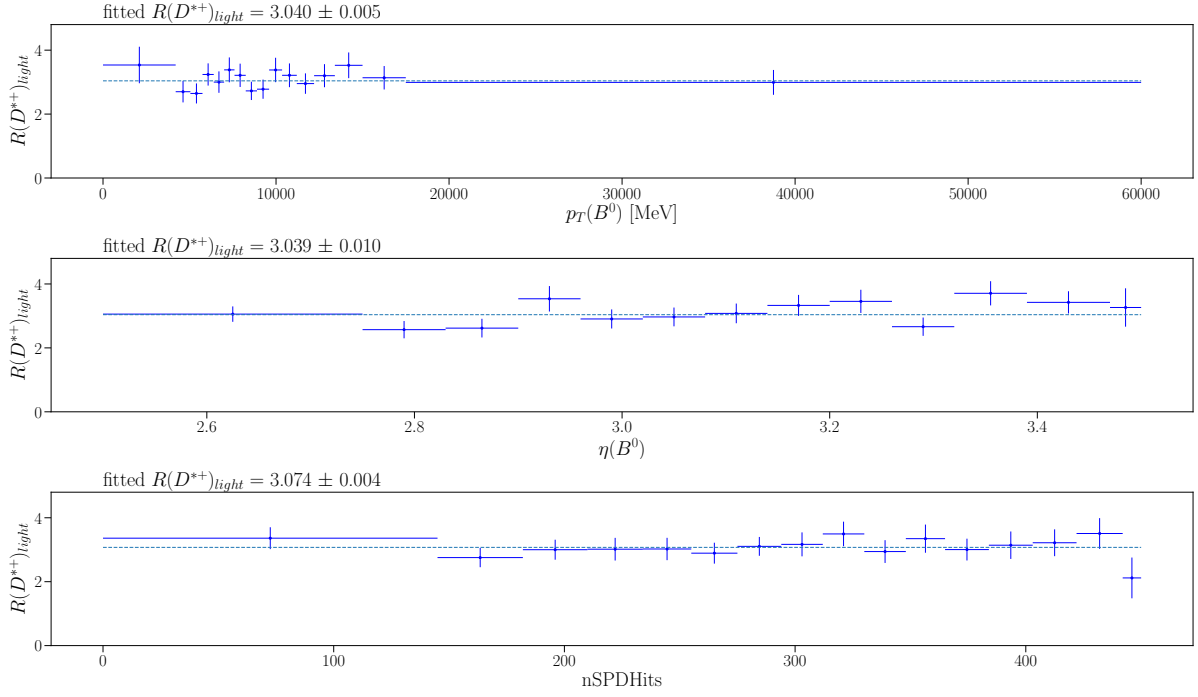


Figure 85: $\bullet R(D^{*+})_{light}$ as function of various quantities for the 2017 sample.

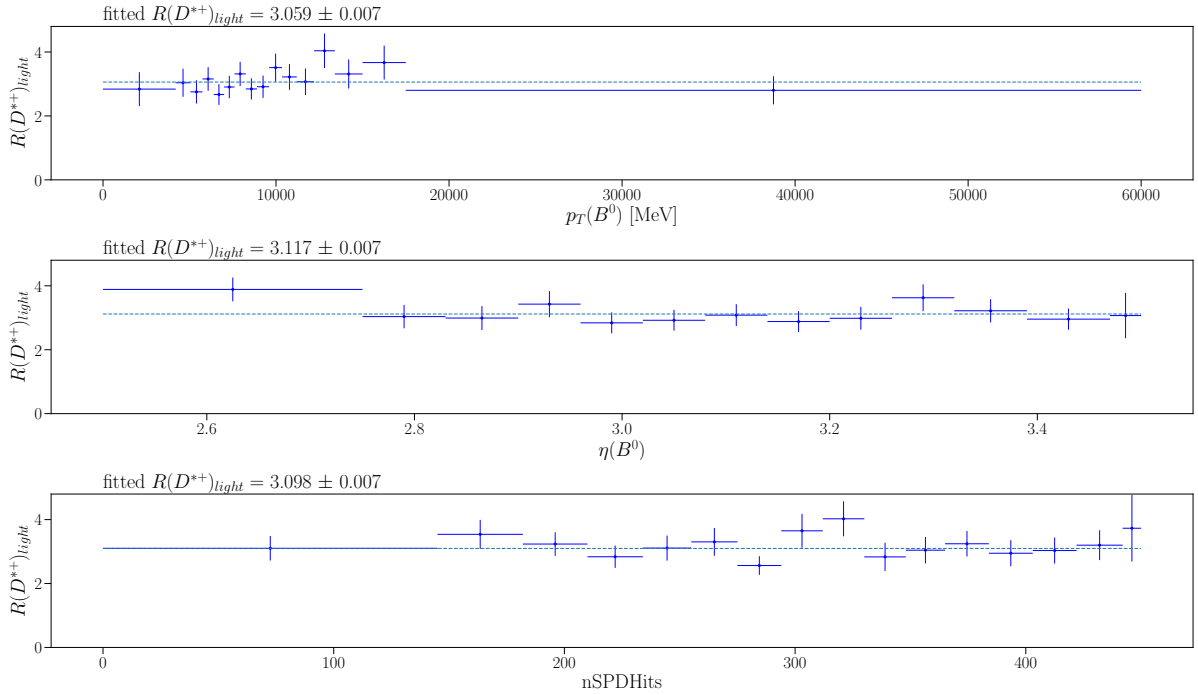


Figure 86: $\bullet R(D^{*+})_{light}$ as function of various quantities for the 2018 sample.

as the uncertainties under consideration are uncorrelated between the years. This results in $R(D^{*+})_{light} = 2.938 \pm 0.064$ where the uncertainty is composed of the statistical uncertainty and the systematic uncertainty due to the simulated sample

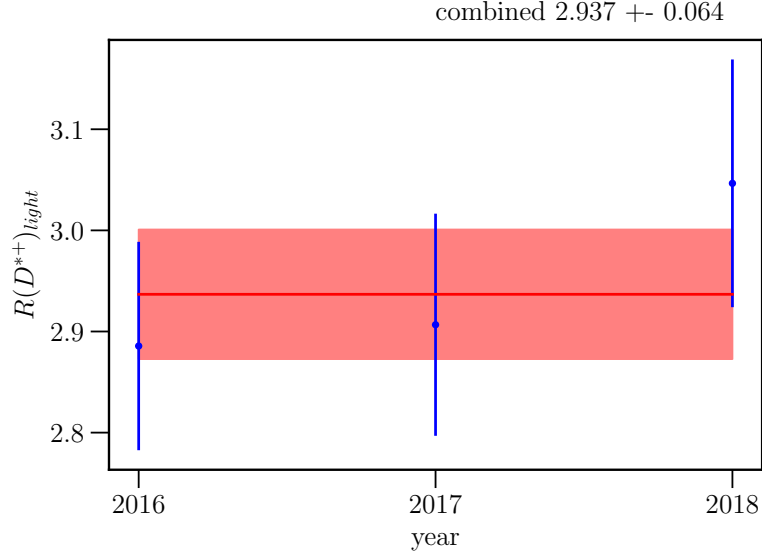


Figure 87: ● Blinded value of $R(D^{*+})_{light}$ obtained for the three years (blue) and the combined value (red). The uncertainties include the statistical uncertainty and the systematic uncertainty due to the size of the simulated samples. The CLN parametrisation is chosen as form factor model.

sizes. The three years are in good agreement with each other and the combined value. In the next chapter additional sources of systematic uncertainties are evaluated and the expected sensitivity is estimated.

Table 19: Sources of systematic uncertainties taken into account and the associated relative systematic uncertainty on $R(D^{*+})_{light}$. Also shown is the total statistical uncertainty.

source	uncertainty $\Delta R(D^{*+})_{light}$ (%)
simulated sample size	2.12
lepton tracking efficiency	0.22
lepton identification	0.16
form factor parametrisation	0.19
total systematic uncertainty	2.15
statistical uncertainty	0.48

III.5 Estimation of systematic uncertainties

As mentioned before, many sources of systematic uncertainty are expected to cancel in the ratio. Of primary interest to estimate the sensitivity of the measurement are the systematic uncertainties, which do not cancel out. These are related to the different behaviour of electrons and muons in the detector. Another source of systematic uncertainty is the size of the available simulation samples. The choice of form factor model is treated as a source of systematic uncertainty. The systematic and statistical uncertainties are summarised in Table 19. This results in the blinded value

$$R(D^{*+})_{light} = X \times (2.937 \pm 0.014(stat.) \pm 0.063(syst.)), \quad (80)$$

with the blinding factor X . The Belle measurement of this ratio cited in the introduction has a relative statistical uncertainty of 0.99% and a relative systematic uncertainty of 2.97%, while this measurement has a relative statistical uncertainty of 0.48% and a relative systematic uncertainty of 2.15%. Exploiting the large LHCb data set the relative statistical uncertainty is smaller than the Belle measurement, even though the fiducial region selection sacrifices many events. The systematic uncertainties evaluated so far show that the different behaviour between electrons and muons in the LHCb detector can be kept under control by this choice of fiducial region. This gives confidence that the sensitivity of this analysis should be at the level of the Belle measurement. While other sources of systematic uncertainty are expected to cancel in the ratio, this needs to be verified to get the final estimate of the total systematic uncertainty. The statistical uncertainty is much smaller than the total systematic uncertainty and the dominant systematic uncertainty is caused by the size of the simulated samples. Increasing the simulation sample sizes by 'simply' generating more events would therefore greatly improve the precision of this analysis.

Table 20: Relative systematic uncertainty due to the lepton tracking efficiency.

year	tracking correction systematic uncertainty (%)		
	muon	electron	total
2016	0.24	0.20	0.31
2017	0.24	0.20	0.31
2018	0.23	0.19	0.30
combined	0.17	0.14	0.22

III.5.1 Tracking efficiency correction

To correct the simulated tracking efficiency for muons and electrons centrally produced correction factors are applied, as described in Section III.2. These correction factors have an associated statistical uncertainty, due to the sizes of the data and simulation samples used, and a systematic uncertainty, due to the procedure of how they were obtained. The statistical and systematic uncertainty on the correction factors needs to be propagated and are a source of systematic uncertainty for this measurement.

To estimate this contribution, a new correction table is created by resampling each correction factor according to its uncertainty, assuming a Gaussian distribution around its central value. Then the newly created table is used for correcting the simulated samples and the efficiency is re-evaluated. By repeating this procedure a distribution of new efficiency values is obtained. The relative systematic uncertainty due to the tracking efficiency corrections is then taken to be the standard deviation of this distribution divided by its mean. The relative uncertainty for the muon and electron tracking efficiencies corrections for each period estimated with 1000 resampled correction tables each is summarised in Table 20. The combined relative systematic uncertainty on the efficiency ratio is also given. It can be seen that the systematic uncertainty is independent of the year and that it is very similar between muons and electrons. This is a result of the choice of fiducial region of the lepton kinematics and demonstrates that final states with electrons can be measured as precisely as final states with muons.

III.5.2 Particle identification efficiency

The particle identification variables used in the selection are replaced by KDE estimates of their distribution obtained from background-subtracted calibration modes. As the systematic uncertainty due to muon and electron identification does not cancel in the ratio its effect needs to be quantified. The sizes of the calibration samples and the kernel-density method are sources of systematic uncer-

Table 21: Relative systematic uncertainty due to the lepton identification efficiency for each year. The uncertainties for each year are combined to obtain the uncertainty on the final result.

year	calibration sample size (%)		KDE method (%)		total (%)	
	μ	e	μ	e	μ	e
2016	0.08	0.17	0.13	0.63	0.15	0.65
2017	0.01	0.26	0.04	0.06	0.05	0.27
2018	0.10	0.24	0.01	0.02	0.10	0.24
combined	0.10	0.12	0.01	0.02	0.10	0.12

tainty on the ratio. The size of the former is estimated by using resampled kernel densities. By using these alternative densities for the PID resampling, the PID efficiency can be estimated for each for them. The expected relative uncertainty is then estimated as the standard deviation divided by the mean of the resulting distribution of efficiencies. This estimation comes with the caveat that only up to ten resampled densities per year and PID variable are available. While this allows for an approximation of the size of the effect, the small number of alternative templates prohibits an exact determination.

The effect of the kernel-density method is estimated by using a KDE with 50% larger kernel width. This results in alternative resampled PID variables. The PID selection efficiency is estimated using these alternative values and the systematic uncertainty is taken as the relative difference between the nominal and alternative PID efficiency. The relative systematic uncertainty due both these effects is summarised for electrons and muons for each year in Table 21.

III.5.3 Size of simulated samples

To estimate the effect of the limited size of the simulated samples on the fit result, the template fit is repeated with and without using the Beeston-Barlow method. In the former case, the uncertainty on R_{raw} obtained from the fit includes the statistical uncertainty and the systematic uncertainty due to the size of the simulation samples used to create the template shapes. In the latter, only the statistical uncertainty is obtained. The systematic contribution can be calculated as the quadratic difference between the respective relative uncertainty. The systematic uncertainty due to the simulation sample sizes and the statistical uncertainty is summarised in Table 22 for all years and the expected uncertainty when combining the three years is given.

This uncertainty also includes the effect of the weighting of the simulated samples, which reduces their effective statistical power. The increase in systematic uncertainty due to the weighting is estimated by repeating the analysis without

applying the simulation correction weights but the form factor weights, which are necessary to obtain reasonable template fit results. The quadratic difference of the relative uncertainties of both scenarios gives the size of the effect of the weighting. It is estimated that the weighting presents a contribution of 1.40% to the relative systematic uncertainty in the quadratic sum in the combination of all years. From this, the effect of choosing the $B^0 \rightarrow J/\psi K^{*0}$ decay mode as control channel rather than the $B^+ \rightarrow K^+ J/\psi$ decay can be approximated. The $B^+ \rightarrow K^+ J/\psi$ yield is observed to be about three times larger than the $B^0 \rightarrow J/\psi K^{*0}$ yield within a given year. Obtaining the weights from the $B^+ \rightarrow K^+ J/\psi$ sample would then reduce the contribution of the weighting procedure to the systematic uncertainty by a factor of $\sqrt{3}$ to 0.81%. The overall systematic uncertainty due to the size of the simulated samples and the weighting procedure then would decrease from 2.12% to 1.78%.

To verify the uncertainty estimate of the Beeston-Barlow method, a toy study is performed, where the data sample is repeatedly refitted 500 times, while the simulated shapes are re-sampled according to their bin uncertainties before each toy fit. These fits are done with the Beeston-Barlow method enabled. The average value and standard deviation of the resulting distribution of fitted ratios are then used to calculate the expected relative uncertainty. This toy study results in uncertainty estimates similar to the ones from the Beeston-Barlow method, but they are up to 8% smaller in each year. Nevertheless this gives confidence that the Beeston-Barlow method gives reasonable estimates of the uncertainty, which in the worst case could be thought of as slightly too conservative.

It is also studied which contribution has the largest influence on the systematic uncertainty. The data fit is repeated with one simulated sample being artificially increased by a factor of 10 by dividing the bin errors of the corresponding histogram by $\sqrt{10}$ before re-doing the fit employing the Beeston-Barlow method. In each fit, the more a sample contributes to the total systematic uncertainty, the more the uncertainty on the fitted ratio will decrease. Using these studies, the total uncertainty due to the limited size of the simulated samples estimated with the Beeston-Barlow method Δ_{MC}^{BB} is then split into the contributions

$$\Delta_{MC}^2 = \Delta_{sig,\mu}^2 + \Delta_{sig,e}^2 + \Delta_{other}^2, \quad (81)$$

where $\Delta_{sig,\mu}$ is the contribution due to the size of the muon signal simulated sample, $\Delta_{sig,e}$ due to the size of the electron signal simulated sample and Δ_{other} due to all other simulated samples, shown in Table 23. As expected, the dominating contribution to this systematic uncertainty is the sample size of the signal modes, as they present the largest yield fraction.

Table 22: Relative systematic uncertainty due to simulated sample sizes obtained with the Beeston-Barlow method and statistical uncertainty for 2016, 2017, 2018 and the combination of the years.

year	Δ_{MC}^{BB} (%)	statistical uncertainty (%)
2016	3.47	0.85
2017	3.68	0.84
2018	3.94	0.81
combined	2.12	0.48

Table 23: Contribution to the relative systematic uncertainty due to simulated sample sizes broken down into the contributions from the muon signal, electron signal and all other simulated sample sizes.

	expected Δ_{MC}^{BB}	$\Delta_{sig,\mu}$	$\Delta_{sig,e}$	Δ_{other}
2016	3.47%	2.51%	1.99%	1.33%
2017	3.68%	2.46%	2.48%	1.16%
2018	3.94%	2.65%	2.63%	1.26%

III.5.4 Form factor model

To estimate the effect of the choice of form factor parametrisation the analysis is repeated with the signal simulation sample weighted to match the BGL model instead of the CLN parametrisation using the measured values of (46). The obtained $R(D^{*+})_{light}$ for each year and the combination is shown in Figure 88, were only the statistical uncertainties and the systematic uncertainties due to simulation sample sizes are considered. Comparing with the results from using the CLN parametrisation, the values for $R(D^{*+})_{light}$ agree within 1% or better for each year. To estimate the systematic uncertainty due to the choice of form factor model, the relative difference between the result for BGL and CLN for each year are combined, as summarised in Table 24. This results in a relative systematic uncertainty of 0.19%.

Table 24: Results for $R(D^{*+})_{light}$ for 2016, 2017 and 2018 using the BGL and CLN form factor parametrisations. The relative difference between the two is also shown and combined to estimate the systematic uncertainty due to form factor model choice.

year	$R(D^{*+})_{light}$ (CLN)	$R(D^{*+})_{light}$ (BGL)	relative difference BGL/CLN
2016	2.886	2.892	0.21%
2017	2.907	2.937	1.03%
2018	3.047	3.033	0.46%
combined uncertainty			0.19%

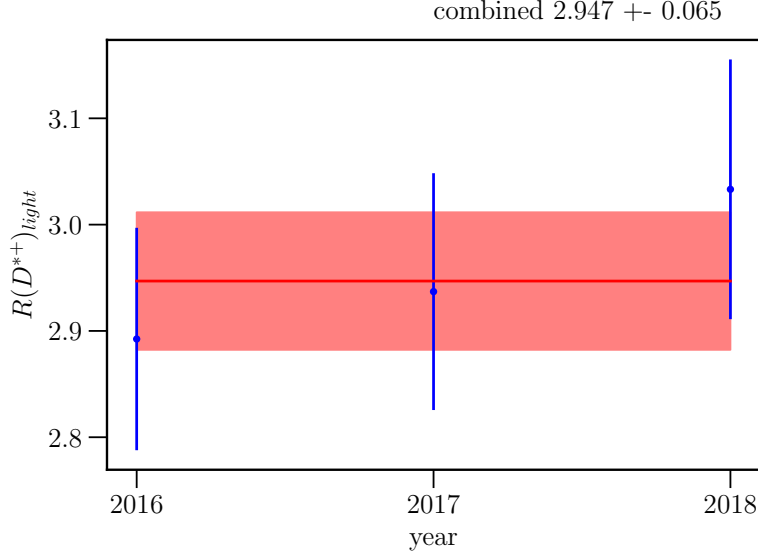


Figure 88: ● Blinded value of $R(D^{*+})_{light}$ obtained for the three years (blue) and the combined value (red). The uncertainties include the statistical uncertainty, the uncertainty on the efficiency calculation and the systematic uncertainty due to the size of the simulated samples. The BGL parametrisation is chosen as form factor model.

III.5.5 Estimation of selection efficiency

The systematic uncertainty due to the signal muon and electron simulation sample sizes is already taken into account by the Beeston-Barlow method when performing the template fit. As the same simulated samples are used to estimate the signal selection efficiencies no further systematic uncertainty is assigned on the efficiency to avoid double-counting this effect. To estimate the systematic uncertainty due to the simulation correction weights the final $R(D^{*+})_{light}$ ratio is obtained without using the weights when calculating the signal selection efficiencies. The difference is of the order $\mathcal{O}(0.1\%)$ and thus presents a negligible contribution to the overall systematic uncertainty.

Chapter IV

Primary vertex reconstruction for the LHCb Upgrade

“Every point of the universe is a fixed point: all you have to do is hang the Pendulum from it.”

Umberto Eco, *Foucault's Pendulum*

For the data-taking period of the LHC set to start in March 2022, the LHCb detector is being upgraded. Not only are many hardware and electronic components replaced, but also the L0 hardware trigger is going to be removed leaving the software-level trigger to deal with the full event rate of 30 MHz. At the same time, the instantaneous luminosity is going to be increased by a factor five to $\mathcal{L} = 2 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$.

The main motivation to completely forgo the hardware trigger is that for many decay channels the signal yield per unit of luminosity saturates at some point if the same trigger strategy as in Run 2 was used, as shown in Figure 89, limiting the effect of increasing the instantaneous luminosity on the collected amount of data. This saturation occurs since above a certain instantaneous luminosity prompt QCD background processes dominate the output rate of the calorimeter triggers. To counteract the increased output rate tighter constraints would have to be introduced at the L0 level, reducing the overall signal selection efficiency. Since many measurements performed at LHCb are limited in statistical precision, collecting more data is necessary. A saturated hardware-level trigger would result in inefficient data-taking and defeat the purpose of the LHCb Upgrade. To overcome this the first stage of the trigger would need to use higher-level objects, such as tracks and PVs, to identify displaced signatures instead of relying on 'raw' detector information.

Removing the L0 trigger puts an enormous strain on the throughput require-

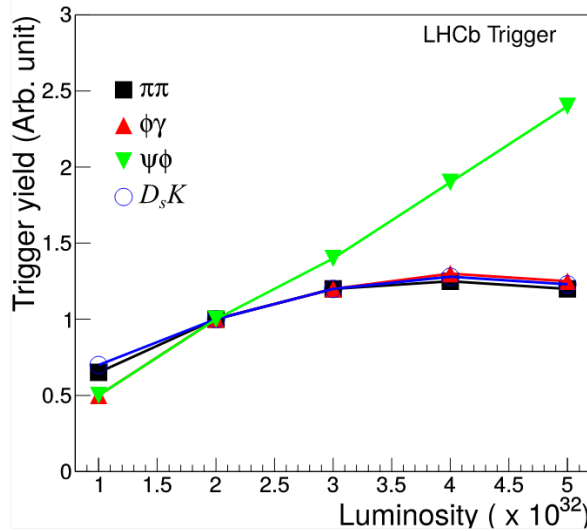


Figure 89: ■ Expected trigger yield per unit of luminosity as function of instantaneous luminosity. The yield for hadronic decays is expected to saturate at large instantaneous luminosity [92].

ment of the software-level trigger, especially its first stage HLT1 [93], as now long tracks and PVs have to be reconstructed at a rate of 30 MHz. Because of this, it was not only necessary to rethink and to improve the reconstruction algorithms already used during Run 2, but also the decision was taken to make use of accelerator hardware by moving the HLT1 sequence to be executed completely on GPUs. This effort was achieved through the *Allen* project [1] [94] within the real-time analysis (RTA) project of LHCb.

The main work performed in the context of the *Allen* project during the course of this thesis consisted of adapting a new PV reconstruction algorithm, originally devised for CPU architectures, to an implementation suitable for GPUs, exploiting its capabilities for parallel execution. Parts of this algorithm, especially concerning the PV fitting part, were modified significantly to achieve the required performance on a GPU.

This chapter first introduces the LHCb Upgrade detector and the boundary conditions HLT1 must fulfil during the Run 3 data-taking period. The motivation for using GPUs to execute the HLT1 sequence is discussed. The adaption of the PV reconstruction algorithm to GPU architectures is described and its physics and timing performance evaluated. Finally a cross-check comparison of the results of compiling the *Allen* software on CPU and GPU hardware is presented.

IV.1 The LHCb Upgrade

While the overall design concept of the LHCb detector remains the same for the Upgrade, many parts are replaced. The opportunity is taken to insert detector elements with finer spatial granularity and more radiation resistance to last until the aimed for 50 fb^{-1} integrated luminosity of data are taken with the Upgrade detector. In case of the VELO, the $R - \phi$ -sensors are replaced by silicon pixel detectors. The previously semicircular detector halves are replaced with two "L"-shapes, as shown in Figure 90. Instead of separate measurements of the r - and ϕ - coordinates the pixel detector delivers the x - and y - coordinates of a hit. A new tracker, called Upstream Tracker (UT), will replace the TT in front of the magnet, while the downstream tracking stations will be replaced by new stations made from scintillating fibres (SciFi). The PS and the first muon station M1 are removed. The optical system and photodetectors of the RICH system are replaced. In addition, a large part of the electronics is changed out due to the L0 being removed. The updated electronics allow for reading out the detector at a maximum data rate of 40 Tbit/s. A schematic view of the Upgrade LHCb detector is shown in Figure 91.

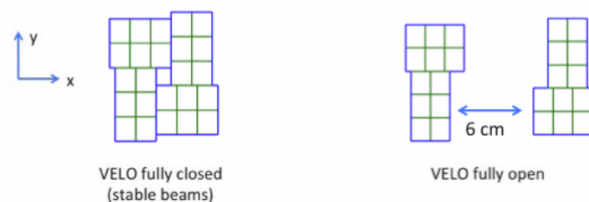


Figure 90: ■ Schematic front view of an Upgrade VELO detector plane in closed (left) and open (right) positions [95]

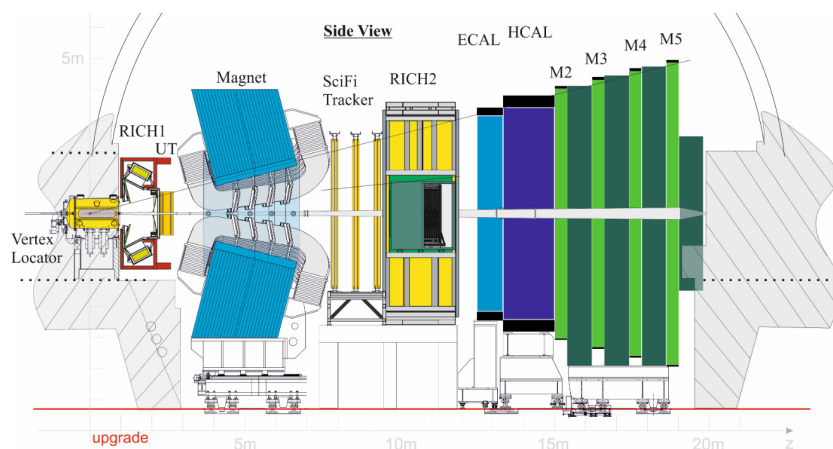


Figure 91: ■ Schematic view of the LHCb Upgrade detector [96].

The software-level trigger for Run 3 and beyond is a particular challenge, as described below. In the effort to address this the LHCb collaboration came up with two viable implementations of the first stage of the software-level trigger HLT1. One was to be executed on CPU architectures, the other on GPU architectures.

IV.1.1 Run 3 trigger

The trigger strategy of the LHCb Upgrade detector for Run 3 is fundamentally changed compared to previous data-taking periods. With the hardware trigger removed and the instantaneous luminosity increased the throughput requirement (in terms of events processed per second) of the software trigger is much harsher, while it also needs to deal with larger occupancies in the detector. This change of paradigm is illustrated in Figure 92 showing an overview of the trigger stages during Run 2 and Run 3. Now the full 30 MHz non-empty event rate is sent directly to the first stage of the software-level trigger, where all sub-detectors have to be read out and the event is partially reconstructed. Similar to the Run 2 HLT1, the first stage of the software-level trigger requires at least the output of the VELO, UT, SciFi and muon stations and the amount of work it has to perform in Run 3 is at least reconstructing charged long tracks using the upstream and downstream tracking stations, reconstructing primary vertices, identifying muons and making selections based on single tracks and two-track combinations. Reaching this minimum baseline for the HLT1 proved a non-trivial challenge, though it was achieved using both CPU- and GPU-based architectures within the budget constraints of the LHCb Upgrade.

Using CPUs or GPUs has consequences on the data flow of the trigger and data acquisition system, as illustrated in Figure 93. The GPUs are placed in the event building servers so that the HLT1 sequence can already be executed at this stage. If the HLT1 decision is positive the event is kept and sent to the event-filter farm. This saves part of the budget which has to be spent on the network as onboard 10 Gb/s network cards are sufficient instead of having to buy custom 100 Gbit/s network cards, as the data rate sent over the network is reduced from a maximum of 40 Tbit/s to 1-2 Tbit/s. An added benefit is that this also frees up slots in the event-builder server to place the GPUs.

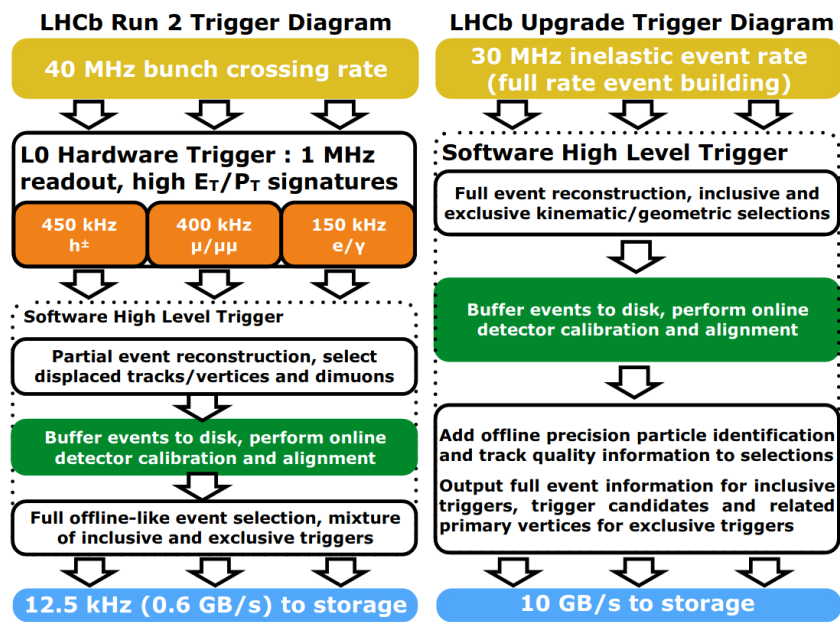


Figure 92: ■ Trigger strategies employed during Run 2 (left) and Run 3 (right) [97].

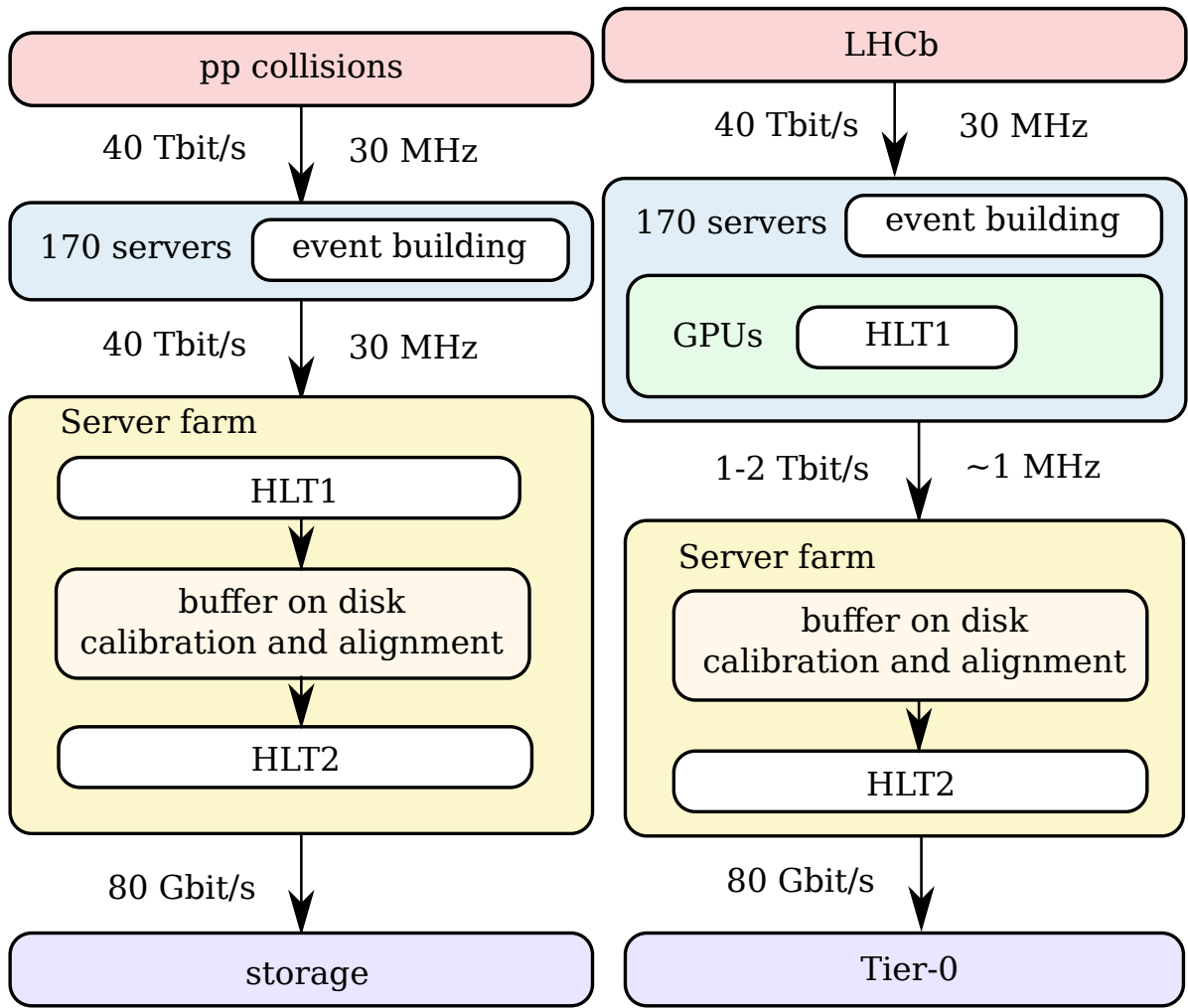


Figure 93: ■ Data flow for the CPU- (left) and GPU- (right) based HLT1. By placing the GPUs in the event-builder server farm, the data rate sent to the event-filter farm can be reduced by an order of magnitude before sending it to the server farm via the network. Adapted from [1] with updated numbers.

IV.2 The Allen project

For the *Allen* project to be chosen as the baseline solution, it had to show that it can address several issues. It needed to be more cost-efficient compared to the CPU solution while reaching the same or better physics and selection performance. As there is only limited space in the event-filter farm servers to place the GPUs, its throughput per GPU must be high enough to meet the overall throughput requirement of 30 MHz. There is space to place two GPUs in each of the 170 servers, which means that a single GPU needs to process events at a rate of about 88 kHz with an appropriate safety margin. To reduce the complexity of the overall system and to avoid having too many data-transfers to and from the GPUs it is a desirable feature that the complete HLT1 sequence runs on GPUs and not only parts of it. The *Allen* software would then take as input the raw detector information of the relevant sub-detectors and deliver as output a report of the decision of each trigger line and which physical objects caused positive decisions.

At the current time, it is not feasible to use GPUs and run *Allen* on them when simulating the HLT1 on the LHC computing grid, where mostly CPU servers are available. For this reason, the *Allen* software must also be compilable and executable on CPU architectures. As compiling software on different architectures can give rise to small differences in the results, for example due to different rounding schemes, it had to be shown that these difference between the CPU and GPU compilations of *Allen* are small, as it otherwise would introduce additional discrepancies between the data and simulation samples.

IV.2.1 Modern processing architectures

Modern CPUs can exploit the SIMD (single instruction, multiple data) paradigm to achieve high throughput. This means that each instruction, e.g. an addition, is not executed on a single entry in memory, but on multiple ones at the same time. The addition of two vectors of the same length can serve as an illustrative example. The corresponding elements of both vectors have to be read from memory, the addition is performed and the result is stored. Without SIMD each element is processed in sequential order, while with SIMD multiple elements are processed in parallel. The number of simultaneously processed elements depends on the vector length available on the employed CPU, which typically is up to 16 on modern hardware.

In contrast, the main paradigm for GPUs is SIMT (single instruction, multiple threads). On a GPU multiple independent threads are launched, which execute the same code. Each thread then operates on a different element in memory. In the

example of vector addition, each thread would perform the addition of a different element. Depending on the length of the input vector and the capabilities of the GPU, one could launch a thread for each element. Alternatively one can loop over the vector in steps given by the number of threads.

The programming language chosen for the *Allen* project is C++ combined with CUDA [98], developed by *Nvidia*, which allows the writing and compiling of code to be executed on *Nvidia* GPUs. The CUDA code is written in a way that by using macros it can be compiled and executed on CPUs. In the CUDA framework threads are organised within 'blocks', which in turn are organised in a 'grid'. As an example of how this would map to the use case at LHCb, one could imagine having a one-dimensional grid with a block associated with each event. Within each event-block threads could be launched to, for example, perform an operation on already found tracks. The number of blocks and threads used for each algorithm and how they are used is non-trivial and an important part of optimizing the code in terms of throughput. An important feature is that threads within the same block can be synchronised and have access to a common memory, called shared memory, which provides a way where the independent threads can exchange information. The GPU 'device' communicates with the CPU 'host' via a global memory, which both can access and write data to. The thread blocks are executed on streaming multiprocessors, which contain many computing cores performing the calculations. Each streaming multiprocessor can handle multiple blocks at the same time depending on its memory resources. Within each block, threads are grouped into so-called 'warps' consisting of 32 threads. The threads in a warp execute the code instructions at exactly the same time.

The GPUs are connected to a server by a PCIe (Peripheral Component Interconnect Express) connection, which puts some limits on the rate at which data can be exchanged between the server and the GPU. To avoid these limitations the whole HLT1 sequence is executed on GPUs so that partial results do not have to be copied to and from the GPU.

IV.3 Primary vertex reconstruction

The first stage of the software-level trigger relies to a large part on displaced signatures. Either high transverse momentum tracks with significant displacement with respect to each PV in an event or a two-track vertex displaced from any PV are used in the selection. For this purpose, PVs must be reconstructed in the HLT1 with high efficiency and a low fake rate. A high efficiency is needed so that prompt tracks are not misclassified as displaced if their associated PV is not reconstructed. A low fake rate ensures that displaced tracks are not misclassified as prompt if they happen to point towards the fake PV. In addition to the general challenges due to the increased instantaneous luminosity and trigger input rate, the PV reconstruction used in the LHCb Upgrade software needs to cope with an increased expected number of PVs per event of about 5, which leads to a higher rate of spatially overlapping PVs, which are more complicated to deal with compared to PVs which are isolated.

The PV reconstruction is based on the tracks found in the VELO, which surrounds the interaction region, as indicated in Figure 94. For Run 3 each event contains on average about 200 VELO tracks and 5.6 'visible' PVs, which produce enough tracks¹ in the VELO so that they can be reconstructed. The PVs are distributed in x and y with a spread of about $\pm 20 \mu\text{m}$ and in z with about $\pm 6 \text{ cm}$. The simulated distribution of x - and z -positions of PVs for Run 3 events is shown in Figure 95. The simulation assumes the beamline to be positioned at $(x, y) = (0, 0)$ and the centre of the pp collisions to be at $z = 0$.

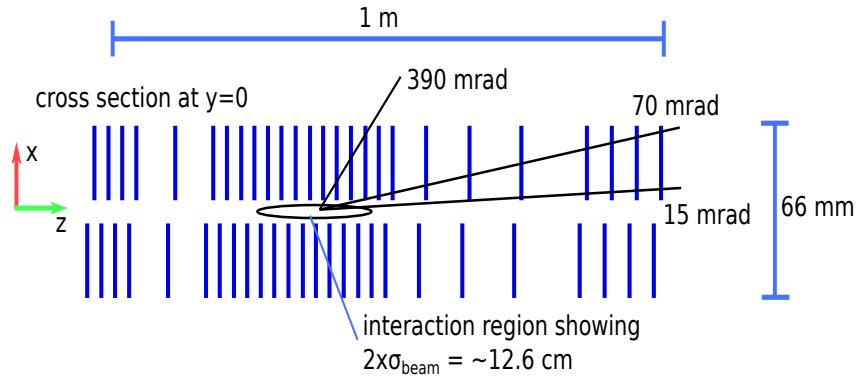


Figure 94: ■ Schematic side view of the Upgrade VELO detector with the pp interaction region [95].

The outputs of the PV reconstruction are the positions of the PVs, the associated uncertainties and the list of tracks associated with each PV. In addition to the throughput requirements, there are several physics performance parameters

¹typically four tracks are required

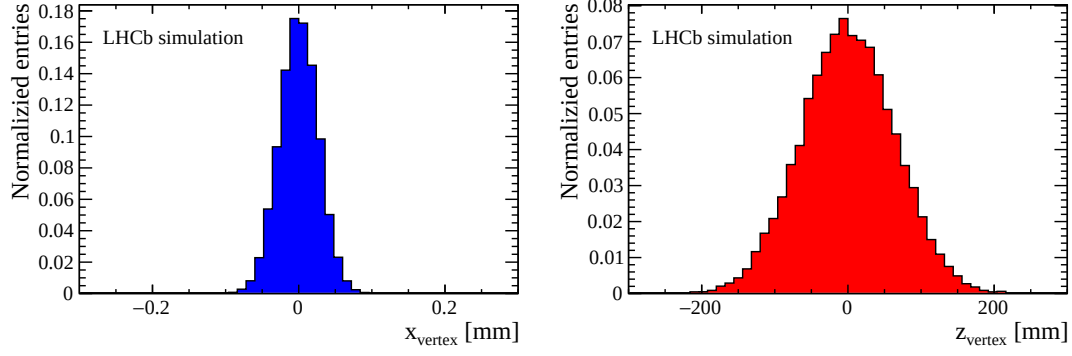


Figure 95: ■● Simulated distribution of PV x - (left) and z - (right) positions during Run 3 [99].

used to judge the viability of a PV reconstruction algorithm as described in Section IV.3.3.

IV.3.1 The Run 3 algorithm ▲

After the tracks in the VELO are found, previous vertex reconstruction algorithms employed at LHCb [100] usually consist of a seeding step, where the initial PV positions are obtained and a vertex fitting part, where the final vertex positions and the corresponding uncertainties are estimated. In the PV reconstruction algorithm used for Run 2, the seeding part was building vertex candidates using track-track combinations. Since this becomes more time-intensive the larger the track multiplicity in the VELO in an event, a different approach – initially implemented on CPU architectures– was developed for Run 3 data-taking. Now the seeding is based on identifying peaks in a histogram to avoid doing such combinations. This histogram is created for each event by first extrapolating the VELO tracks from their first measurement to their point of closest approach to the beamline $\vec{x}_{POCA}$. This requires that the position of the beamline is known with sufficient precision. It is obtained during the VELO closing procedure as straight-line tracking and vertexing of the beam-beam vertices can be performed while the VELO is closing. This information must be accessible to the GPU. The histogram is filled with density distributions around those points taking into account the uncertainty of the track states. In the following, the basic steps of the algorithm are discussed as it was implemented on CPU architectures. In Section IV.3.2 the changes made to obtain a fast implementation on GPUs are discussed.

The main motivation of the histogramming approach can be understood by comparing the distribution of the PVs in x , y and z with the distance of the track x, y, z -position to the PV position in simulated data. In case of the x - and y -coordinates, the distance is defined as the distance of the track x - and y -position

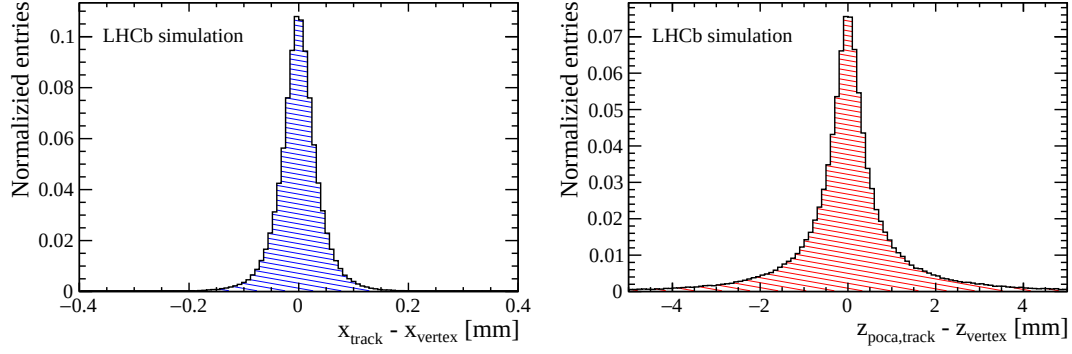


Figure 96: ■● Distance between the x -position of the track when extrapolated to the z -position of its origin PV and the true z -position of the PV (left). Difference between the true z -position of a PV and the z -position of the point of closest approach to the beamline of a track originating from the PV. Both distributions are obtained from simulation of Run 3 conditions [99].

after extrapolating the track state to the true z -position of its origin PV. For the z -coordinate, the z -position z_{POCA} of the point of closest approach (POCA) of the track to the beamline is compared with the simulated MC PV z -position.

Since the distributions in x and y are very similar, only x and z distributions are shown in Figure 96, to be compared with Figure 95. For x (and y), the distribution of the PVs is similar to the distance to the tracks meaning that it is not possible to infer the PV x -position from x_{track} . For the z -coordinate on the other hand, the PVs are spread out over a wide range compared to the distance of the track to the PV. From this, it seems that it should be possible to get a reasonable estimate of the PV z -position by looking at the z_{poca} of its tracks. This can be done by filling a histogram with the values of z_{poca} . Since the tracks originating from a certain PV should have similar values of z_{poca} , peaking distributions in the histogram should indicate the presence of a PV, roughly at the z -position of the peak. When this algorithm was developed, the novelty of also taking into account the uncertainty on z_{poca} when filling the histogram was introduced.

With this in mind the algorithm can be roughly divided into the following steps:

1. extrapolation of the VELO track states to the point of closest approach $\vec{x}_{POCA}$ to the beamline
2. filling a histogram with a density distribution around z_{POCA} of the extrapolated tracks
3. searching for peaks in this histogram
4. assigning tracks to the identified peaks

5. fitting each vertex using the assigned tracks

VELO track extrapolation

Before the PV reconstruction algorithm is executed, the track state (x, y, t_x, t_y) at a certain z -position, where $t_x = \frac{dx}{dz}$ and $t_y = \frac{dy}{dz}$ are the track slopes, is estimated with a simplified Kalman filter [101], where the x and y components are treated independently. From this the covariance matrix C , describing the uncertainties and correlations of the track parameters, is obtained

$$C = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{xt_x} & \sigma_{xt_y} \\ \sigma_{xy} & \sigma_y^2 & \sigma_{yt_x} & \sigma_{yt_y} \\ \sigma_{xt_x} & \sigma_{yt_x} & \sigma_{t_x}^2 & \sigma_{t_x t_y} \\ \sigma_{xt_y} & \sigma_{yt_y} & \sigma_{t_x t_y} & \sigma_{t_y}^2 \end{pmatrix}. \quad (82)$$

Since x and y are treated independently, the corresponding correlation terms are zero and the covariance matrix simplifies to

$$C = \begin{pmatrix} \sigma_x^2 & 0 & \sigma_{xt_x} & 0 \\ 0 & \sigma_y^2 & 0 & \sigma_{yt_y} \\ \sigma_{xt_x} & 0 & \sigma_{t_x}^2 & 0 \\ 0 & \sigma_{yt_y} & 0 & \sigma_{t_y}^2 \end{pmatrix}. \quad (83)$$

The track states are extrapolated from the first-measured point to their point of closest approach to the beamline by propagating the track state and covariance matrix by the distance Δz with

$$\Delta z = \frac{t_x(x_{beamline} - x) + t_y(y_{beamline} - y)}{t_x^2 + t_y^2}. \quad (84)$$

Histogram filling

To take into account the uncertainty of the track extrapolation when filling the histogram of z_{POCA} values, each track contributes to multiple histogram bins around its estimated z_{POCA} value. The contribution to each bin is calculated according to a Gaussian density distribution with mean value z_{POCA} and standard deviation $\sigma_{z_{POCA}}$ defined by

$$\sigma_{z_{POCA}} = \left[\begin{pmatrix} t_x \\ t_y \end{pmatrix}^T \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}^{-1} \begin{pmatrix} t_x \\ t_y \end{pmatrix} \right]^{-\frac{1}{2}}. \quad (85)$$

The uncertainty $\sigma_{z_{POCA}}$ is defined this way so that it increases when the uncertainty on the track position, σ_x and σ_y , is larger or when the track slopes are small. A small track slope means that its angle with respect to the z -axis is small, so that it provides less precise information about z_{POCA} . The contribution of a track to a bin i in the histogram with the upper and lower limits $z_{i,max}$ and $z_{i,min}$ can then be obtained by integrating the Gaussian density in this range

$$\frac{1}{\sqrt{2\pi}\sigma_{z_{POCA}}} \int_{z_{i,min}}^{z_{i,max}} \exp\left(-\frac{(z - z_{POCA})^2}{2\sigma_{z_{POCA}}^2}\right) dz. \quad (86)$$

To avoid having to calculate these integrals of Gaussian distributions, which might slow down the algorithm, template distributions are used instead, which are created and saved in advance. The template distribution approximates the contribution to each bin for a track with a $\sigma_{z_{POCA}}$ in a certain range.

The histogramming approach requires the definition of several parameters, like the size of the bins, which can influence the physics and timing performance. A too small bin size could make it harder to identify peaks in the histogram, negatively affecting the reconstruction efficiency, as statistical fluctuations could be misidentified as peaks. A too large bin size on the other hand could result in two distinct PVs being merged into one peak, reducing the overall reconstruction efficiency.

A typical histogram covering the full z -range where PVs are expected is shown in Figure 97. Peaking structures, which likely correspond to the presence of a PV at that z -position, can already be seen 'by eye'.

Peak search

The peak finding algorithm first identifies 'proto-clusters', which are a continuous range of bins with bin content larger than a certain threshold, by 'scanning' the histogram from left to right. These clusters are then subdivided into significant peaks and dips. The exact definition of 'significant' depends on a configurable parameter. The peaks identified in the example histogram are shown in Figure 98, marked by blue lines. The central region of the histogram is shown in Figure 99, also marking the edges of each cluster by red lines. Each peak corresponds to a vertex candidate for the subsequent fit if it has enough associated tracks. Also shown are the z -positions of reconstructible and non-reconstructible simulated PVs by green and yellow lines, respectively. It can be seen that for most cases the peak corresponds to a true PV. In this example, a broader structure is observed around $z = -40$ mm caused by a true PV. This broad structure is not identified as a peak and the corresponding PV is not reconstructed, which is a source of

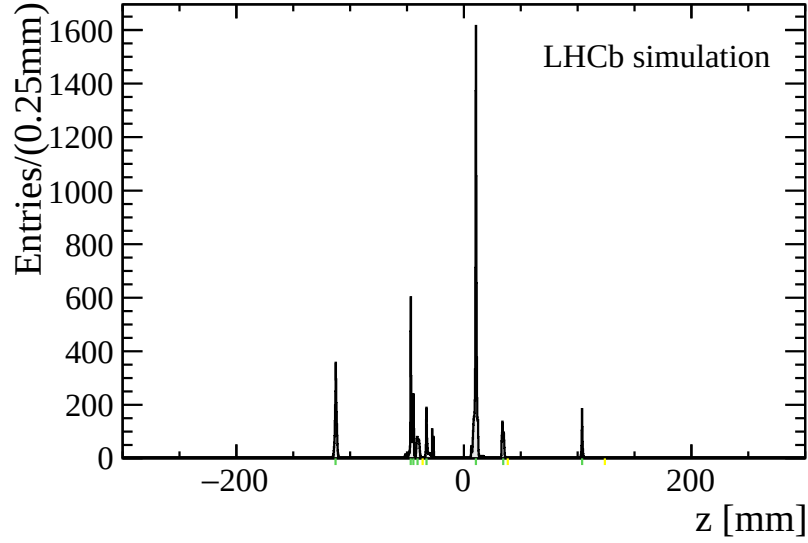


Figure 97: ■● Typical histogram filled by the PV reconstruction algorithm. The histogram is filled with the z_{POCA} values of all VELO tracks taking into account the uncertainty. The green and yellow lines indicate the position of reconstructible and non-reconstructible simulated PVs, respectively. Adapted from [99].

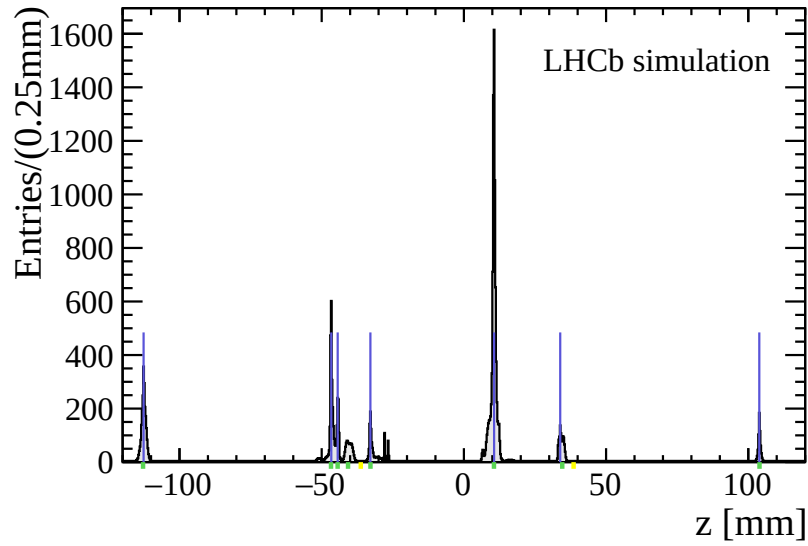


Figure 98: ■● Typical histogram filled by the PV reconstruction algorithm zoomed on central z -values. The histogram is filled with the z_{POCA} values of all VELO tracks taking into account the uncertainty. Identified peaks are shown as blue vertical lines. The green and yellow lines indicate the position of reconstructible and non-reconstructible simulated PVs, respectively. Adapted from [99].

inefficiency of the algorithm. Other regions where the histogram has non-zero entries, but no true PV is presented are likely due to tracks from secondary PVs.

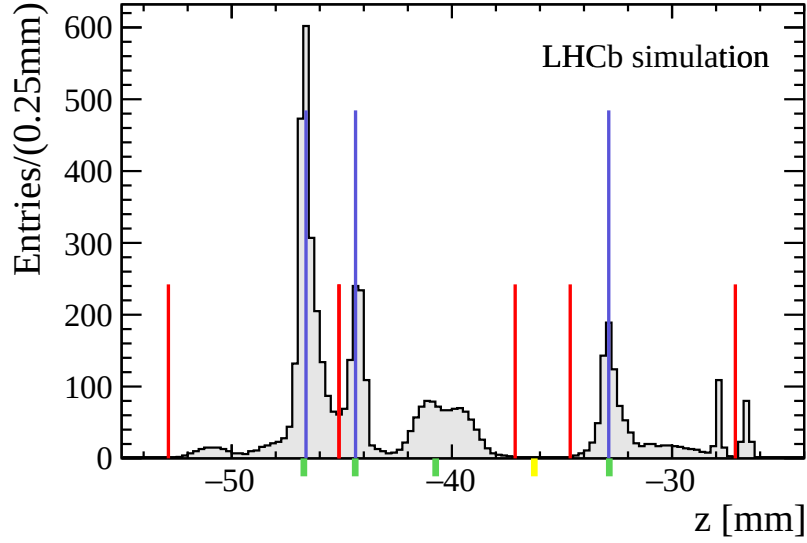


Figure 99: ■● Typical histogram filled by the PV reconstruction algorithm zoomed on some of the identified peaks. The histogram is filled with the z_{POCA} values of all VELO tracks taking into account the uncertainty. Identified peaks are shown as blue vertical lines and the borders of the peaks as vertical red lines. The green and yellow lines indicate the position of reconstructible and non-reconstructible simulated PVs, respectively. Adapted from [99].

Track association

After the peaks at positions z_i have been found, the VELO tracks are distributed among them. This is done by assigning a track to the closest PV candidate in z . Each peak is required to have a minimum amount of associated tracks to be used as vertex candidate in the fitting step. If a vertex candidate has too few tracks assigned, the tracks are re-distributed among the other neighbouring candidates. In this way, each track is assigned to exactly one vertex candidate and can only contribute to the fit of this vertex as explained in the next step. This is a fundamental difference to the GPU implementation as explained in the next section.

Vertex fit

The PV reconstruction algorithm needs to provide an estimation of the position of the vertex

$$\vec{q}_{vertex} = \begin{pmatrix} x_{vertex} \\ y_{vertex} \\ z_{vertex} \end{pmatrix} \quad (87)$$

and the associated uncertainties. As starting point for the vertex fit the initial position of the vertex candidate is defined using the known (x, y) -position of the

beamline and the z -positions of the found peaks as

$$\vec{q}_{vertex,initial} = \begin{pmatrix} x_{beamline} \\ y_{beamline} \\ z_{peak} \end{pmatrix}. \quad (88)$$

The problem of fitting a single vertex using a number i of associated tracks can be solved by finding the vertex position minimizing the χ^2 -sum defined as

$$\chi^2 = \sum_{tracks} r_i^T W_i r_i, \quad (89)$$

where r_i are the residuals of the track extrapolated to the current vertex z -position with respect to the current vertex x - and y -positions,

$$r_i = \begin{pmatrix} x_{res} \\ y_{res} \end{pmatrix}_i = \begin{pmatrix} x_{vtx} + t_x(z_{trk_i} - z_{vtx}) - x_{trk,i} \\ y_{vtx} + t_y(z_{trk_i} - z_{vtx}) - y_{trk,i} \end{pmatrix}, \quad (90)$$

and W_i is the inverted reduced track covariance matrix of the x and y -coordinates

$$W_i = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}_i^{-1}. \quad (91)$$

It is advantageous to multiply the contribution of each track by a Tukey weight w_i [102], which reduces the impact of outliers or unrelated tracks from secondary vertices on the PV fit

$$\chi^2 = \sum_{tracks} w_i r_i^T W_i r_i. \quad (92)$$

The Tukey weights are defined as a function of the χ_i^2 of a track by

$$w(\chi_i^2) = \begin{cases} \sqrt{1 - \frac{\chi_i^2}{\chi_{max}^2}}, & \chi_i^2 < \chi_{max}^2, \\ 0, & \chi_i^2 \geq \chi_{max}^2 \end{cases}, \quad (93)$$

where χ_{max}^2 is a cut-off value and χ_i^2 is defined as

$$\chi_i^2 = \begin{pmatrix} x_{res} \\ y_{res} \end{pmatrix}_i^T \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}_i^{-1} \begin{pmatrix} x_{res} \\ y_{res} \end{pmatrix}_i, \quad (94)$$

which simplifies to

$$\chi_i^2 = \frac{x_{res,i}^2}{\sigma_{x,i}^2} + \frac{y_{res,i}^2}{\sigma_{y,i}^2} \quad (95)$$

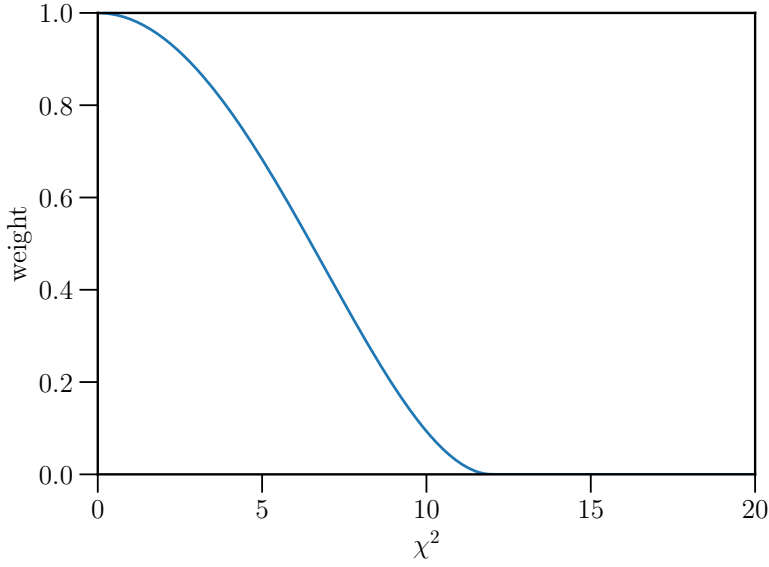


Figure 100: • Tukey weight as function of the χ^2 of a track with a cut-off at $\chi_{max}^2 = 12$.

using that x and y are independent and the off-diagonal element of the reduced covariance matrix are zero. An example with $\chi_{max}^2 = 12$ is shown in Figure 100.

Minimizing the χ^2 -sum means finding the vertex position q_{final} for which

$$\left. \frac{d\chi^2}{dq} \right|_{q_{final}} = 0. \quad (96)$$

This is done using the Newton–Raphson method, which provides an estimation of the position of the zero of a function $f(X)$ starting from an initial guess at x_0 . A better approximation x_1 of the zero is then obtained by

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} \quad (97)$$

By repeating this a more precise estimation can be reached.

The quantities that need to be calculated are

$$\frac{d\chi^2}{dq} = \sum_{tracks} 2w_i H_i^T W_i r_i \quad (98)$$

and

$$\frac{d^2\chi^2}{dq^2} = \sum_{tracks} 2w_i H_i^T W_i H_i, \quad (99)$$

where H_i is the derivative of the residuals

$$H_i = \frac{dr_i}{dq} = \begin{pmatrix} 1 & 0 & -t_{x,i} \\ 0 & 1 & -t_{y,i} \end{pmatrix}. \quad (100)$$

The vertex position $\vec{q} = (x_{vtx}, y_{vtx}, z_{vtx})^T$ is then updated by

$$\Delta q = - \left(\frac{d^2 \chi^2}{dq^2} \right)^{-1} \frac{d\chi^2}{dq}. \quad (101)$$

The covariance matrix C_{vtx} describing the uncertainty on the reconstructed vertex position and the correlation between the vertex parameters is calculated by

$$C_{vtx} = 2 \left(\frac{d^2 \chi^2}{dq^2} \right)^{-1}. \quad (102)$$

The vertex fit is executed iteratively, updating all the weights in each step, until it has converged. The convergence criterion used is that the change in vertex position is small or a maximum number of iterations have been performed.

IV.3.2 Implementation on GPUs

The GPU implementation for the *Allen* project is based on the algorithm devised for CPU architectures described in the previous section and follows a similar logic. The challenge is to identify the parts of the algorithm which can make use of the thread-level parallelism offered by GPUs to achieve the necessary throughput. Especially for the vertex fit part, this meant choosing a different approach as explained below.

The first step, the track extrapolation, can be performed in parallel by assigning one thread to a track since the track states are independent of each other and are read from and saved to distinct places in memory. The histogram can be filled in a similar manner, where one thread is assigned to an extrapolated track, looks up its z_{POCA} -position and increases the corresponding histogram bins, again taking into account the uncertainty on z_{POCA} . However one must be careful to avoid so-called 'race conditions', where two threads access and write to the same memory location at the same time. For example, two threads could read the content of the same bin at the same time, increase it by the respective amount, but one thread first writes the result to memory, which is overwritten by the second thread. One way to address this is the usage of 'atomic' functions, which are built-in functions which block other threads from accessing the memory until a thread is finished with increasing the bin content. The drawback is that this could slow down the

execution of the algorithm if a large number of threads try to access the same memory location at the same time, but this is not observed to be a significant problem here.

The peak search is at the current moment not yet optimised for execution on GPUs. It still follows the sequential logic of the initial CPU implementation. The reason for this is that other parts of the algorithm, especially the vertex fitting, contributed more to the overall execution time of the PV finding and so got more attention first. In the future, one could devise more suitable implementations of the peak finding procedure. For example, one could subdivide the histogram into different, possibly overlapping regions, where within every region a thread is assigned to identify peaks.

The next two steps –the association of tracks to PV candidates and the PV fitting– were changed the most compared to the initial CPU algorithm and in a sense merged into one part. The changes were motivated by the track assignment algorithm not necessarily being suitable for GPUs. As tracks get re-distributed if a vertex candidate turns out to have too few associated tracks this creates dependencies between the vertex candidates which might make an efficient GPU implementation, which ideally has independent tasks acting on independent sets of data, more difficult. Another issue being addressed with the changed algorithm is how to assign a track to a certain PV candidate. With multiple PVs per event, two or more PVs may be close to each other. In these cases, it might make more sense to 'share' the track between the PVs instead of strictly assigning it to one candidate. Therefore, an alternative approach was developed which mapped well to the capabilities for parallel execution offered by GPUs by fitting all PVs simultaneously.

This is achieved by changing the definition of the weights used in the fitting procedure. The weights used in the GPU implementation now not only depend on the χ_i^2 of a track with respect to the current vertex candidate being fitted, but also on the presence of other nearby vertex candidates and instead of assigning a track to a single vertex candidate, it can contribute to the fit of each candidate using this weight. This approach is inspired by multi-vertex fitter algorithms and the same definition of the weights is used [103]. The weight w_{ij} with which the track i contributes to the fit of vertex candidate j is then given by

$$w_{ij} = \frac{\exp(-\chi_{ij}^2/2)}{\exp(-\chi_{cut}^2/2) + \sum_{k=vertices} \exp(-\chi_{ik}^2/2)}, \quad (103)$$

where χ_{cut}^2 is a cut-off value and the χ_{ij}^2 of a track i with respect to a PV j is defined as in (95). This means that a track, which is close to two vertex candidates,

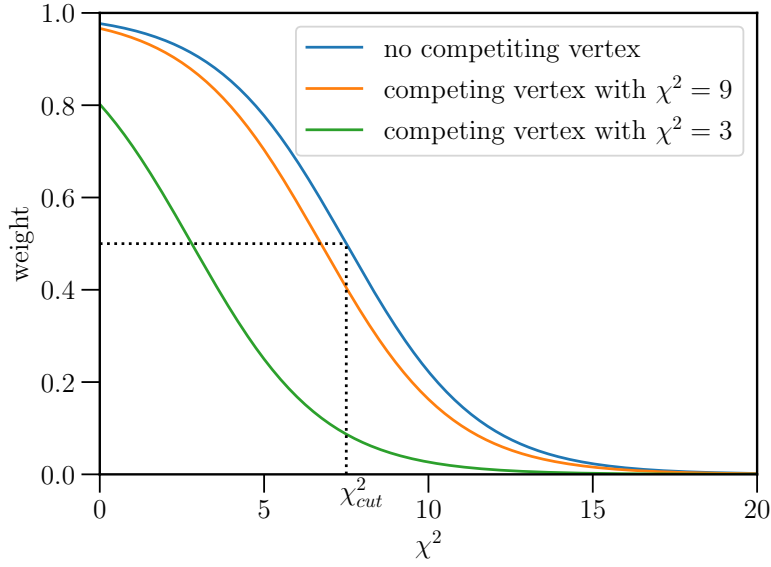


Figure 101: ● Multi-vertex track weights as function of χ^2 for a vertex candidate with no competing vertices, a competing vertex with $\chi^2=3$ and $\chi^2 = 9$, all with $\chi^2_{cut} = 7.5$.

will contribute to both in the respective fits, but with a smaller weight, reducing its impact on both fits. As an illustration of this, Figure 101 shows the weight as function of χ^2 in the case where no other competing vertices are in proximity and the cases where there are other PVs with $\chi^2 = 9$ and $\chi^2 = 3$ with respect to that track. The parameter controlling the steepness of the curve is χ^2_{cut} . The smaller its value, the faster the weight falls to zero with increasing χ^2 . It is not a strict cut-off like in the case of the Tukey weight introduced earlier but can be understood as the χ^2 value at which the weight is 0.5 in the case of no other competing vertices. If there are no other consistent vertices nearby, this weight function becomes very similar to the Tukey weight distribution. To keep the fit of a certain vertex candidate independent of the other vertices in the event, the weight is always calculated using the initial position of the other vertices.

Each vertex fit can be done in parallel, where within each the χ^2 -sum over all tracks and derived quantities can also be computed in parallel. To speed up the calculations and to prevent completely unrelated tracks, whose weights would be in any case almost zero, from contributing to the vertex fit, only tracks within a certain χ^2_i with respect to the vertex candidate are considered. To speed up the matrix calculations they are explicitly written out exploiting the fact that many elements of the matrices are zero.

To reduce the amount of duplicate PVs, where one 'true' PV is reconstructed as two separate PVs, an additional step after the fitting part searches for PVs within

close proximity of each other and rejects the one with fewer associated tracks. The amount of 'fake' reconstructed PVs which do not correspond to a simulated PV are reduced by a radial requirement around the beamline.

IV.3.3 Performance evaluation

The physics performance is evaluated using simulated samples of the LHCb Upgrade detector under the expected Run 3 conditions. Since the quantities of the true simulated PVs is known in simulation, they can be matched to the reconstructed ones and the performance parameters can be determined. A true PV is matched to a reconstructed PV by checking if the reconstructed z -position is consistent with the true value taking into account the uncertainty on the reconstructed value. If a simulated PV could be matched to more than one reconstructed PV only the closest PV is considered. An alternative matching approach uses the tracks associated with the reconstructed PVs. Using that tracks can be matched to its corresponding simulated particle and each simulated particle to its PV of origin it is possible to match reconstructed PVs to the true simulated PVs by comparing these lists of tracks and particles. However, when two reconstructed PVs are close together it is not straightforward to associate a track to one PV over the other and this approach does not perform well.

Efficiency

The PV reconstruction efficiency ϵ is defined as the number of matched reconstructed PVs divided by the total number of reconstructible PVs

$$\epsilon = \frac{N_{matched}^{rec}}{N_{reconstructible}^{simulated}}. \quad (104)$$

The typical criterion for a PV to be considered as reconstructible is that it produces at least four reconstructible VELO tracks. Since in the trigger selection displaced tracks are identified by their minimum IP with respect to any PV, this can only work properly if the PV reconstruction efficiency is as high as possible. Otherwise, tracks produced by a PV could be misclassified as displaced tracks if their PV was not reconstructed.

Fake rate

The fake rate f is the number of reconstructed PVs, which are not matched to a true simulated PV, divided by the total number of reconstructed PVs

$$f = \frac{N_{total}^{rec} - N_{matched}^{rec}}{N_{total}^{rec}} = \frac{N_{notmatched}^{rec}}{N_{total}^{rec}}. \quad (105)$$

A large fake rate could change the minimum IP of a track since it could accidentally point to a fake PV. Furthermore, SVs from long-lived particles could be misidentified as PVs, which would reduce the efficiency of selecting that decay since now its decay products seem like prompt tracks.

Position resolution

The resolution of the (x, y, z) -position of the vertex is defined as the standard deviation of the distribution of the differences between the positions of the true simulated PV and its matched reconstructed PV in one of the variables. The resolution impacts the precision of the impact parameter and its significance, which are the quantities used to select displaced particles.

Physics performance

The PV reconstruction efficiency is shown in Figure 102 as a function of the number of associated tracks and z -position of the simulated PV. The PV reconstruction efficiency is expected to be lower for PVs with a smaller number of associated tracks, for example if a peak resulting from such PVs is not significant enough to be identified by the peak finding procedure. It can also be seen that the PV reconstruction efficiency is slightly reduced at the centre. This can be explained by the observation that at the centre the PVs are more densely populated and are more likely to spatially overlap. Such PVs are hard to distinguish and they may be reconstructed as a single PV instead of two distinct ones. The figures also show the performance of the initial implementation of the algorithm on CPUs. It can be seen that the *Allen* implementation achieves a similar or even slightly better reconstruction efficiency performance.

The resolution of the x - and z -positions of the reconstructed PVs is shown in Figure 103. It can be seen that the resolution strongly depends on the number of tracks, as more tracks contributing to a PV allow it to be reconstructed with larger precision. The resolution is degraded in the region of about $-150 < z < -100$ mm. This can be understood by the spacing of the VELO layers chosen for the Upgrade,

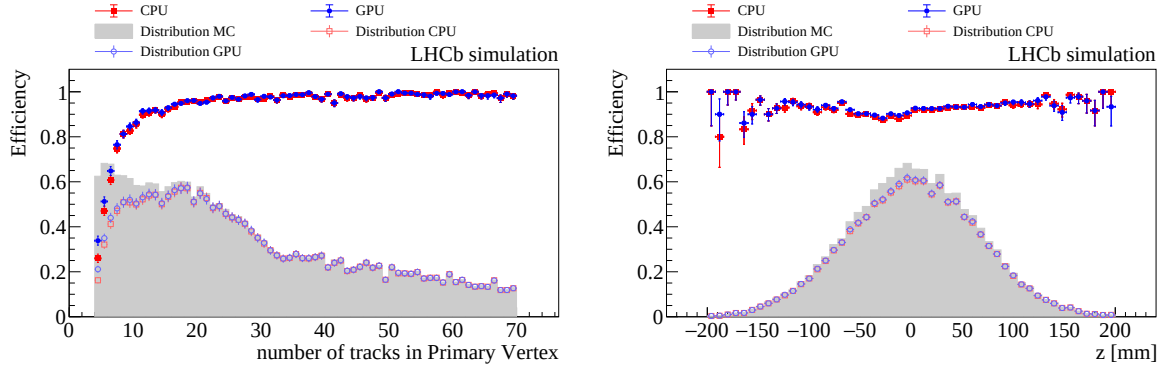


Figure 102: ● Efficiency to reconstruct PVs as function of the tracks associated to the PV (left) and true z -position of the simulated PV (right) obtained from simulation of Run 3 conditions. The red points are obtained from the CPU implementation of the PV reconstruction algorithm, the blue points from the ported and changed GPU implementation. The distribution of reconstructible simulated PVs is shown as grey histogram. The hollow red and blue points show the distribution of reconstructed PVs in the CPU and GPU cases, respectively.

as shown in Figure 104. This region falls within a gap between the detector layers, where the distance to the closest layer is larger compared to other regions in z where PVs are found. This means that the distance the track state has to be extrapolated to the PV position is larger resulting in larger associated uncertainties. This leads to a reconstruction of the PV position with a larger uncertainty. While this degradation in resolution is quite significant it should be noted that in this z -region only a small fraction of the total number of PVs are expected so the overall effect on the selection of displaced tracks is expected to be small.

IV.3.4 Timing performance

The relative timing of the different parts of the GPU implementation of the PV reconstruction algorithm is broken down in Figure 105. A large part is taken by the peak finding. This is because other parts of the algorithm, especially the fitting, have received more attention and have been optimised for efficient execution on GPUs². The peak finding procedure on the other hand still follows the sequential logic of the initial CPU implementation. It is clear that this part should be looked at next for speeding up the PV reconstruction algorithm. This demonstrates the importance of devising implementations optimised for GPU architectures. In a previous version of the algorithm the fitting part, before it was optimised, took almost twice as long as the peak finding, as shown in Figure 106.

Within the whole GPU HLT1 sequence the PV reconstruction algorithm takes

²a lot of the credit for the improvements in speed, especially for the fitting part, goes to collaborators, not the author

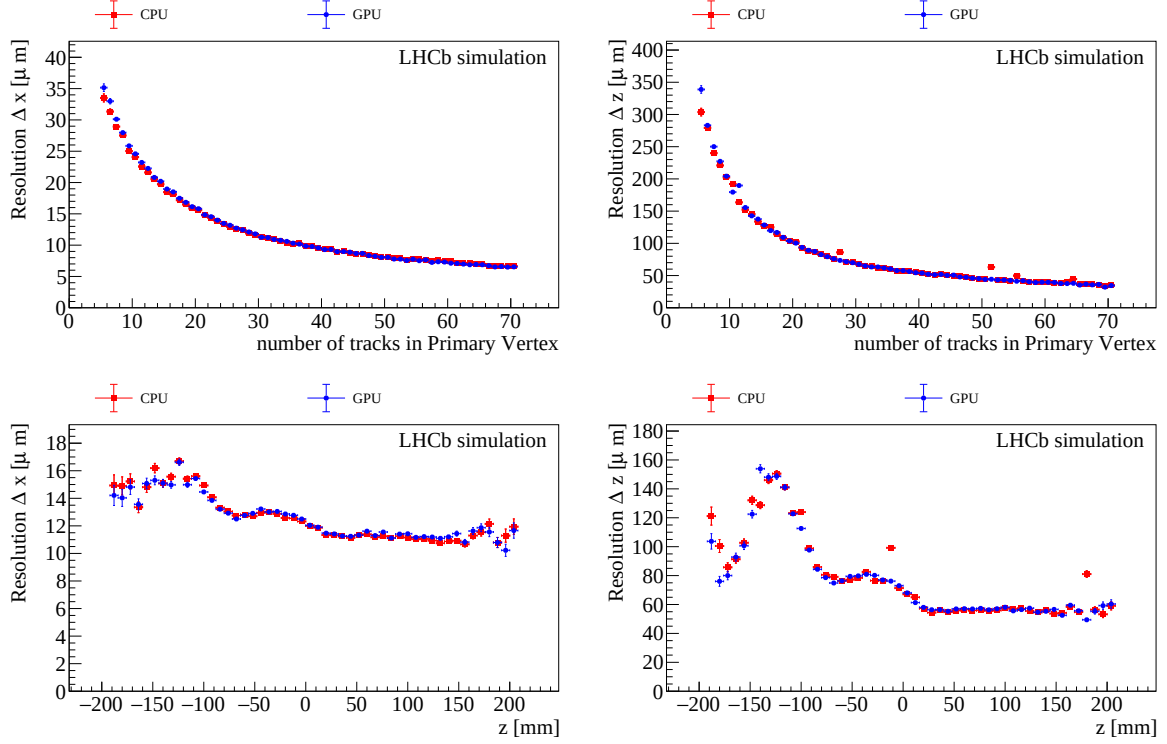


Figure 103: ● Resolution of the x -position (left) and z -position (right) of the reconstructed PVs as function of the number of tracks in the PV (top) and true z -position of the simulated PV (right) obtained from simulation of Run 3 conditions. The red points are obtained from the CPU implementation of the PV reconstruction algorithm, the blue points from the ported and changed GPU implementation.

about 13% of the total, keeping it well within the timing budget. It is seen that the whole HLT1 sequence achieves a throughput of over 100 kHz on a consumer-grade GPU, which means that the HLT1 can be run using less than 300 GPUs, which easily fit within the 340 available slots in the event-building farm.

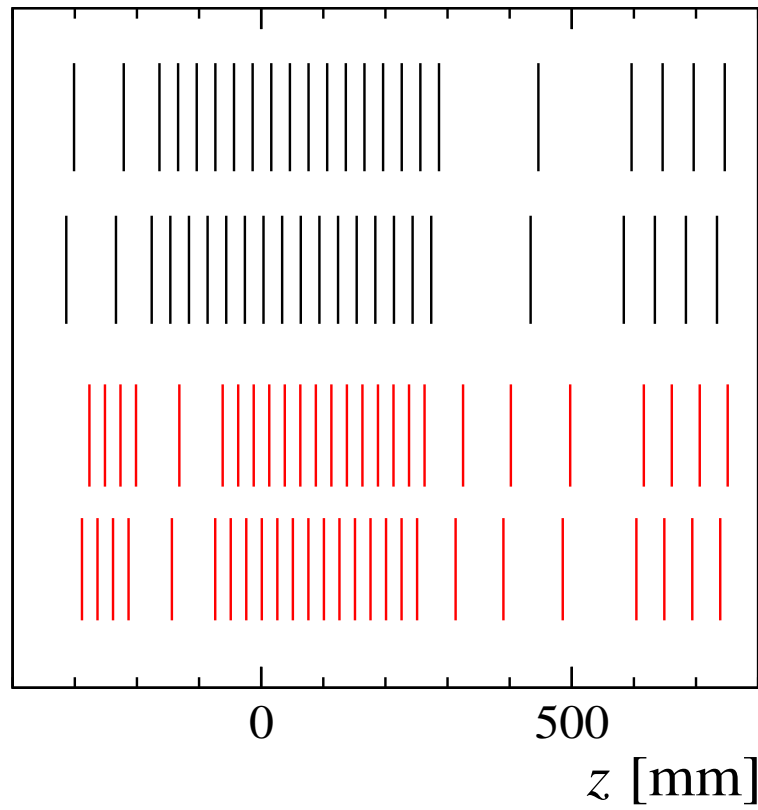


Figure 104: ■ Comparison of current (top, black) and Upgrade (bottom, red) VELO layer z -positions [95].

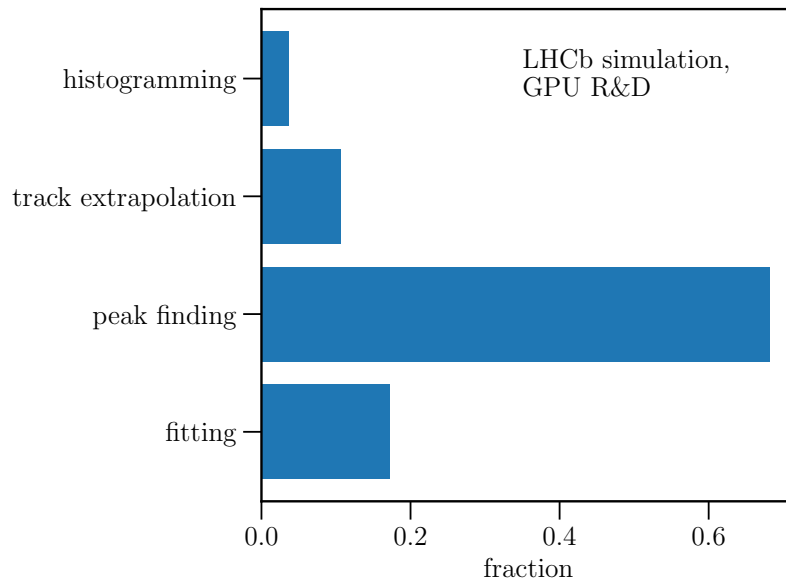


Figure 105: ● Timing fractions of the different step of the PV reconstruction of the GPU implementation. The fractions were determined with the version of the software in September 2020.

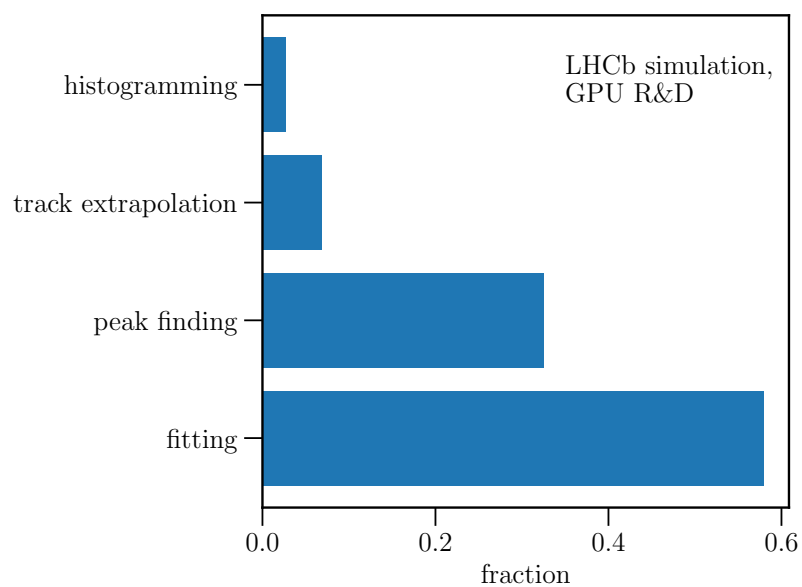


Figure 106: ■● Timing fractions of the different step of the PV reconstruction of the GPU implementation. The fractions were determined with the version of the software in May 2020. [99].

IV.4 Compilation on CPU

It is a desirable feature to be able to compile the *Allen* software also on CPU architectures. Not only is this a requirement for simulating events processed by *Allen* in the trigger and to re-run the HLT1 steps in HLT2 if necessary, it is also useful from a practical point of view if one wants to work on the *Allen* software, but does not have a GPU available for testing. For this purpose, it has to be demonstrated that the compilation for GPU or CPU based architectures does not significantly change the result. Potential differences could arise for example due to different floating-point number rounding. Also the compilers processing the code are different. As they internally optimise the code instructions, it can lead to different results. This is not unique when comparing compilations for CPU and GPU but can also happen when compiling the same code on different CPU architectures³. While this might not be significant for intermediate quantities as e.g. the covariance matrix element of a certain track state, the differences might cumulate in the final objects used in the selection, for example the vertices reconstructed using multiple tracks as input, slightly changing the result when compiling and executing the *Allen* application on GPUs and CPUs. It should be stressed that this is not a comparison of the implementation of similar algorithms on CPU or GPU architectures, as was the case in the previous section for the PV reconstruction procedure, but the compilation of the same code on different hardware. In an ideal world, this would result in exactly the same results, but this is rarely the case for the previously stated reasons.

Potential differences in various quantities are checked for using simulated $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decays generated with Upgrade conditions. As an example that the final quantities show deviations at the per-mile level Figures 107 and 108 show the differences in the track and PV reconstruction efficiencies when compiling *Allen* on CPU and GPU architectures. The observed differences are small and so it can be guaranteed that executing the HLT1 sequence in simulation processes running on CPU reproduce well what happens when *Allen* runs on GPUs during Run 3 data-taking.

As a final comparison, the selection rates of the main HLT1 trigger lines are compared in Table 25 showing again consistent results between the compilations for the two architectures. Performing these cross-checks also turned out to be useful for spotting bugs in the code or detecting non-deterministic behaviour when executing on GPUs caused for example by race conditions, where different threads simultaneously read and write to the same memory location.

³for example X86 and ARM CPUs

Table 25: ■● Comparison of the selection rates between the compilation of *Allen* for GPU and x86 architectures. Differences are at the permille level or smaller [94].

Trigger line	GPU compilation rate (kHz)	CPU compilation rate (kHz)	$\Delta(\text{GPU-CPU})$ (kHz)
One track	4086	4074	12
Two track	13410	13404	6
Single muon	516	516	0
Displaced dimuon	8808	8808	0
High mass dimuon	6924	6930	-6
Dimuon Soft	480	480	0
Inclusive	17232	17238	-6

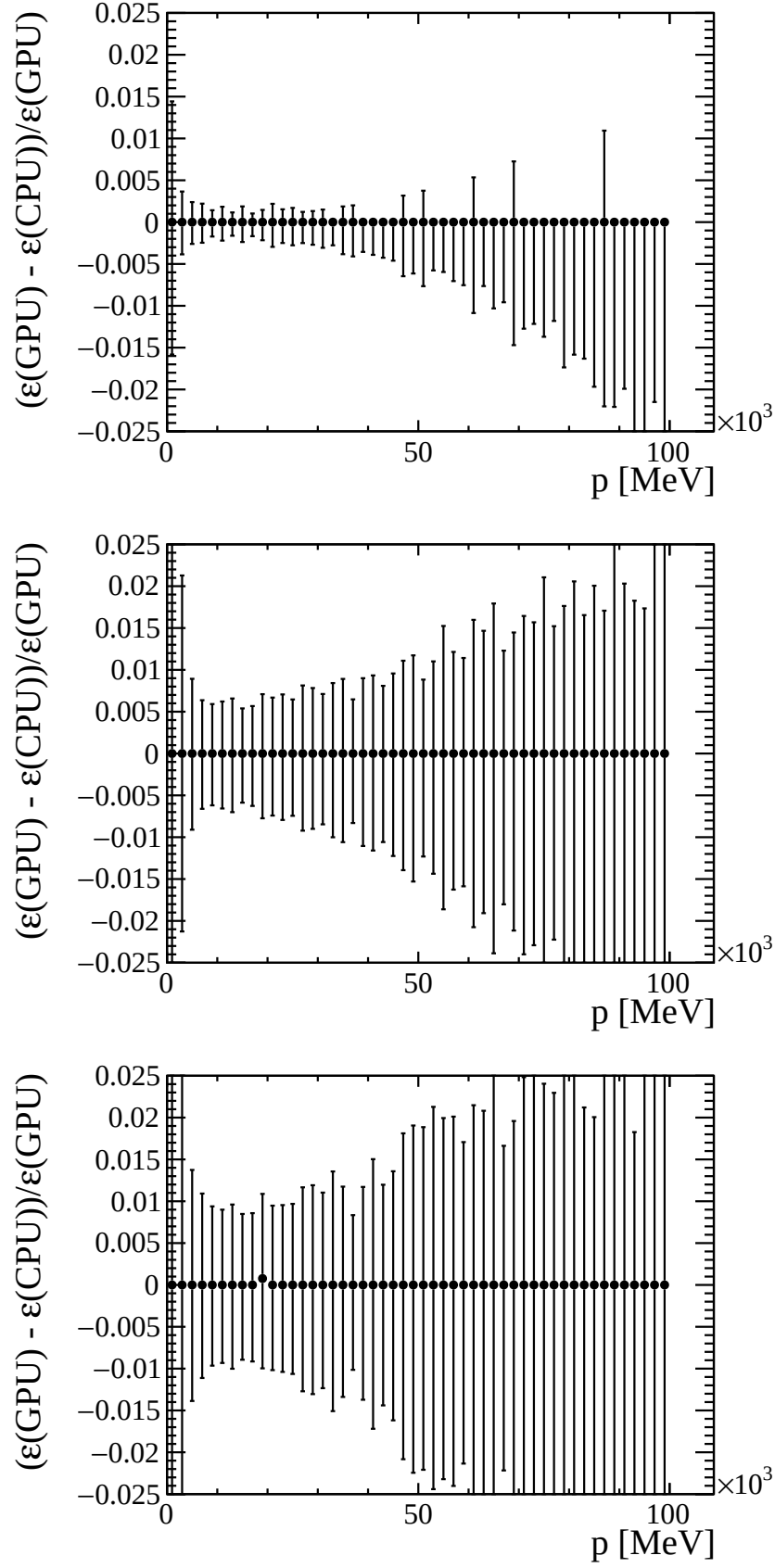


Figure 107: ■ ● Relative difference in reconstruction efficiency between CPU and GPU compilations of *Allen* for the (top) Velo, (middle) VeloUT and (bottom) Forward reconstruction sequences as function of momentum. The efficiencies are shown for long tracks from B hadrons in the range $2 < \eta < 5$ excluding electron tracks [94].

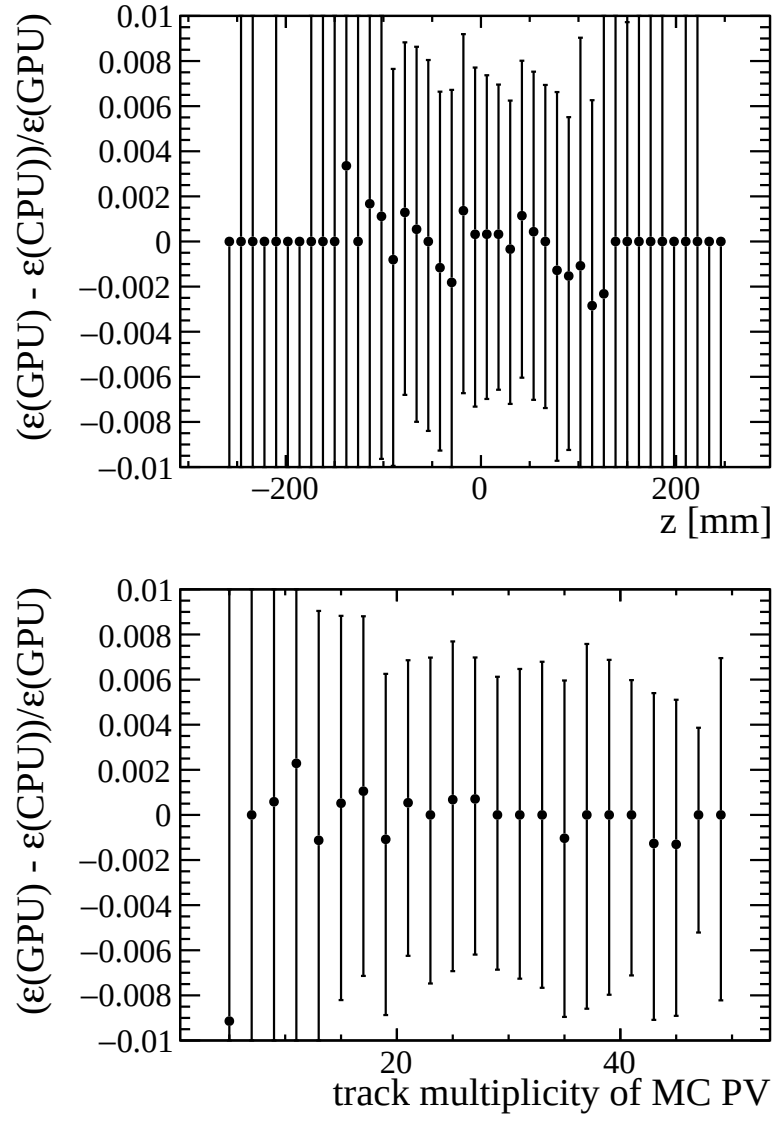


Figure 108: ■● Relative difference in PV reconstruction efficiency between CPU and GPU compilations of Allen [94].

Chapter V

Conclusion

“Verrückt, aber weise.
Gefangen, aber frei.
Physiker, aber unschuldig.”

Friedrich Dürrenmatt, *Die Physiker*

Experimental tests of lepton universality have in recent times received a lot of attention. In many cases, the measured observables differ significantly from the prediction by the SM. While no single measurement showed the magical 5σ deviation yet, considering all the experimental evidence points to the violation of lepton universality. This would be a clear observation of an effect not described by the SM and would motivate the introduction of BSM particles. Examining the results using EFT approaches the introduction of BSM contributions can yield in a better description of the data. This thesis documents a test of lepton universality using semileptonic $\bar{B}^0$ decays by measuring the relative branching fraction to a final state containing an electron or a muon. This is the first time this has been measured with pp collision data and is complementary to measurements performed at e^+e^- colliders, where no sign of lepton universality violation has been observed in this decay. The blinded value is determined to be

$$\begin{aligned} R(D^{*+})_{light} &= \frac{\mathcal{B}(\bar{B}^0 \rightarrow D^{*+} e^- \bar{\nu}_e)}{\mathcal{B}(\bar{B}^0 \rightarrow D^{*+} \mu^- \bar{\nu}_\mu)} \\ &= X \times (2.937 \pm 0.014(stat.) \pm 0.063(syst.)), \end{aligned} \tag{106}$$

where X is the blinding factor. The total uncertainty is at a comparable level compared to the Belle measurement, which demonstrates that precision measurement can be performed at hadron colliders using semileptonic decays to a single electron.

This measurement can also be understood as a first step towards a measurement

at LHCb of the relative branching fraction

$$\frac{\mathcal{B}(B^0 \rightarrow D^{*-} \tau^+ \nu_\tau)}{\mathcal{B}(B^0 \rightarrow D^{*-} e^+ \nu_\mu)}, \quad (107)$$

where the tau lepton decays to an electron and neutrinos. This would be complementary and a cross-check to the already performed measurement of the ratio with muons instead of electrons. As the expected systematic uncertainties between using muons or electrons are mostly independent this would increase the sensitivity of the $R(D^*)$ measurement.

To estimate the increase in sensitivity of the $R(D^*)$ measurement by LHCb when using the electron final states in addition to the muon states, the following is considered. From the data and simulation samples used in this analysis, it is seen that the muon signal yield is about 10 – 20% larger than the electron signal yield after the selection is applied. As only the L0I trigger category is used, which presents about 30% of the total muon and 45% of the total electron yields, the overall ratio of muon to electron yield is about 1.8. This means that it would be expected that a measurement of $R(D^*)$ using electron final states would have a $\sqrt{1.8}$ larger statistical uncertainty compared to a measurement using muon final states. As the main systematic uncertainties should be independent between the two measurements and as it should be possible to reach a similar systematic uncertainty using electrons instead of muons, the systematic uncertainty would decrease by a factor $\sqrt{2}$ when adding the electron measurement. With these assumptions, the overall uncertainty of the $R(D^*)$ measurement would be reduced by about 25% when adding the measurement using the electron final states to the muon final states if the muonic measurement has a statistical and systematic uncertainty of similar size, as was the case for the $R(D^*)$ measurement using Run 1 data. In the end, all final states, with tauons decaying to electrons or muons, could be included in a combined analysis.

The *Allen* project implements the full HLT1 sequence used in the Upgrade LHCb trigger during Run 3 on GPUs. With the removal of the hardware-level trigger, the input rate to the HLT1 is 30 MHz. To reach this rate, given the external constraints, the HLT1 sequence is implemented on GPUs optimizing the usage of their capabilities for parallel execution. During this thesis, the PV reconstruction step, necessary to identify displaced signatures, was ported from the initial CPU implementation to GPUs. For this, the existing code not only had to be rewritten, but major changes to the PV fitting were made. The change allows tracks to contribute to several fits, if they are consistent with multiple PV candidates, by defining an appropriate weight. The GPU implementation performs equally well

as the initial CPU algorithm in terms of physics performance and is fast enough so that the whole sequence can be executed within the timing constraints.

Appendix A

Meaning of selection requirements

m or **M** Invariant mass of the candidate.

χ^2_{DOCA} Distance of closest approach significance.

χ^2_{vtx}/DOF χ^2 of the vertex fit divided by the number of degrees of freedom.

DIRA Direction angle defined as cosine between the momentum vector of the particle and the direction vector between its decay and origin vertices.

χ^2_{VD} Significance of the vertex distance with respect to the associated PV.

χ^2_{tr}/DOF χ^2 of track fit divided by the number of degrees of freedom.

IP impact parameter of track with respect to a (primary) vertex

χ^2_{IP} impact parameter significance

min. χ^2_{IP} Minimum χ^2_{IP} with respect to all PVs in the event

FAKEPROB Probability that the track is a fake track.

Appendix B

Control channel selection

Table 26: Stripping selection in `Bu2LLK_mmLine`, for the control channel $B^0 \rightarrow J/\psi K^{*0}$, $J/\psi \rightarrow \mu^+ \mu^-$.

Particle	Selection
J/ψ	$M < 5500 \text{ MeV}$ $\chi^2_{\text{vtx}}/\text{ndf} < 9$ $\chi^2_{VD} > 16$ $\chi^2_{\text{IP}} > 0$ $\chi^2_{\text{DOCA}} < 30$
μ	$PT > 300 \text{ MeV}$ $\chi^2_{\text{IP}} > 9$ ISMUON HASMUON
K^{*0}	$M < 2600 \text{ MeV}$ $\chi^2_{\text{vtx}} < 25$ $PT > 400 \text{ MeV}$ $\chi^2_{\text{doca}} < 30$ $APT > 500 \text{ MeV}$ $ AM - m(K^{*0}) < 300 \text{ MeV}$ $(\chi^2_{\text{IP}} > 9 \text{ or } \chi^2_{\text{IP}}(\text{both daughters}) > 9)$
K	$PIDK > -5$
B^0	$ M - m(B^0) < 1.5 \text{ GeV}$ $\chi^2_{\text{vtx}}/\text{ndf} < 9$ $\chi^2_{\text{IP}} < 25.$ $\chi^2_{VD} > 100$ DIRA > 0.9995

Table 27: Stripping selection in **Bu2LLK_eeLine**, for the control channel $B^0 \rightarrow J/\psi K^{*0}$, $J/\psi \rightarrow e^+ e^-$.

Particle	Selection
J/ψ	$M < 5500 \text{ MeV}$ $\chi^2_{\text{vtx}}/\text{ndf} < 9$ $\chi^2_{VD} > 16$ $\chi^2_{\text{IP}} > 0$ $\chi^2_{\text{DOCA}} < 30$
e	$PT > 500 \text{ MeV}/c$ $\chi^2_{\text{IP}} > 9$ $PIDe > 0$
K^{*0}	$M < 2600 \text{ MeV}$ $\chi^2_{\text{vtx}} < 25$ $p_T > 400 \text{ MeV}$ $\chi^2_{\text{doqa}} < 30$ $\text{APT} > 500 \text{ MeV}$ $ AM - m(K^{*0}) < 300 \text{ MeV}$ $(\chi^2_{\text{IP}} > 9 \text{ or } \chi^2_{\text{IP}}(\text{both daughters}) > 9)$
K	$PIDK > -5$
B^0	$ M - m(B^0) < 1.5 \text{ GeV}$ $\chi^2_{\text{vtx}}/\text{ndf} < 9$ $\chi^2_{\text{IP}} < 25.$ $\chi^2_{VD} > 100$ $\text{DIRA} > 0.9995$

Table 28: First part of tuple-level selections for the control channel $B^0 \rightarrow J/\psi K^{*0}$, $J/\psi \rightarrow \mu^+ \mu^-$.

Particle	Selection
J/ψ	$3048.9 \text{ MeV} < M < 3139.9 \text{ MeV}$
K^{*0}	$792. \text{ MeV} < M < 992. \text{ MeV}$
μ	$PID_{mu} > 2.0$ $p > 2600. \text{ MeV}$ $p < 200000. \text{ MeV}$ $2.5 < \eta < 3.5$ $p_T > 150 \text{ MeV}$ $p_T < 50000 \text{ MeV}$
K	$PID_K - PID_p > -10$ $\text{ProbNNk} > 0.2$ $p > 2600. \text{ MeV}$ $p < 200000. \text{ MeV}$ $1.5 < \eta < 5.$ $\chi_{IP}^2 > 4$
π	$PID_K < 10$ $\text{ProbNNpi} > 0.2$ $p > 2600. \text{ MeV}$ $p < 20000. \text{ MeV}$ $1.5 < \eta < 5.$ $\chi_{IP}^2 > 4.$

Table 29: First part of tuple-level selections for the control channel $B^0 \rightarrow J/\psi K^{*0}$, $J/\psi \rightarrow e^+ e^-$.

Particle	Selection
J/ψ	$3048.9 \text{ MeV} < M < 3139.9 \text{ MeV}$
K^{*0}	$792. \text{ MeV} < M < 992. \text{ MeV}$
e	$PID_e > 2.0$ $p > 2600. \text{ MeV}$ $p < 200000. \text{ MeV}$ $2.5 < \eta < 3.5$ $p_T > 150 \text{ MeV}$ $p_T < 50000 \text{ MeV}$
K	$PID_K - PID_p > -10$ $\text{ProbNNk} > 0.2$ $p > 2600. \text{ MeV}$ $p < 200000. \text{ MeV}$ $1.5 < \eta < 5.$ $\chi_{IP}^2 > 4.$
π	$PID_K < 10$ $\text{ProbNNpi} > 0.2$ $p > 2600. \text{ MeV}$ $p < 200000. \text{ MeV}$ $1.5 < \eta < 5.$ $\chi_{IP}^2 > 4.$

Appendix C

Matrix elements of misidentification matrix used for lepton misID background

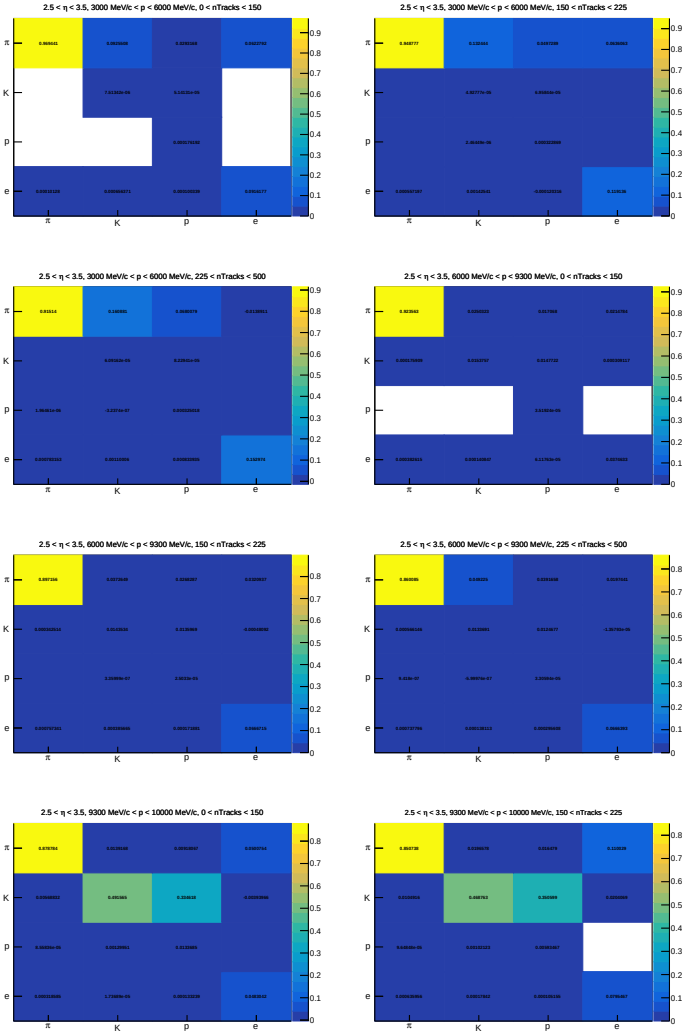


Figure 109: Matrix of (mis-)identification probability of different particle species in the bins of p and η bin using the muon signal mode PID requirement for the 2016 sample.

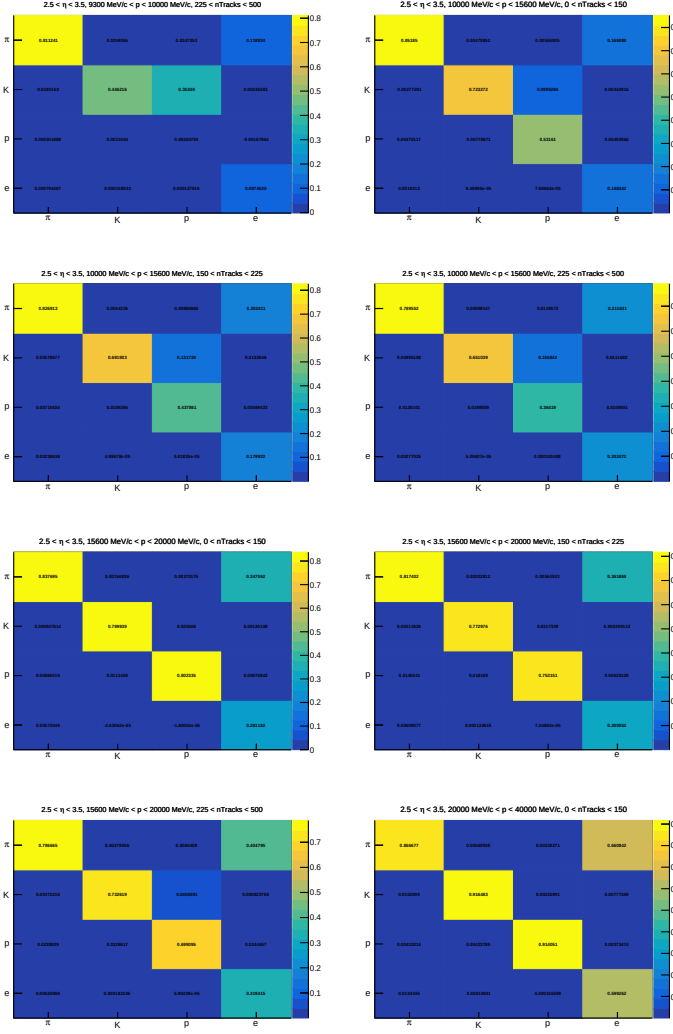


Figure 110: Matrix of (mis-)identification probability of different particle species in the bins of p and η bin using the muon signal mode PID requirement for the 2016 sample.

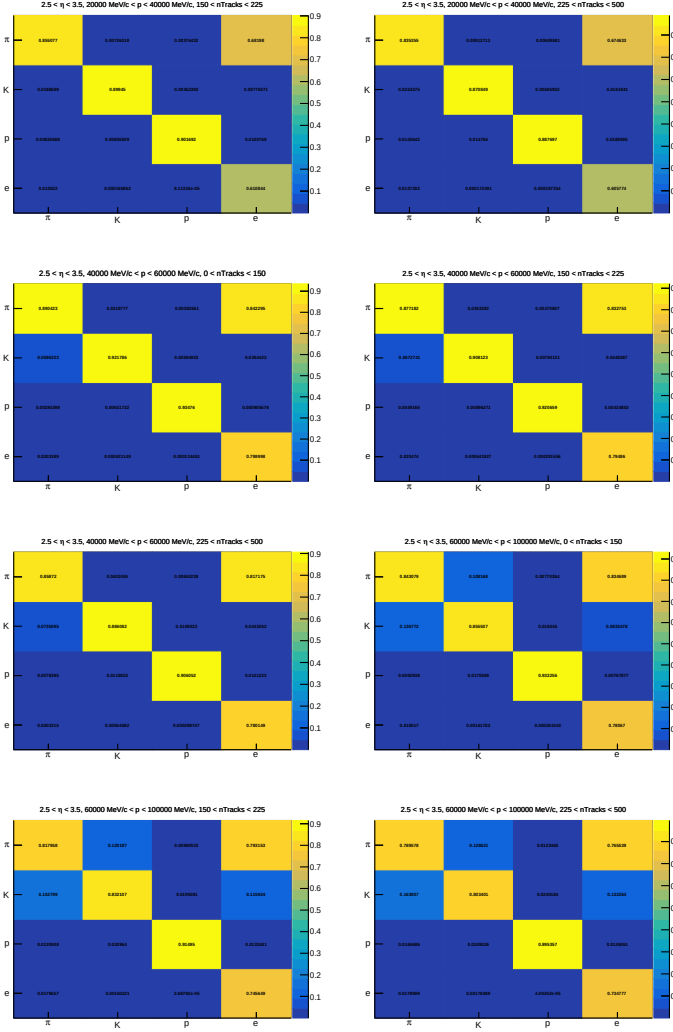


Figure 111: Matrix of (mis-)identification probability of different particle species in the bins of p and η bin using the muon signal mode PID requirement for the 2016 sample.

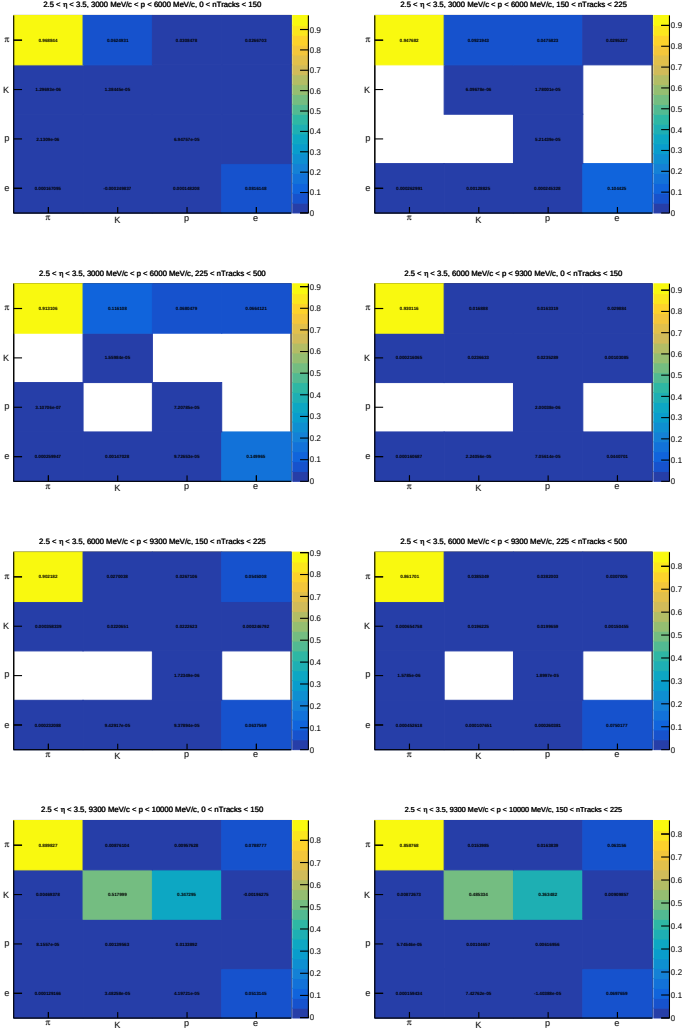


Figure 112: Matrix of (mis-)identification probability of different particle species in the bins of p and η bin using the muon signal mode PID requirement for the 2017 sample.

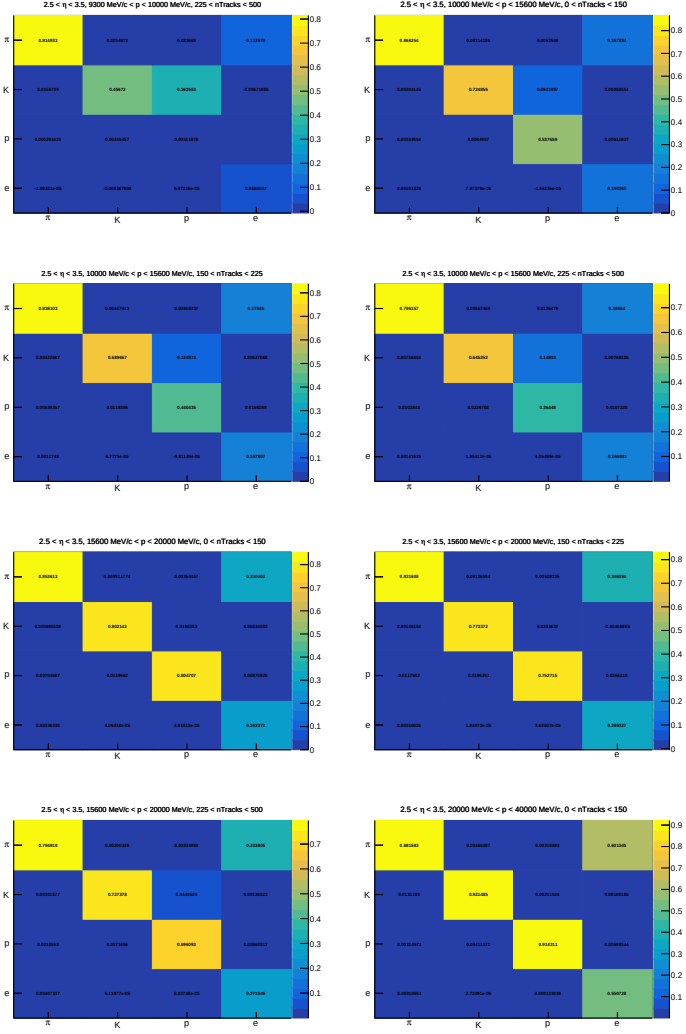


Figure 113: Matrix of (mis-)identification probability of different particle species in the bins of p and η bin using the muon signal mode PID requirement for the 2017 sample.

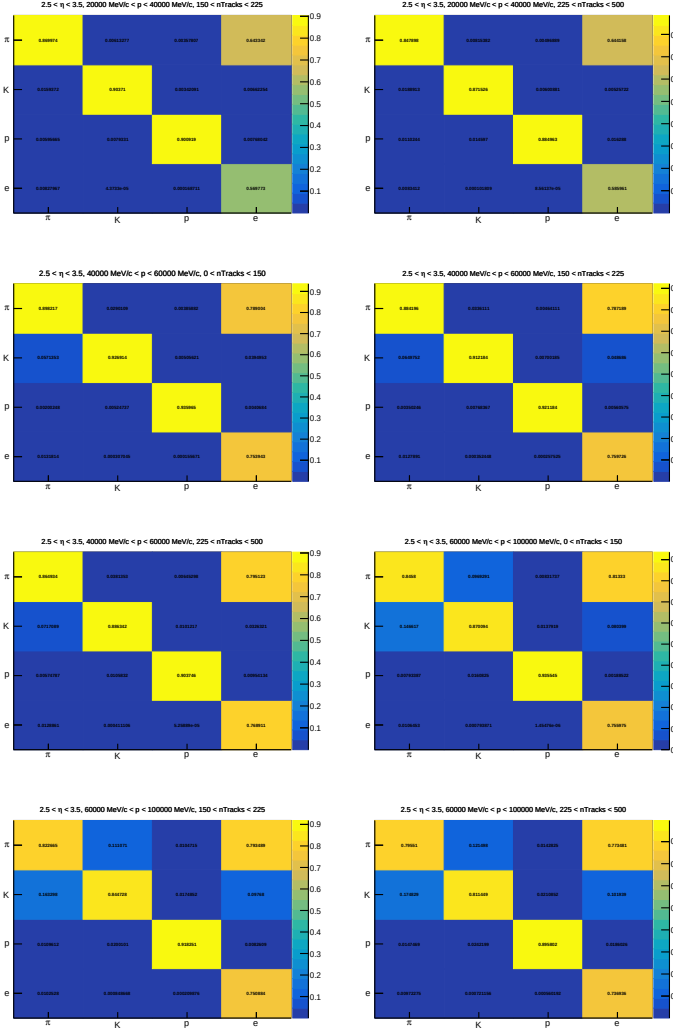


Figure 114: Matrix of (mis-)identification probability of different particle species in the bins of p and η bin using the muon signal mode PID requirement for the 2017 sample.

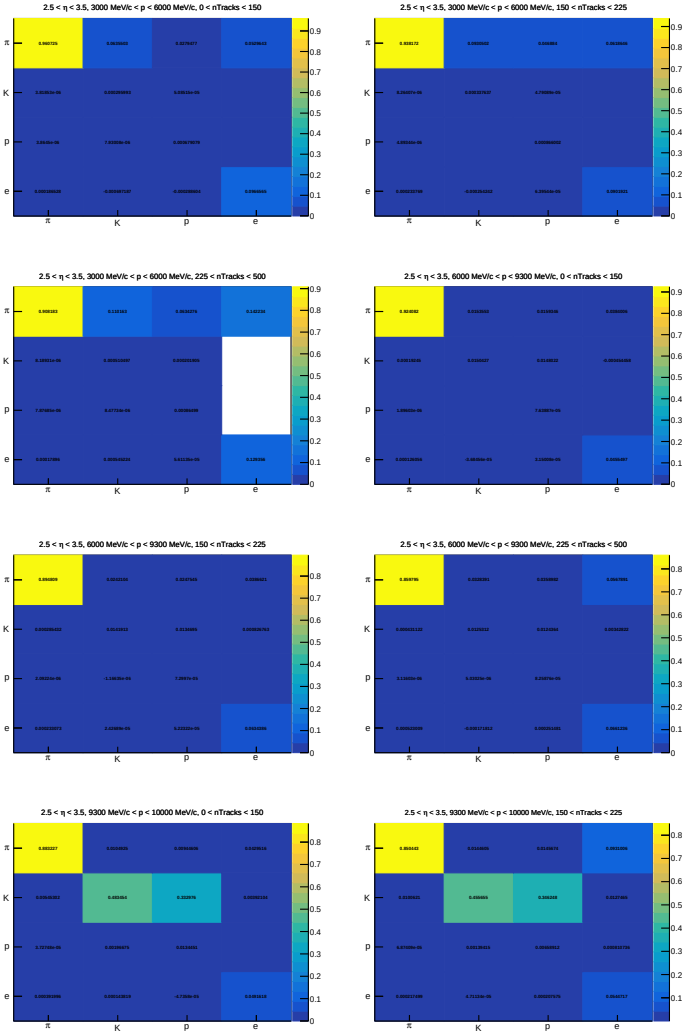


Figure 115: Matrix of (mis-)identification probability of different particle species in the bins of p and η bin using the muon signal mode PID requirement for the 2018 sample.

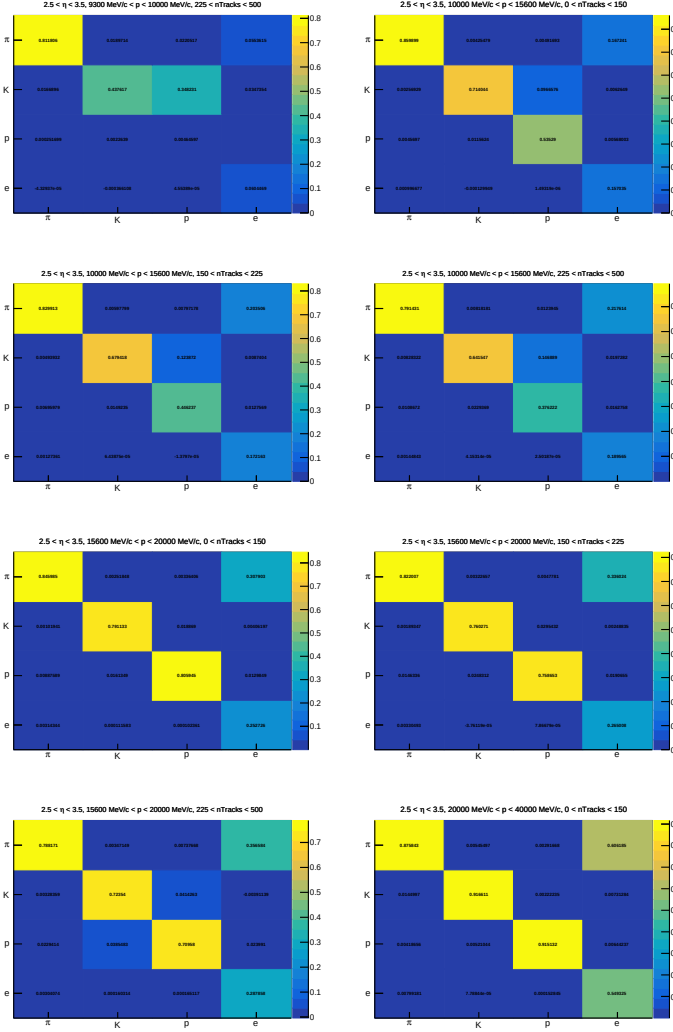


Figure 116: Matrix of (mis-)identification probability of different particle species in the bins of p and η bin using the muon signal mode PID requirement for the 2018 sample.

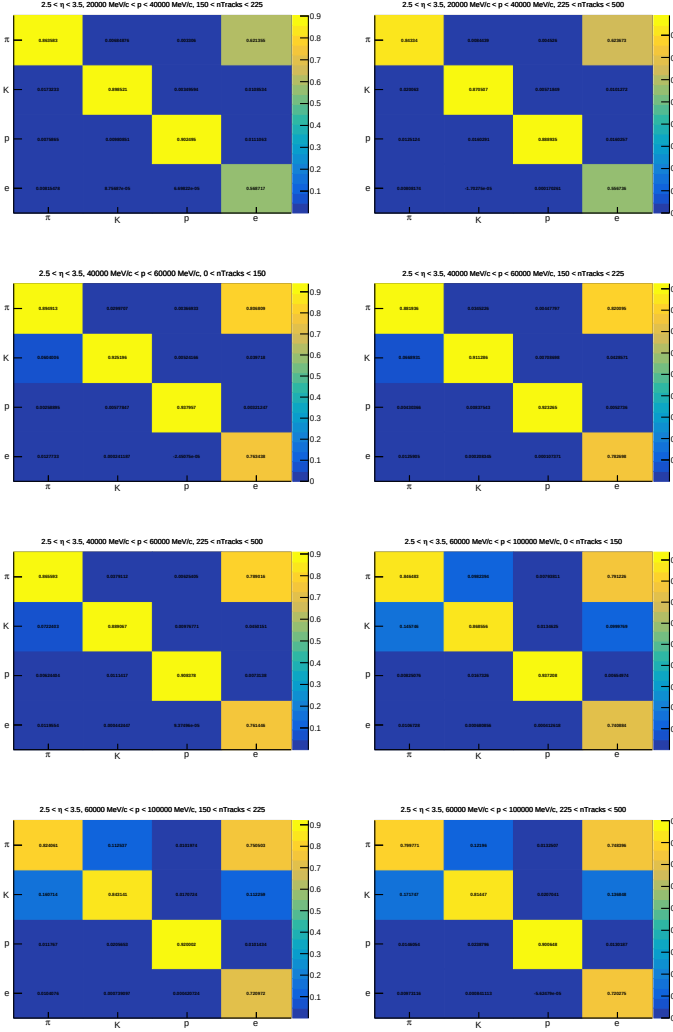


Figure 117: Matrix of (mis-)identification probability of different particle species in the bins of p and η bin using the muon signal mode PID requirement for the 2018 sample.

Appendix D

Background studies

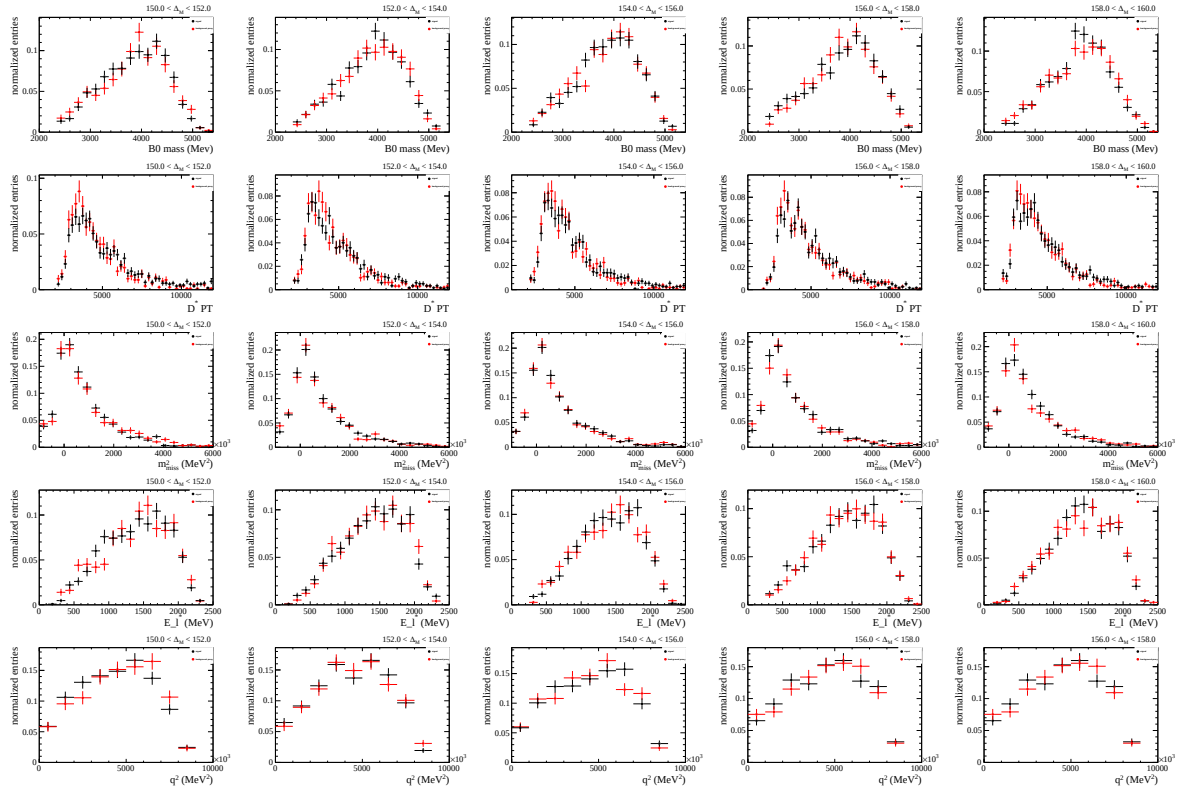


Figure 118: ● Comparison of B invariant mass, p_T of the D^* and the fit variables between real and fake D^* templates in bins of the $\Delta(M_{D^{*+}} - M_{D^0})$ upper sideband from left to right for the muon final state for 2017 data.

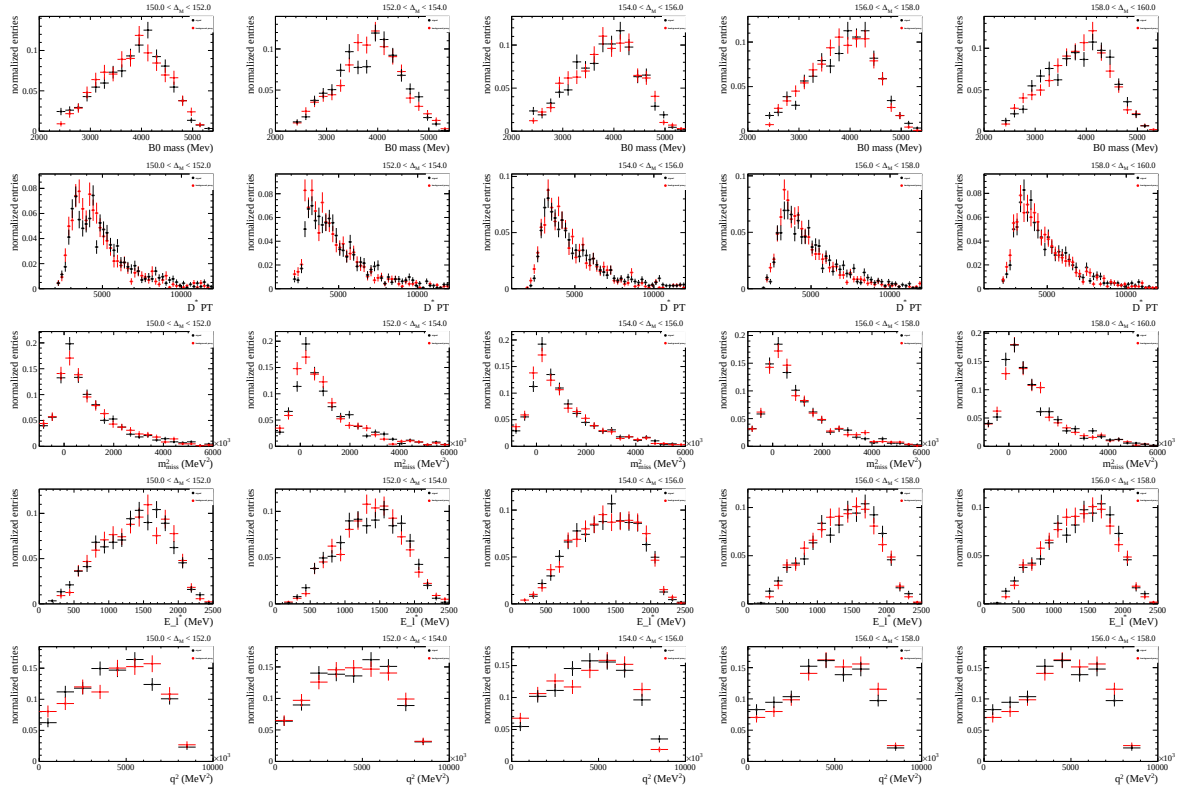


Figure 119: ● Comparison of B invariant mass, p_T of the D^* and the fit variables between real and fake D^* templates in bins of the $\Delta(M_{D^{*+}} - M_{D^0})$ upper sideband from left to right for the electron final state for 2017 data.

Appendix E

TISTOS efficiencies on data and simulation ▲

E.1 L0 control mode

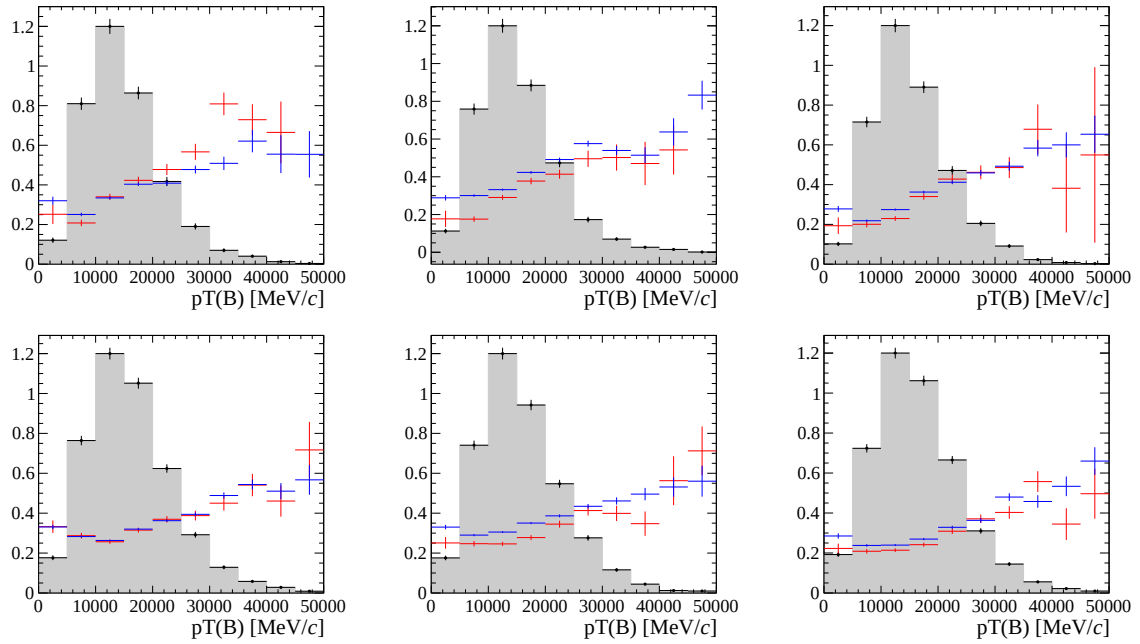


Figure 120: ▲ Values of $\epsilon_{L0,data}^{tistos}$ (red) and $\epsilon_{L0,MC}^{tistos}$ (blue) as a function of $p_T(B^0)$. The background-subtracted $p_T(B^0)$ distribution of data with $L0TOS$ requirement is also overlaid. The top row shows the dielectron control mode and the bottom row the dimuon control mode with the 2016, 2017 and 2018 samples from left to right.

E.2 HLT control mode

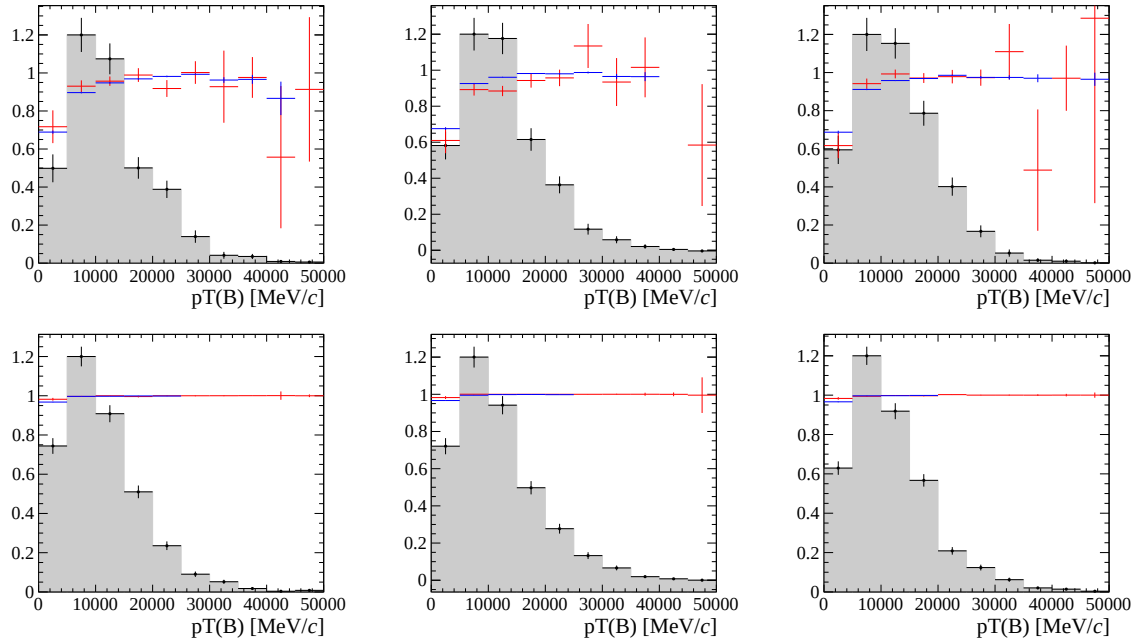


Figure 121: ▲ Values of $\epsilon_{HLT,con,data}^{tistos}$ (red) and $\epsilon_{HLT,con,MC}^{tistos}$ (blue) as a function of $p_T(B^0)$. The background-subtracted $p_T(B^0)$ distribution of data with *HLTTIS* requirement is also overlaid. The top row shows the dielectron control mode and the bottom row the dimuon control mode with the 2016, 2017 and 2018 samples from left to right.

E.3 HLT signal mode

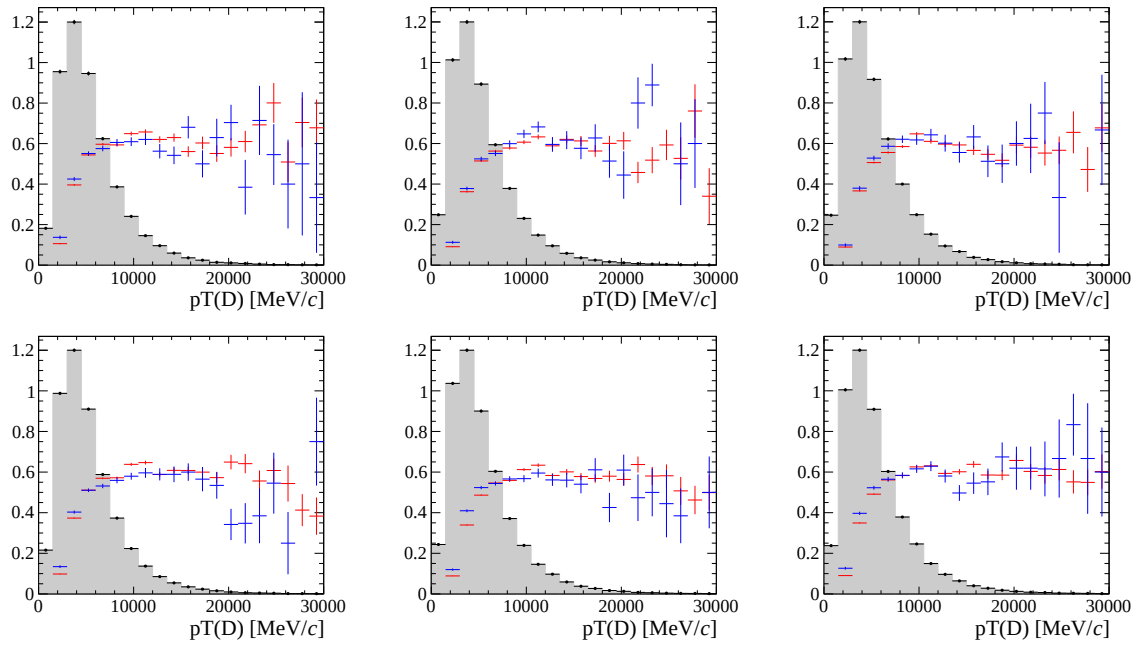


Figure 122: $\epsilon_{HLT,sig,data}^{tistos}$ (red) and $\epsilon_{HLT,sig,MC}^{tistos}$ (blue) as a function of $p_T(D^0)$. The background-subtracted $p_T(D^0)$ distribution of data with *HLTTIS* requirement is also overlaid. The top row shows the dielectron control mode and the bottom row the dimuon control mode with the 2016, 2017 and 2018 samples from left to right.

Appendix F

Template projections

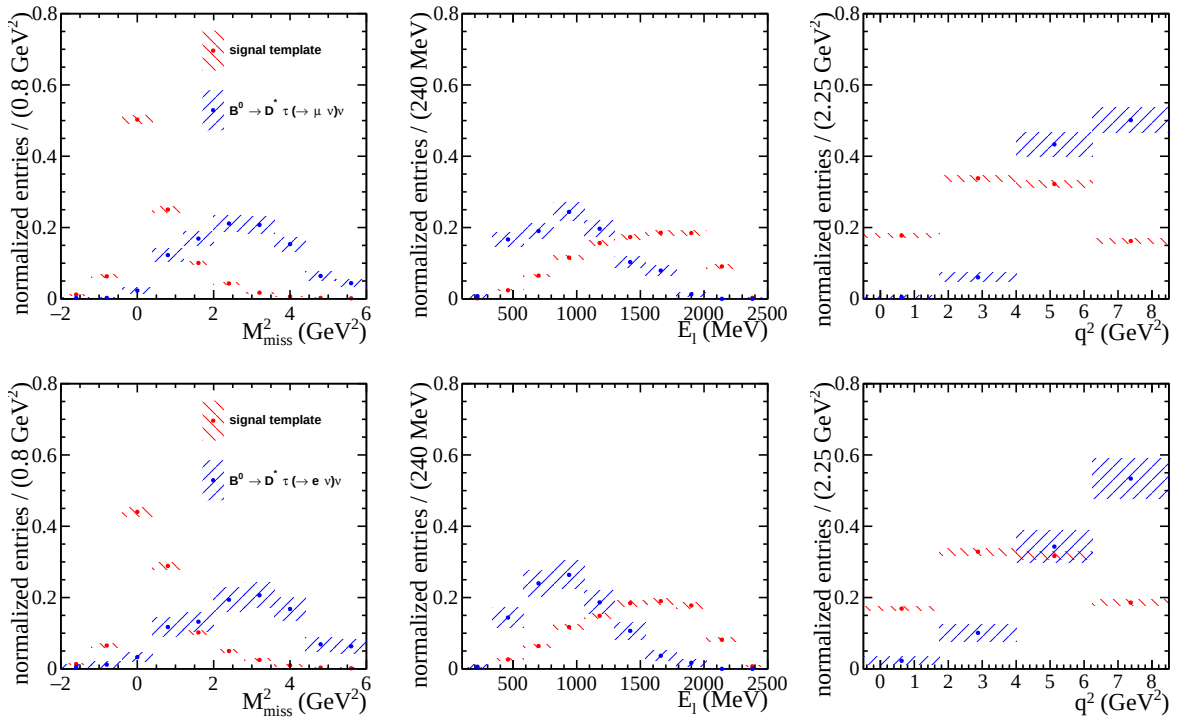


Figure 123: ● Comparison of the signal shape (red) with the background template (blue) from $B^0 \rightarrow D^{*-} \tau^+ \nu_\tau$ with $\tau \rightarrow \mu \nu \bar{\nu}$ (top) and $\tau \rightarrow e \nu \bar{\nu}$ (bottom) for 2016.

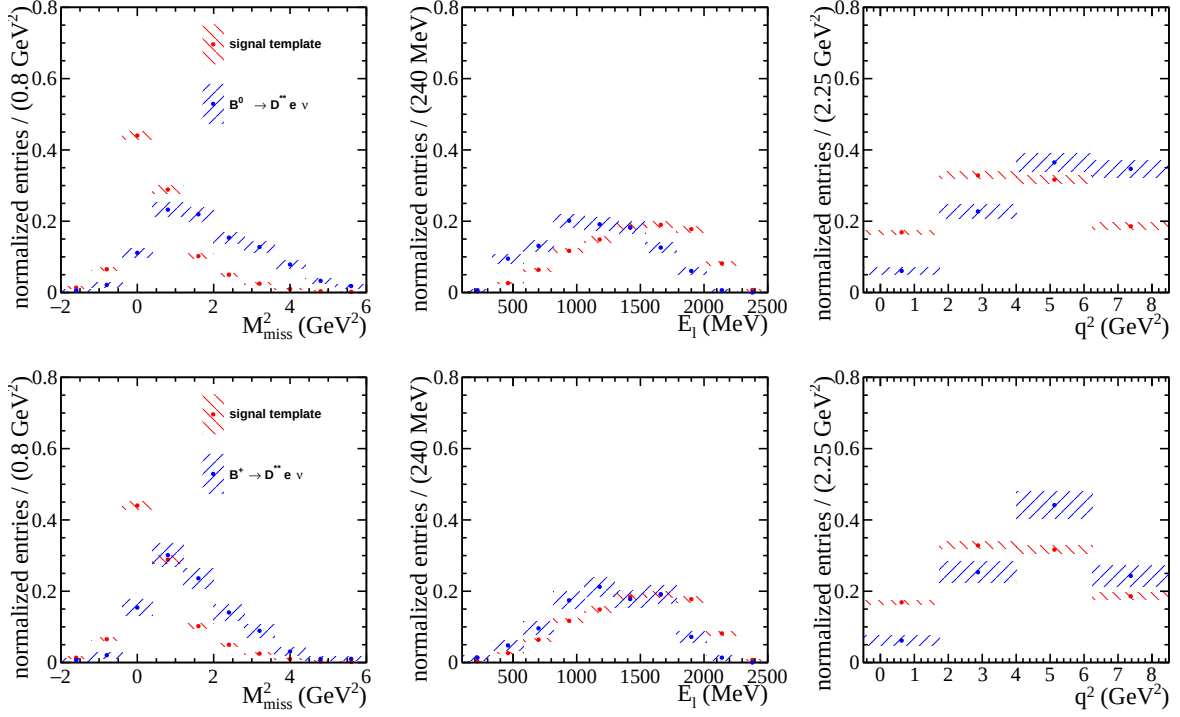


Figure 124: ● Comparison of the signal shape (red) with the background template (blue) from $B^0 \rightarrow D^{**} e^+ \nu_e$ (top) and $B^+ \rightarrow D^{**} e^+ \nu_e$ (bottom) for 2016.

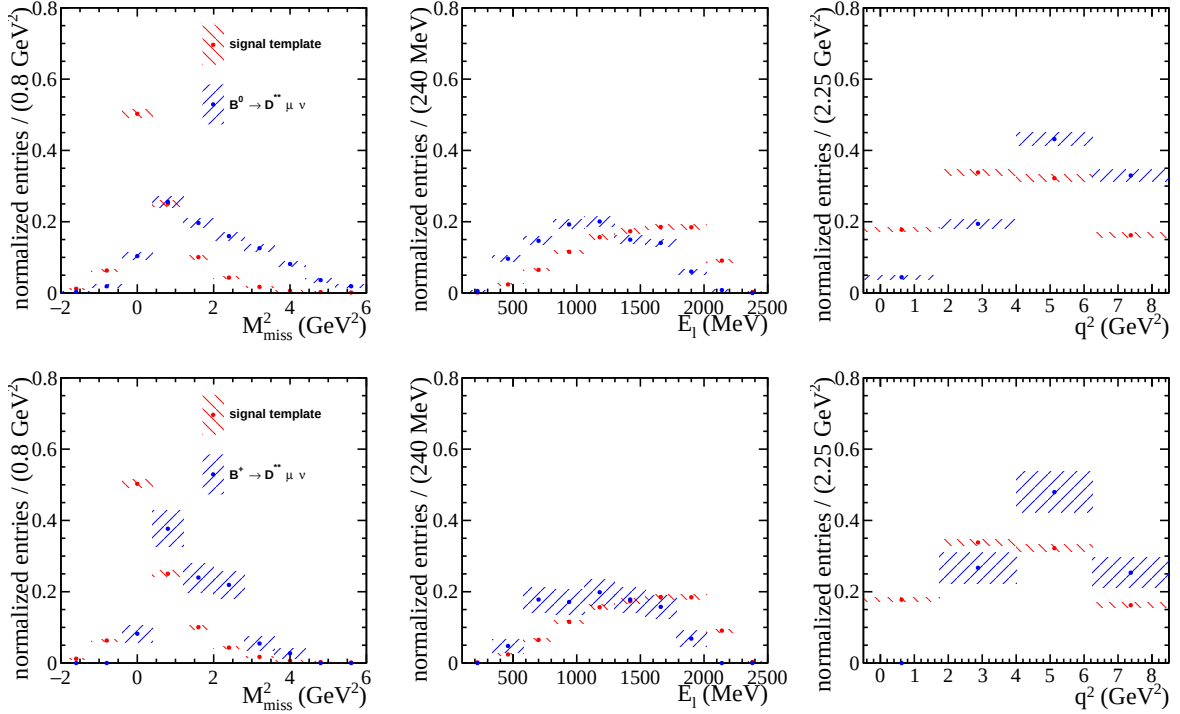


Figure 125: ● Comparison of the signal shape (red) with the background template (blue) from $B^0 \rightarrow D^{**} \mu^+ \nu_\mu$ (top) and $B^+ \rightarrow D^{**} \mu^+ \nu_\mu$ (bottom) for 2016

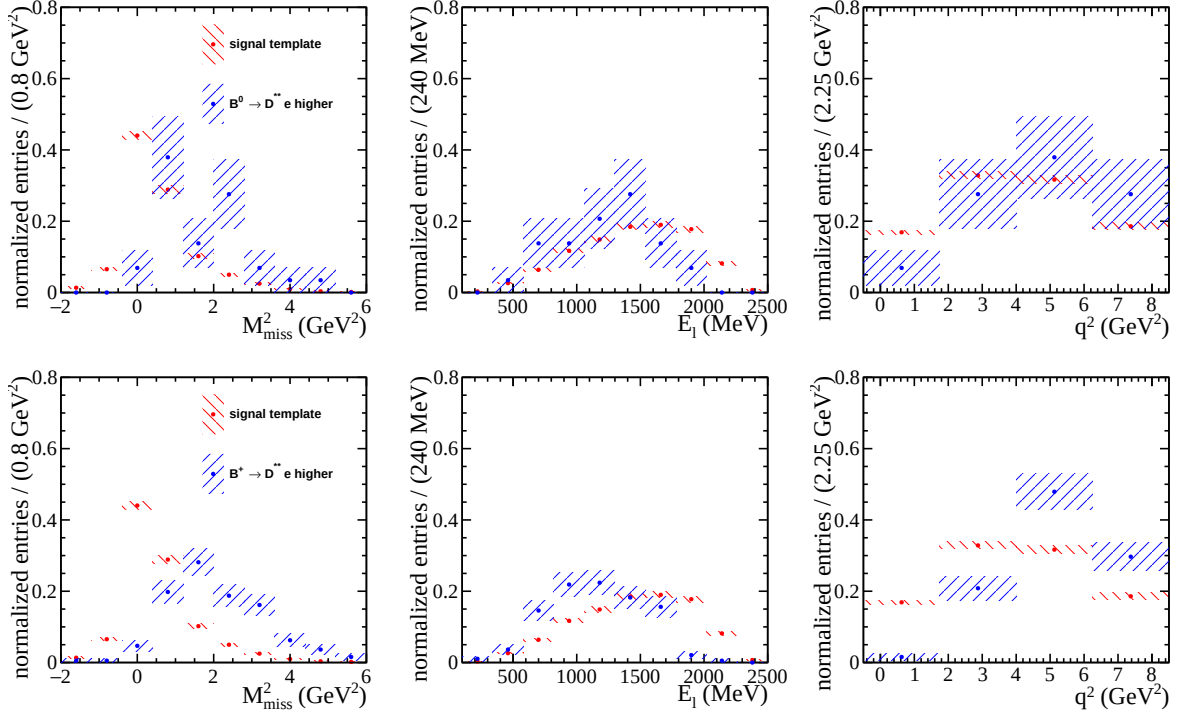


Figure 126: ● Comparison of the signal shape (red) with the background template (blue) from higher $B^0 \rightarrow D^{**} e^+ \nu_e$ (top) and $B^+ \rightarrow D^{**} e^+ \nu_e$ (bottom) for 2016.

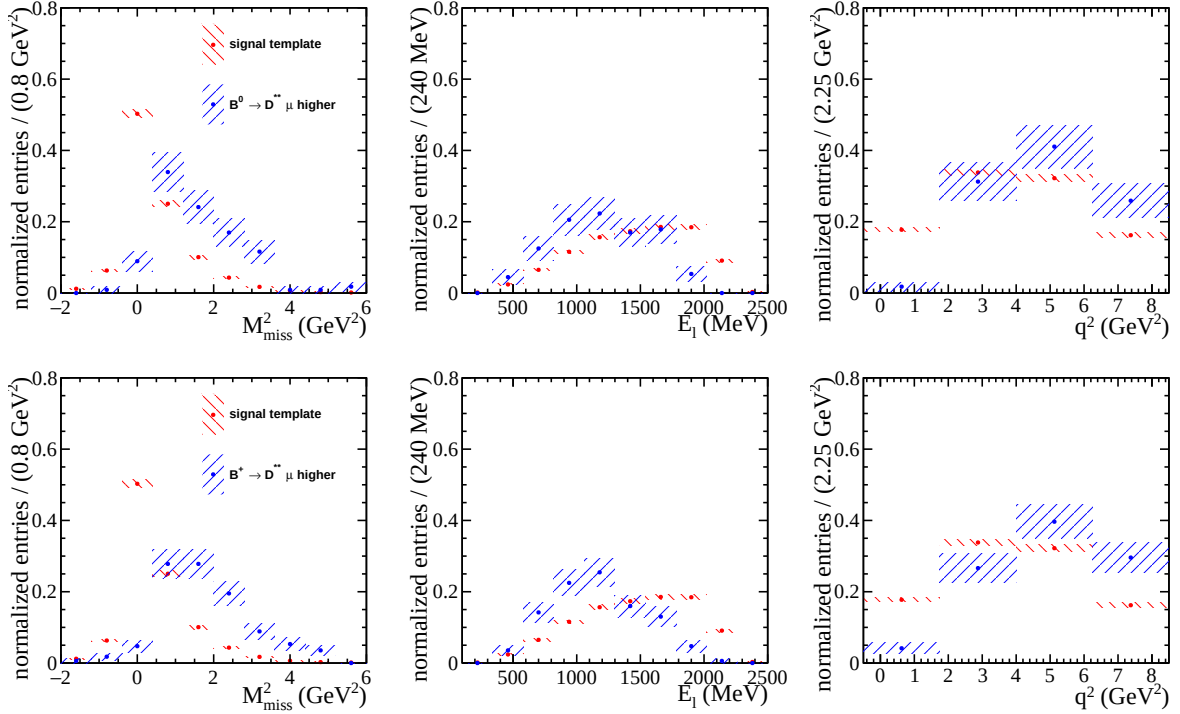


Figure 127: ● Comparison of the signal shape (red) with the background template (blue) from higher $B^0 \rightarrow D^{**} \mu^+ \nu_\mu$ (top) and $B^+ \rightarrow D^{**} \mu^+ \nu_\mu$ (bottom) for 2016.

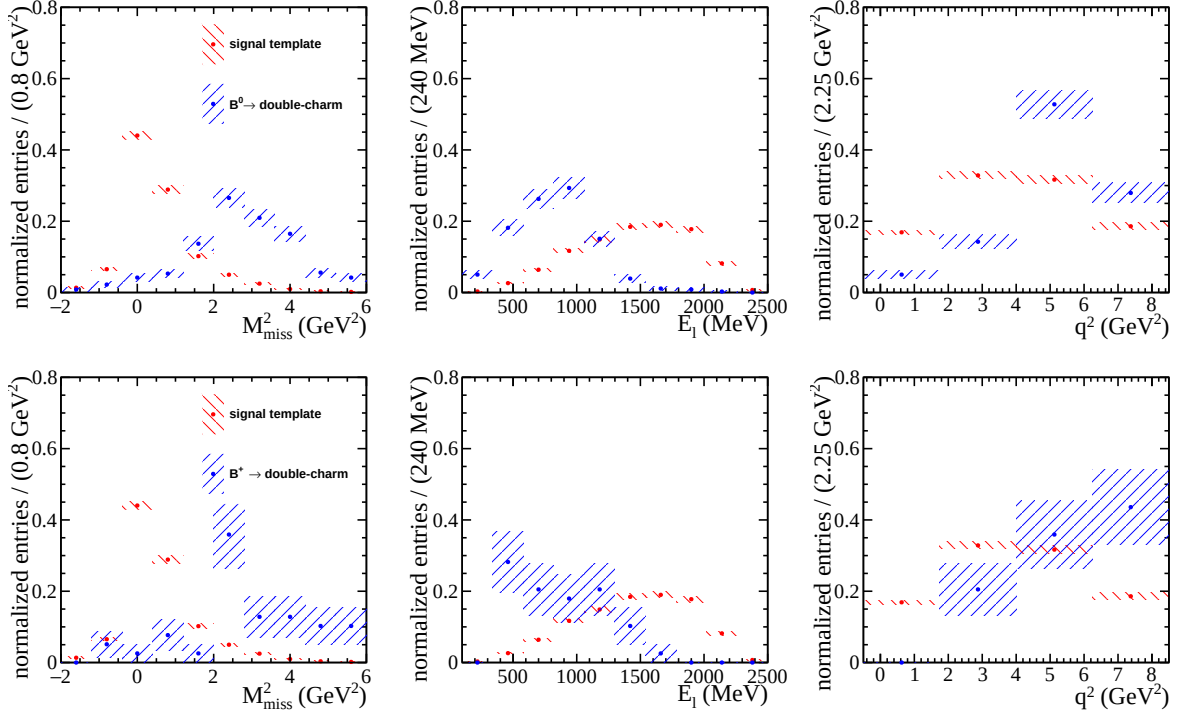


Figure 128: ● Comparison of the signal shape (red) with the background template (blue) from $B^0 \rightarrow D^* D(\rightarrow e^+ \nu_e) X$ (top) and $B^+ \rightarrow D^* D(\rightarrow e^+ \nu_e) X$ (bottom) for 2016.

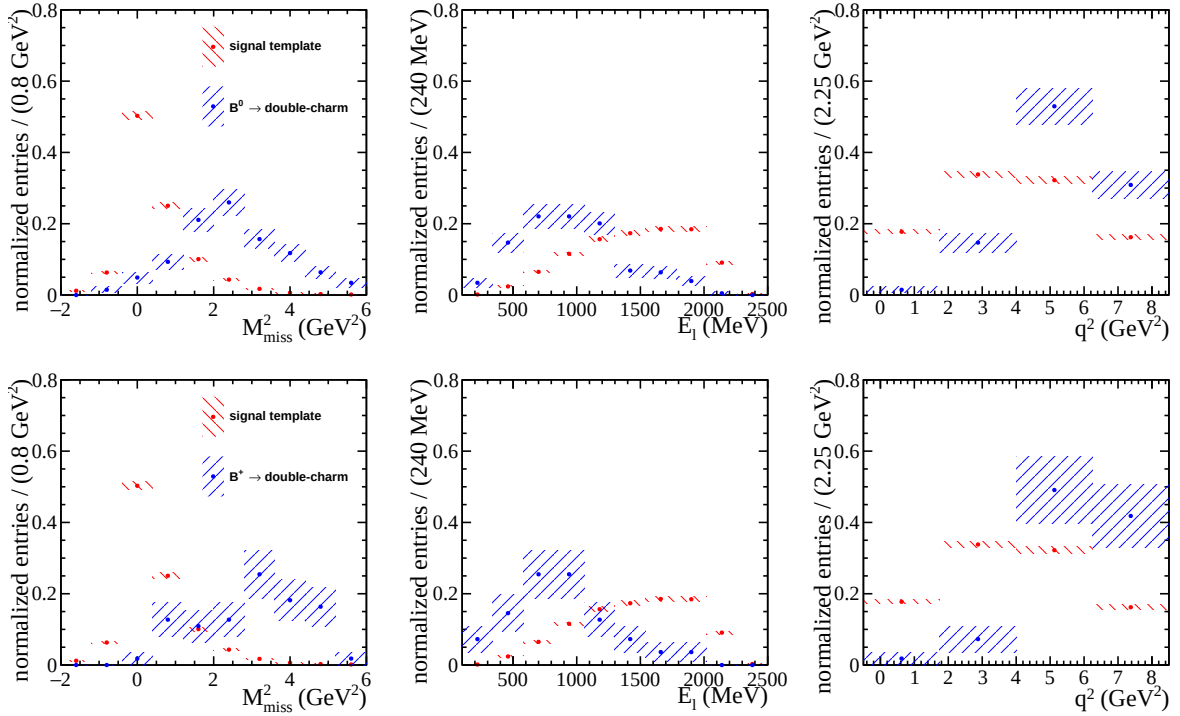


Figure 129: ● Comparison of the signal shape (red) with the background template (blue) from $B^0 \rightarrow D^* D(\rightarrow \mu^+ \nu_\mu) X$ (top) and $B^+ \rightarrow D^* D(\rightarrow \mu^+ \nu_\mu) X$ (bottom) for 2016.

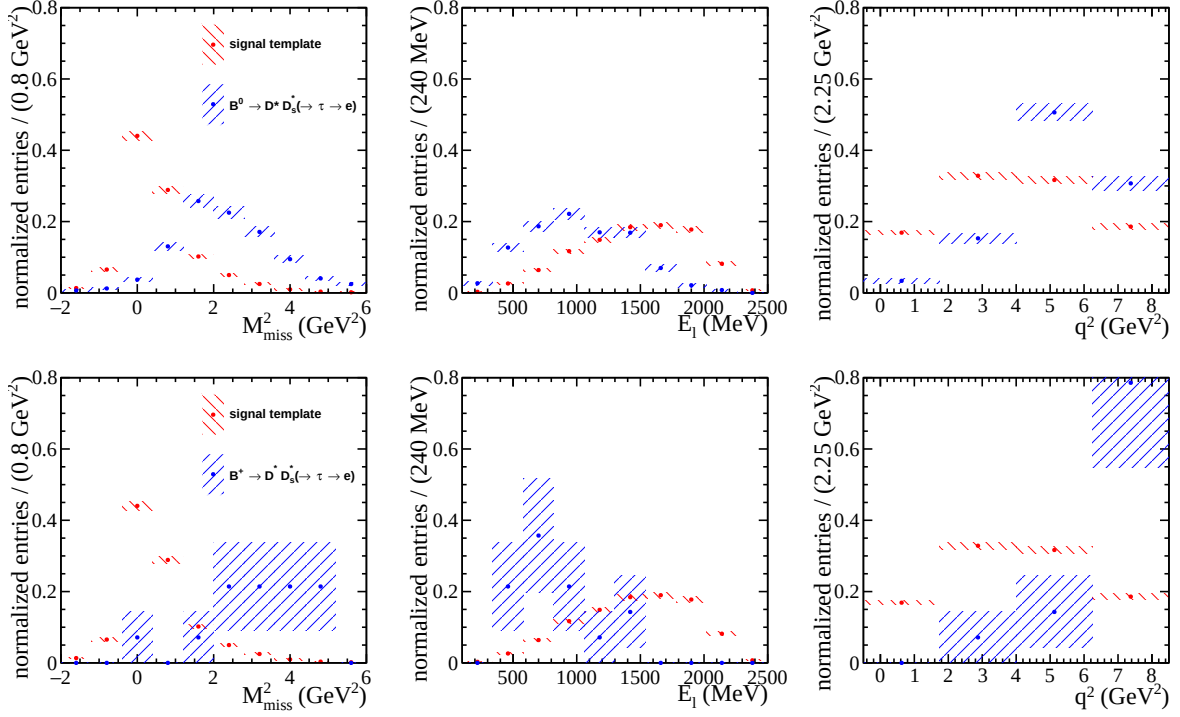


Figure 130: • Comparison of the signal shape (red) with the background template (blue) from $B^0 \rightarrow D^* D_s^+ (\rightarrow \tau^+ \nu_\tau) X$ (top) and $B^+ \rightarrow D^* D_s^+ (\rightarrow \tau^+ \nu_\tau) X$ (bottom) with $\tau \rightarrow e \nu \bar{\nu}$ for 2016

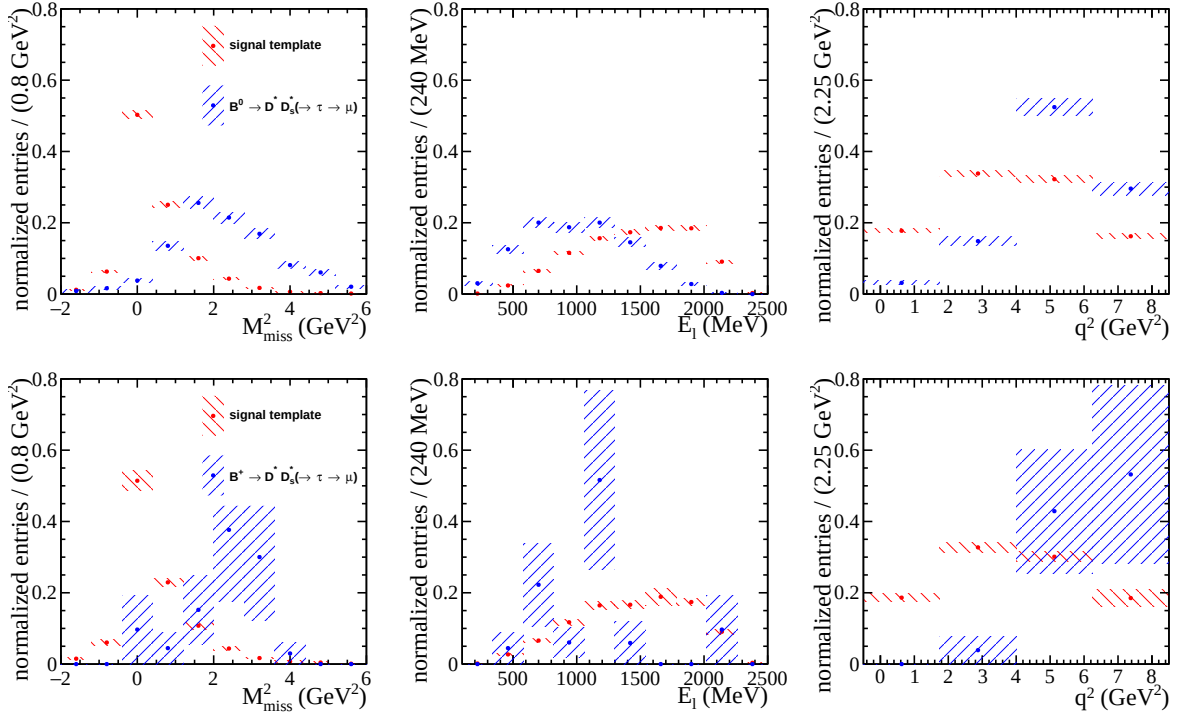


Figure 131: • Comparison of the signal shape (red) with the background template (blue) from $B^0 \rightarrow D^* D_s^+ (\rightarrow \tau^+ \nu_\tau) X$ (top) and $B^+ \rightarrow D^* D_s^+ (\rightarrow \tau^+ \nu_\tau) X$ (bottom) with $\tau \rightarrow \mu \nu \bar{\nu}$ for 2016.

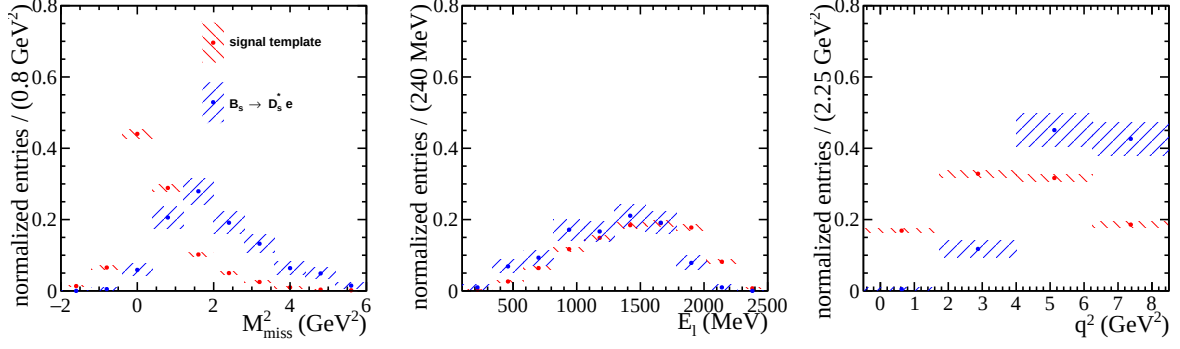


Figure 132: ● Comparison of the signal shape (red) with the background template (blue) from $B_s^0 \rightarrow D_s^{*+} e^- \nu_e$ for 2016

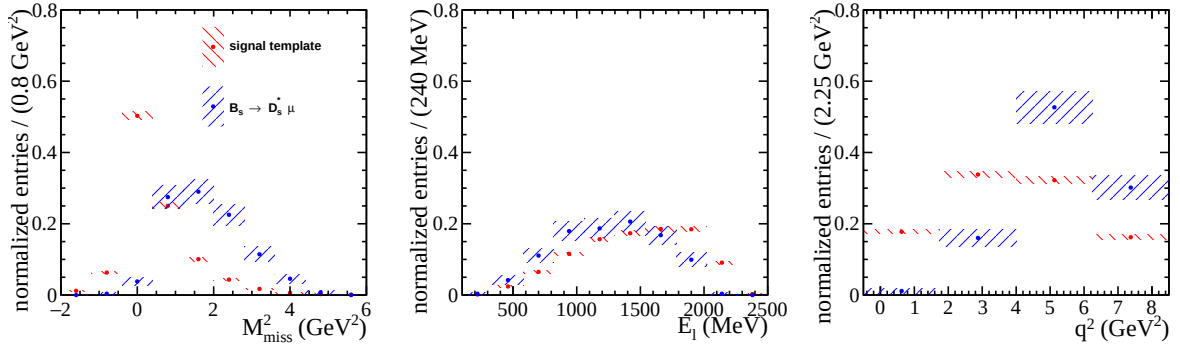


Figure 133: ● Comparison of the signal shape (red) with the background template (blue) from $B_s^0 \rightarrow D_s^{*+} \mu^- \nu_\mu$ for 2016.

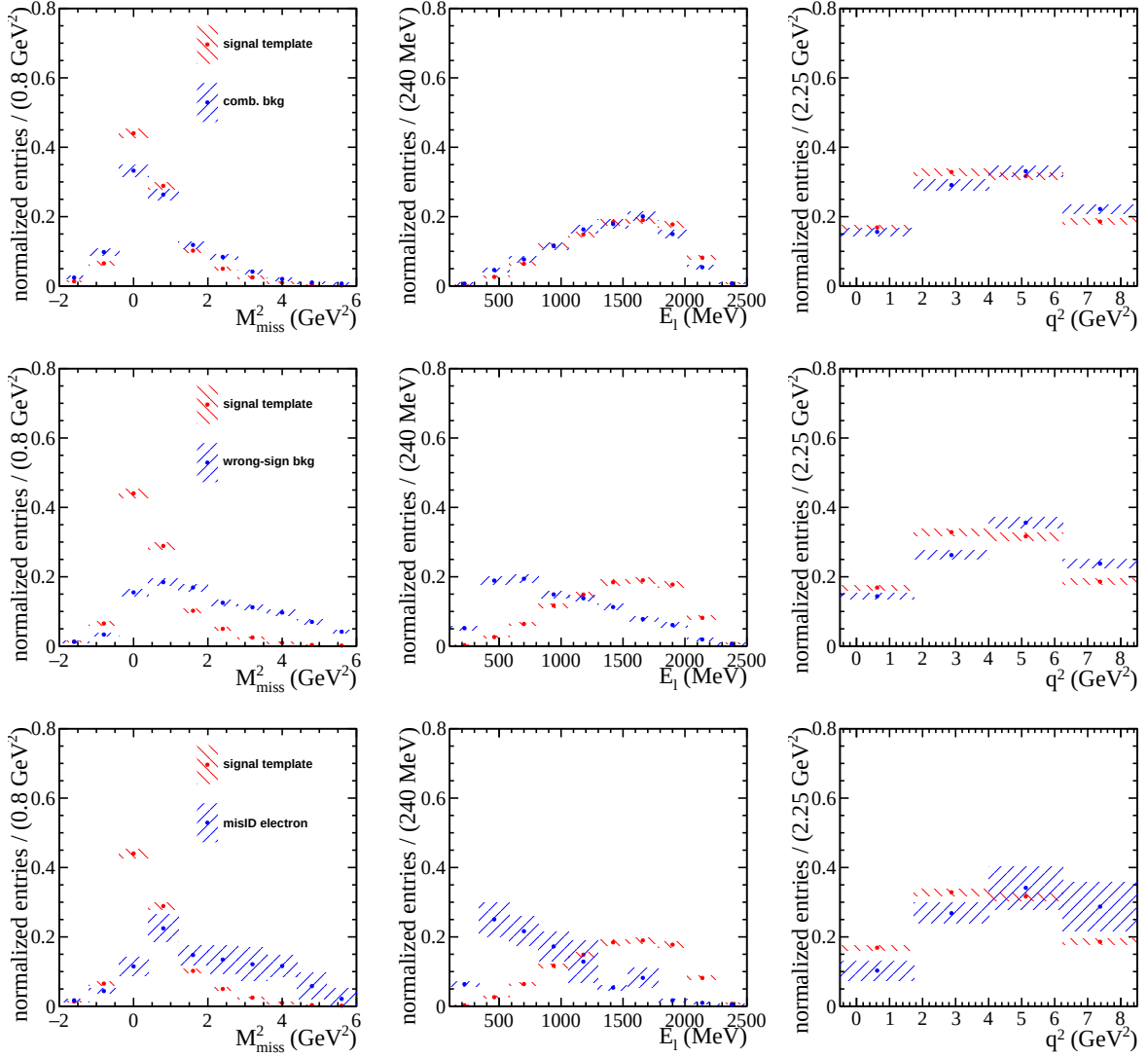


Figure 134: ● Comparison of the signal shape (red) with the combinatorial background templates (blue) for 'fake' D^* and 'real' electron combinations (top), 'real' D^* and 'real' electron combinations (middle), and 'real' D^* and 'fake' electron (bottom) combinations for 2016.

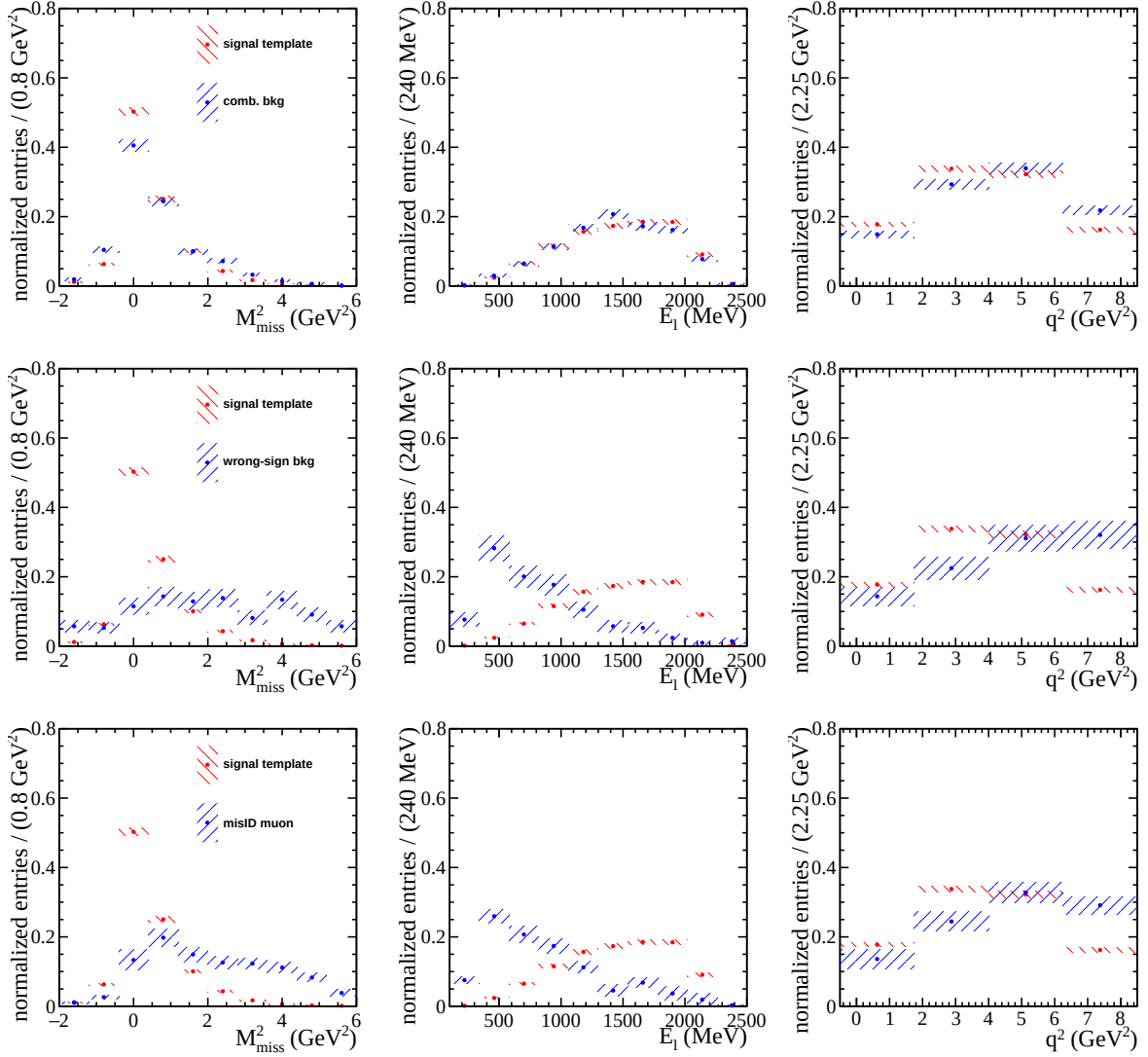


Figure 135: ● Comparison of the signal shape (red) with the combinatorial background templates for 'fake' D^* and 'real' muon combinations (top), 'real' D^* and 'real' muon combinations (middle), and 'real' D^* and 'fake' muon (bottom) combinations for 2016.

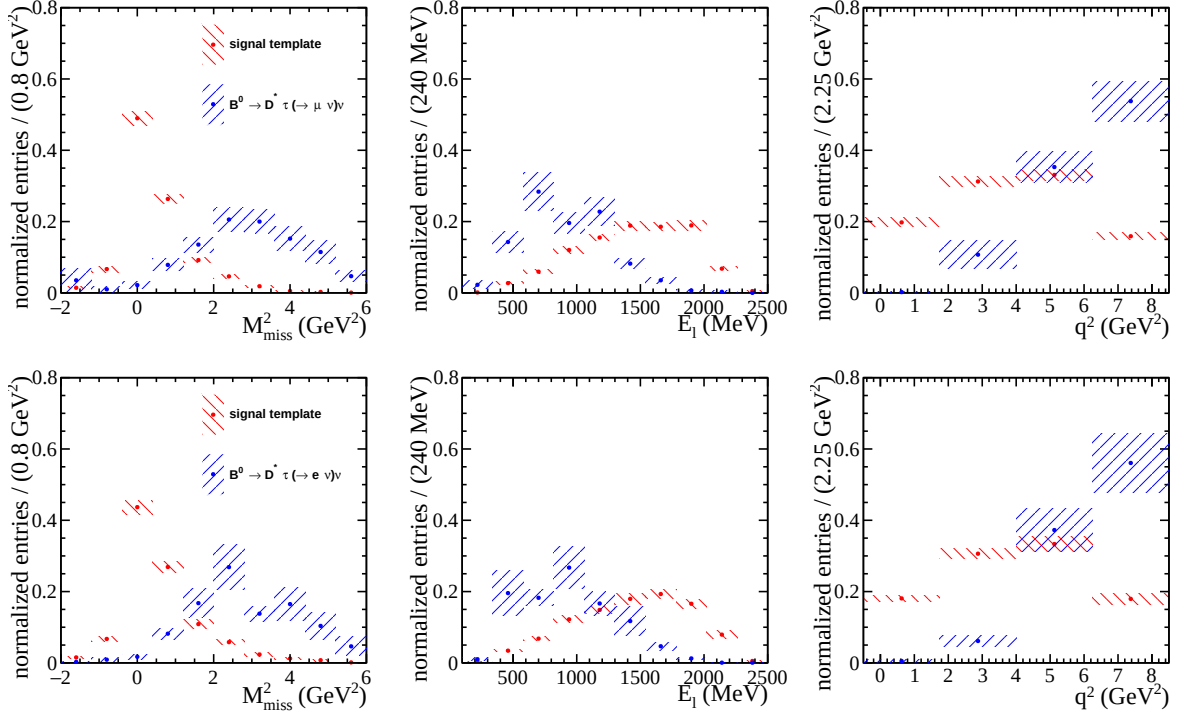


Figure 136: • Comparison of the signal shape (red) with the background template (blue) from $B^0 \rightarrow D^{*+} \tau^+ \nu_\tau$ with $\tau \rightarrow \mu \nu \bar{\nu}$ (top) and $\tau \rightarrow e \nu \bar{\nu}$ (bottom) for 2017.

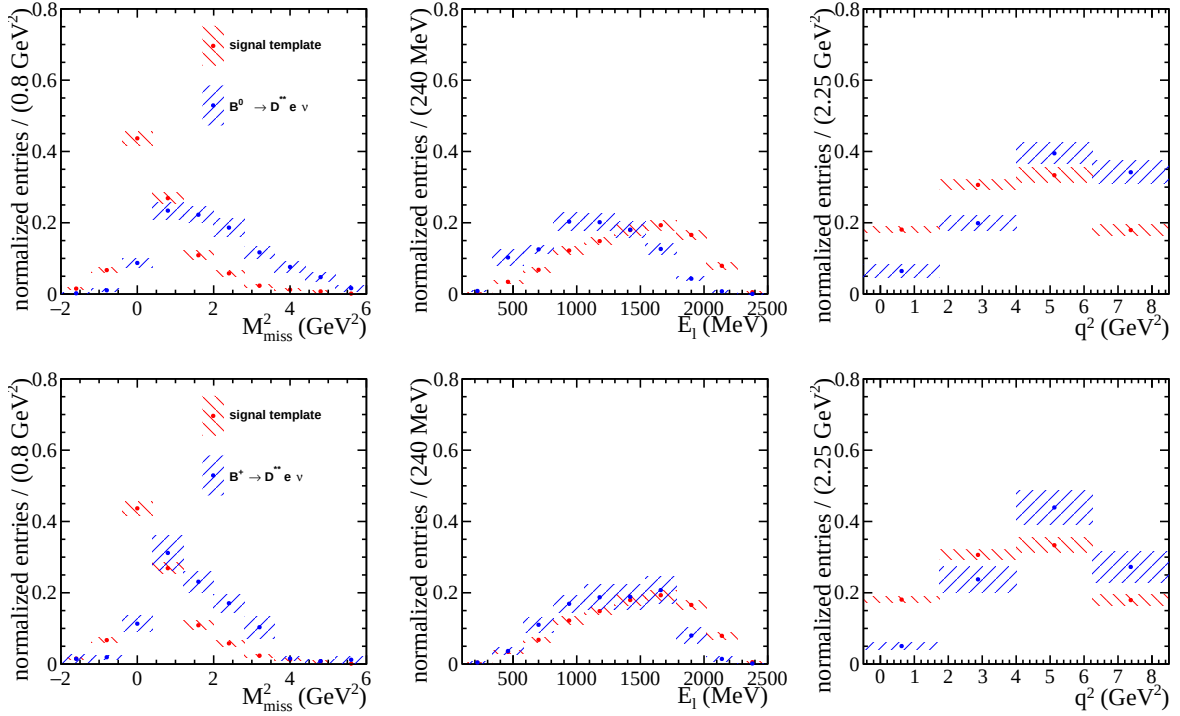


Figure 137: • Comparison of the signal shape (red) with the background template (blue) from $B^0 \rightarrow D^{*+} e^+ \nu_e$ (top) and $B^+ \rightarrow D^{*+} e^+ \nu_e$ (bottom) for 2017.

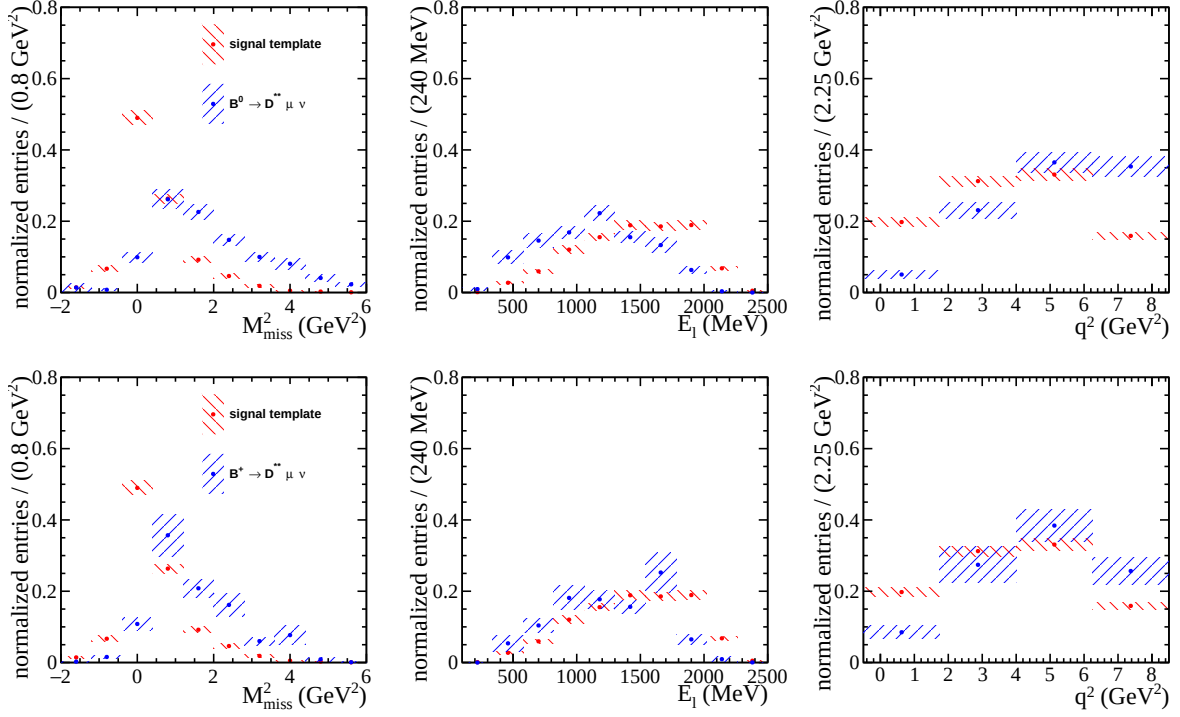


Figure 138: ● Comparison of the signal shape (red) with the background template (blue) from $B^0 \rightarrow D^{**} \mu^+ \nu_\mu$ (top) and $B^+ \rightarrow D^{**} \mu^+ \nu_\mu$ (bottom) for 2017

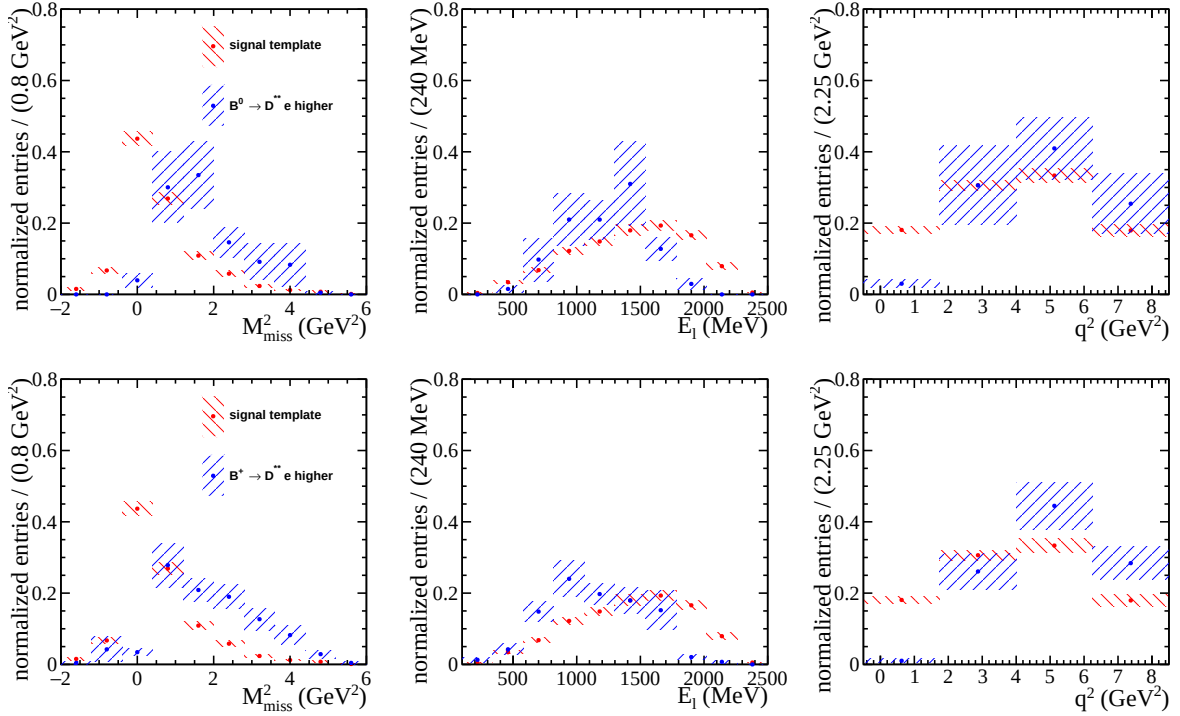


Figure 139: ● Comparison of the signal shape (red) with the background template (blue) from higher $B^0 \rightarrow D^{**} e^+ \nu_e$ (top) and $B^+ \rightarrow D^{**} e^+ \nu_e$ (bottom) for 2017.

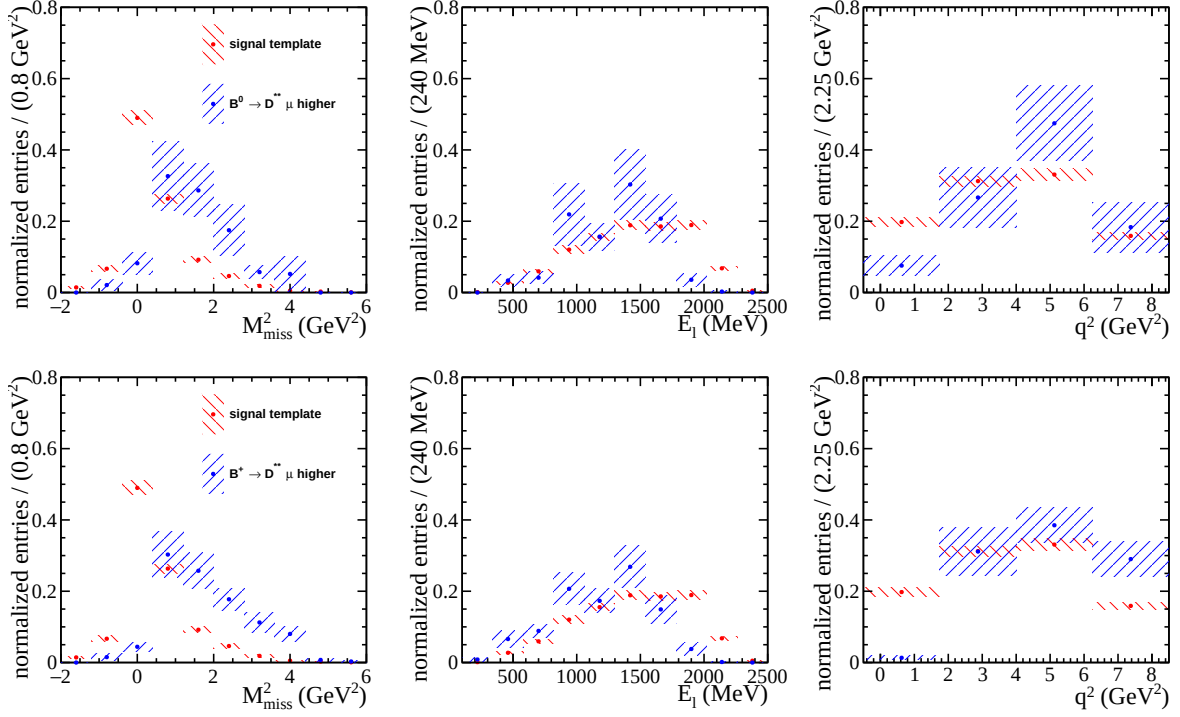


Figure 140: ● Comparison of the signal shape (red) with the background template (blue) from higher $B^0 \rightarrow D^{**} \mu^+ \nu_\mu$ (top) and $B^+ \rightarrow D^{**} \mu^+ \nu_\mu$ (bottom) for 2017.

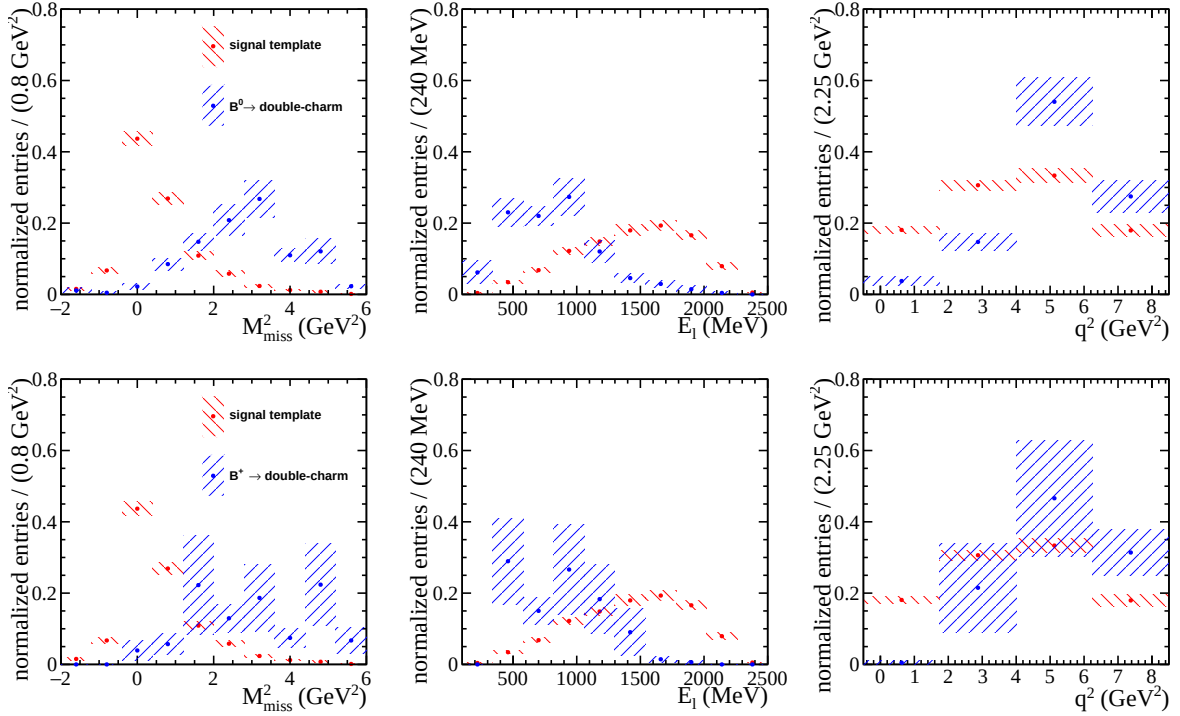


Figure 141: ● Comparison of the signal shape (red) with the background template (blue) from $B^0 \rightarrow D^* D(\rightarrow e^+ \nu_e) X$ (top) and $B^+ \rightarrow D^* D(\rightarrow e^+ \nu_e) X$ (bottom) for 2017.

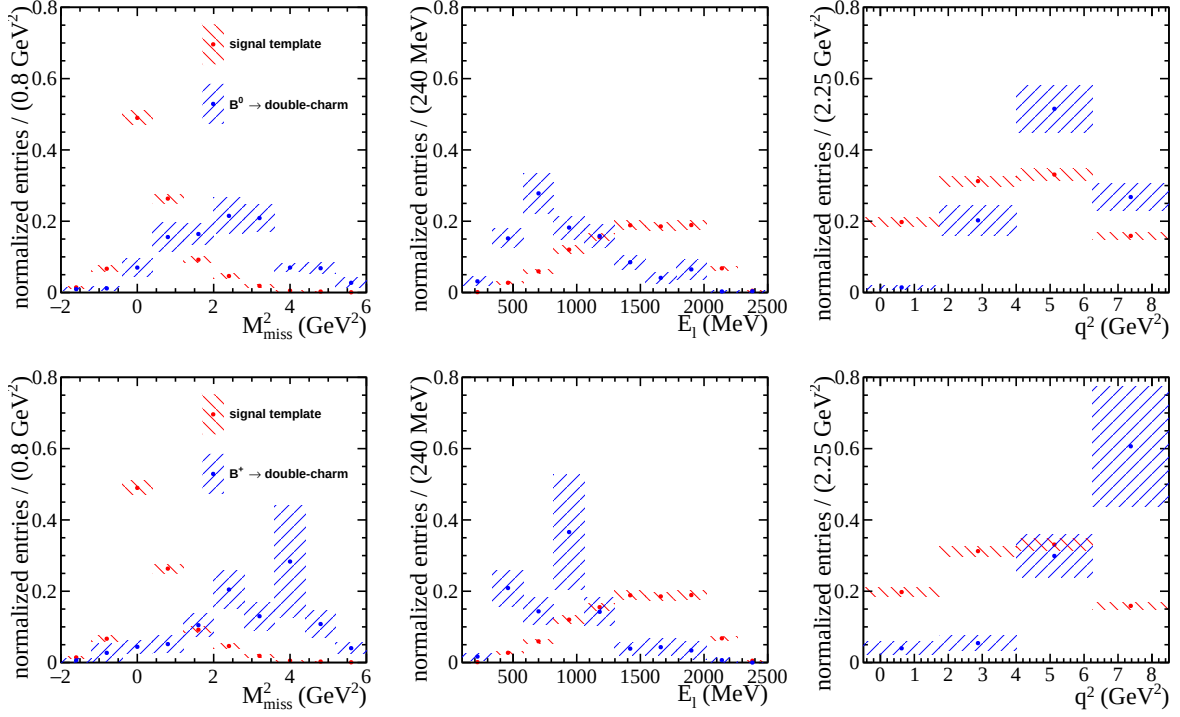


Figure 142: ● Comparison of the signal shape (red) with the background template (blue) from $B^0 \rightarrow D^* D(\rightarrow \mu^+ \nu_\mu) X$ (top) and $B^+ \rightarrow D^* D(\rightarrow \mu^+ \nu_\mu) X$ (bottom) for 2017.

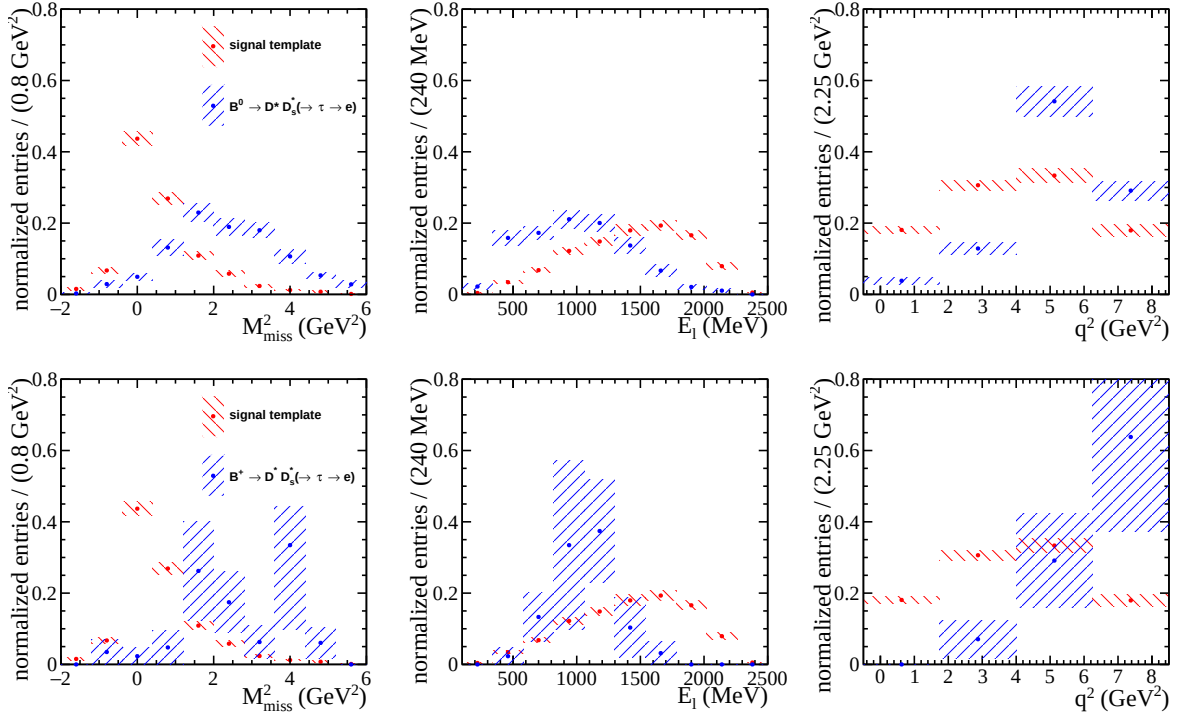


Figure 143: ● Comparison of the signal shape (red) with the background template (blue) from $B^0 \rightarrow D^* D_s^+(\rightarrow \tau^+ \nu_\tau) X$ (top) and $B^+ \rightarrow D^* D_s^+(\rightarrow \tau^+ \nu_\tau) X$ (bottom) with $\tau \rightarrow e \nu \bar{\nu}$ for 2017

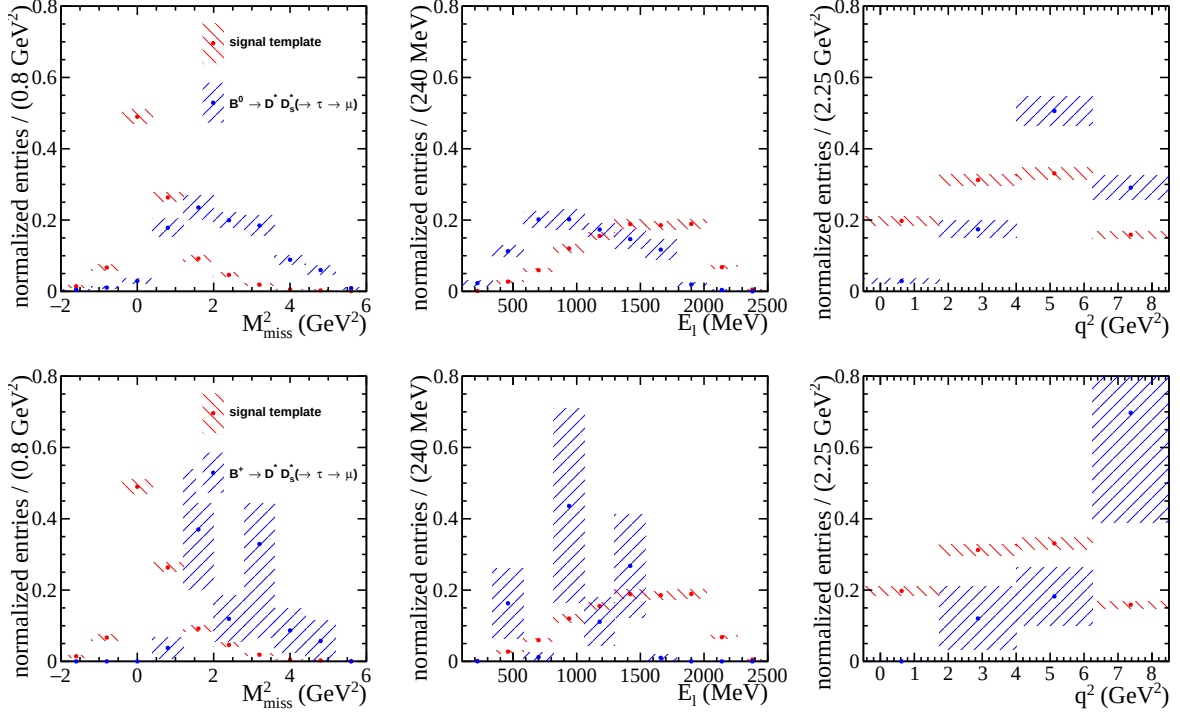


Figure 144: • Comparison of the signal shape (red) with the background template (blue) from $B^0 \rightarrow D^* D_s^-(\tau \rightarrow \tau \mu)$ (top) and $B^+ \rightarrow D^* D_s^-(\tau \rightarrow \tau \mu)$ (bottom) with $\tau \rightarrow \mu \nu \bar{\nu}$ for 2017.

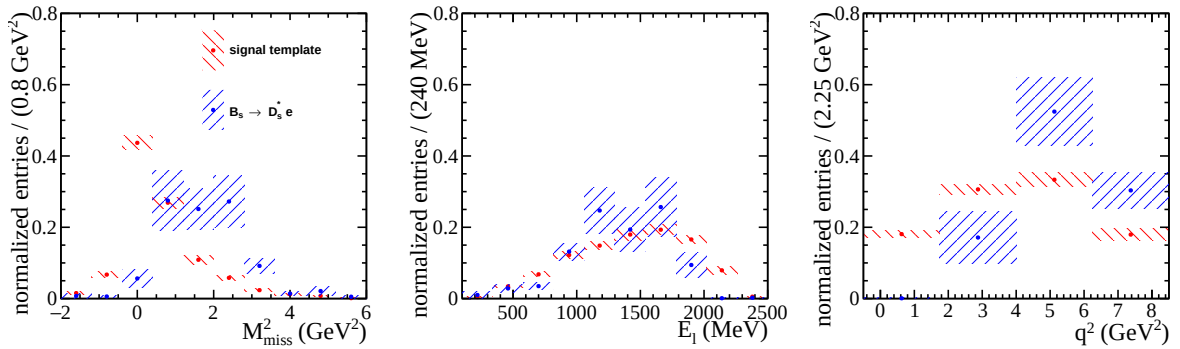


Figure 145: • Comparison of the signal shape (red) with the background template (blue) from $B^0 \rightarrow D_s^{**} e \nu_e$ for 2017

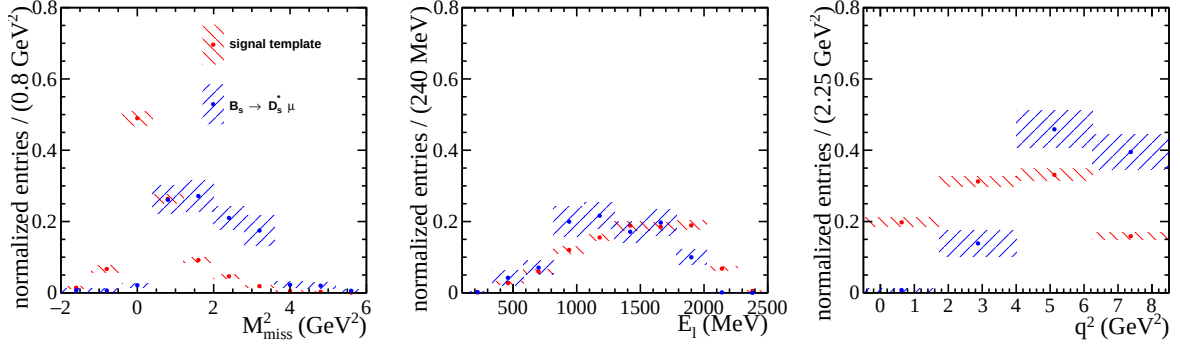


Figure 146: ● Comparison of the signal shape (red) with the background template (blue) from $B_s^0 \rightarrow D_s^{*0} \mu \nu_\mu$ for 2017.

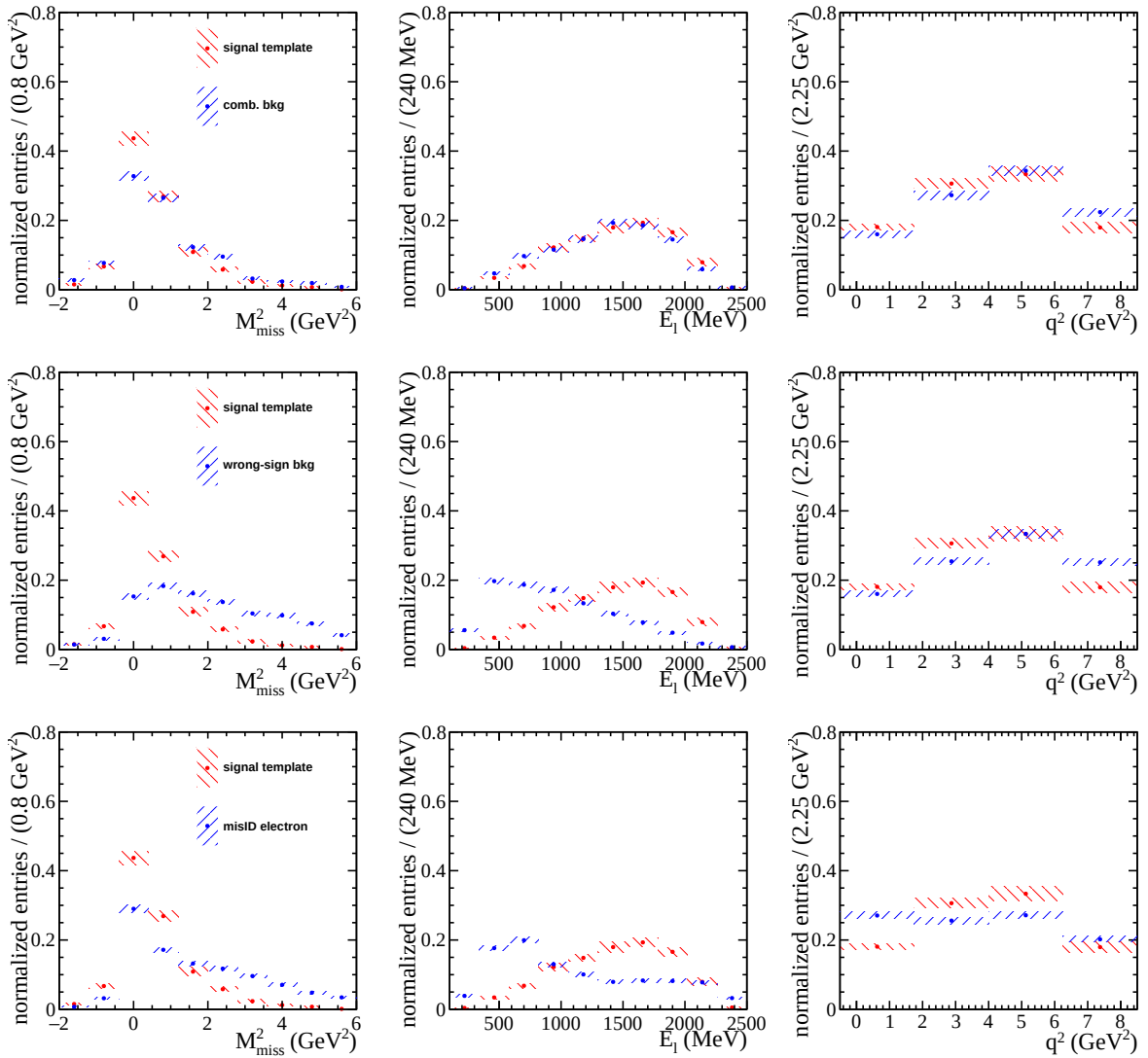


Figure 147: ● Comparison of the signal shape (red) with the combinatorial background templates (blue) for 'fake' D^* and 'real' electron combinations (top), 'real' D^* and 'real' electron combinations (middle), and 'real' D^* and 'fake' electron (bottom) combinations for 2017.

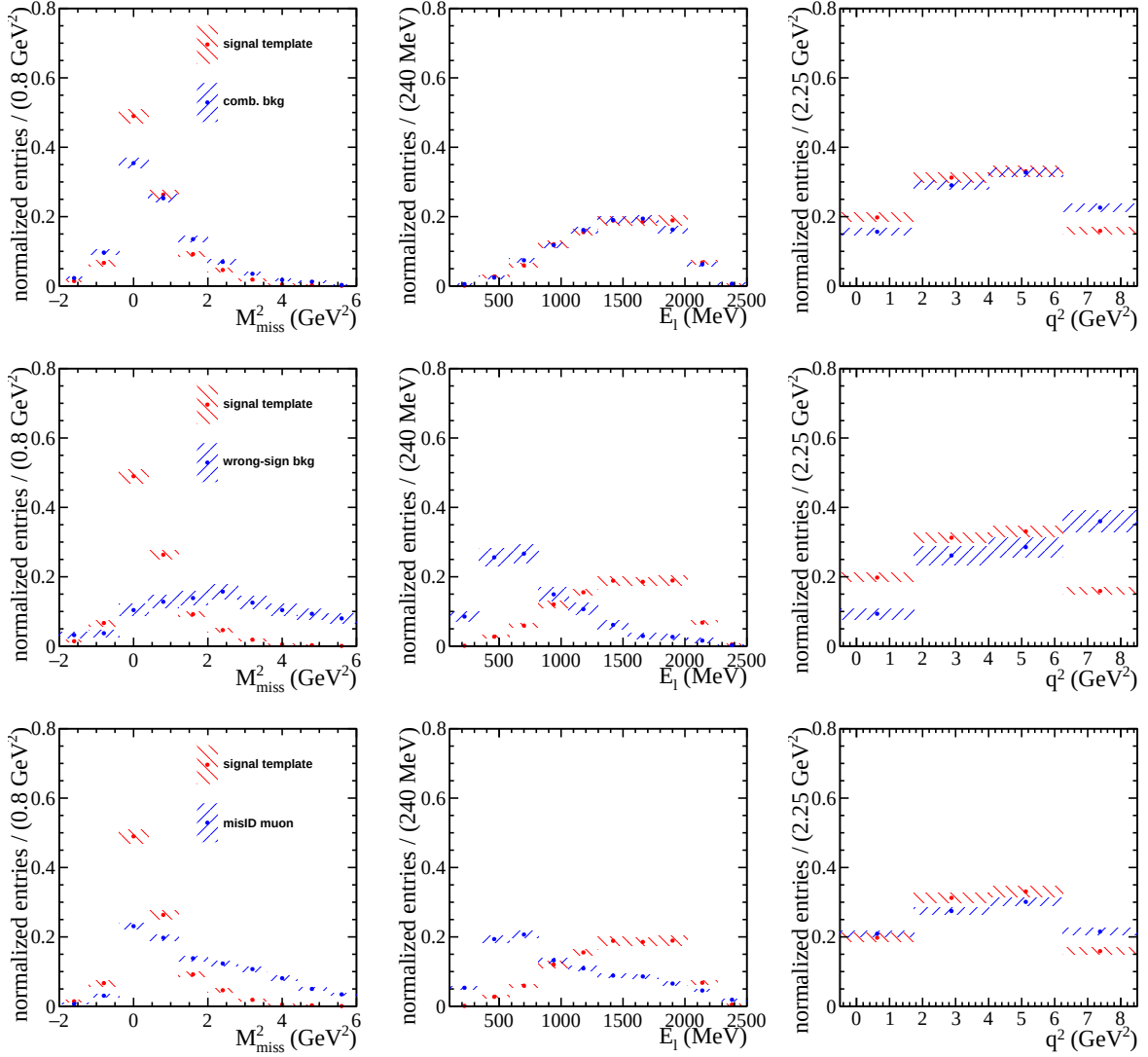


Figure 148: ● Comparison of the signal shape (red) with the combinatorial background templates for 'fake' D^* and 'real' muon combinations (top), 'real' D^* and 'real' muon combinations (middle), and 'real' D^* and 'fake' muon (bottom) combinations for 2017.

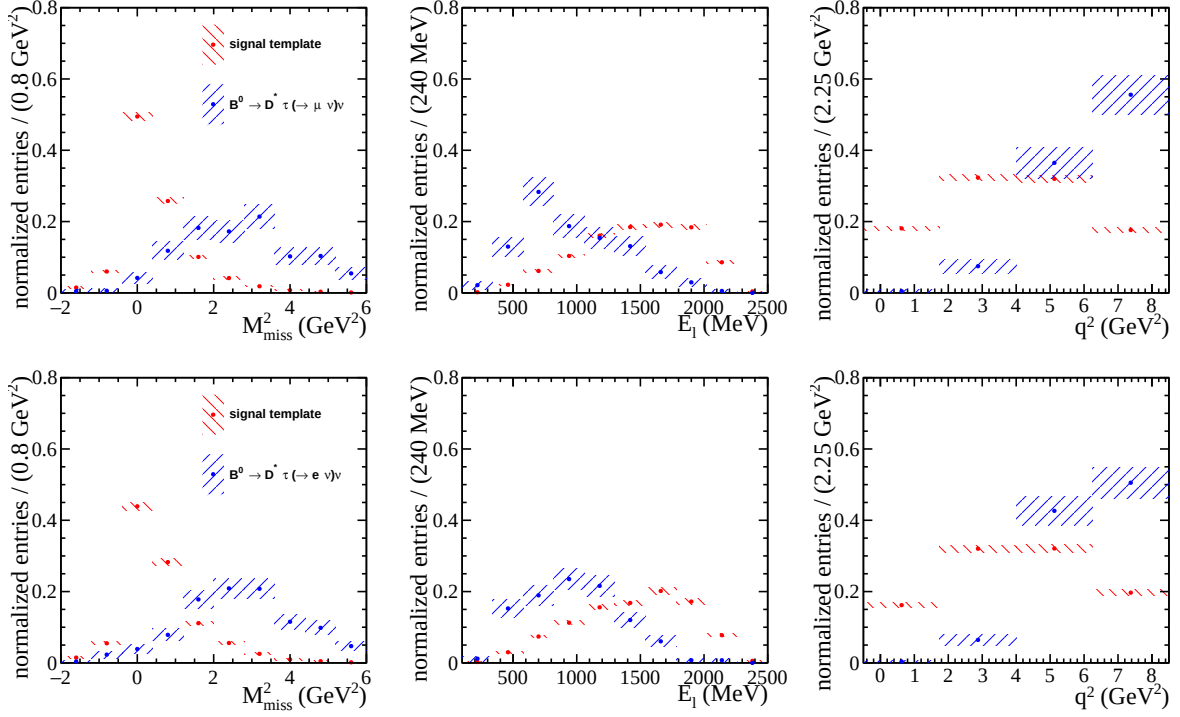


Figure 149: • Comparison of the signal shape (red) with the background template (blue) from $B^0 \rightarrow D^{*+} \tau^+ \nu_\tau$ with $\tau \rightarrow \mu \nu \bar{\nu}$ (top) and $\tau \rightarrow e \nu \bar{\nu}$ (bottom) for 2018.

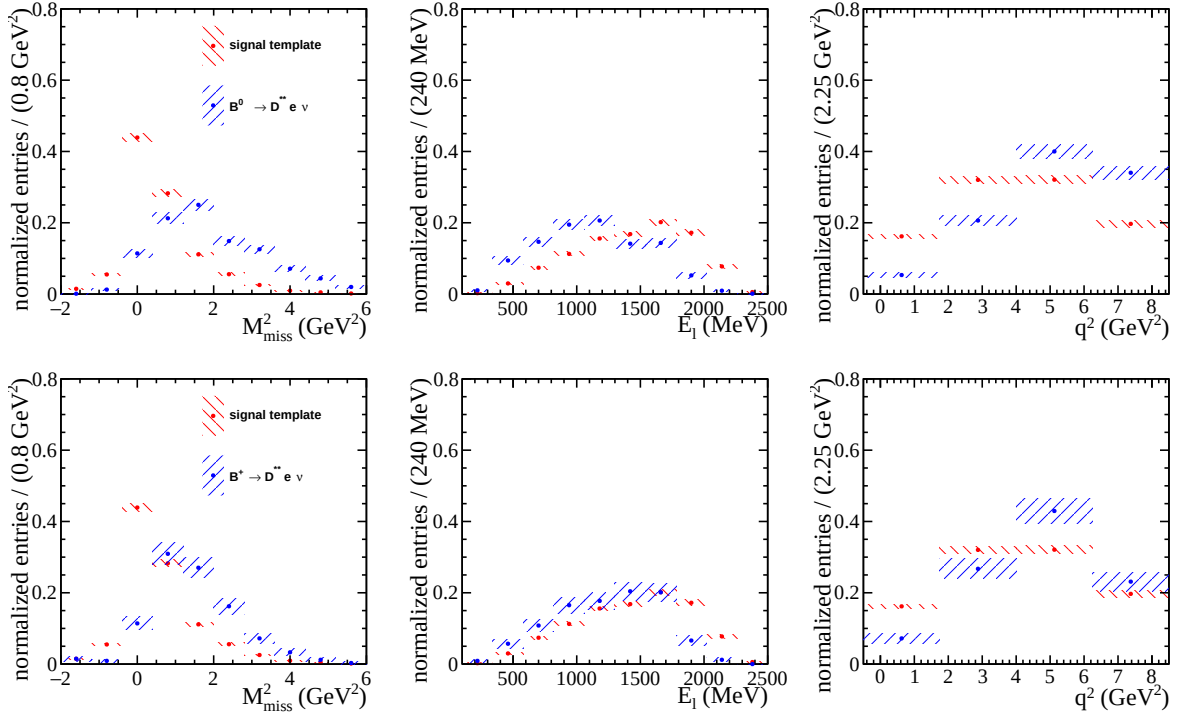


Figure 150: • Comparison of the signal shape (red) with the background template (blue) from $B^0 \rightarrow D^{*+} e^+ \nu_e$ (top) and $B^+ \rightarrow D^{*+} e^+ \nu_e$ (bottom) for 2018.

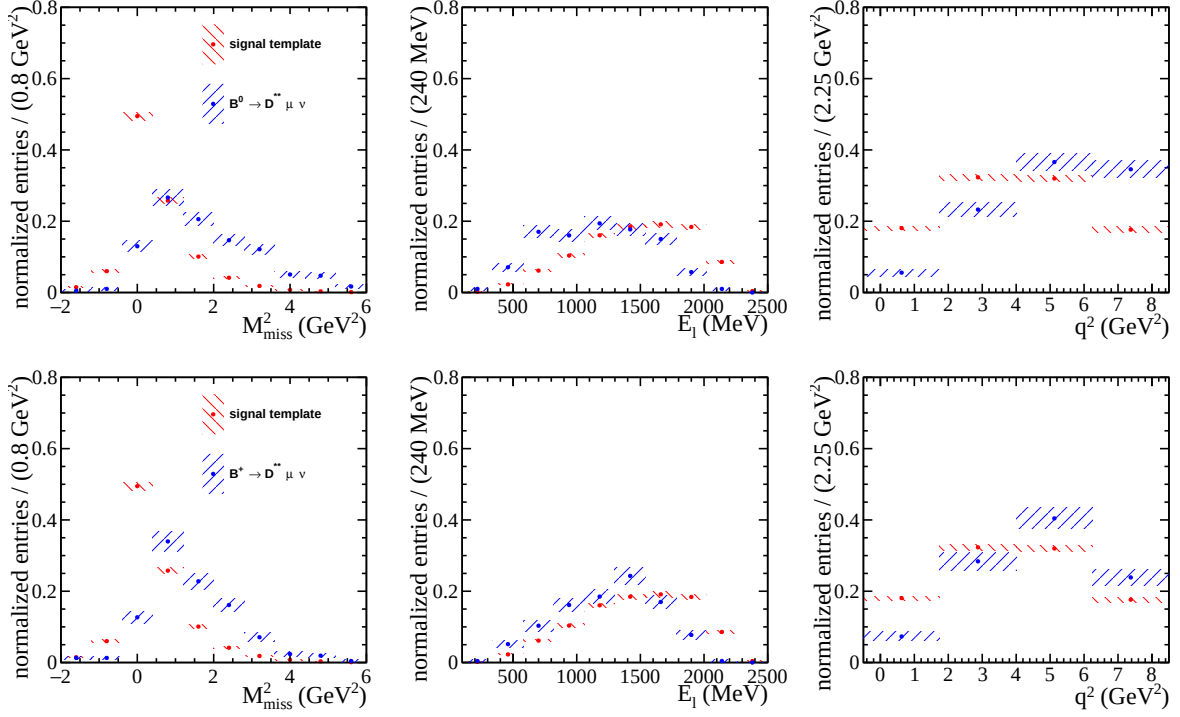


Figure 151: ● Comparison of the signal shape (red) with the background template (blue) from $B^0 \rightarrow D^{**} \mu^+ \nu_\mu$ (top) and $B^+ \rightarrow D^{**} \mu^+ \nu_\mu$ (bottom) for 2018

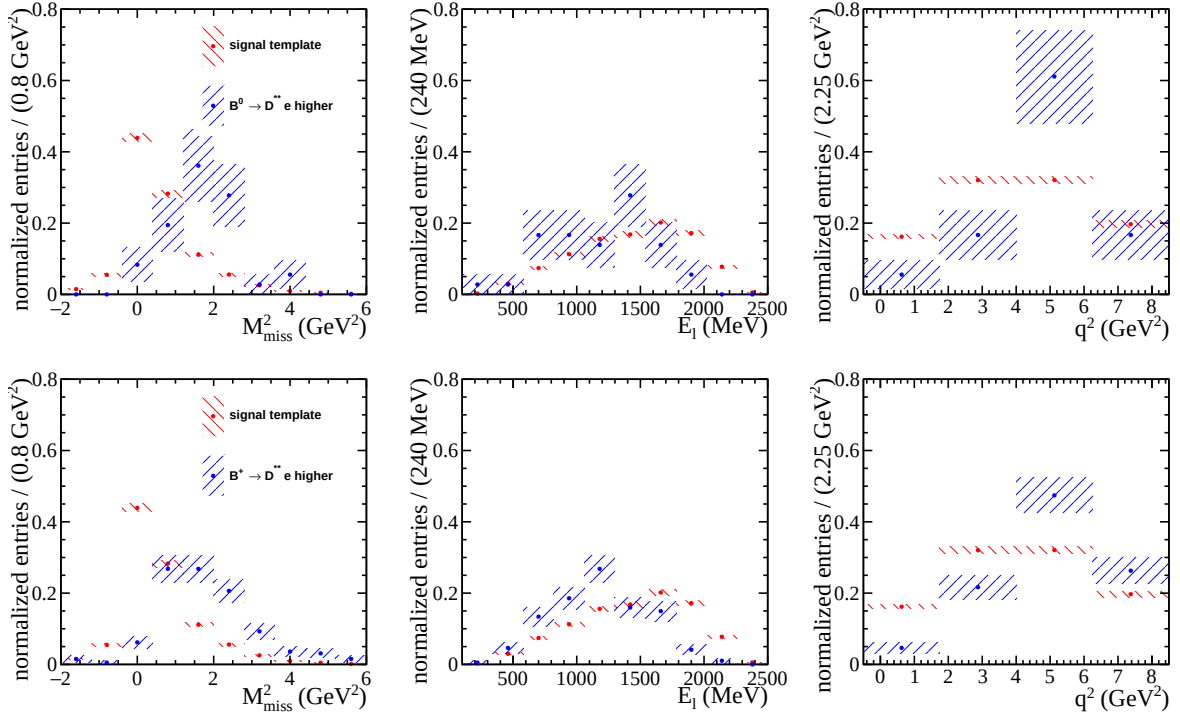


Figure 152: ● Comparison of the signal shape (red) with the background template (blue) from higher $B^0 \rightarrow D^{**} e^+ \nu_e$ (top) and $B^+ \rightarrow D^{**} e^+ \nu_e$ (bottom) for 2018.

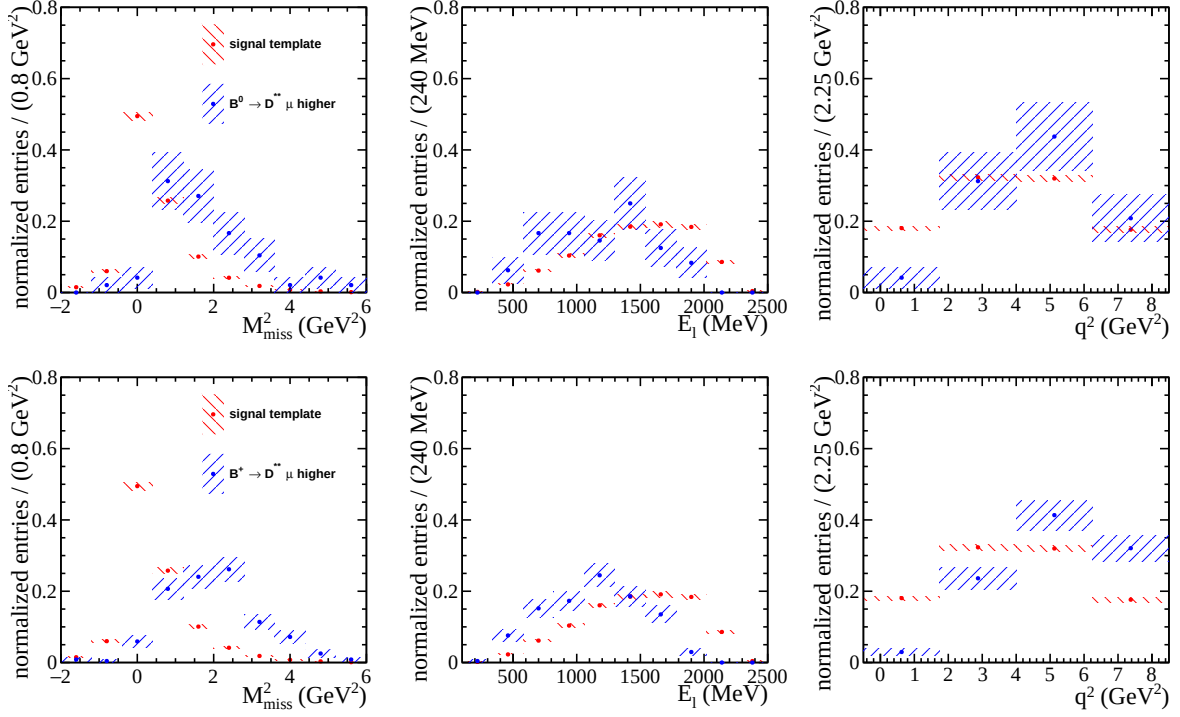


Figure 153: ● Comparison of the signal shape (red) with the background template (blue) from higher $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu$ (top) and $B^+ \rightarrow D^{*-} \mu^+ \nu_\mu$ (bottom) for 2018.

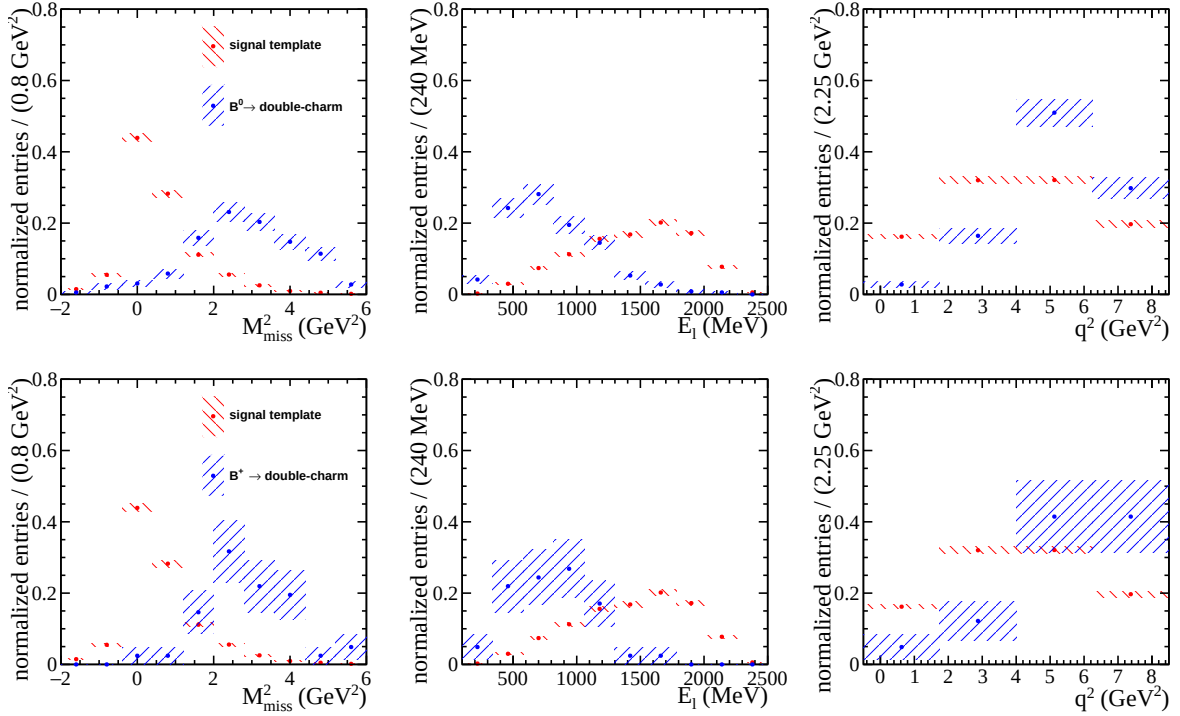


Figure 154: ● Comparison of the signal shape (red) with the background template (blue) from $B^0 \rightarrow D^* D (\rightarrow e^+ \nu_e) X$ (top) and $B^+ \rightarrow D^* D (\rightarrow e^+ \nu_e) X$ (bottom) for 2018.

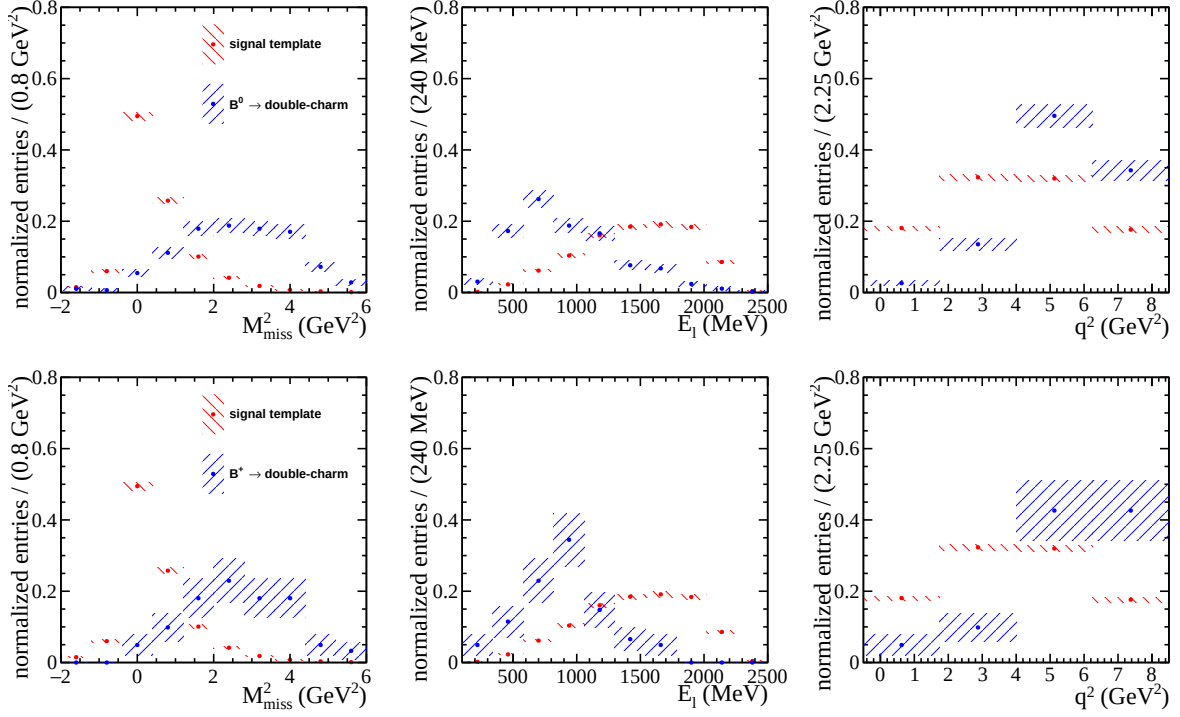


Figure 155: ● Comparison of the signal shape (red) with the background template (blue) from $B^0 \rightarrow D^* D(\rightarrow \mu^+ \nu_\mu) X$ (top) and $B^+ \rightarrow D^* D(\rightarrow \mu^+ \nu_\mu) X$ (bottom) for 2018.

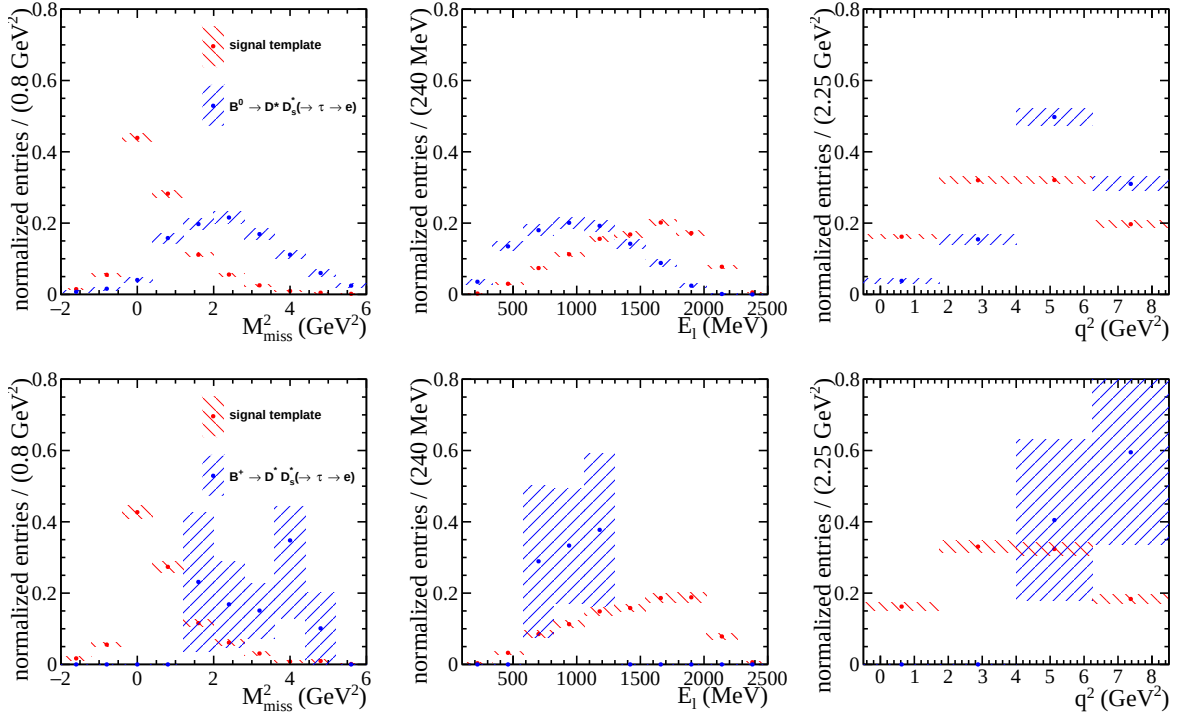


Figure 156: ● Comparison of the signal shape (red) with the background template (blue) from $B^0 \rightarrow D^* D_s^+(\rightarrow \tau^+ \nu_\tau) X$ (top) and $B^+ \rightarrow D^* D_s^+(\rightarrow \tau^+ \nu_\tau) X$ (bottom) with $\tau \rightarrow e \nu \bar{\nu}$ for 2018

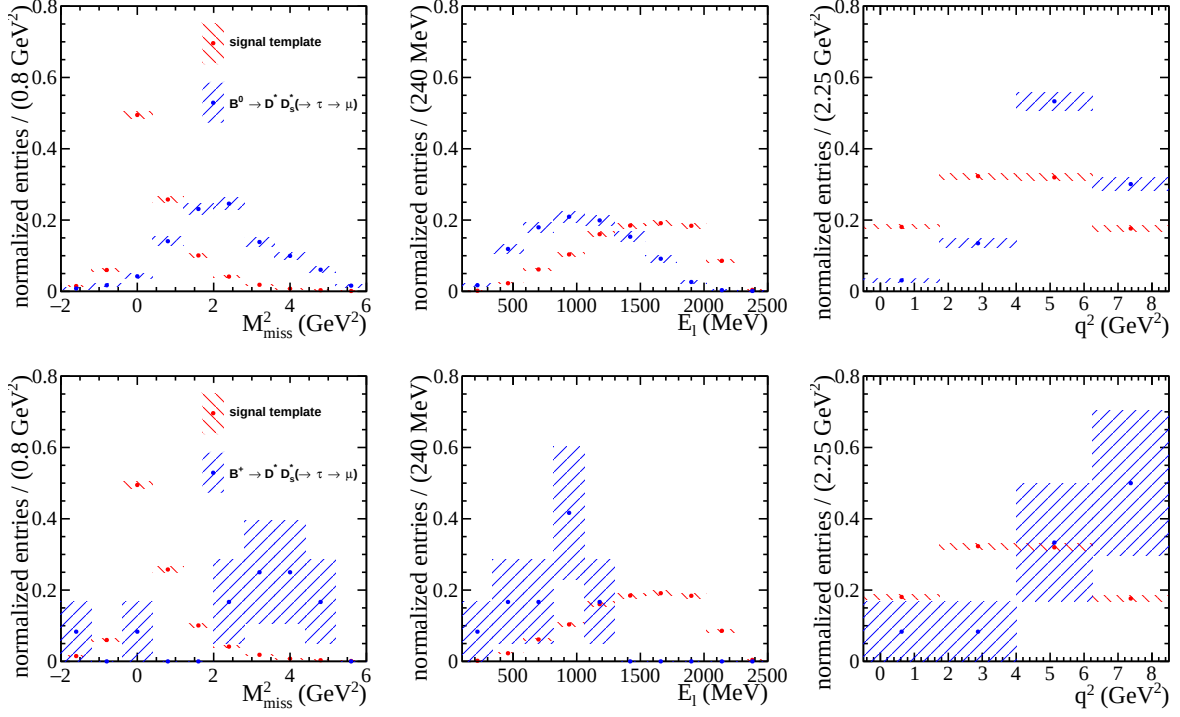


Figure 157: ● Comparison of the signal shape (red) with the background template (blue) from $B^0 \rightarrow D^* D_s^+ (\rightarrow \tau^+ \nu_\tau) X$ (top) and $B^+ \rightarrow D^* D_s^+ (\rightarrow \tau^+ \nu_\tau) X$ (bottom) with $\tau \rightarrow \mu \nu \bar{\nu}$ for 2018.

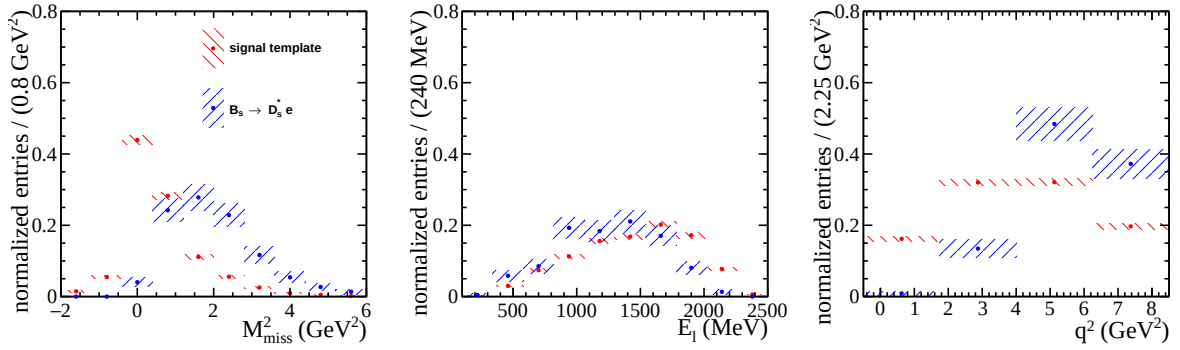


Figure 158: ● Comparison of the signal shape (red) with the background template (blue) from $B_s^0 \rightarrow D_s^{**} e \nu_e$ for 2018

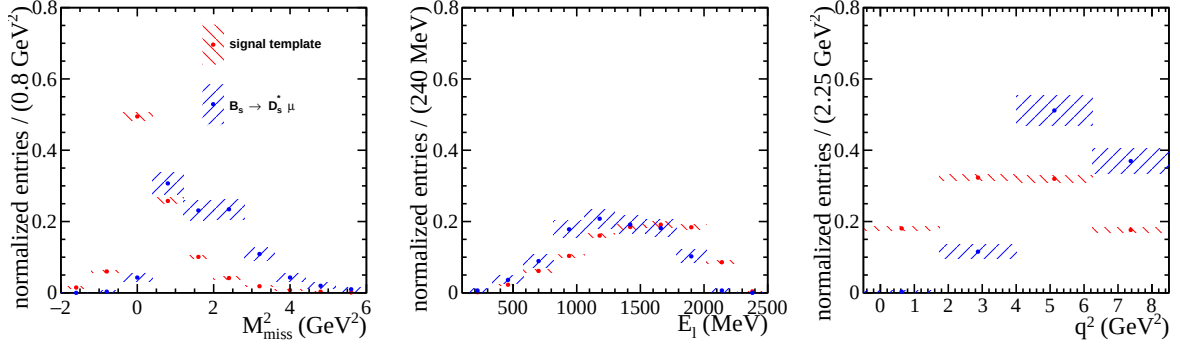


Figure 159: ● Comparison of the signal shape (red) with the background template (blue) from $B_s^0 \rightarrow D_s^{*0} \mu \nu_\mu$ for 2018.

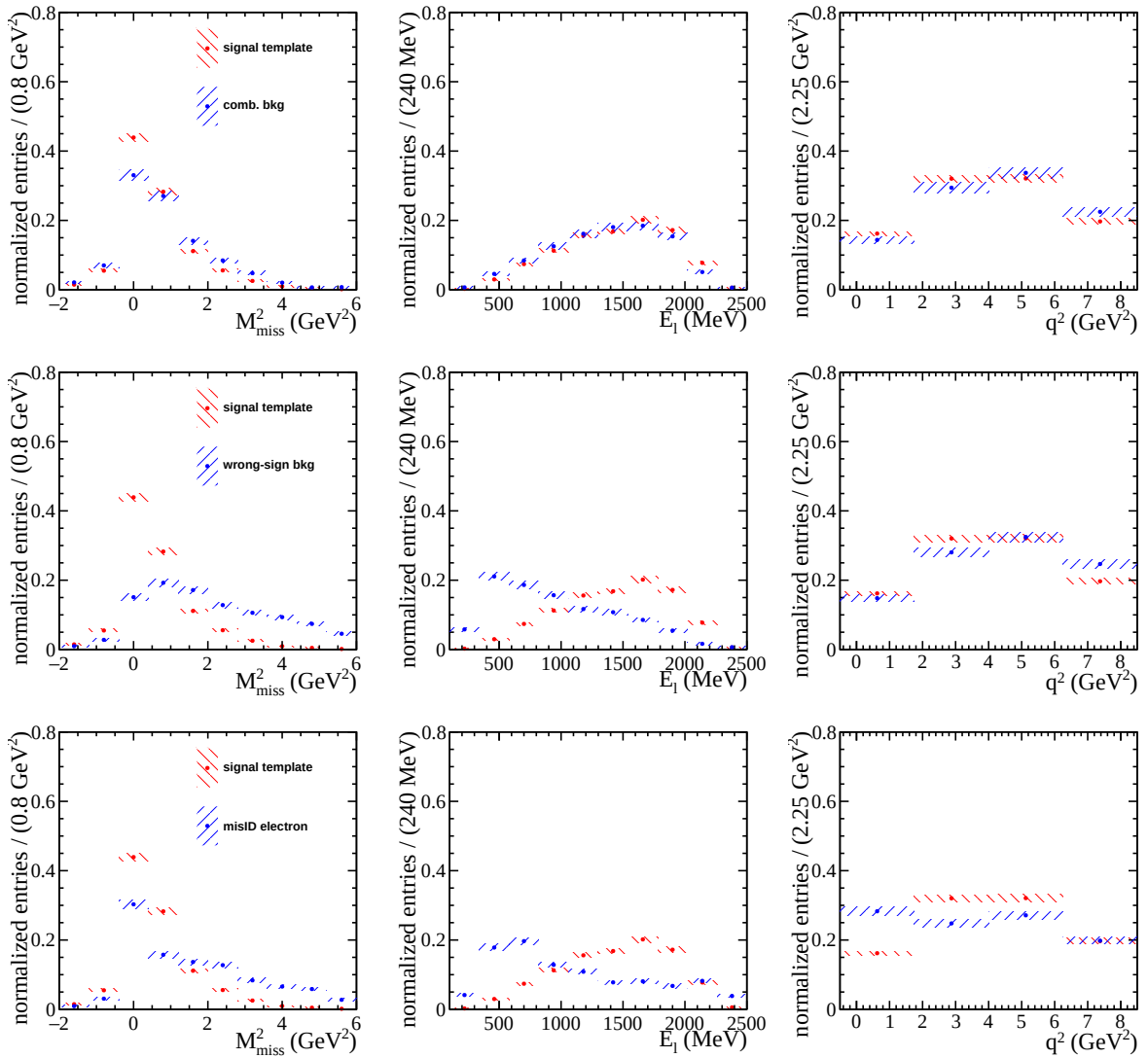


Figure 160: ● Comparison of the signal shape (red) with the combinatorial background templates (blue) for 'fake' D^* and 'real' electron combinations (top), 'real' D^* and 'real' electron combinations (middle), and 'real' D^* and 'fake' electron (bottom) combinations for 2018.

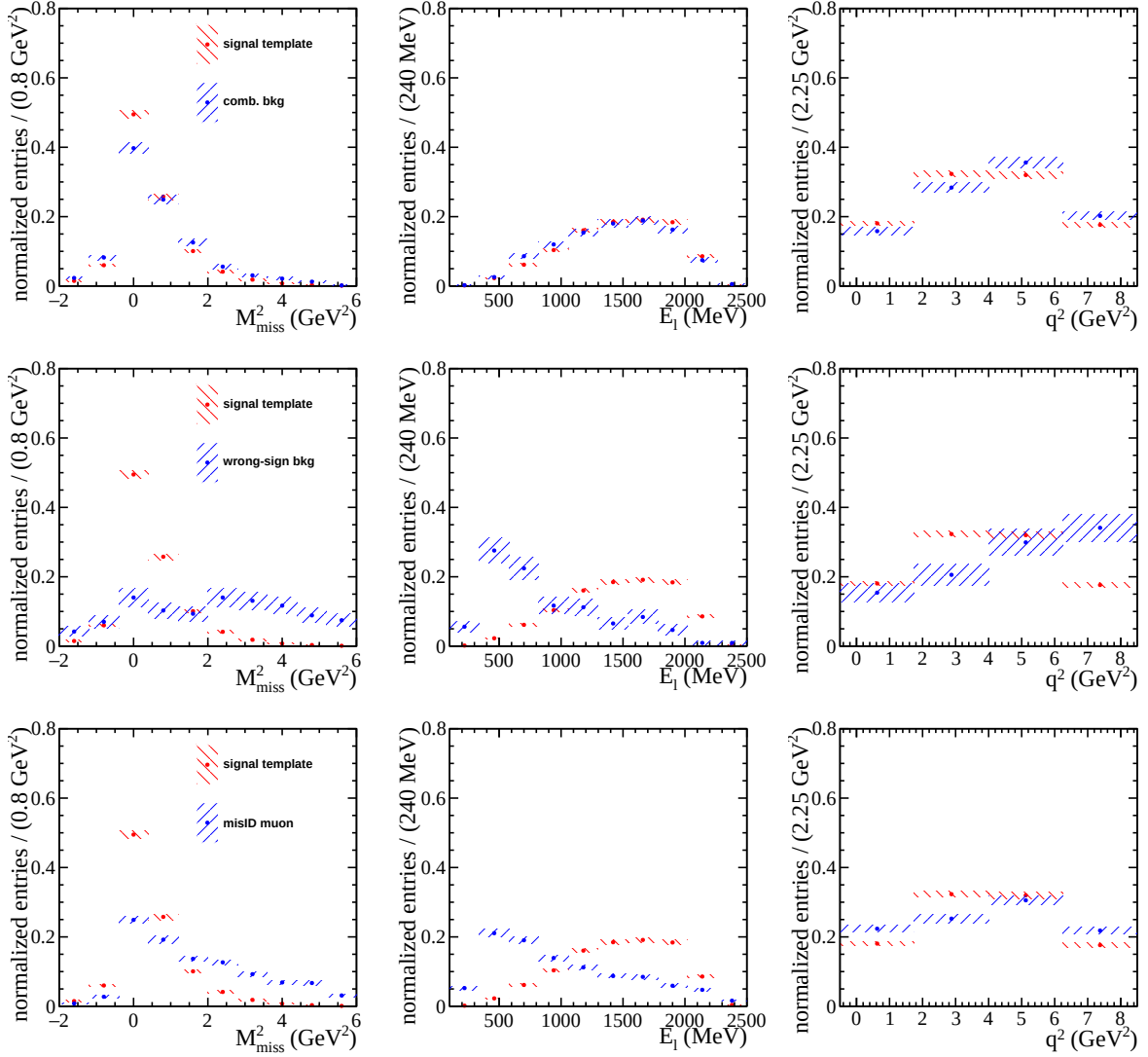


Figure 161: ● Comparison of the signal shape (red) with the combinatorial background templates for 'fake' D^* and 'real' muon combinations (top), 'real' D^* and 'real' muon combinations (middle), and 'real' D^* and 'fake' muon (bottom) combinations for 2018.

Appendix G

Template fit projections in q^2 bins

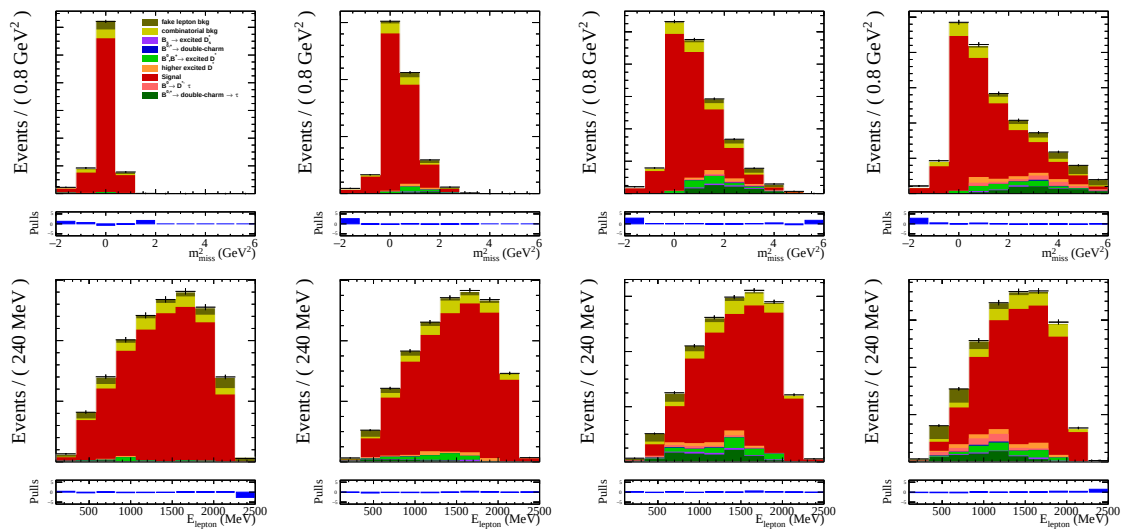


Figure 162: ● One-dimensional projections in m_{miss}^2 (top) and E_{μ}^* (bottom) in bins of q^2 of the fit result of the model to data for the muon final state for 2016.

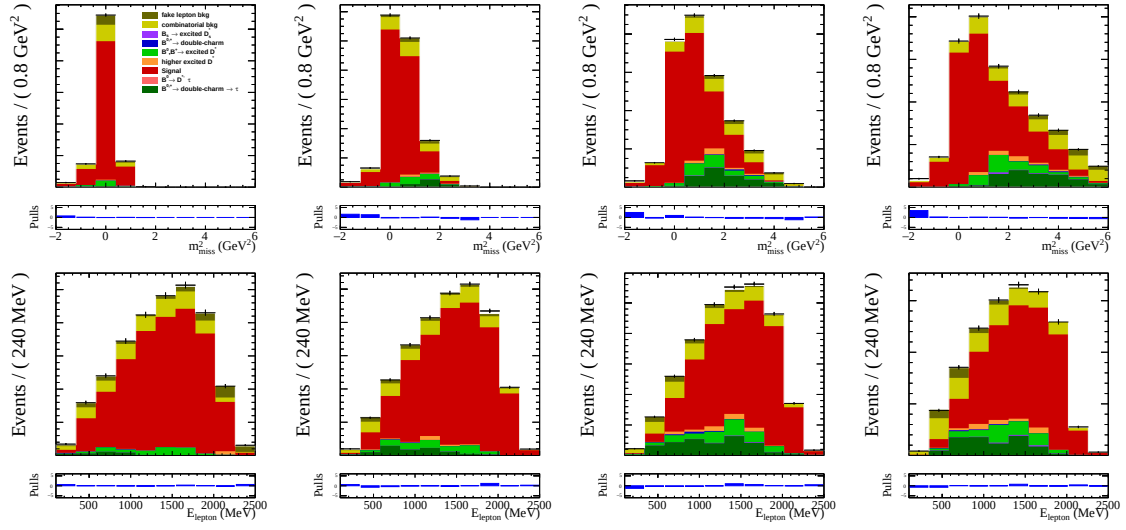


Figure 163: ● One-dimensional projections in m_{miss}^2 (top) and E_μ^* (bottom) in bins of q^2 of the fit result of the model to data for the electron final state for 2016.

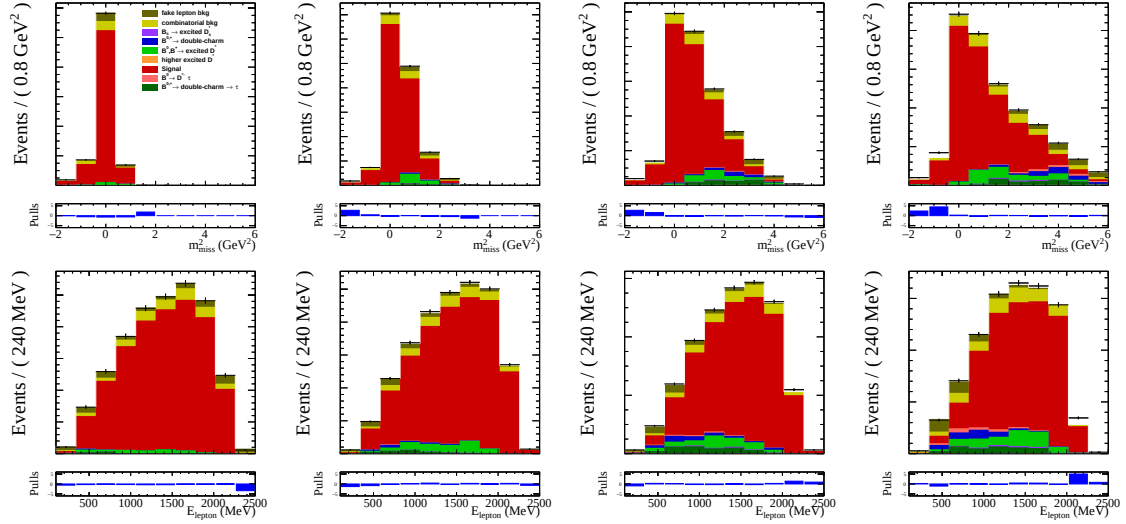


Figure 164: ● One-dimensional projections in m_{miss}^2 (top) and E_μ^* (bottom) in bins of q^2 of the fit result of the model to data for the muon final state for 2017.

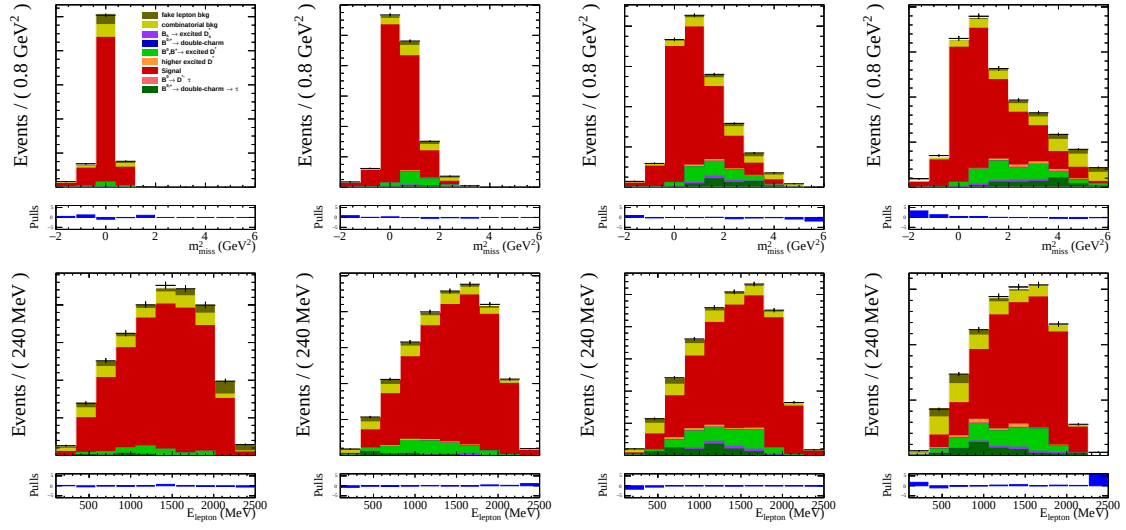


Figure 165: ● One-dimensional projections in m_{miss}^2 (top) and E_{μ}^* (bottom) in bins of q^2 of the fit result of the model to data for the electron final state for 2017.

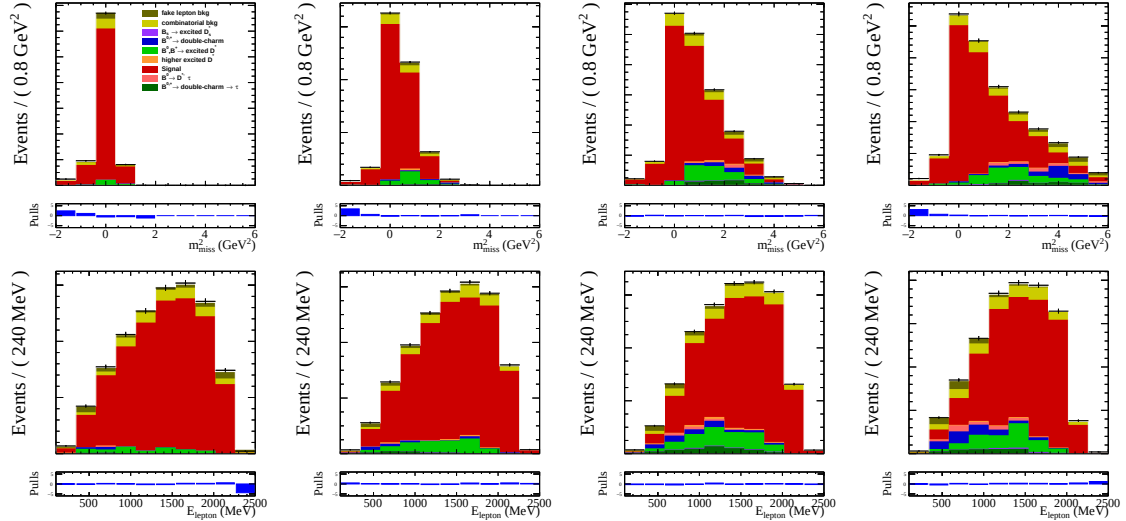


Figure 166: ● One-dimensional projections in m_{miss}^2 (top) and E_{μ}^* (bottom) in bins of q^2 of the fit result of the model to data for the muon final state for 2018.

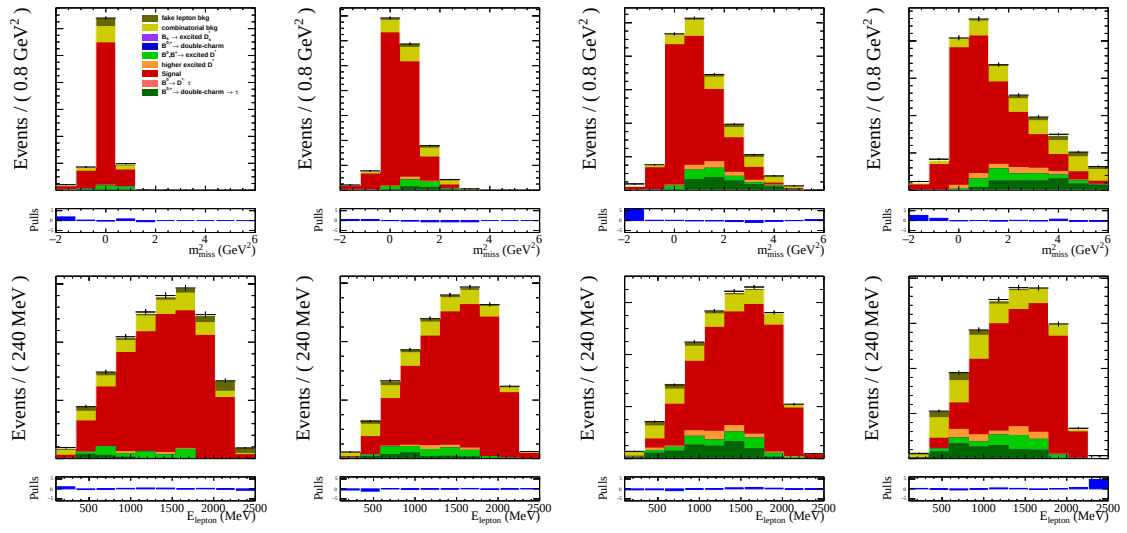


Figure 167: ● One-dimensional projections in m_{miss}^2 (top) and E_μ^* (bottom) in bins of q^2 of the fit result of the model to data for the electron final state for 2018.

Appendix H

Template fit projections without MC shape variations

To give an impression how the templates are varied by the Beeston-Barlow method, the same projections, but with the shape variation set to unity are shown in Figures 168-179 for the 2016, 2017 and 2018 samples. It should be noted that the pulls shown in this case do not take into account the uncertainty on the MC templates.

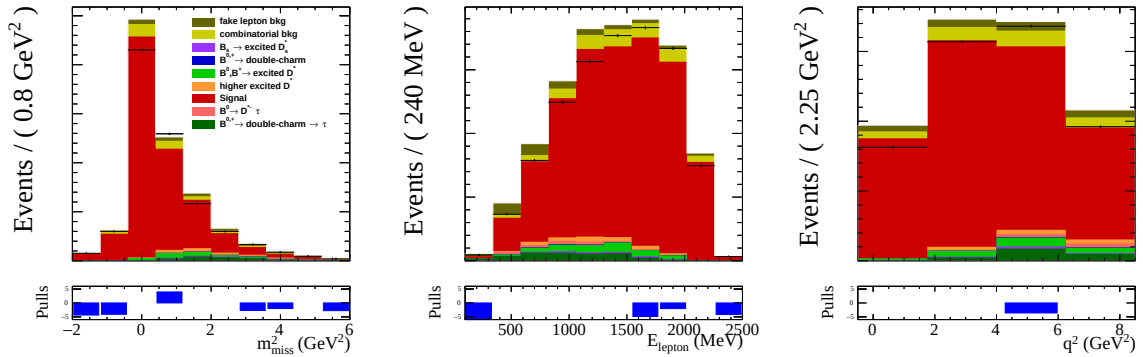


Figure 168: • One-dimensional projections in m^2_{miss} (left), E^*_μ (middle) and q^2 (right) of the fit result of the model to data for the muon final state for 2016 with the shape variations introduced by the Beeston-Barlow method set to unity.

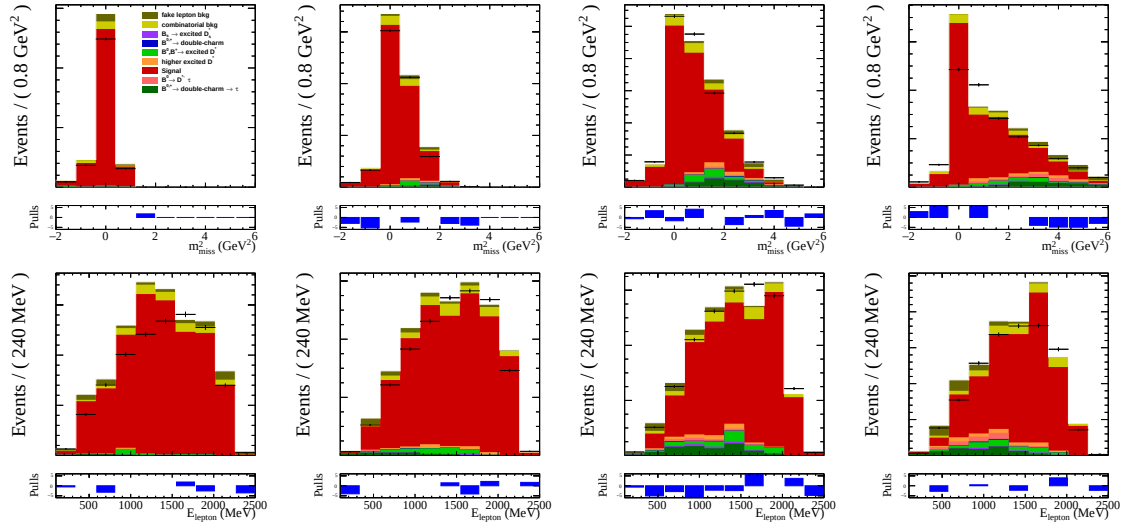


Figure 169: • One-dimensional projections in m_{miss}^2 (top) and E_{μ}^* (bottom) in bins of q^2 of the fit result of the model to data for the muon final state for 2016 with the shape variations introduced by the Beeston-Barlow method set to unity.

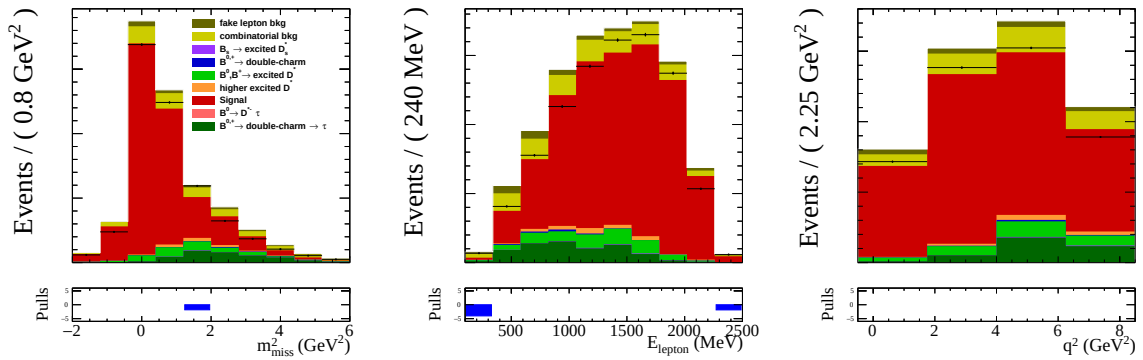


Figure 170: • One-dimensional projections in m_{miss}^2 (left), E_{μ}^* (middle) and q^2 (right) of the fit result of the model to data for the electron final state for 2016 with the shape variations introduced by the Beeston-Barlow method set to unity.

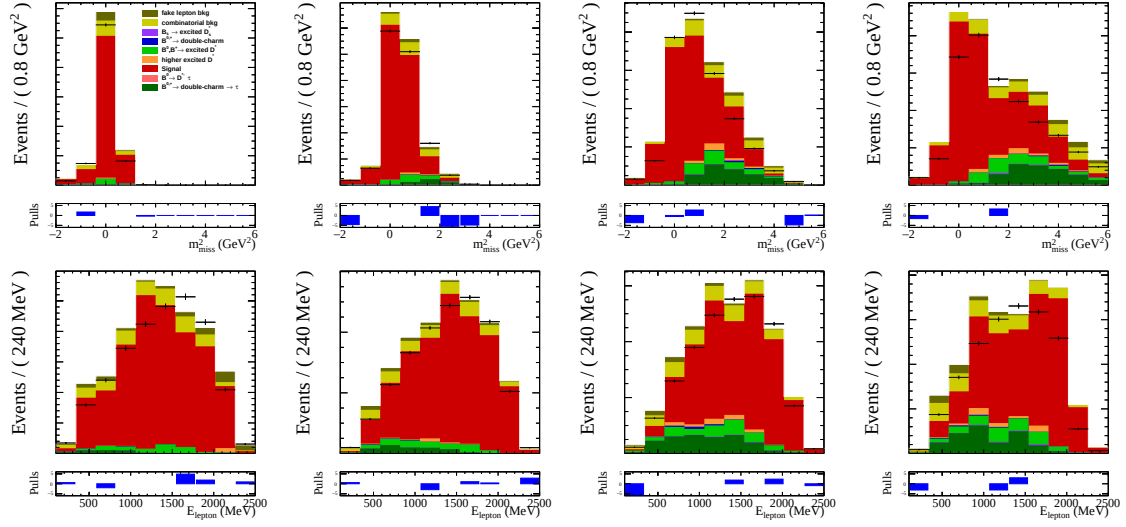


Figure 171: ● One-dimensional projections in m_{miss}^2 (top) and E_{μ}^* (bottom) in bins of q^2 of the fit result of the model to data for the electron final state for 2016 with the shape variations introduced by the Beeston-Barlow method set to unity.

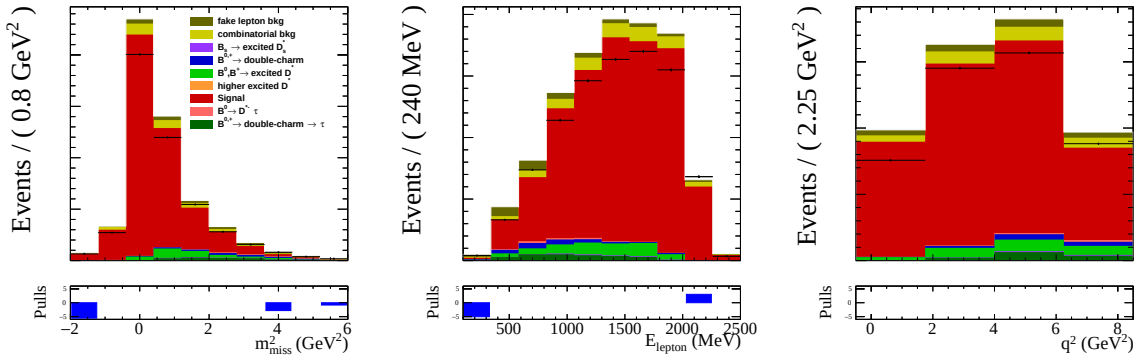


Figure 172: ● One-dimensional projections in m_{miss}^2 (left), E_{μ}^* (middle) and q^2 (right) of the fit result of the model to data for the muon final state for 2017 with the shape variations introduced by the Beeston-Barlow method set to unity.

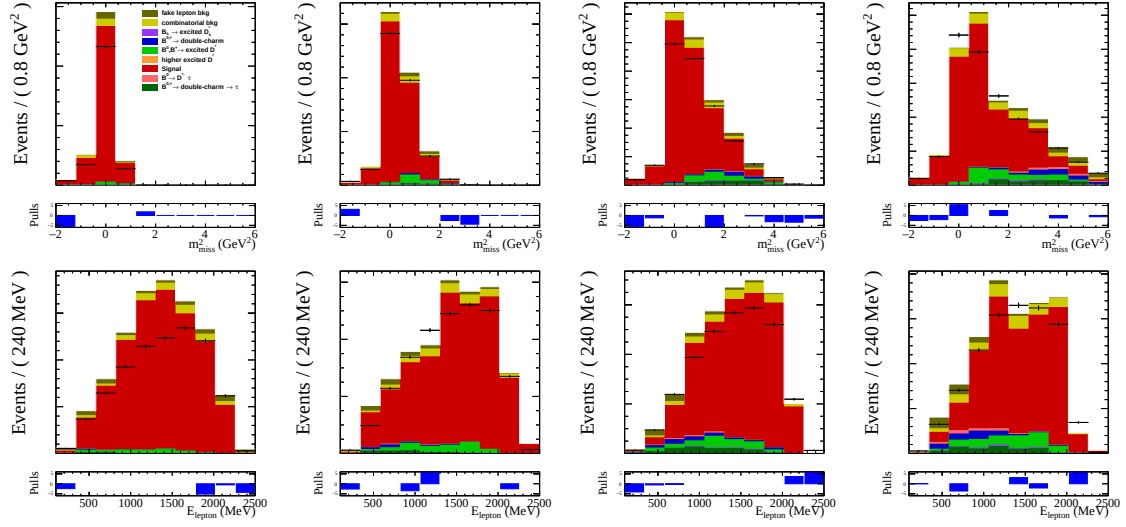


Figure 173: ● One-dimensional projections in m^2_{miss} (top) and E^*_μ (bottom) in bins of q^2 of the fit result of the model to data for the muon final state for 2017 with the shape variations introduced by the Beeston-Barlow method set to unity.

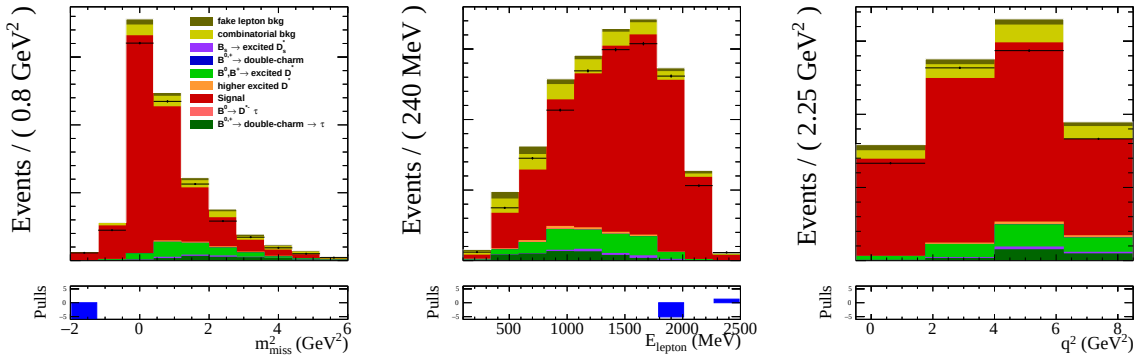


Figure 174: ● One-dimensional projections in m^2_{miss} (left), E^*_μ (middle) and q^2 (right) of the fit result of the model to data for the electron final state for 2017 with the shape variations introduced by the Beeston-Barlow method set to unity.

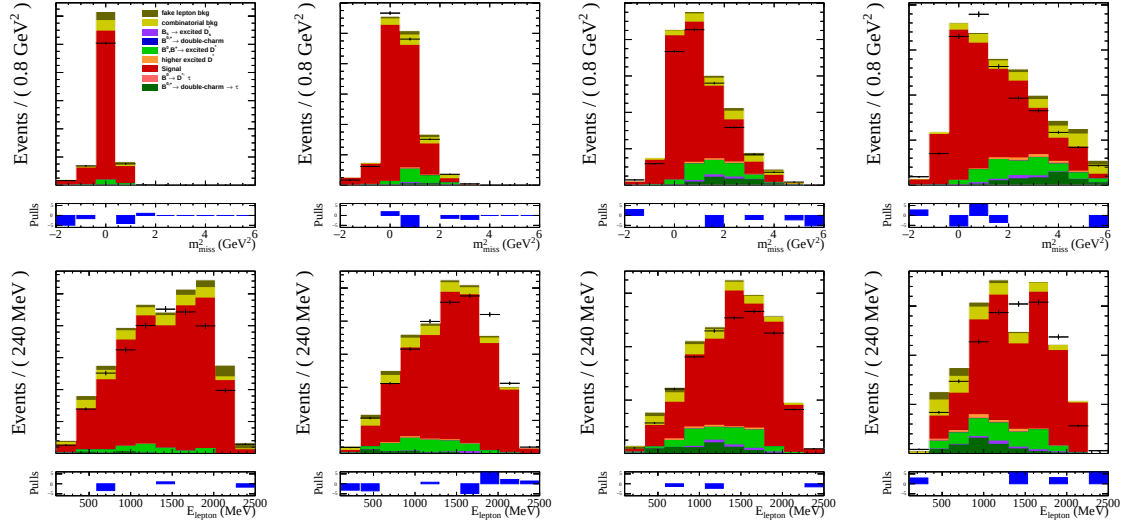


Figure 175: ● One-dimensional projections in m_{miss}^2 (top) and E_{μ}^* (bottom) in bins of q^2 of the fit result of the model to data for the electron final state for 2017 with the shape variations introduced by the Beeston-Barlow method set to unity.

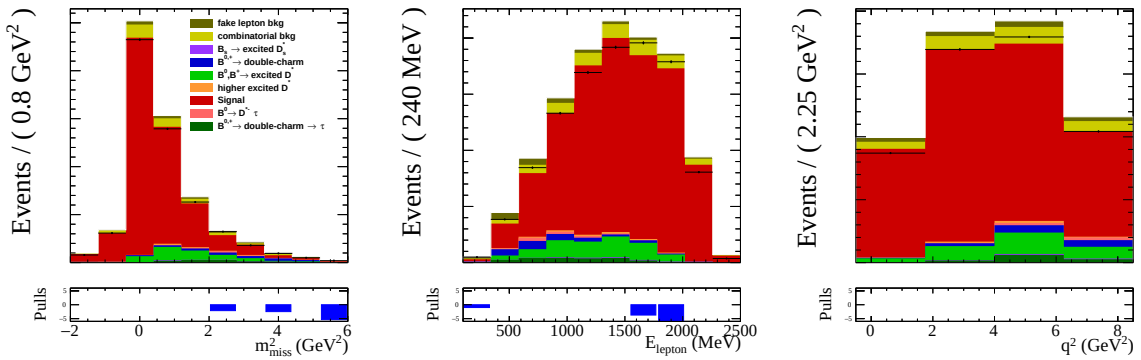


Figure 176: ● One-dimensional projections in m_{miss}^2 (left), E_{μ}^* (middle) and q^2 (right) of the fit result of the model to data for the muon final state for 2018 with the shape variations introduced by the Beeston-Barlow method set to unity.

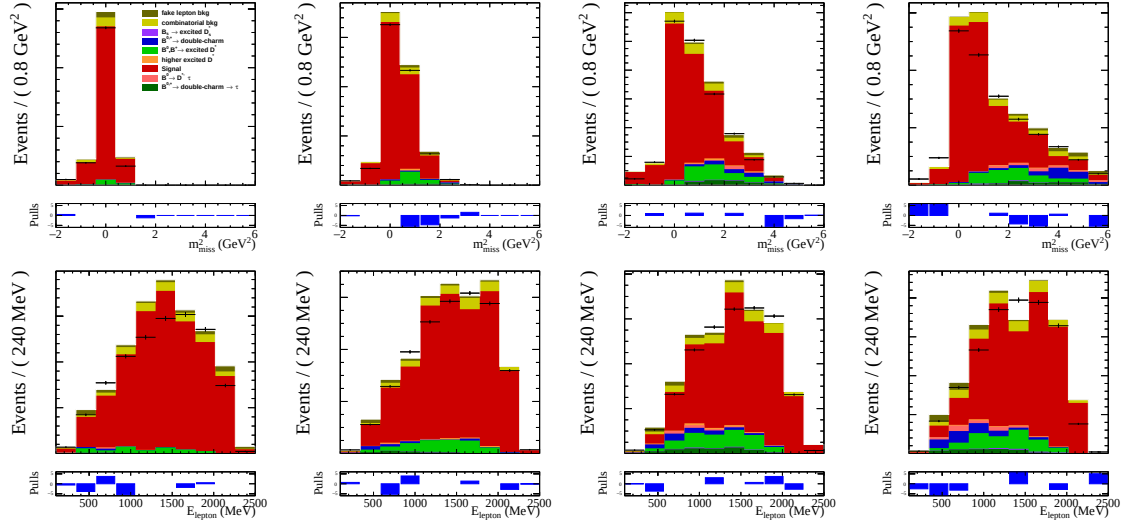


Figure 177: • One-dimensional projections in m_{miss}^2 (top) and E_{μ}^* (bottom) in bins of q^2 of the fit result of the model to data for the muon final state for 2018 with the shape variations introduced by the Beeston-Barlow method set to unity.

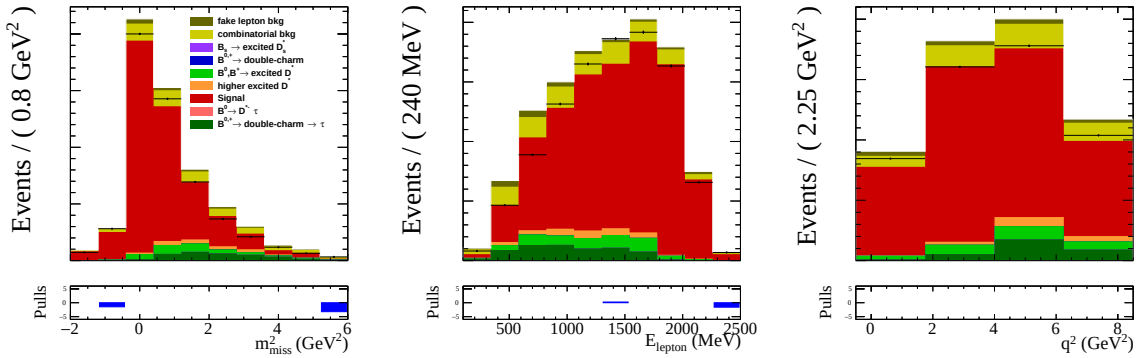


Figure 178: • One-dimensional projections in m_{miss}^2 (left), E_{μ}^* (middle) and q^2 (right) of the fit result of the model to data for the electron final state for 2018 with the shape variations introduced by the Beeston-Barlow method set to unity.

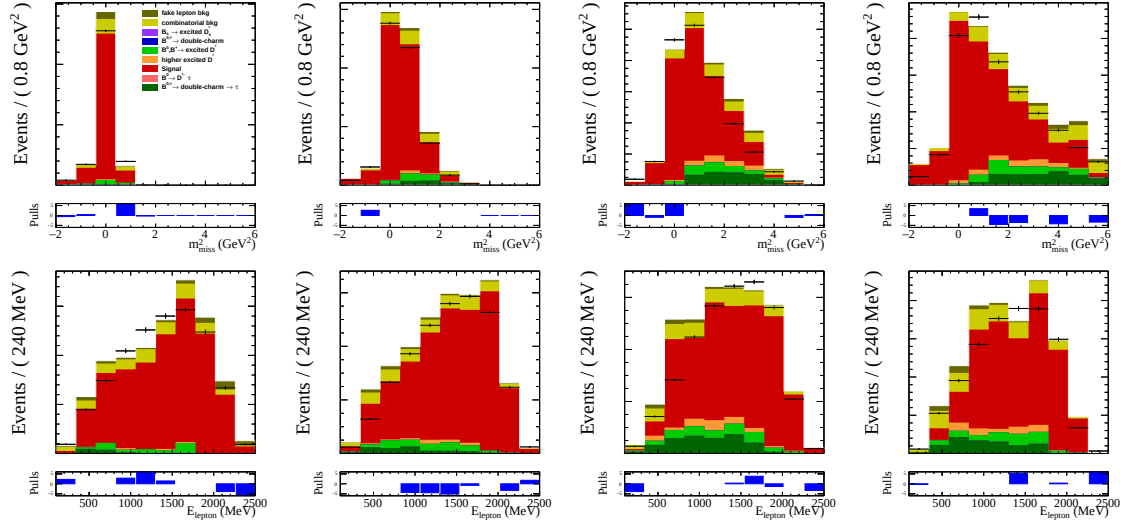


Figure 179: ● One-dimensional projections in m_{miss}^2 (top) and E_{μ}^* (bottom) in bins of q^2 of the fit result of the model to data for the electron final state for 2018 with the shape variations introduced by the Beeston-Barlow method set to unity.

Bibliography

- [1] R. Aaij *et al.*, *Allen: A high level trigger on GPUs for LHCb*, Comput. Softw. Big Sci. **4** (2020) 7, [arXiv:1912.09161](#).
- [2] https://upload.wikimedia.org/wikipedia/commons/thumb/0/00/Standard_Model_of_Elementary_Particles.svg/2000px-Standard_Model_of_Elementary_Particles.svg.png. Accessed: 2020-09-14.
- [3] ATLAS, G. Aad *et al.*, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1, [arXiv:1207.7214](#).
- [4] CMS, S. Chatrchyan *et al.*, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, Phys. Lett. B **716** (2012) 30, [arXiv:1207.7235](#).
- [5] Particle Data Group, M. Tanabashi *et al.*, *Review of Particle Physics*, Phys. Rev. D **98** (2018) 030001.
- [6] T. Aoyama *et al.*, *The anomalous magnetic moment of the muon in the standard model*, Physics Reports **887** (2020) 1 .
- [7] Muon $g-2$, G. W. Bennett *et al.*, *Final Report of the Muon E821 Anomalous Magnetic Moment Measurement at BNL*, Phys. Rev. D **73** (2006) 072003, [arXiv:hep-ex/0602035](#).
- [8] CKMfitter group, J. Charles *et al.*, *CP violation and the CKM matrix: Assessing the impact of the asymmetric B factories*, Eur. Phys. J. **C41** (2005) 1, [arXiv:hep-ph/0406184](#), updated results and plots available at <http://ckmfitter.in2p3.fr/>.
- [9] ATLAS Collaboration, G. Aad *et al.*, *Combined measurements of higgs boson production and decay using up to 80 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13 \text{ TeV}$ collected with the atlas experiment*, Phys. Rev. D **101** (2020) 012002.

- [10] SNO, Q. R. Ahmad *et al.*, *Measurement of the rate of $\nu_e + d \rightarrow p + p + e^-$ interactions produced by ^8B solar neutrinos at the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **87** (2001) 071301, [arXiv:nuc1-ex/0106015](#).
- [11] F. Iocco, M. Pato, and G. Bertone, *Evidence for dark matter in the inner Milky Way*, Nature Phys. **11** (2015) 245, [arXiv:1502.03821](#).
- [12] Planck, N. Aghanim *et al.*, *Planck 2018 results. VI. Cosmological parameters*, [arXiv:1807.06209](#).
- [13] A. D. Sakharov, *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe*, Sov. Phys. Usp. **34** (1991) 392.
- [14] ALEPH, DELPHI, L3, OPAL, SLD, LEP Electroweak Working Group, SLD Electroweak Group, SLD Heavy Flavour Group, S. Schael *et al.*, *Precision electroweak measurements on the Z resonance*, Phys. Rept. **427** (2006) 257, [arXiv:hep-ex/0509008](#).
- [15] ALEPH, DELPHI, L3, OPAL, LEP Electroweak, S. Schael *et al.*, *Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP*, Phys. Rept. **532** (2013) 119, [arXiv:1302.3415](#).
- [16] ATLAS Collaboration, *Test of the universality of τ and μ lepton couplings in W boson decays from $t\bar{t}$ events at 13 TeV with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2020-014, CERN, Geneva, 2020.
- [17] ATLAS, M. Aaboud *et al.*, *Precision measurement and interpretation of inclusive W^+ , W^- and Z/γ^* production cross sections with the ATLAS detector*, Eur. Phys. J. C **77** (2017) 367, [arXiv:1612.03016](#).
- [18] Particle Data Group, P. A. Zyla *et al.*, *Review of Particle Physics*, PTEP **2020** (2020) 083C01.
- [19] KEDR, V. M. Aulchenko *et al.*, *Measurement of the ratio of the leptonic widths $\Gamma_{ee}/\Gamma_{\mu\mu}$ for the J/ψ meson*, Phys. Lett. B **731** (2014) 227, [arXiv:1311.5005](#).
- [20] LHCb, R. Aaij *et al.*, *Search for lepton-universality violation in $B^+ \rightarrow K^+ \ell^+ \ell^-$ decays*, Phys. Rev. Lett. **122** (2019) 191801, [arXiv:1903.09252](#).
- [21] LHCb, R. Aaij *et al.*, *Test of lepton universality with $\Lambda_b^0 \rightarrow p K^- \ell^+ \ell^-$ decays*, [arXiv:1912.08139](#).
- [22] LHCb, R. Aaij *et al.*, *Test of lepton universality with $B^0 \rightarrow K^{*0} \ell^+ \ell^-$ decays*, JHEP **08** (2017) 055, [arXiv:1705.05802](#).

- [23] LHCb, R. Aaij *et al.*, *Angular analysis and differential branching fraction of the decay $B_s^0 \rightarrow \phi \mu^+ \mu^-$* , JHEP **09** (2015) 179, [arXiv:1506.08777](#).
- [24] LHCb, R. Aaij *et al.*, *Measurement of CP -Averaged Observables in the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ Decay*, Phys. Rev. Lett. **125** (2020) 011802, [arXiv:2003.04831](#).
- [25] LHCb, R. Aaij *et al.*, *Test of Lepton Flavor Universality by the measurement of the $B^0 \rightarrow D^{*-} \tau^+ \nu_\tau$ branching fraction using three-prong τ decays*, Phys. Rev. D **97** (2018) 072013, [arXiv:1711.02505](#).
- [26] LHCb, R. Aaij *et al.*, *Measurement of the ratio of the $B^0 \rightarrow D^{*-} \tau^+ \nu_\tau$ and $B^0 \rightarrow D^{*-} \mu^+ \nu_\mu$ branching fractions using three-prong τ -lepton decays*, Phys. Rev. Lett. **120** (2018) 171802, [arXiv:1708.08856](#).
- [27] LHCb, R. Aaij *et al.*, *Measurement of the ratio of branching fractions $\mathcal{B}(B_c^+ \rightarrow J/\psi \tau^+ \nu_\tau)/\mathcal{B}(B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu)$* , Phys. Rev. Lett. **120** (2018) 121801, [arXiv:1711.05623](#).
- [28] HFLAV, Y. S. Amhis *et al.*, *Averages of b -hadron, c -hadron, and τ -lepton properties as of 2018*, [arXiv:1909.12524](#), updated results and plots available at <https://hflav.web.cern.ch/>.
- [29] NA62, C. Lazzeroni *et al.*, *Precision Measurement of the Ratio of the Charged Kaon Leptonic Decay Rates*, Phys. Lett. B **719** (2013) 326, [arXiv:1212.4012](#).
- [30] V. Cirigliano and I. Rosell, *Two-loop effective theory analysis of π (K) $\rightarrow e$ anti- ν_e/γ branching ratios*, Phys. Rev. Lett. **99** (2007) 231801, [arXiv:0707.3439](#).
- [31] PiENU, A. Aguilar-Arevalo *et al.*, *Improved Measurement of the $\pi \rightarrow e \nu$ Branching Ratio*, Phys. Rev. Lett. **115** (2015) 071801, [arXiv:1506.05845](#).
- [32] BaBar, B. Aubert *et al.*, *Measurements of Charged Current Lepton Universality and $|V_{us}|$ using Tau Lepton Decays to $e^- \bar{\nu}_e \nu_\tau$, $\mu^- \bar{\nu}_\mu \nu_\tau$, $\pi^- \nu_\tau$, and $K^- \nu_\tau$* , Phys. Rev. Lett. **105** (2010) 051602, [arXiv:0912.0242](#).
- [33] Belle, E. Waheed *et al.*, *Measurement of the CKM matrix element $|V_{cb}|$ from $B^0 \rightarrow D^{*-} \ell^+ \nu_\ell$ at Belle*, Phys. Rev. D **100** (2019) 052007, [arXiv:1809.03290](#).
- [34] M. Neubert, *Heavy quark effective theory*, Subnucl. Ser. **34** (1997) 98, [arXiv:hep-ph/9610266](#).
- [35] S. Fajfer, J. F. Kamenik, and I. Nisandzic, *On the $B \rightarrow D^* \tau \bar{\nu}_\tau$ Sensitivity to New Physics*, Phys. Rev. D **85** (2012) 094025, [arXiv:1203.2654](#).

- [36] N. Isgur and M. B. Wise, *Weak decays of heavy mesons in the static quark approximation*, Physics Letters B **232** (1989) 113 .
- [37] N. Isgur and M. B. Wise, *Weak transition form factors between heavy mesons*, Phys. Lett. B **237** (1990) 527.
- [38] I. Caprini, L. Lellouch, and M. Neubert, *Dispersive bounds on the shape of anti- $B \rightarrow D^{(*)}$ lepton anti-neutrino form-factors*, Nucl. Phys. B **530** (1998) 153, [arXiv:hep-ph/9712417](#).
- [39] C. G. Boyd, B. Grinstein, and R. F. Lebed, *Precision corrections to dispersive bounds on form-factors*, Phys. Rev. D **56** (1997) 6895, [arXiv:hep-ph/9705252](#).
- [40] S. Jaiswal, S. Nandi, and S. K. Patra, *Extraction of $|V_{cb}|$ from $B \rightarrow D^{(*)}\ell\nu_\ell$ and the Standard Model predictions of $R(D^{(*)})$* , JHEP **12** (2017) 060, [arXiv:1707.09977](#).
- [41] F. U. Bernlochner, Z. Ligeti, M. Papucci, and D. J. Robinson, *Combined analysis of semileptonic B decays to D and D^* : $R(D^{(*)})$, $|V_{cb}|$, and new physics*, Phys. Rev. D **95** (2017) 115008, [arXiv:1703.05330](#), [Erratum: Phys.Rev.D 97, 059902 (2018)].
- [42] D. Bigi, P. Gambino, and S. Schacht, *$R(D^*)$, $|V_{cb}|$, and the Heavy Quark Symmetry relations between form factors*, JHEP **11** (2017) 061, [arXiv:1707.09509](#).
- [43] S. L. Glashow, D. Guadagnoli, and K. Lane, *Lepton Flavor Violation in B Decays?*, Phys. Rev. Lett. **114** (2015) 091801, [arXiv:1411.0565](#).
- [44] A. K. Alok, A. Dighe, S. Gangal, and D. Kumar, *Continuing search for new physics in $b \rightarrow s\mu\mu$ decays: two operators at a time*, JHEP **06** (2019) 089, [arXiv:1903.09617](#).
- [45] M. Freytsis, Z. Ligeti, and J. T. Ruderman, *Flavor models for $\bar{B} \rightarrow D^{(*)}\tau\bar{\nu}$* , Phys. Rev. D **92** (2015) 054018, [arXiv:1506.08896](#).
- [46] A. K. Alok *et al.*, *New physics solutions for R_D and R_{D^*}* , JHEP **09** (2018) 152, [arXiv:1710.04127](#).
- [47] A. G. Akeroyd and C.-H. Chen, *Constraint on the branching ratio of $B_c \rightarrow \tau\bar{\nu}$ from LEP1 and consequences for $R(D^{(*)})$ anomaly*, Phys. Rev. D **96** (2017) 075011, [arXiv:1708.04072](#).

- [48] A. K. Alok, D. Kumar, S. Kumbhakar, and S. Uma Sankar, *Resolution of R_D/R_{D^*} puzzle*, Phys. Lett. B **784** (2018) 16, [arXiv:1804.08078](#).
- [49] A. K. Alok, D. Kumar, S. Kumbhakar, and S. Uma Sankar, *Solutions to $R_D-R_{D^*}$ in light of Belle 2019 data*, Nucl. Phys. B **953** (2020) 114957, [arXiv:1903.10486](#).
- [50] C. Lefevre, *LHC guide book (French version)*. LHC guide book (version français), <http://cds.cern.ch/record/1164451/>. Accessed: 2020-10-22.
- [51] LHCb, J. Alves, A. Augusto *et al.*, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.
- [52] LHCb collaboration, C. Elsässer, *$\bar{b}b$ production angle plots*, https://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb_ProductionAngles.html. Accessed: 2020-06-02.
- [53] C. Abellan Beteta *et al.*, *Monitoring radiation damage in the LHCb Tracker Turicensis*, [arXiv:1809.05063](#).
- [54] C. A. Beteta *et al.*, *Calibration and performance of the LHCb calorimeters in Run 1 and 2 at the LHC*, [arXiv:2008.11556](#).
- [55] J. Alves, A. A. *et al.*, *Performance of the LHCb muon system*, JINST **8** (2013) P02022, [arXiv:1211.1346](#).
- [56] R. E. Kalman, *A new approach to linear filtering and prediction problems*, Transactions of the ASME—Journal of Basic Engineering **82** (1960) 35.
- [57] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Meth. A **262** (1987) 444.
- [58] LHCb collaboration, *Track types for the lhcb run 1 and 2*, <https://twiki.cern.ch/twiki/pub/LHCb/ConferencePlots/trackTypesRunIAndII.pdf>. Accessed: 2020-10-22.
- [59] LHCb, R. Aaij *et al.*, *Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC*, JINST **14** (2019) P04013, [arXiv:1812.10790](#).
- [60] LHCb, R. Aaij *et al.*, *LHCb Detector Performance*, Int. J. Mod. Phys. A **30** (2015) 1530022, [arXiv:1412.6352](#).
- [61] LHCb Collaboration, *Muon track reconstruction efficiency at LHCb*, .

- [62] LHCb, R. Aaij *et al.*, *Measurement of the electron reconstruction efficiency at LHCb*, **arXiv:1909.02957**.
- [63] LHCb collaboration, *Kaon identification efficiency and pion misidentification rate*, https://twiki.cern.ch/twiki/pub/LHCb/PIDConferencePlots/Plot_2017Validation_KPi_P.pdf. Accessed: 2020-10-22.
- [64] F. Archilli *et al.*, *Performance of the Muon Identification at LHCb*, JINST **8** (2013) P10020, **arXiv:1306.0249**.
- [65] LHCb Collaboration, C. Marin Benito, *PID strategy and performance at LHCb in Run 2*, PoS **ICHEP2018** (2019) 687. 4 p.
- [66] R. Aaij *et al.*, *A comprehensive real-time analysis model at the LHCb experiment*, JINST **14** (2019) P04006, **arXiv:1903.01360**.
- [67] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, *Data driven trigger efficiency determination at LHCb*, Tech. Rep. LHCb-PUB-2014-039. CERN-LHCb-PUB-2014-039, CERN, Geneva, 2014.
- [68] R. Aaij *et al.*, *Optimization of the muon reconstruction algorithms for LHCb Run 2*, Tech. Rep. LHCb-PUB-2017-007. CERN-LHCb-PUB-2017-007, CERN, Geneva, 2017.
- [69] V. V. Gligorov and M. Williams, *Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree*, JINST **8** (2013) P02013, **arXiv:1210.6861**.
- [70] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. A **389** (1997) 81.
- [71] W. D. Hulsbergen, *Decay chain fitting with a Kalman filter*, Nucl. Instrum. Meth. A **552** (2005) 566, **arXiv:physics/0503191**.
- [72] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, **arXiv:0710.3820**.
- [73] I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, J. Phys. Conf. Ser. **331** (2011) 032047.
- [74] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152.

- [75] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A precision tool for QED corrections in Z and W decays*, Eur. Phys. J. **C45** (2006) 97, [arXiv:hep-ph/0506026](#).
- [76] Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.
- [77] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270.
- [78] F. U. Bernlochner *et al.*, *Das ist der HAMMER: Consistent new physics interpretations of semileptonic decays*, Eur. Phys. J. C **80** (2020) 883, [arXiv:2002.00020](#).
- [79] LHCb, R. Aaij *et al.*, *Measurement of the ratio of branching fractions $\mathcal{B}(\bar{B}^0 \rightarrow D^{*+} \tau^- \bar{\nu}_\tau)/\mathcal{B}(\bar{B}^0 \rightarrow D^{*+} \mu^- \bar{\nu}_\mu)$* , Phys. Rev. Lett. **115** (2015) 111803, [arXiv:1506.08614](#), [Erratum: Phys.Rev.Lett. 115, 159901 (2015)].
- [80] LHCb Collaboration, *Measurement of the electron reconstruction efficiency at LHCb*, Journal of Instrumentation **14** (2019) P11023.
- [81] J. Therhaag, *TMVA Toolkit for multivariate data analysis in ROOT*, PoS **ICHEP2010** (2010) 510.
- [82] M. Pivk and F. R. Le Diberder, *SPlot: A Statistical tool to unfold data distributions*, Nucl. Instrum. Meth. A **555** (2005) 356, [arXiv:physics/0402083](#).
- [83] T. Skwarnicki, *A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances*, PhD thesis, Institute of Nuclear Physics, Krakow, 1986, DESY-F31-86-02.
- [84] Particle Data Group, M. Tanabashi *et al.*, *Review of particle physics*, Phys. Rev. **D98** (2018) 030001.
- [85] L. Anderlini *et al.*, *The PIDCalib package*, Tech. Rep. LHCb-PUB-2016-021. CERN-LHCb-PUB-2016-021, CERN, Geneva, 2016.
- [86] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, eConf **C0303241** (2003) MOLT007, [arXiv:physics/0306116](#).
- [87] R. Aaij *et al.*, *Selection and processing of calibration samples to measure the particle identification performance of the lhcb experiment in run 2*, EPJ Techniques and Instrumentation **6** (2019) .

- [88] A. Poluektov, *Kernel density estimation of a multidimensional efficiency profile*, Journal of Instrumentation **10** (2015) P02011.
- [89] LHCb, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, JINST **10** (2015) P02007, [arXiv:1408.1251](#).
- [90] ROOT Collaboration, K. Cranmer *et al.*, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, Tech. Rep. CERN-OPEN-2012-016, New York U., New York, 2012.
- [91] R. J. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, Comput. Phys. Commun. **77** (1993) 219.
- [92] J. Albrecht, C. Fitzpatrick, V. Gligorov, and G. Raven, *The upgrade of the LHCb trigger system*, JINST **9** (2014) C10026, [arXiv:1410.5012](#).
- [93] LHCb collaboration, *LHCb Trigger and Online Upgrade Technical Design Report*, CERN-LHCC-2014-016, 2014.
- [94] LHCb Collaboration, *LHCb Upgrade GPU High Level Trigger Technical Design Report*, Tech. Rep. CERN-LHCC-2020-006. LHCb-TDR-021, CERN, Geneva, 2020.
- [95] LHCb Collaboration, *LHCb VELO Upgrade Technical Design Report*, Tech. Rep. CERN-LHCC-2013-021. LHCb-TDR-013, 2013.
- [96] LHCb, *LHCb Tracker Upgrade Technical Design Report*, CERN-LHCC-2014-001, LHCb-TDR-015, 2014.
- [97] LHCb Collaboration, *RTA and DPA dataflow diagrams for Run 1, Run 2, and the upgraded LHCb detector*, <http://cds.cern.ch/record/2730181>. LHCb-FIGURE-2020-016.
- [98] J. Nickolls, I. Buck, M. Garland, and K. Skadron, *Scalable parallel programming with cuda*, Queue **6** (2008) 40–53.
- [99] LHCb Collaboration, *Run 3 Primary Vertex reconstruction description and performance*, <http://cds.cern.ch/record/2714851>. LHCb-FIGURE-2020-005.
- [100] M. Kucharczyk, P. Morawski, and M. Witek, *Primary Vertex Reconstruction at LHCb*, Tech. Rep. LHCb-PUB-2014-044. CERN-LHCb-PUB-2014-044, CERN, Geneva, 2014.

- [101] P. Billoir, M. De Cian, and S. Stemmle, *A parametrized Kalman filter for fast track fitting at LHCb*, , in preparation.
- [102] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions*, John Wiley & Sons, NewYork, 1986.
- [103] R. Frühwirth and W. Waltenberger, *Adaptive Multi-vertex fitting*, Tech. Rep. CMS-CR-2004-062, CERN, Geneva, 2004. doi: 10.5170/CERN-2005-002.280.