

Testing lepton universality in penguin decays of beauty mesons using the LHCb detector

Thibaud Humair

High Energy Physics
Blackett Laboratory
Imperial College London

CERN-THESIS-2019-044
02/05/2019



A thesis submitted to Imperial College London
for the degree of Doctor of Philosophy.

March 2019

Abstract

Lepton flavour universality is tested by measuring the ratio of branching fractions of $B^+ \rightarrow K^+ \mu^+ \mu^-$ and $B^+ \rightarrow K^+ e^+ e^-$ decays, R_K . The measurement is performed in the di-lepton mass squared region $1.1 < q^2 < 6.0 \text{ GeV}^2$, using approximately 5 fb^{-1} of data collected with the LHCb detector between 2011 and 2016 at centre-of-mass energies of 7, 8 and 13 TeV. The value of R_K is measured to be

$$R_K = 0.846^{+0.060}_{-0.054} (\text{stat.})^{+0.016}_{-0.014} (\text{syst.}),$$

where the first uncertainty is statistical and the second is systematic. This result is compatible with the Standard Model expectation at the level of 2.5 standard deviations. This analysis has an uncertainty 36% smaller than previous measurements of this quantity.

The analysis is performed using $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ and $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ decays as control and normalisation channels. Innovative calibration techniques and cross-checks are employed. The cross-checks demonstrate that the electron detection efficiencies are well understood, and that the LHCb detector can be used to study precisely and reliably the decays of beauty mesons to final states with electrons.

Declaration of originality

The analysis presented in this thesis is the result of the research I performed between October 2015 and December 2018, with the support of the Imperial College High Energy Physics group and members of the LHCb Collaboration. All the analysis work presented in this thesis was performed by myself, with two exceptions. Firstly, although I developed the fit strategy presented in Sec. 8.4, Dr Paula Álvarez Cartelle contributed to implementing the fit, performed the fit validation shown in Sec. 8.6 and extracted the confidence interval on R_K . Secondly, Dr Álvarez Cartelle also evaluated the systematic uncertainties related to the decay model in Sec. 9.9 and to the fit shape in Sec. 9.10. All results and figures presented in this thesis that were not the product of my own work are appropriately referenced.

This thesis has not been submitted for any other qualification.

Thibaud Humair, March 2019

Copyright declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Contents

List of figures	16
List of tables	18
1 Introduction	19
I Theoretical and experimental overview	20
2 Flavour in the Standard Model and beyond	21
2.1 Fields and Symmetries in the Standard Model	21
2.2 The weak interaction and electroweak unification	24
2.3 Higgs mechanism and origin of mass	26
2.4 Quantum Chromodynamics	30
2.5 Indirect searches for New Physics	31
2.6 Lepton flavour and lepton flavour violation	31
2.7 Flavour changing neutral currents	33
2.8 Flavour changing neutral currents of B hadrons	34
2.9 Computation of observables in $B \rightarrow K\ell^+\ell^-$ decays	35
2.10 Review of $b \rightarrow s\mu^+\mu^-$ experimental results	38
2.11 The R_K and R_{K^*} measurements	39
2.12 Anomalies in tree level beauty decays	41
2.13 Complete models explaining the B anomalies	42
2.14 Purpose of the thesis	45
3 Experimental apparatus	46
3.1 The Large Hadron Collider	46
3.2 Interaction of charged particles with matter	47
3.2.1 Energy loss by ionisation and excitation	47
3.2.2 Energy loss by bremsstrahlung	48
3.2.3 Electromagnetic showers	49
3.2.4 Hadronic interaction and showers	49
3.2.5 Cherenkov radiation	49

3.3	Layout of the LHCb detector	50
3.4	Tracking and vertexing	51
3.4.1	Vertex locator	51
3.4.2	Tracking system	51
3.4.3	Magnet	52
3.4.4	Track reconstruction and performance	53
3.5	Particle identification	53
3.6	Cherenkov detectors	54
3.7	Calorimeter system	55
3.8	Muon system	56
3.9	Bremsstrahlung recovery	57
3.10	Trigger system	58
3.10.1	Low level trigger	58
3.10.2	High level trigger	59
3.11	Material in the LHCb detector	59
II	The R_K analysis	61
4	Analysis strategy	62
4.1	$K^+e^+e^-$ and $K^+\mu^+\mu^-$ final states at LHCb	62
4.2	Normalisation and control modes	65
4.3	Outline	68
5	Data and simulation samples	69
5.1	Data samples	69
5.2	Simulation samples	69
6	Selection	70
6.1	Types of backgrounds	70
6.1.1	Charmed and strange partially-reconstructed backgrounds	71
6.1.2	Peaking backgrounds with a single mis-identified track	71
6.1.3	Peaking backgrounds with two mis-identified tracks	71
6.1.4	Cascade backgrounds	72
6.1.5	Backgrounds with converted photons	72
6.2	The selection chain	72

6.3	Decay quality requirements	73
6.4	Definition of q^2 and mass ranges	75
6.5	Fiducial cuts	76
6.6	Cascade background vetoes	77
6.7	Particle identification	79
6.8	Trigger	79
6.8.1	L0 trigger requirements	79
6.8.2	HLT1 trigger requirements	80
6.8.3	HLT2 trigger requirements	81
6.9	Multivariate selection	82
6.9.1	Multivariate selection training	82
6.9.2	Multivariate selection optimisation and performance	83
6.10	Fit window	86
6.11	Multiple candidates	88
7	Efficiencies and corrections to the simulation	90
7.1	Truth-matching, ghosts, and the definition of q^2_{true}	92
7.2	Fit to the preselected control modes	94
7.3	Corrections to particle identification efficiencies	94
7.3.1	Kaon, muon and pion (mis-)identification efficiencies	94
7.3.2	Electron identification efficiencies	97
7.3.3	Combination of PID efficiencies and biases	99
7.4	Calibration of L0 trigger efficiencies	100
7.4.1	L0 efficiency measurements on simulated events	101
7.4.2	L0 efficiency measurements on data and data/simulation ratio . . .	105
7.4.3	Trigger efficiency combination and factorisation bias	106
7.4.4	Hidden variables in L0 efficiency parametrisation	108
7.5	Calibration of HLT efficiencies	109
7.6	Kinematic corrections	111
7.6.1	Correction to the occupancy distribution	112
7.7	Kinematic distributions in simulation and data	114
7.8	q^2 and B mass resolution in the electron modes	117
7.9	Summary and list of efficiencies	119

8	Yields and fits	121
8.1	Fit method and shapes	121
8.1.1	Background shapes	122
8.1.2	Signal shapes	122
8.2	Fit to resonant modes	123
8.2.1	Fit to $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$	123
8.2.2	Fit to $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$	124
8.2.3	Fit to $B^+ \rightarrow K^+ \psi(2S)(\mu^+ \mu^-)$	126
8.2.4	Fit to $B^+ \rightarrow K^+ \psi(2S)(e^+ e^-)$	127
8.3	Fit to rare modes	128
8.3.1	Fit to $B^+ \rightarrow K^+ \mu^+ \mu^-$	128
8.3.2	Fit to $B^+ \rightarrow K^+ e^+ e^-$	129
8.4	Extraction of R_K from the fits to the rare modes	131
8.5	Statistical and systematic uncertainties on R_K , $R_{\psi(2S)}$ and $r_{J/\psi}$	133
8.6	Fit validation and expected statistical precision	134
9	Systematic uncertainties	136
9.1	Total systematic uncertainties	137
9.2	Size of the calibration and simulation samples	139
9.3	Kinematic reweighting scheme	140
9.4	PID calibration	142
9.5	Trigger calibration	143
9.5.1	Tag bias impacting the muon, electron and hadron trigger calibration	143
9.5.2	Paramatisation of the electron trigger efficiencies	143
9.5.3	Biases caused by the TIS calibration histograms	144
9.6	Occupancy proxy	146
9.7	Material budget and tracking efficiency	147
9.8	Electron q^2 and mass resolution	149
9.8.1	Right tail of the mass and q^2 distributions	149
9.8.2	Trigger bias on the mass and q^2 resolution	150
9.8.3	Momentum dependency of the mass and q^2 resolution	151
9.9	Decay model	152
9.10	Fit shape	153
9.11	Partially reconstructed background constraint	154

10 Cross-checks	155
10.1 Integrated measurement of $r_{J/\psi}$ and $R_{\psi(2S)}$	155
10.2 Differential measurement of $r_{J/\psi}$	155
10.3 $R_{J/\psi}^{\text{bin}}$ test	161
10.4 Branching fraction of the $B^+ \rightarrow K^+ \mu^+ \mu^-$ decay	167
10.5 Contamination from backgrounds with two mis-identified tracks	168
11 Result	170
12 Conclusions and implications	177
References	180
 III Appendix	 190
A Fit to the J/ψ mode in all data-taking conditions	191
B Supplementary kinematic distributions	195
C Complete list of efficiencies	199
D $B^+ \rightarrow \pi^+ J/\psi(e^+e^-)$ mis-ID mass variable transformation	206
E $r_{J/\psi}$ differential in the $h\text{TOS!}$ and TIS! categories	208

List of Figures

2.1	Feynman diagrams for the short-distance contribution to $b \rightarrow s\ell^+\ell^-$ decays.	34
2.2	Schematic of the $B \rightarrow K\ell^+\ell^-$ differential branching fraction as a function of q^2 .	35
2.3	Feynman diagrams for $b \rightarrow s\ell^+\ell^-$ decays in the effective theory.	36
2.4	Theory prediction (coloured bands) and LHCb measurement (black points with error bars) of the $B^+ \rightarrow K^+\mu^+\mu^-$ and $B^0 \rightarrow K^{*0}\mu^+\mu^-$ differential branching fractions.	38
2.5	Theory prediction, as well as LHCb measurement of the angular observable P'_5 in $B^0 \rightarrow K^{*0}\mu^+\mu^-$ decays.	39
2.6	Values of R_K and R_{K^*} when varying the Wilson coefficients of the muon mode individually, and fit to the $b \rightarrow s\ell^+\ell^-$ observables with floating δC_9^μ and δC_9^e .	41
2.7	Combination of the R_D and R_{D^*} measurements and comparison with the theory prediction.	42
2.8	NP Feynman diagram for a $b \rightarrow s\ell^+\ell^-$ decay mediated by a LQ.	42
2.9	Feynman diagrams for B_s^0 mixing in the SM and in NP models.	43
2.10	Search for a LQ coupling to a τ lepton and b quark using the CMS detector.	44
3.1	Energy lost per distance travelled in iron for muons and electrons as a function of momentum.	48
3.2	Schematic of the LHCb detector.	50
3.3	Schematic of the two kinds of VELO sensors and schematic illustrating how the sensors are placed around the beam pipe	52
3.4	Cherenkov angle as a function of momentum.	55
3.5	Schematic of the calorimeter system.	56
3.6	Amount of material and resolution on the impact parameter as a function of the polar angle.	60
4.1	Two-dimensional distributions of $m(K^+\ell^+\ell^-), q^2$ from the muon and electron datasets.	62
4.2	Distributions of $m(\ell^+\ell^-)$ for $B^+ \rightarrow K^+J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow K^+J/\psi(e^+e^-)$ simulated Run 1 events.	64

4.3	Distributions of kinematic variables for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$, $B^+ \rightarrow K^+ e^+e^-$, $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$, and $B^+ \rightarrow K^+ \mu^+\mu^-$ modes.	67
6.1	Schematic of a B^+ decay defining quantities used for the selection.	73
6.2	Reconstructed q^2 shapes for simulated $B^+ \rightarrow K^+ \phi(\rightarrow e^+e^-)$ and $B^0 \rightarrow K^{*0} \gamma$ events.	76
6.3	K^+e^- mass distributions from signal and cascade simulation samples. . . .	78
6.4	Distribution of $p_T(e)$, $\log \chi^2_{IP}(e)$ for electrons selected from $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ candidates from the Run 2 data.	81
6.5	ROC curves for the rare modes in different trigger conditions for the Run 2 data.	83
6.6	Illustration of the exponential fit to the $K^+e^+\mu^-$ sample to extrapolate the number of events from the sideband to the signal region	84
6.7	Expected significance as a function of the BDT working point for the $B^+ \rightarrow K^+e^+e^-$ decay in the e TOS category, and for the $B^+ \rightarrow K^+\mu^+\mu^-$ decay.	85
6.8	Constrained B mass distribution for partially-reconstructed backgrounds to the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode and to the rare $B^+ \rightarrow K^+e^+e^-$ mode. . . .	88
7.1	Distribution of the reconstructed q^2 , true post-FSR q^2 and true pre-FSR q^2 for simulated $B^+ \rightarrow K^+e^+e^-$ and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events.	93
7.2	Fit to preselected $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the data in Run 2.	95
7.3	Efficiencies of the PID requirements applied to muons and kaons as a function of their momentum.	97
7.4	Efficiency of the PID requirement $DLL_e > 3$ measured from the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ calibration sample in the 2016 data.	98
7.5	Efficiency of the kaon and electron PID cut as a function of momentum from simulated $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events, illustrating the possible biases.	100
7.6	Efficiencies of the L0Muon, L0Electron, L0Hadron and TIS trigger algorithms.	103
7.7	Data over simulation ratio of efficiencies of the L0Muon trigger, L0Electron, L0Hadron and TIS trigger algorithms.	107
7.8	Efficiency of the L0Muon trigger as a function of the transverse momentum of a muon probe and efficiency of the L0Electron trigger as a function of the transverse energy of an electron probe.	108

7.9	Efficiency of the L0Hadron trigger and efficiency of the TIS trigger for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulated events.	109
7.10	Efficiency of the L0Muon, L0Electron, L0Hadron and TIS requirements for $B^+ \rightarrow K^+ J/\psi(\ell^+\ell^-)$, $B^+ \rightarrow K^+ \psi(2S)(\ell^+\ell^-)$ and $B^+ \rightarrow K^+ \ell^+ \ell^-$ simulated events.	110
7.11	Data over simulation ratios of HLT trigger efficiencies as a function of the transverse momentum of the B^+ candidate.	111
7.12	Distributions for various occupancy proxies for $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events.	113
7.13	Kinematic distributions for preselected $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events.	115
7.14	Kinematic distributions for preselected $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events.	116
7.15	Fit functions to the $m(e^+e^-)$ distributions.	118
8.1	Fit to $m_{\text{DTF}}^{J/\psi}$ for $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events in the Run 2 data.	124
8.2	Fit to $m_{\text{DTF}}^{J/\psi}$ for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the Run 2 data selected in the $e\text{TOS}$ trigger category.	125
8.3	Fit to $m_{\text{DTF}}^{\psi(2S)}$ for $B^+ \rightarrow K^+ \psi(2S)(\mu^+\mu^-)$ events in the Run 2 data.	126
8.4	Fit to $m_{\text{DTF}}^{\psi(2S)}$ for $B^+ \rightarrow K^+ \psi(2S)(e^+e^-)$ events in the Run 2 data.	127
8.5	Fit to $m(K^+\mu^+\mu^-)$ for $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events in the Run 2 data.	129
8.6	Fit to $m(K^+e^+e^-)$ for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the Run 2 data.	130
9.1	Efficiency-corrected yield as a function of several kinematic variables for the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ modes.	141
9.2	Efficiency of either of the two electrons in $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ triggering the L0Electron line.	144
9.3	Efficiency $\varepsilon(\mu\text{TIS} \parallel h\text{TIS})$ as a function of the maximum electron p_{T}	145
9.4	Efficiency $\varepsilon(\mu\text{TIS} \parallel h\text{TIS})$ as a function of the $p_{\text{T}}(B^+)$	146
9.5	Distribution for the maximum electron rapidity for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events.	148
9.6	Data over simulation ratios of the two-dimensional electron rapidity distributions for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the Run 2 conditions.	149
9.7	Fit to the mass $m(e^+e^-)$ for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ data events using an alternative model.	151
9.8	Data-over-simulation ratio of $m(e^+e^-)$ resolution for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the Run 2 data-taking conditions.	152

9.9	Differential branching fractions predictions from the flavio software for $B^+ \rightarrow K^+ e^+ e^-$ and $B^+ \rightarrow K^+ \mu^+ \mu^-$	153
10.1	Distributions of $r_{J/\psi}$ for the preselected $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ mode in the e TOS category and the preselected $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ mode in the μ TOS category, in the Run 2 data-taking conditions.	158
10.2	Illustration of the variables used to parametrise the $B^+ \rightarrow K^+ \ell^+ \ell^-$ decay. .	161
10.3	Result of the $R_{J/\psi}^{\text{bin}}$ test in two-dimensional bins, where the $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ mode is in the e TOS category and the $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ mode is in the μ TOS category, both in the Run 2 data-taking conditions. .	164
10.4	Differential branching fraction of the $B^+ \rightarrow K^+ \mu^+ \mu^-$ decay compared with the measurement from Ref. [31].	167
10.5	Distribution of the $m(K^+ e^+_{\rightarrow \pi^+} e^-_{\rightarrow \pi^-})$ for $B^+ \rightarrow K^+ e^+ e^-$ selected candidates in the combined Run 1 and Run 2 sample.	168
10.6	Distribution of the $m_{\text{DTF}}^{J/\psi}(K^+_{\rightarrow e^+} e^+_{\rightarrow K^+} e^-)$ invariant mass in simulation and data.	169
11.1	Profiled likelihood for R_K	171
11.2	Fit to $m(K^+ \mu^+ \mu^-)$ for $B^+ \rightarrow K^+ \mu^+ \mu^-$ events in the Run 1 and Run 2 data.	173
11.3	Fit to $m(K^+ e^+ e^-)$ for $B^+ \rightarrow K^+ e^+ e^-$ events in the e TOS trigger category and in the Run 1 and Run 2 data.	174
11.4	Fit to $m(K^+ e^+ e^-)$ for $B^+ \rightarrow K^+ e^+ e^-$ events in the h TOS! trigger category and in the Run 1 and Run 2 data.	175
11.5	Fit to $m(K^+ e^+ e^-)$ for $B^+ \rightarrow K^+ e^+ e^-$ events in the TIS! trigger category and in the Run 1 and Run 2 data.	176
12.1	Theoretical prediction for R_K (left) and R_{K^*} (right) for different values of the Wilson coefficients.	178
A.1	Fit to $m_{\text{DTF}}^{J/\psi}$ for $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ and $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ events in the Run 1 data-taking conditions.	192
A.2	Fit to $m_{\text{DTF}}^{J/\psi}$ for $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ events in the Run 1 and Run 2 data-taking conditions selected in the h TOS! trigger category.	193
A.3	Fit to $m_{\text{DTF}}^{J/\psi}$ for $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ events in the Run 1 and Run 2 data-taking conditions selected in the TIS! trigger category.	194

B.1	Kinematic distributions for preselected $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ events in the μ TOS category and in the Run 2 data-taking conditions.	195
B.2	Kinematic distributions for preselected $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ events in the e TOS category and in the Run 2 data-taking conditions.	196
B.3	Kinematic distributions for preselected $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ events in the μ TOS category and in the Run 1 data-taking conditions.	197
B.4	Kinematic distributions for preselected $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ events in the e TOS category and in the Run 1 data-taking conditions.	198
D.1	Comparison of mass distributions obtained with $B^+ \rightarrow \pi^+ J/\psi(\mu^+ \mu^-)$ simulated events with that obtained using $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ events through the transformation explained in the text.	207
E.1	Plots of $r_{J/\psi}$ for the preselected $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ mode in the h TOS! category and the preselected $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ mode in the μ TOS category, in the Run 2 data-taking condition.	208
E.2	Plots of $r_{J/\psi}$ for the preselected $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ mode in the TIS! category and the preselected $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ mode in the μ TOS category, in the Run 2 data-taking condition.	211

List of Tables

2.1	Particle content of the Standard Model.	21
2.2	Examples of mesons and their quark content.	30
3.1	Centre-of-mass energy, $b\bar{b}$ production cross-section, bunch spacing and integrated luminosity for the four years of LHC data.	47
6.1	Offline selection cuts.	74
6.2	Signal and control samples and their corresponding q^2 and mass ranges. . .	76
6.3	Expected level of cascade and mis-ID backgrounds.	78
6.4	Trigger thresholds for the electron, hadron and muon L0 triggers for each year of data-taking.	80
6.5	Variables used by the BDT classifier.	83
6.6	Expected signal and combinatorial yields in the fit window for the electron and muon rare modes in different trigger and data-taking conditions. . . .	86
6.7	Fraction of events for which more than one $B^+ \rightarrow K^+ J/\psi$ candidates are reconstructed, when the full selection chain is applied.	89
7.1	Relative fractions of simulated events classified as ghosts in fully selected $B^+ \rightarrow K^+ \ell^+ \ell^-$ and $B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-)$ events, after removing multiple candidates.	93
7.2	Number of $B^+ \rightarrow K^+ e^+ e^-$ events migrating in and out of the range $1.1 < q^2 < 6.0 \text{ GeV}^2$	119
7.3	Total efficiencies for $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$, $B^+ \rightarrow K^+ e^+ e^-$, $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ and $B^+ \rightarrow K^+ \mu^+ \mu^-$ decays.	120
8.1	Values of the parameters c_K^{rt} and $c_{\psi(2S)}^{rt}$ and constraint on the partially-reconstructed background fractions r_{prc}^{rt}	133
8.2	Expected sensitivity on the determination of R_K	135
9.1	Fractional error matrices σ_c , $\sigma_c^{\psi(2S)}$, and $\sigma^{J/\psi}$	137
9.2	List of systematic uncertainties on $r_{J/\psi}$, R_K , and $R_{\psi(2S)}$	138
9.3	Fractional error matrix on the r_{prc}^{rt} factors.	154
10.1	Integrated values of $r_{J/\psi}$ and $R_{\psi(2S)}$	156

11.1	Values for R_K measured separately in the different trigger categories and data.	172
C.1	Efficiencies for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$, $B^+ \rightarrow K^+ e^+ e^-$, $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow K^+ \mu^+ \mu^-$ decays in the Run 1 conditions, as well as efficiency ratios. Efficiencies are computed with no corrections to the simulation.	200
C.2	Efficiencies for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$, $B^+ \rightarrow K^+ e^+ e^-$, $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow K^+ \mu^+ \mu^-$ decays in the Run 2 conditions, as well as efficiency ratios. Efficiencies are computed with no corrections to the simulation.	201
C.3	Efficiencies for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$, $B^+ \rightarrow K^+ e^+ e^-$, $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow K^+ \mu^+ \mu^-$ decays in the Run 1 conditions, as well as efficiency ratios. Efficiencies are computed including the L0, HLT and PID corrections to the simulation.	202
C.4	Efficiencies for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$, $B^+ \rightarrow K^+ e^+ e^-$, $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow K^+ \mu^+ \mu^-$ decays in the Run 2 conditions, as well as efficiency ratios. Efficiencies are computed including the L0, HLT and PID corrections to the simulation.	203
C.5	Efficiencies for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$, $B^+ \rightarrow K^+ e^+ e^-$, $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow K^+ \mu^+ \mu^-$ decays in the Run 1 conditions, as well as efficiency ratios. The efficiencies are computed with all corrections applied to the simulation.	204
C.6	Efficiencies for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$, $B^+ \rightarrow K^+ e^+ e^-$, $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow K^+ \mu^+ \mu^-$ decays in the Run 2 conditions, as well as efficiency ratios. The efficiencies are computed with all corrections applied to the simulation.	205

1. Introduction

This thesis presents a test of the Standard Model of particle physics consisting of the measurement of the observable R_K . This observable is defined as the ratio of branching fractions of the decays $B^+ \rightarrow K^+ \mu^+ \mu^-$ and $B^+ \rightarrow K^+ e^+ e^-$. The branching fractions of the two processes are predicted to be equal in the Standard Model, yielding $R_K = 1$. The decays $B^+ \rightarrow K^+ \mu^+ \mu^-$ and $B^+ \rightarrow K^+ e^+ e^-$ are very rare, making them sensitive to the presence of new interactions beyond the Standard Model. In the case where these new interactions couple differently to electrons and muons, R_K could however deviate from unity. Some previous measurements related to decays of B hadrons show some hints of discrepancies between the experimental values and the Standard Model expectations, the so-called “ B anomalies”. A precise measurement of R_K could settle the question of whether these anomalies are genuine or mere statistical fluctuations. The measurement performed in this thesis uses data collected with the LHCb detector between 2011 and 2016, and enables R_K to be measured with unprecedented precision.

This thesis is subdivided into three parts. Part I consists of two introductory chapters. In Chpt. 2, some theoretical concepts are introduced, focusing on lepton flavour and on beauty mesons, and the B anomalies are discussed. The LHCb detector is introduced in Chpt. 3, where an emphasis is put on the parts of the detector that are most relevant for the measurement of R_K . Part II is the central part of the thesis and presents the measurement of R_K , explaining in detail the various analysis techniques that allow R_K to be measured reliably at LHCb. Part III contains some supplementary material.

Natural units, in which $c = \hbar = 1$, are used throughout the thesis. Whenever a particle or a particle decay is mentioned, charge conjugation is implied.

Part I

Theoretical and experimental overview

2. Flavour in the Standard Model and beyond

This first chapter introduces some theoretical concepts, and some recent experimental developments that motivate the research presented in this thesis. This chapter is based on Ref. [1–4].

2.1 Fields and Symmetries in the Standard Model

The Standard Model (SM) of particle physics [1–3] is the minimal theory describing the fundamental particles that constitute matter, and their interactions. The SM is a quantum field theory, describing particles as quantum oscillations of fields. The different particles contained in the SM are shown in Tab. 2.1, together with their charges and masses. Each charged particle comes with an anti-particle partner of equal mass but opposite charge. There are three types of fields contained in the SM:

Table 2.1: Particle content of the Standard Model. The mass of each particle (taken from Ref. [5]) is indicated below its symbol.

Fermions				Bosons	
generation	1	2	3	vector	scalar
charged leptons ($q = +1$)	e 511 keV	μ 106 MeV	τ 1.78 GeV	γ 0	h 125 GeV
neutrinos ($q = 0$)	ν_e < 2 eV	ν_μ < 2 eV	ν_τ < 2 eV	W^+ 80.4 GeV	
up-type quarks ($q = +2/3$)	u 2.2 MeV	c 1.30 GeV	t 170 GeV	Z^0 91.2 GeV	
down-type quarks ($q = -1/3$)	d 4.7 MeV	s 96 MeV	b 4.6 GeV	g 0	

- The spinor fields correspond to particles of spin one half that constitute matter.

These particles are called fermions and are of two types: quarks and leptons. Quarks appear in six flavours (see Tab. 2.1): up u , down d , charm c , strange s , top t and bottom b . Leptons come in three flavours. Each flavour has two leptons: a charged lepton $\ell^\pm$ and a neutrino ν . The three lepton flavours are electron (e^- and ν_e), muon (μ^- and ν_μ) and tau (τ^- and ν_τ). All fermions are also classified in three generations, which have increasing mass, as shown in Tab. 2.1. In the SM, each fermion can have two chirality states: they can be either right-handed or left-handed, which, in the massless limit, correspond to the spin being either parallel or anti-parallel to the momentum of the particle;

- The vector fields correspond to particles of spin 1. Such particles with integer spin are also called bosons. They are the carriers of the forces through which matter particles interact. The photons γ , quanta of the electromagnetic field A , interact with all charged particles. The bosons Z and $W^\pm$ are the carriers of the weak force and interact with all fermions. Finally, the gluons g are the strong force carriers and couple only to quarks;
- Finally, the Higgs field is a scalar field, which corresponds to spinless particles. The Higgs field allows particles to acquire mass, as described in Sec. 2.3.

The Lagrangian formalism is used to describe the dynamics of the fields and enables measurable quantities to be computed. For example, the part of the SM Lagrangian describing the electromagnetic interaction between electrons is

$$\mathcal{L}^{\text{E.M.}} = \underbrace{-i\bar{e}\gamma_\mu\partial^\mu e - m\bar{e}e - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}}_{\mathcal{L}^{\text{kin}}} - \underbrace{g_e A_\mu \bar{e}\gamma^\mu e}_{\mathcal{L}^{\text{int}}}, \quad (2.1)$$

where A_μ is the electromagnetic vector field, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the electromagnetic field tensor, and e is the electron spinor field ¹. The symbols γ_μ are the Dirac matrices. The first three terms on the right-hand side of Eq. 2.1 constitute the kinetic part of the electromagnetic Lagrangian $\mathcal{L}^{\text{kin}}$, describing how free electrons and photons behave when they do not interact with each other. The last term is the interaction part $\mathcal{L}^{\text{int}}$, describing how the photons can couple to an electron current $\bar{e}\gamma^\mu e$. The coupling constant g_e sets the strength of the interaction.

¹In this chapter, the same symbol is used to describe a field and the particle it corresponds to, except in the case of the photon (γ) and the electromagnetic field (A_μ).

Observable quantities, such as a particle lifetime or a differential cross-section, can be computed from the Lagrangian. The computation of an observable involves the modulus-squared of an amplitude $\mathcal{A}$ given by

$$|\mathcal{A}|^2 = |\langle f | \mathcal{O} | i \rangle|^2, \quad (2.2)$$

where $|f\rangle$ and $|i\rangle$ describe the final and initial states of the fields. For example, $|f\rangle = |\gamma\gamma\rangle$ and $|i\rangle = |e^+e^-\rangle$ for an electron pair annihilating into two photons. The symbol $\mathcal{O}$ represents an operator. This operator is a function of the interaction Lagrangian that describes the interaction allowing the transition from $|i\rangle$ to $|f\rangle$ to happen. In general, $\mathcal{A}$ cannot be computed analytically and approximate methods are used. For example, if the coupling constant is small, as is the case for the electroweak force, observables can be computed as a perturbative expansion in powers of the coupling constant. The different terms appearing in the expansion are pictorially represented by Feynman diagrams.

A fundamental concept in particle physics is that of invariance under symmetry transformations. The first type of transformations are the spatial transformations: the SM being a relativistic theory, its Lagrangian is invariant under the Lorentz group, *i.e.* under rotations and Lorentz boosts. Another kind of transformations are three discrete transformations: charge conjugation C that reverse the electric charges of particles, parity conjugation P that reverses the three spatial coordinates, and time reversal T . Whereas the part of the Lagrangian describing the strong and electromagnetic forces is symmetric under C , P , and T , the weak part breaks P and C maximally, and also breaks CP conjugation.

Another type of transformation is the gauge transformation. They are related to the fact that the complex phases of the field operators have no physical meaning, as they do not appear in the expressions of physical observables. Two observers at different coordinates in space time are able to make the same physical predictions using any phase convention they want. In the case of electromagnetism, the gauge group is $U(1)$, and the electron field transforms as

$$e(x) \rightarrow e'(x) = \exp(ig_e\theta(x))e, \quad (2.3)$$

where θ is the local phase shift. The Lagrangian for free electrons is not invariant under this symmetry, but the Lagrangian $\mathcal{L}^{\text{E.M.}}$, which includes the electromagnetic interaction, is invariant under the condition that A_μ transforms as

$$A_\mu \rightarrow A'_\mu = A_\mu - \partial_\mu\theta. \quad (2.4)$$

Hence, in a gauge invariant theory, invariance under a local gauge transformation is intimately linked to the existence of a so-called gauge field and hence of a force. More precisely, there are as many gauge bosons as there are generators in the gauge group. The full SM gauge group is $SU(3) \times SU(2) \times U(1)$. The group $SU(3)$ is related to the strong force, whereas $SU(2) \times U(1)$ describes the electromagnetic and weak force, as explained in more detail in the following sections.

2.2 The weak interaction and electroweak unification

The research presented in this thesis belongs to the domain of Flavour Physics, which is related to the quark flavours and how they interact via the weak force. This section is a summary of how the weak interaction is described in the SM and how it is unified with the electromagnetic interaction in a gauge invariant theory.

The first simple model describing the weak interaction is that of Fermi [6]. It was first published in 1933 and designed to explain beta decays, that is the decay of the neutron to a proton, an electron, and a neutrino $n^0 \rightarrow p^+ e^- \nu$. Its interaction Lagrangian is that of a point-like interaction between four fields,

$$\mathcal{L}^{\text{Fermi}} = G_F (\bar{p} \gamma_\mu n) (\bar{e} \gamma^\mu \nu), \quad (2.5)$$

where p , n , e , ν are the field operators of the different particles and the coupling G_F is the Fermi constant.

In 1956, Wu studied the decay of neutrons in polarised cobalt nuclei. She observed that the electrons do not decay isotropically, but in a direction opposite to the spin of the nucleus [7]. This can be explained in a theory which breaks parity, *i.e.* which is not invariant under reversal of space coordinates $x^i \rightarrow -x^i$. In the case of the weak interaction specifically, this is explained by the fact that only left-handed spinors interact with the weak force. To take this into account, Eq. 2.5 is modified to become

$$\mathcal{L}^{\text{V-A}} = 2\sqrt{2}G_F (\bar{p}_L \gamma_\mu n_L) (\bar{e}_L \gamma^\mu \nu_L) \quad (2.6)$$

$$= \frac{1}{\sqrt{2}} G_F (\bar{p} \gamma_\mu (1 - \gamma^5) n) (\bar{e} \gamma^\mu (1 - \gamma^5) \nu), \quad (2.7)$$

where the subscript L indicates the left-handed component of a spinor. A spinor is projected onto its left-handed component using the projector $\frac{1-\gamma^5}{2}$, *e.g.* $p_L = \frac{1-\gamma^5}{2} p$.

In the 1960s, Glashow, Weinberg and Salam (GWS) developed a gauge-invariant theory that describes the electromagnetic and weak interactions in a unified way [8–10]. The GWS

theory is invariant under an $SU(2)$ symmetry. In connection to the spin, this symmetry is referred to as ‘weak isospin’, with its third component denoted I_3 . To keep the theory chiral, the left-handed and right-handed components of the fermions belong to two different representations of $SU(2)$. The left-handed fermions are grouped into six $SU(2)$ doublets, three doublets denoted L for left-handed leptons and three doublets Q for left-handed quarks,

$$\begin{aligned} L^i &= \begin{pmatrix} \nu^i \\ \ell^i \end{pmatrix}_L = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L, \\ Q^i &= \begin{pmatrix} u^i \\ d^i \end{pmatrix}_L = \begin{pmatrix} u \\ d \end{pmatrix}_L, \begin{pmatrix} c \\ s \end{pmatrix}_L, \begin{pmatrix} t \\ b \end{pmatrix}_L, \end{aligned} \quad (2.8)$$

with $i = 1, 2, 3$. The right-handed fermions are all weak isospin singlets, *i.e.* have $I_3 = 0$.

In addition to the weak isospin, the GWS theory is also invariant under a $U(1)$ symmetry. The charge conserved under this symmetry is referred to as the weak hypercharge Y . Left-handed and right-handed fermions have different hypercharge. The weak hypercharge, weak isospin and the electric charge q (in unit of the elementary charge) are related to each other according to²

$$Y = q - I_3, \quad (2.9)$$

meaning that the weak hypercharge corresponds to the electric charge for right-handed fermions but not for left-handed ones.

The GLS theory is hence invariant under a local $SU(2) \times U(1)$ phase transformation of each fermionic field ψ , which reads

$$\psi(x) \rightarrow \psi'(x') = \exp(ig_w \theta^a(x)t^a + ig' Y \alpha(x)) \psi(x), \quad (2.10)$$

where g_w and g' are the coupling strengths of the gauge bosons, t^a are the three generators of $SU(2)$ ($t^a = 0$ for the singlet representation, and $t^a = \frac{1}{2}\sigma^a$ for the doublet representation, where σ^a are the Pauli matrices).

There are four bosons associated to the four generators of $SU(2) \times U(1)$: W^a related to the weak isospin and B related to the weak hypercharge. These are related to the physical bosons of the weak interaction, the photon A , the W^+ , W^- and Z according to

²This equation follows the convention of Ref. [1]. In other textbooks, the hypercharge is defined such that $Y = 2(q - I_3)$.

the relations

$$\begin{pmatrix} Z \\ A \end{pmatrix} = \begin{pmatrix} \cos \theta_W & -\sin \theta_W \\ \sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} W^3 \\ B \end{pmatrix}, \quad (2.11)$$

$$W^\pm = \frac{1}{2}(W^1 \pm W^2).$$

The weak mixing angle θ_W also relates the couplings g and g' to the electromagnetic coupling g_e ,

$$g_e = g_w \sin \theta_W = g' \cos \theta_W. \quad (2.12)$$

The interaction terms between the quarks and the bosons then read

$$\begin{aligned} \mathcal{L}_{\text{quark}}^I = & \frac{g_e}{\sqrt{2} \sin \theta_W} [W_\mu^+ \bar{u}_L^i \gamma^\mu d_L^j + W_\mu^- d_L^i \gamma^\mu u_L^j] \\ & + g_e A_\mu \left[\frac{2}{3} \bar{u}_i \gamma^\mu u_i - \frac{1}{3} \bar{d}_i \gamma^\mu d_i \right] \\ & - g_e \tan \theta_W Z_\mu \left[\frac{2}{3} \bar{u}_i \gamma^\mu u_i - \frac{1}{3} \bar{d}_i \gamma^\mu d_i \right] \\ & + \frac{g_e}{\sin \theta_W \cos \theta_W} Z_\mu \left[\frac{1}{2} \bar{u}_L \gamma^\mu u_L - \frac{1}{2} \bar{d}_L \gamma^\mu d_L \right]. \end{aligned} \quad (2.13)$$

The interaction terms between the bosons and the leptons look similar,

$$\begin{aligned} \mathcal{L}_{\text{lepton}}^I = & \frac{g_e}{\sqrt{2} \sin \theta_W} [W_\mu^+ \bar{\nu}_L^i \gamma^\mu \ell_L^i + W_\mu^- \bar{\ell}_L^i \gamma^\mu \nu_L^i] \\ & + g_e A_\mu [\bar{\ell}^i \gamma^\mu \ell^i] \\ & + \frac{g_e}{\tan \theta_W} Z_\mu \bar{\ell}_i \gamma^\mu \ell_i \\ & + \frac{g_e}{\sin \theta_W \cos \theta_W} Z_\mu \left[\frac{1}{2} \bar{\nu}_L^i \gamma^\mu \nu_L^i - \frac{1}{2} \bar{\ell}_L^i \gamma^\mu \ell_L^i \right] \end{aligned} \quad (2.14)$$

where the charged currents, leading to the neutron decay, appear on the first line, the electric current appears on the second line, and an extra neutral current appears on the third and fourth lines.

2.3 Higgs mechanism and origin of mass

The Z and W bosons, as well as quarks and charged leptons, are massive. In the GWS theory presented here, neutrinos are considered massless. Although it is known today that neutrinos have a small mass [11], this has a negligible influence on the physics discussed

in this thesis (see Sec. 2.6). A mass term for a gauge field such as the Z is of the form $m^2 Z_\mu Z^\mu$. For a fermionic field ψ it is of the form

$$\mathcal{L} = m\bar{\psi}\psi = m(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L). \quad (2.15)$$

Such mass terms do not respect the $SU(2) \times U(1)$ symmetry, neither in the case of the gauge bosons, nor for fermions, as their left-handed component transforms differently to their right-handed one.

The phenomenon giving rise to fermion and boson masses is the Higgs mechanism. It is based on the idea, developed by Higgs, Brout and Englert in the 1960s, that the symmetry is indeed broken but only at low energy [12, 13]. This is realised by adding to the theory an extra weak isospin doublet of complex scalar fields ϕ , with

$$\phi = \begin{pmatrix} \phi^+(x) \\ \phi^0(x) \end{pmatrix} \quad (2.16)$$

whose dynamics are governed by the $SU(2)$ -invariant Lagrangian

$$\mathcal{L} = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - V(\phi), \quad (2.17)$$

with the potential

$$V(\phi) = \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2, \quad (2.18)$$

where μ and λ are real and positive coefficients. At low energy, the field ϕ will settle around a particular direction in the minimum of the potential. By convention, this particular direction is chosen to be

$$\phi_{\min} = \begin{pmatrix} 0 \\ \frac{\mu}{\sqrt{2\lambda}} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \quad (2.19)$$

The constant $v = \frac{\mu}{\sqrt{\lambda}}$ is the value of the potential at its minimum and is referred to as the vacuum expectation value. The arbitrary choice of direction for the minimum breaks the symmetry of the Lagrangian spontaneously. A low energy Lagrangian, valid when ϕ oscillates close to its minimum, can be written. It can be shown that such a low-energy Lagrangian is only $U(1)$ invariant (corresponding the symmetry of the electromagnetic field). Moreover, the kinetic terms related to the W and Z do not need to respect the broken symmetry, and these bosons can acquire a mass. A real scalar field h is also present in this low energy Lagrangian, corresponding to the real oscillations of ϕ^0 around the minimum of the potential. This can be schematically summarised as

$$\begin{aligned}
\mathcal{L}^{\text{full}} &\longrightarrow \mathcal{L}^{\text{broken}} \\
\text{SU}(2) \times \text{U}(1) &\longrightarrow \text{U}(1) \\
\text{massless } W^1, W^2, W^3, B &\longrightarrow \begin{array}{l} \text{massless } A \\ \text{massive } W^+, W^-, Z^0 \end{array} \\
\phi(x) = \begin{pmatrix} \phi^+(x) \\ \phi^0(x) \end{pmatrix} &\longrightarrow \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}
\end{aligned} \tag{2.20}$$

where $\mathcal{L}^{\text{full}}$ is the Lagrangian of the full theory, and $\mathcal{L}^{\text{broken}}$ the low energy Lagrangian with the broken symmetries. The scalar field $h(x)$ corresponds to the Higgs boson observed in 2012 by the ATLAS and CMS collaborations [14, 15].

The mechanism by which fermions acquire mass is very relevant for flavour physics and is therefore presented here in more detail. The electron weak isospin doublet, as well as the right-handed electron weak isospin singlet interact with the scalar field ϕ as

$$\mathcal{L}_e^{\text{Yuk}} = -Y^e \left[(\bar{\nu}_e \ \bar{e})_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} e_R + \bar{e}_R (\phi^{+*} \ \phi^{0*}) \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \right], \tag{2.21}$$

where Y_e is the coupling of this interaction, referred to as a Yukawa interaction. This term is $\text{SU}(2) \times \text{U}(1)$ invariant. Substituting the field ϕ by its expression after spontaneous symmetry breaking from Eq. 2.20, this term becomes

$$\mathcal{L}_e^{\text{Yuk}} = -\frac{Y^e}{\sqrt{2}} h (\bar{e}_L e_R + \bar{e}_R e_L) - \frac{Y^e}{\sqrt{2}} v (\bar{e}_L e_R + \bar{e}_R e_L), \tag{2.22}$$

where the first term represents an interaction between the electron and the Higgs field, and the second term gives the electron a mass

$$m_e = \frac{Y^e v}{\sqrt{2}}. \tag{2.23}$$

This illustrates how charged leptons acquire mass via the spontaneous symmetry breaking mechanism. Down-type quarks can acquire mass with a term of the same form. Up-type quarks, however, couple to the scalar field with a term of the form:

$$\mathcal{L}_u^{\text{Yuk}} = Y^u \left[(\bar{u} \ \bar{d})_L (-i\sigma_2) \begin{pmatrix} \phi^{+*} \\ \phi^{0*} \end{pmatrix} u_R + \bar{u}_R (\phi^0 \ \phi^+) (i\sigma_2^\dagger) \begin{pmatrix} u \\ d \end{pmatrix}_L \right]. \tag{2.24}$$

The above term is $\text{SU}(2) \times \text{U}(1)$ invariant and gives rise, after symmetry breaking, to a mass term for the up quark $m_u = \frac{Y_u v}{\sqrt{2}}$.

The total Yukawa term, which generates mass terms for quarks and charged leptons (but not to neutrinos), reads, after symmetry breaking,

$$\mathcal{L}^{\text{Yuk}} = \frac{v}{\sqrt{2}} \left[-Y_{ij}^\ell \bar{e}_L^i e_R^j - Y_{ij}^d \bar{d}_L^i d_R^j - Y_{ij}^u \bar{u}_L^i u_R^j + \text{h.c.} \right], \quad (2.25)$$

where “h.c.” stands for the hermitian conjugate of the whole expression. The Yukawa couplings are grouped into three matrices Y_{ij}^ℓ , Y_{ij}^d and Y_{ij}^u , which can contain non-diagonal terms. In this section, all quark fields are expressed in the “flavour basis”, where the coupling terms with the gauge bosons are diagonal. It is however possible to rotate the quark and lepton fields to move to the “mass basis”, where all the Yukawa matrices are diagonal. First in the case of the quarks, this can be done with two unitary matrices: U_{ij}^d acting on d_L , d_R and U_{ij}^u acting on u_L , u_R such that $U^{d\dagger} Y^d U^d$ and $U^{u\dagger} Y^u U^u$ are diagonal. When rotating the quark fields in the weak boson interaction expression Eq. 2.13, the matrices cancel out in the interaction term with the Z and A , as these do not mix-up up and down quarks. However, the coupling with the W now contains an extra factor $V = U^{u\dagger} U^{d\dagger}$ and becomes

$$\frac{e}{\sqrt{2} \sin \theta_W} \left[W_\mu^+ \bar{u}_L^i \gamma^\mu V^{ij} d_L^j + W_\mu^- d_L^i \gamma^\mu (V^\dagger)^{ij} u_L^j \right]. \quad (2.26)$$

In the case of the leptons, because the Yukawa term giving mass to the neutrinos is absent, both the neutrino and charged lepton fields can be rotated by a single matrix U_{ij}^ℓ such that $U^{\ell\dagger} Y^\ell U^\ell$ is diagonal.

In summary, it is possible to find a basis in which both the mass term and gauge interactions terms are diagonal for the three lepton families. This implies the conservation of the separate lepton numbers discussed in Sec. 2.6. For the quarks, it is not possible to diagonalise both the mass and interaction terms at the same time. When expressing the Lagrangian in the mass basis, the couplings with the W bosons mix the different quark flavours. The matrix V , that describes the mixing between the different flavours, is called the CKM matrix. Its elements contain a single complex phase, which is the only source of CP violation in the SM. The moduli of the CKM matrix elements are very close to 1 on the diagonal and smaller off-diagonal, meaning that each up-type quark couples preferably to the down-type quark in the same $SU(2)$ doublet. The study of the CKM matrix and its CP violating phase is one of the main aims of Flavour Physics.

2.4 Quantum Chromodynamics

In the SM, quarks do not only interact via the electroweak force, but also via the strong force. This force is described in a theory referred to as Quantum Chromodynamics (QCD). This theory is based on the symmetry group $SU(3)$, and all quarks appear as triplets of this group. The QCD gauge bosons are called ‘gluons’. The QCD symmetry group remains unbroken, such that the gluons are massless and quarks in a given triplet remain undistinguishable from each other.

The QCD coupling constant g_s depends on the scale of the interaction. It is stronger at low energy scales or large distance scales, and weaker at high energy scales and short distance scales. The potentially large value of g_s means that the QCD interaction term cannot be treated as a small perturbation to the Lagrangian. Hence, QCD effects cannot in general be computed in powers of g_s , but other techniques have to be used. The increasing strength of the interaction with distance means that QCD is confining: quarks cannot propagate freely, but are confined into multiple quark states that are singlet under the QCD gauge group. They can form baryons, which are particles formed of three quarks, like the proton (uud) or neutron (udd). They can also form mesons, which are made of a quark-antiquark pair. Of particular importance in this thesis are the beauty mesons, containing a b quark. Other states mentioned in this thesis are summarised in Tab. 2.2. The parameter which characterises the energy below which g_s is strong is $\Lambda_{\text{QCD}} \approx 200$ MeV. Hence, the size of a QCD bound state is of the order of $1/\Lambda_{\text{QCD}}$.

Table 2.2: Examples of mesons and their quark content.

name	quark content
light mesons	$\pi^+ = u\bar{d}$, $K^+ = u\bar{s}$, $K_{L/S} = d\bar{s}$
charm mesons	$D^+ = c\bar{d}$, $D^0 = c\bar{u}$
beauty mesons	$B^+ = u\bar{b}$, $B^0 = d\bar{b}$, $B_s^0 = s\bar{b}$
quarkonium	$\phi = s\bar{s}$, $J/\psi = c\bar{c}$, $\psi(2S) = c\bar{c}$

2.5 Indirect searches for New Physics

The SM is an extremely successful and precise theory. Using this model, the existence of the Higgs boson was predicted almost five decades before its discovery. The SM also predicts the mass and widths of the Z and W bosons with sub-permille precision [16], and the gyromagnetic moment of the electron with a precision better than one part per billion [17]. However, it has limitations. For example, it does not generate enough CP violation to explain the large asymmetry between matter and anti-matter observed in the universe. Moreover, it contains no particle candidate for dark matter.

In order to resolve the shortcomings of the SM, new types of fields and interactions are expected to exist in nature. There are two main ways to search for such new fields. The first one is often referred to as “direct searches”, which consists of trying to produce the new particles associated with such fields in a collider and observing their decay products. The second one is “indirect searches”, which consist of looking for the effect of the new fields in the decay of known particles, by measuring whether observables associated to the decay are modified with respect to their SM expectation. Unlike direct searches, indirect searches have the advantage of being sensitive to a scale much higher than the centre-of-mass energy of a given collider.

The following sections discuss searches for New Physics (NP) in the decays of beauty hadrons. Typical observables related to these decays are the probability of such a decay occurring, the so-called branching fraction $\mathcal{B}$, or the angular distribution of the decay products. An observable Q is in general expressed in terms of a sum of an amplitude A_{SM} , coming from the SM interaction Lagrangian, and an amplitude A_{NP} , coming from interactions beyond the SM, *i.e.*

$$Q \sim |A_{\text{SM}} + A_{\text{NP}}|^2. \quad (2.27)$$

From this expression, it is clear that the sensitivity of the observable Q to NP is greater if A_{SM} is small. Hence, it is useful to look for decays that are predicted to be suppressed in the SM. The following sections discuss in more detail indirect searches with decays suppressed by lepton and quark flavour symmetries.

2.6 Lepton flavour and lepton flavour violation

Each lepton flavour family has a lepton flavour number L_ℓ : L_e , L_μ and L_τ . This number takes the value $L_\ell = 1$ for leptons and $L_\ell = -1$ for antileptons. In the SM, these three

numbers are separately conserved, as the Lagrangian contains no term that mixes the three lepton flavours. For example, the observed muon decay $\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$ has $L_\mu = 1$ and $L_e = 0$ in the initial and final state. This conservation is sometimes referred to as ‘accidental’, in the sense that it does not result from any symmetry imposed when constructing the Lagrangian, but from the fact that neutrinos are massless. It is now known, because of neutrino oscillations, that neutrinos have very small masses. Oscillating neutrinos can generate charged lepton flavour violations, but at a negligible level. Indeed, the branching fraction of such decays, in the SM extended with neutrino mass, is suppressed by a factor $(m_\nu/m_W)^4 \sim 10^{-49}$ [18].

Observing a decay violating the conservation of charged lepton flavour would therefore indicate the presence of NP. Lepton flavour violating decays correspond to the limit case in Eq. 2.27 where $A_{\text{SM}} = 0$. Such decays have been searched for in many experiments but never observed. Instead, limits have been placed on electroweak and Higgs boson decays such as [5]

$$Z^0 \rightarrow e^\pm \mu^\mp < 7.5 \times 10^{-7}, \quad Z^0 \rightarrow \tau^\pm \mu^\mp < 1.2 \times 10^{-5}, \quad H^0 \rightarrow \tau^\pm \mu^\mp < 1.43\% \quad (2.28)$$

at 95% C.L.. Stringent limits are also placed on flavour-violating decays of leptons or light mesons, for example

$$\mu^- \rightarrow e^- \gamma < 4.2 \times 10^{-13}, \quad K^+ \rightarrow \pi^+ e^- \mu^+ < 1.3 \times 10^{-11}, \quad \tau^- \rightarrow \mu^- \phi < 8.4 \times 10^{-8} \quad (2.29)$$

at 90% C.L.. The flavour violating decays of beauty mesons are somewhat less constrained, especially those having a $\mu\tau$ pair in the final state, for example

$$B^+ \rightarrow K^+ e^\pm \mu^\mp < 9.1 \times 10^{-8}, \quad B^+ \rightarrow K^+ \mu^\pm \tau^\mp < 4.8 \times 10^{-5} \quad (2.30)$$

at 90% C.L..

In addition to lepton flavour conservation, the absence of lepton mixing terms also implies that the couplings of the gauge bosons to the leptons do not depend on the lepton flavour, *e.g.* the coupling of the Z to $\mu^+ \mu^-$ is the same as the coupling to $e^+ e^-$. This is referred to as lepton flavour universality (LFU). This has been verified to hold in the decay of the W and Z bosons by the LEP experiment [19, 20]. Other LFU tests are performed by measuring the ratio of branching fractions between two processes with different flavours of leptons. The two branching fractions are not necessarily equal because of phase space effects and helicity suppression (*i.e.* because of the different masses of the leptons), but this difference can be computed more or less precisely depending on the decay. For example,

LFU holds to a very good precision in the decays of light mesons *e.g.* [21, 22]

$$\frac{\mathcal{B}(\pi^+ \rightarrow e^+ \nu_e)}{\mathcal{B}(\pi^+ \rightarrow \mu^+ \nu_\mu)} = (1.2344 \pm 0.0029) \times 10^{-4} \quad \text{SM} = (1.2352 \pm 0.0002) \times 10^{-4}, \quad (2.31)$$

$$\frac{\mathcal{B}(K^+ \rightarrow e^+ \nu_e)}{\mathcal{B}(K^+ \rightarrow \mu^+ \nu_\mu)} = (2.488 \pm 0.011) \times 10^{-5} \quad \text{SM} = (2.477 \pm 0.001) \times 10^{-5}, \quad (2.32)$$

where SM indicates the prediction³. It is also known that LFU also holds in the decays of charmonium resonances to a very good precision, *e.g.* [23]

$$\frac{\mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-)}{\mathcal{B}(J/\psi \rightarrow e^+ e^-)} = 1.0022 \pm 0.0065 \quad \text{SM} = 1.0. \quad (2.33)$$

In summary, LFU and the absence of LFV is measured to hold well in the decays of the gauge bosons, as well as in the decays of leptons and light mesons. The limits set on LFV decays of beauty mesons are somewhat less stringent. Moreover, some anomalies are observed in LFU tests with beauty mesons. These will be discussed further in Sec. 2.11, 2.12.

2.7 Flavour changing neutral currents

Unlike leptons, quarks of different flavours mix in the SM Lagrangian. Recalling Eq. 2.13 and 2.26, there is only one term mixing the quark flavours: the coupling of the quarks to the W boson. The photon and the Z , however, couple only to two quarks of the same flavour, *i.e.* there exists no flavour changing neutral current (FCNC) at tree level in the SM. However, quark flavour is not protected by any symmetry and FCNC's do occur in the SM at loop level. They are mediated by box or so-called penguin diagrams, as illustrated in Fig. 2.1 for a beauty quark turning into a strange quark emitting two leptons, $b \rightarrow s \ell^+ \ell^-$. Due to the loop suppression, the branching fractions of FCNC-mediated decays are very small in the SM. Some NP models however predict FCNC at tree level. This means that, in Eq. 2.27, A_{SM} and A_{NP} could potentially be of the same order, making FCNC an excellent probe to search for NP. The decays of beauty mesons mediated by a $b \rightarrow s \ell^+ \ell^-$ transition are the focus of this thesis. Before reviewing the experimental studies of decays mediated by a $b \rightarrow s \ell^+ \ell^-$ transition in Sec. 2.10, some theoretical concepts related to these decays are discussed in Sec. 2.8, and 2.9.

³These ratios are very far from 1 because the electron decays are helicity suppressed.

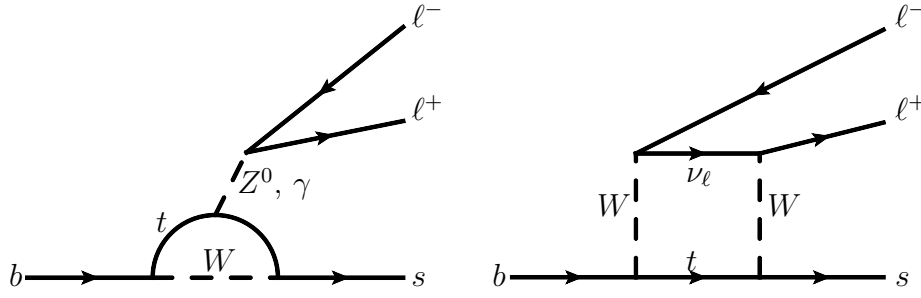


Figure 2.1: Feynman diagrams for the short-distance contribution to $b \rightarrow s \ell^+ \ell^-$ decays: penguin diagram (left) and neutrino box diagram (right).

2.8 Flavour changing neutral currents of B hadrons

Experimentally, $b \rightarrow s \ell^+ \ell^-$ transitions are observed in the decay of a beauty hadron into a strange hadron and a pair of leptons, $B \rightarrow K \ell^+ \ell^-$. A $B \rightarrow K \ell^+ \ell^-$ decay is however not only mediated by the $b \rightarrow s \ell^+ \ell^-$ transitions shown in Fig. 2.1, — hereafter referred to as the rare decay — but can also proceed via intermediate resonances. It is useful to study the differential branching fraction of $B \rightarrow K \ell^+ \ell^-$ decays as a function of the square of the dilepton mass, q^2 . Fig. 2.2 shows the q^2 distribution for a $B^+ \rightarrow K^+ \ell^+ \ell^-$ decay (the symbols C are explained in Sec. 2.9). This distribution is characterised by four distinct regimes:

1. The very low q^2 region, $q^2 \lesssim 1 \text{ GeV}^2$. The shape of the differential branching fractions depends on the spin of the strange hadron. For a scalar K ($K = K^+, K_S$), the rare decay at low q^2 interferes with the tree level transition $b \rightarrow s \bar{s} s$, where an $s \bar{s}$ pair hadronises into a resonance such as the ϕ , which subsequently decays electromagnetically into two leptons. For a vector K ($K = K^{*0}, \phi$), the very low q^2 region is dominated by the radiative decay $b \rightarrow s \gamma, \gamma \rightarrow \ell^+ \ell^-$;
2. The low q^2 region extending up to $q^2 \approx 6 \text{ GeV}^2$. This region is dominated by the rare decay;
3. The resonant region, starting at $q^2 \approx (2m(c))^2 \approx 6 \text{ GeV}^2$ and extending up to $q^2 \approx 15 \text{ GeV}^2$. In this region, the $B^+ \rightarrow K^+ \ell^+ \ell^-$ decay is dominated by a tree level $b \rightarrow c \bar{c} s$ transition, as shown in Fig. 2.2. The charm quark pair hadronises into a J/ψ or $\psi(2S)$ resonance, which subsequently decays electromagnetically into two leptons. In this region, the branching fraction of the resonant decay is approximately

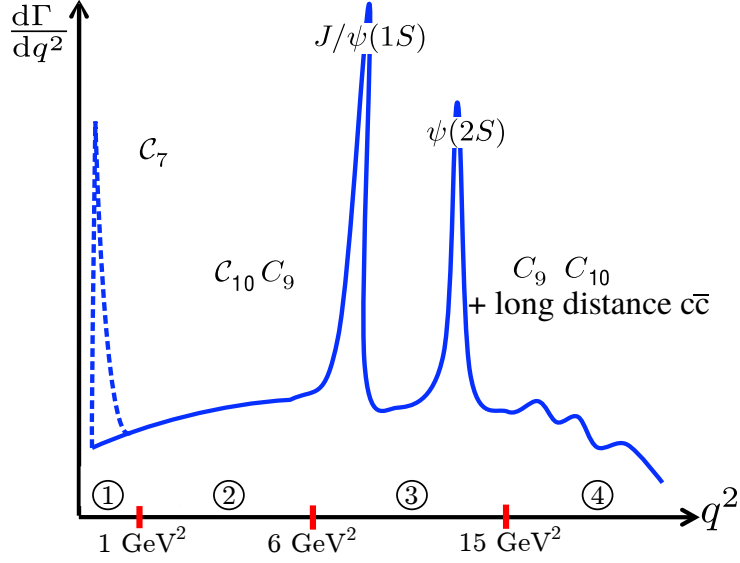


Figure 2.2: Schematic of the $B \rightarrow K\ell^+\ell^-$ differential branching fraction as a function of q^2 . The four different regimes are indicated with numbers on the x axis. The radiative pole in the very low q^2 region, shown with the dashed line, does not appear for scalar K .

two orders of magnitude higher than that of the rare decay;

4. The region above $q^2 \approx 15 \text{ GeV}^2$ is characterised by the interference between the rare decay and charmonium resonances that are broader and smaller than the J/ψ and $\psi(2S)$.

In summary, the low q^2 region $1 \lesssim q^2 \lesssim 6 \text{ GeV}^2$ is the cleanest region to study the rare $b \rightarrow s\ell^+\ell^-$ transition, in the sense that this region is free from interference with resonant transitions. The next section summarises the theoretical tools that allow the calculation of the observables related to $B \rightarrow K\ell^+\ell^-$ decays.

2.9 Computation of observables in $B \rightarrow K\ell^+\ell^-$ decays

Two very different kinds of physics are at play in a $B \rightarrow K\ell^+\ell^-$ decay. The electroweak physics which characterises the quark level transition $b \rightarrow s\ell^+\ell^-$, and the QCD that holds the strange and beauty hadrons together. These two phenomena are characterised by two very different distance scales. The $b \rightarrow s\ell^+\ell^-$ transition shown in Fig. 2.1 is characterised

by a distance on the order of $1/m(W) \approx 10^{-2} \text{ GeV}^{-1}$, whereas the size of the hadrons is of the order of $1/\Lambda_{\text{QCD}} \approx 5 \text{ GeV}^{-1}$. Hence, to a good approximation, the short-distance electroweak physics decouples from the long distance QCD effects.

In practice, this decoupling of scales means that the SM interaction Lagrangian that characterises $B \rightarrow K\ell^+\ell^-$ transitions can be rewritten as a sum of effective coupling constants, also called Wilson coefficients C_i , times operators $\mathcal{O}_i$ [24],

$$\mathcal{L}_{\text{eff}}^I = \frac{4G_F}{\sqrt{2}} V_{tb} V_{ts}^* \frac{g_e^2}{4\pi} \sum_i C_i \mathcal{O}_i, \quad (2.34)$$

where V_{tb} and V_{ts} are CKM matrix elements. In this effective Lagrangian, heavy particle fields, such as the top quark, Z or W do not appear explicitly, but their effect is integrated out in the Wilson coefficients C_i . Lighter fields appear in the operators $\mathcal{O}_i$. This effective Lagrangian integrates out the short distance effects into the coefficients C_i , where they are separated from the long distance effects described by the operators $\mathcal{O}_i$. Two examples of effective field theories were shown in Eq. 2.5 and in Eq. 2.6, which integrate out the W boson but are nevertheless a good model for the weak decay of the neutron. An example of an effective Feynman diagram for a $b \rightarrow s\ell^+\ell^-$ decay is shown in Fig. 2.3.

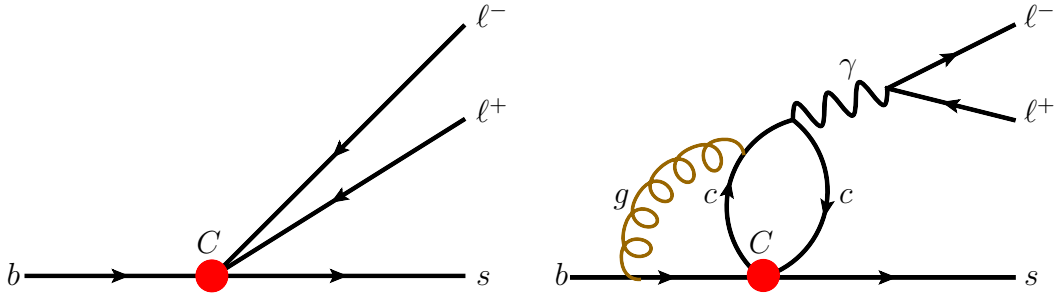


Figure 2.3: Feynman diagrams for $b \rightarrow s\ell^+\ell^-$ decays in the effective theory of Eq. 2.34: short distance contribution (left) and long distance charm loop contribution (right) where the charm loop exchanges a gluon with the b quark.

For $B \rightarrow K\ell^+\ell^-$ in the low q^2 region, the operators $\mathcal{O}_9$ and $\mathcal{O}_{10}$ are the main contributions to the decay,

$$\mathcal{O}_9 = \bar{s}\gamma_\mu b_L \bar{\ell}\gamma^\mu \ell, \quad \mathcal{O}_{10} = \bar{s}\gamma_\mu b_L \bar{\ell}\gamma^\mu \gamma^5 \ell. \quad (2.35)$$

They describe the neutrino box and Z penguin diagrams. In the SM, $C_9 \approx -C_{10} \approx 4.9$. The operator describing the photonic penguin decay is $\mathcal{O}_7$, which is responsible for the pole in the very low q^2 region for vector K . In other q^2 regions, other operators become

significant, such as those describing the transition $b \rightarrow c\bar{c}s$ that mediates the resonant decay.

The value of the Wilson coefficients remains the same for any $b \rightarrow s\ell^+\ell^-$ mediated decay, *e.g.* $B^+ \rightarrow K^+\mu^+\mu^-$, $B^0 \rightarrow K^{*0}\mu^+\mu^-$, $B_s^0 \rightarrow \mu^+\mu^-$, because only the long distance part of these decays differ, but the short distance physics remains the same. The hypothetical presence of NP would manifest itself by shifting the value of some of the Wilson coefficients by a quantity δC_i , *i.e.* $C_i \rightarrow C_i + \delta C_i$. This illustrates the advantage of using the effective Lagrangian $\mathcal{L}_{\text{eff}}^I$ rather than the full SM Lagrangian to describe B decays: NP can be searched for in a model independent way by extracting the Wilson coefficients from observables measured in several different decays.

The computation of observables, such as the branching fraction of a $B \rightarrow K\ell^+\ell^-$ decay in a given q^2 range, involves the computation of some amplitude A of operators from the Lagrangian in Eq. 2.34,

$$A = \langle K\ell^+\ell^- | \mathcal{O}_i | B \rangle \quad (2.36)$$

$$= \underbrace{\langle \ell^+\ell^- | \mathcal{O}_{i\ell} | 0 \rangle}_{A_{i\ell}} \cdot \underbrace{\langle K | \mathcal{O}_{ih} | B \rangle}_{A_{ih}} + \text{corrections}. \quad (2.37)$$

The amplitude is factorised into two independent parts, a leptonic part $A_{i\ell}$, and a hadronic part A_{ih} . This is referred to as “naïve factorisation” [25]. Long distance contributions such as $b \rightarrow c\bar{c}s$ break the factorisation hypothesis, as illustrated in Fig. 2.3 where a charm quark exchanges a gluon with the B hadron. Such contributions are largest in the resonance region, but can also induce a small correction to the naïve factorisation even at low q^2 . This can be understood as coming from the fact that the J/ψ or $\psi(2S)$ resonances have long tails that span over the whole q^2 range [26, 27].

The computation of the Wilson coefficients is perturbative and precise. The hadronic amplitude A_{ih} is parametrised as a function of so-called form factors, which require non-perturbative computations. Long distance contributions such as $b \rightarrow c\bar{c}s$ also require non-perturbative techniques. Different methods are employed to perform these computations — the light cone sum rules and QCD factorisation at low q^2 [25, 28], and lattice computations as well as heavy quark effective theories at high q^2 [25, 29, 30] — and are associated with larger theoretical uncertainties.

In summary, the effective theory describing $b \rightarrow s\ell^+\ell^-$ decays allows searches for NP to be made in a model-independent way by extracting the value of the Wilson coefficients from various observables and checking whether they are shifted from their SM expectation.

However, this procedure is not free from theoretical uncertainties, as the expressions for the observables as a function of the Wilson coefficients involve complex QCD computations.

2.10 Review of $b \rightarrow s\mu^+\mu^-$ experimental results

At the time of writing, the most precise measurements of $b \rightarrow s\ell^+\ell^-$ transitions are the measurements of angular observables and branching fractions made with the data recorded by the LHCb detector in 2011 and 2012. The LHCb detector is described in Chpt. 3.

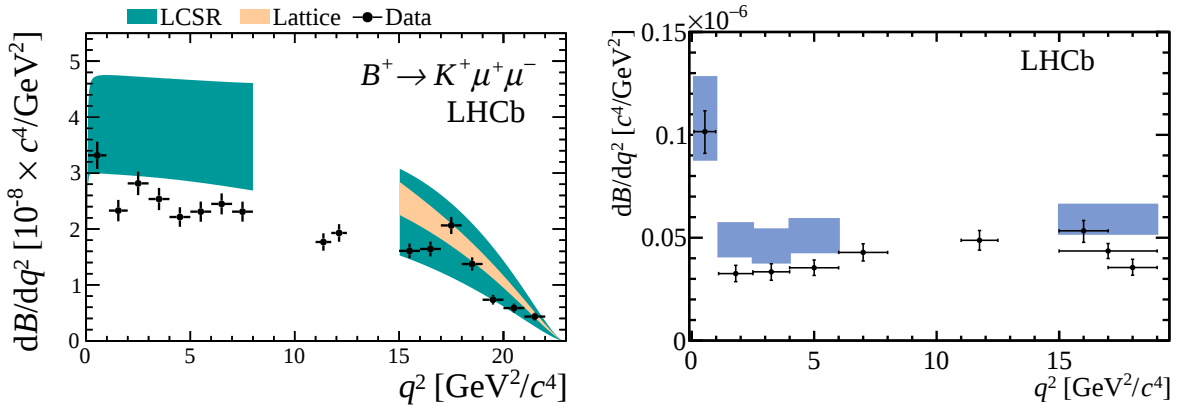


Figure 2.4: Theory prediction and LHCb measurement of the $B^+ \rightarrow K^+ \mu^+ \mu^-$ differential branching fraction (left) and of the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ differential branching fraction (right). Taken from Ref. [31, 32]

The measurements of the differential branching fractions for $B^+ \rightarrow K^+ \mu^+ \mu^-$ and $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decays are shown in Fig. 2.4. For both decays, the experimental results undershoot the theory predictions in the low q^2 region. This discrepancy is also observed in other $b \rightarrow s\mu^+\mu^-$ decays such as $B_s^0 \rightarrow \phi \mu^+ \mu^-$ [33] or $B^0 \rightarrow K_S \mu^+ \mu^-$ [31].

The q^2 -differential measurement of the angular observable P'_5 [34] is shown in Fig. 2.5. The definition of this observable can be found in Ref. [35]. It is measured by analysing the angular distribution of all four decay products in the $B^0 \rightarrow K^{*0}(\rightarrow K^+ \pi^-) \mu^+ \mu^-$ decay. The variable P'_5 is defined in such a way that it has reduced hadronic uncertainties compared to other angular observables. A discrepancy is again visible in the low q^2 region: the LHCb measurement in the fourth and fifth bin in Fig. 2.5 shows a discrepancy with respect to the SM prediction at the level of 3.7 standard deviations.

Using the framework presented in Sec. 2.9, the above experimental measurements, as well as other $b \rightarrow s\ell^+\ell^-$ related results not discussed here, can be interpreted in terms of

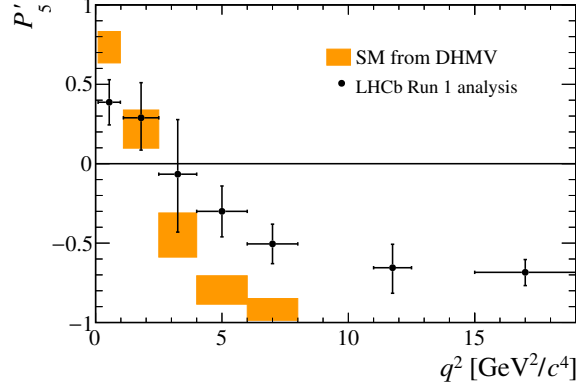


Figure 2.5: Theory prediction, as well as LHCb measurement of the angular observable P'_5 in $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decays. The measurements are from Ref. [34].

the Wilson coefficients. Ref. [36–40] have performed such an analysis and have shown that all the discrepancies in the measurements of both the differential branching fractions and angular observables together point to the same effect. In fact, all tensions can be resolved simultaneously by introducing a shift in either C_9 , or in $C_9 - C_{10}$. The significance of this shift is at the level of 4 to 5 standard deviations.

2.11 The R_K and R_{K^*} measurements

The large shift of the Wilson coefficients is very intriguing. It is however difficult to draw clear conclusions about whether this shift comes entirely from NP, given the theoretical uncertainties in the computations that relate the Wilson coefficients to the $b \rightarrow s \mu^+ \mu^-$ observables. Measurements of observables with very precise theory predictions are therefore well motivated. One instance of such precise observable is the ratio of branching fractions of a $b \rightarrow s \mu^+ \mu^-$ over $b \rightarrow s e^+ e^-$ decay. This defines the observables R_K and R_{K^*} ,

$$R_K = \frac{\int_{q_{\min}^2}^{q_{\max}^2} \frac{d\mathcal{B}(B^+ \rightarrow K^+ \mu^+ \mu^-)}{dq^2} dq^2}{\int_{q_{\min}^2}^{q_{\max}^2} \frac{d\mathcal{B}(B^+ \rightarrow K^+ e^+ e^-)}{dq^2} dq^2}, \quad (2.38)$$

$$R_{K^*} = \frac{\int_{q_{\min}^2}^{q_{\max}^2} \frac{d\mathcal{B}(B^0 \rightarrow K^{*0} \mu^+ \mu^-)}{dq^2} dq^2}{\int_{q_{\min}^2}^{q_{\max}^2} \frac{d\mathcal{B}(B^0 \rightarrow K^{*0} e^+ e^-)}{dq^2} dq^2}. \quad (2.39)$$

As QCD has nothing to do with the leptons, all the hadronic uncertainties completely cancel in the R_K and R_{K^*} ratios. In the SM, these ratios can deviate slightly from unity

due to the fact that electrons and muons have different masses. In particular, in the very low q^2 region, the $B^+ \rightarrow K^+ \mu^+ \mu^-$ branching fraction is reduced as the larger muon mass reduces the phase space available to the $B^+ \rightarrow K^+ \mu^+ \mu^-$ decay. At all q^2 , radiative corrections also affect the two modes differently. In particular, the leptons may interact in the electromagnetic field surrounding the decay and emit radiation. This type of radiation is known as Final State Radiation (FSR) and induces a small correction on the branching fraction which depends on the mass of the leptons. However, these corrections are predicted to be at the $\mathcal{O}(1\%)$ level in the SM [41]. Therefore, measuring a significant deviation of R_K or R_{K^*} from unity would constitute an unambiguous sign of (lepton flavour non-universal) NP. The measurements of R_K and R_{K^*} are not only motivated by the $b \rightarrow s \mu^+ \mu^-$ anomalies, but also by the fact that LFU is less constrained in decays of heavy quarks, as discussed in Sec. 2.6.

The two observables R_K and R_{K^*} have been measured at LHCb in the low q^2 region and, in the case of R_{K^*} , in the very low q^2 region. The results are [42, 43]

$$\begin{aligned} R_K &= 0.745_{-0.074}^{+0.090} \pm 0.036 & \text{for } 1.0 < q^2 < 6.0 \text{ GeV}^2, \\ R_{K^*} &= 0.69_{-0.07}^{+0.11} \pm 0.05 & \text{for } 1.1 < q^2 < 6.0 \text{ GeV}^2, \\ R_{K^*} &= 0.66_{-0.07}^{+0.11} \pm 0.03 & \text{for } 0.045 < q^2 < 1.1 \text{ GeV}^2. \end{aligned} \quad (2.40)$$

All three measurements undershoot the SM prediction by $\sim 30\%$, each of them corresponding to a mild tension at a level of ~ 2.5 standard deviations.

The breaking of lepton universality requires a lepton flavour dependent shift of one or several Wilson coefficients, denoted δC_i^e for $b \rightarrow s e^+ e^-$ and δC_i^μ for $b \rightarrow s \mu^+ \mu^-$. Because the K^+ and K^{*0} have different spins and intrinsic parities, the two R_K and R_{K^*} measurements give complementary information about the Wilson coefficients. This is illustrated in Fig. 2.6, which shows how the two measurements taken together are sufficient to discriminate between a NP contribution in C_9^μ , C_{10}^μ , and $C_9^\mu - C_{10}^\mu$ to $b \rightarrow s \mu^+ \mu^-$ (assuming no NP contribution in $b \rightarrow s e^+ e^-$). The analysis in Ref. [37] shows that the R_K and R_{K^*} measurements, together with the $b \rightarrow s \mu^+ \mu^-$ measurements, favour a scenario in which the NP contribution to $b \rightarrow s e^+ e^-$ is negligible ($\delta C_i^e \approx 0$), but the NP contribution to $b \rightarrow s \mu^+ \mu^-$ is significant. This is illustrated in Fig. 2.6, which shows that a fit to the $b \rightarrow s \ell^+ \ell^-$ measurements with free-floating δC_9^μ and δC_{10}^e favours a displacement from the SM value δC_9^μ at a level of 5 to 6 standard deviations. A fit with free-floating $\delta C_9^\mu - \delta C_{10}^\mu$ also shows a significant pull from the SM expectation [37].

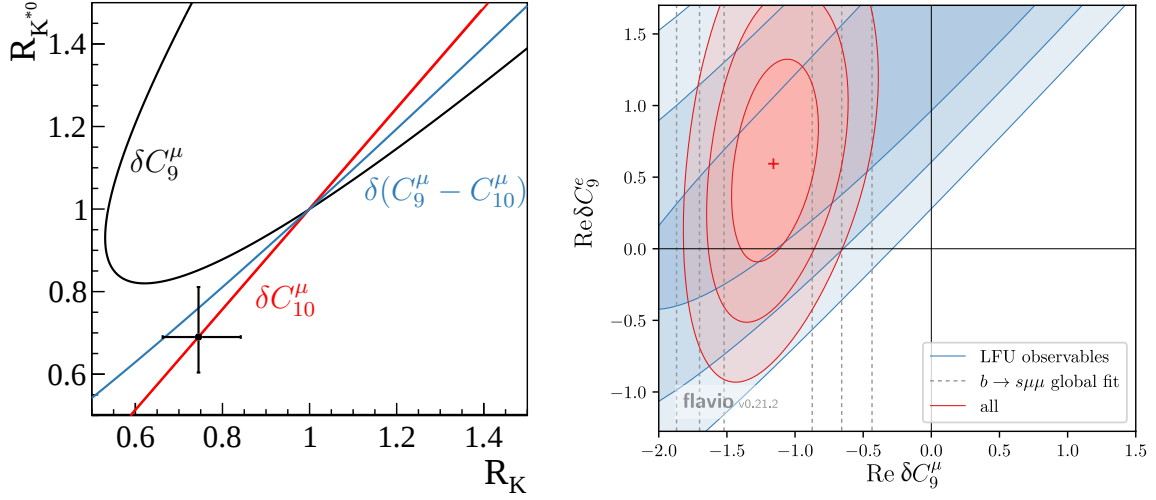


Figure 2.6: Values of R_K and R_{K^*} when varying the Wilson coefficients of the muon mode individually (left). The LHCb results in the low q^2 range are shown with the black error bars. Based on Ref. [38]. Fit to the $b \rightarrow s\ell^+\ell^-$ observables with floating δC_9^μ and δC_9^e (right), including or not the LFU observables R_K and R_{K^*} . Taken from Ref. [37].

2.12 Anomalies in tree level beauty decays

In addition to $b \rightarrow s\ell^+\ell^-$, LFU is also studied in the decays of beauty mesons mediated by $b \rightarrow c\ell\nu$ transitions. In this case, the ratios of branching fractions of interest are denoted R_D and R_{D^*} ,

$$R_D = \frac{\mathcal{B}(B \rightarrow D\tau^+\nu_\tau)}{\mathcal{B}(B \rightarrow D\mu^+\nu_\mu)}, \quad R_{D^*} = \frac{\mathcal{B}(B \rightarrow D^*\tau^+\nu_\tau)}{\mathcal{B}(B \rightarrow D^*\mu^+\nu_\mu)}, \quad (2.41)$$

where the B or $D^{(*)}$ denote either a charged or neutral beauty or charmed meson. There are two important differences between the $R_{D^{(*)}}$ and the $R_{K^{(*)}}$ observables. First, hadronic uncertainties do not perfectly cancel in the $R_{D^{(*)}}$ result, and the uncertainty on the theory prediction is at the level of 1%. Second, $b \rightarrow c\ell\nu$ is a tree level transition with a much higher branching fraction than $b \rightarrow s\ell^+\ell^-$. Following the arguments of Sec. 2.5, NP contributions, if they exist, are expected to be relatively smaller in the $R_{D^{(*)}}$ ratios than in the $R_{K^{(*)}}$ ratios.

The experimental measurements of R_D and R_{D^*} , performed using the LHCb, Belle and BaBar detectors [45], are shown in Fig. 2.7. Their combination shows a discrepancy at the level of four standard deviations with respect to the SM expectation. This constitutes another hint of LFU breaking in the decays of beauty hadron.

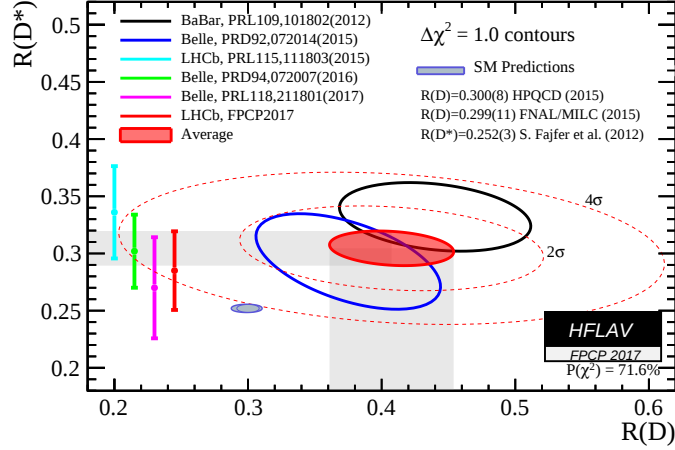


Figure 2.7: Combination of the R_D and R_{D^*} measurement performed by LHCb, BaBar and Belle, and comparison with the theory prediction. Figure taken from Ref. [44].

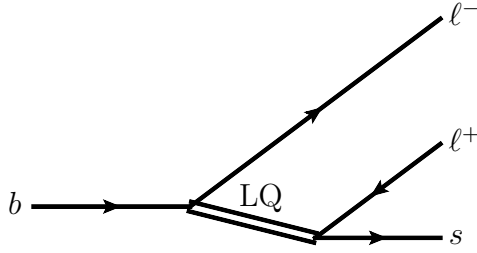


Figure 2.8: NP Feynman diagram for a $b \rightarrow s\ell^+\ell^-$ decay mediated by a LQ.

2.13 Complete models explaining the B anomalies

The effective field theory approach is useful to analyse the tensions appearing in different observables in a model-independent way. It highlights how the $b \rightarrow s\ell^+\ell^-$ anomalies together point to a single modification of the Wilson coefficients. Eventually, these new values must originate from a full theory beyond the SM with new types of interactions. It is hence useful to consider whether theories exist that are able to explain the B anomalies while respecting the constraints for all other measurements, not necessarily concerning $b \rightarrow s\ell^+\ell^-$ decays.

A possible candidate for such a theory involves a new mediator that can couple to both a quark and a lepton, a so-called leptoquark (LQ). This mediator can in principle contribute to $b \rightarrow s\ell^+\ell^-$ decays with the Feynman diagram shown in Fig. 2.8. Some

examples of experimental measurements that place stringent constraints on LQ models are given below.

Lepton flavour: An ideal model that can explain the $b \rightarrow s\ell\ell$ anomalies should also ease the $R(D^{(*)})$ tensions. However, it must at the same time respect the strong constraints in lepton or light meson decays such as those discussed in Sec. 2.6. This means that the LQ must couple very weakly to the first generation of quark and leptons, mildly to the second generation and strongly to the third generation.

Proton decay: The proton has a lifetime of at least $\sim 10^{30}$ years [5]. In the SM, baryon number (*i.e.* number of quarks) is conserved⁴, which forbids the proton to decay. A LQ could in principle break baryon number conservation and generate a decay of the type $p^+ \rightarrow \pi^0 \ell^+$.

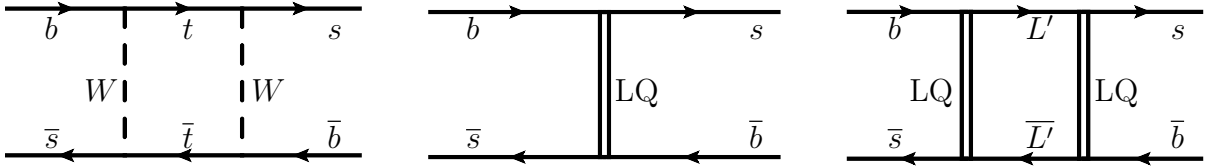


Figure 2.9: Feynman diagrams for B_s^0 mixing in the SM (left), in a NP model where a LQ can induce FCNC at tree level (centre), and in a NP model where quarks couple to a LQ and new heavy leptons L' .

B_s^0 and B^0 mixing: When propagating, neutral beauty mesons B_s^0 and B^0 oscillate between their matter B and anti-matter $\bar{B}$ states. This mixing is mediated by a FCNC and occurs at loop level in the SM, as shown in Fig. 2.9. The frequency at which B_s^0 and B^0 mixing occurs has been measured by several experiments with a sub-percent precision [5]. A new mediator like a LQ can have an impact on the oscillation frequency. This impact can be particularly large in the case where the LQ can induce mixing at tree level, *e.g.* if it can couple to a $b - u$ or $b - s$ quark pair. The theory prediction for the mixing frequency is made using lattice QCD and has a 5% precision [47]. This is a worse precision than the experimental result and hence limits the ability to search for NP. This means that a NP

⁴There exists however a solution to the SM equations which constitutes an exception to the baryon number conservation. This configuration of the SM fields might hypothetically have existed in the early universe [46].

model predicting a difference in the mixing frequency with respect to the SM up to a level of a few percent is compatible with existing experimental constraints.

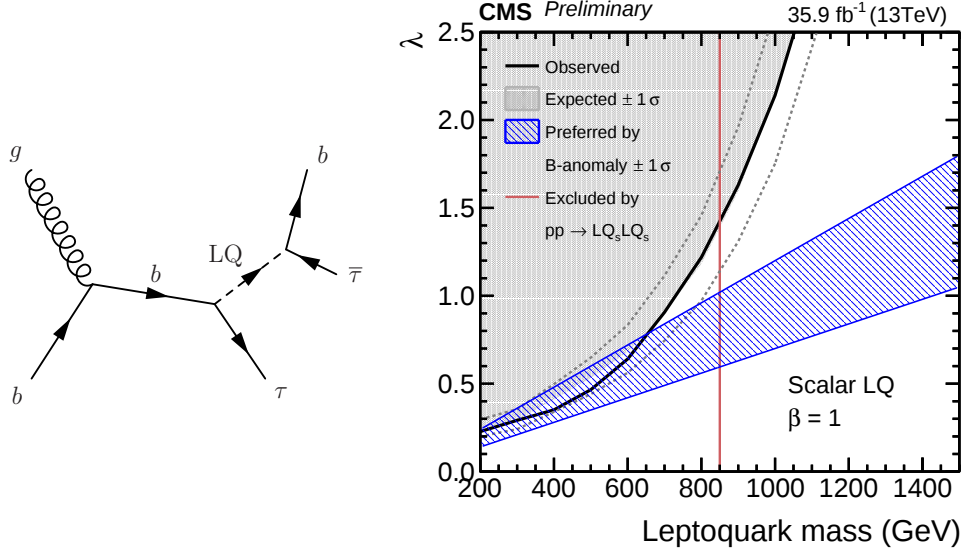


Figure 2.10: Search for a LQ coupling to a τ lepton and b quark using the CMS detector. Feynman diagram for the production and decay of the LQ in proton collisions (left), and exclusion region in the mass, coupling (λ) plane (right). Superimposed on the mass, coupling plane is the region preferred by the $b \rightarrow s\ell^+\ell^-$ anomalies in a simplified scalar LQ model. Figures taken from Ref. [48].

Direct searches: Any new particle predicted by a NP model, including the LQ itself, must be compatible with the limits derived from direct searches. An example of such a direct LQ search performed using the CMS detector [49] is shown in Fig. 2.10.

A class of NP models able to explain the flavour anomalies while evading these constraints are the so-called Pati-Salam models [50]. The basic idea of such models is to extend the $SU(3)$ symmetry of QCD to $SU(4)$. Fermions are quadruplets of this group, with three entries corresponding to quarks of the three colours, and one entry being a lepton. This symmetry is then broken down to $SU(3)$, with quarks remaining charged under this group, and leptons being singlets of that group. While the unbroken generators correspond to the gluons, the broken ones will give rise to new massive bosons, including a LQ. These models are attractive, from the theoretical point of view, because they unify quark and leptons. However, Pati-Salam models with the same fermion content as the SM must

have a very massive LQ to evade the LFV constraints in light meson decays, and are then unable to explain the B anomalies.

Some models [51–53] propose to extend the fermionic content of Pati-Salam models to include new, heavy fermions. These models are constructed in such a way that only the new heavy fermions couple to the LQ. However, Yukawa couplings induce a mixing between the new heavy fermions and the light SM fermions, in the same way as the SM Yukawa couplings including a mixing between quark flavours. The mixing terms are chosen to be very small for light leptons and quarks, but larger for quarks and leptons in the second and third generation. This means that the LQ acquires an effective coupling to fermions which is larger for higher generations. Using a LQ of a mass at the TeV scale allows the $b \rightarrow s\ell\ell$ anomalies to be explained, eases the tensions in $R(D^{(*)})$, while evading the constraints discussed in Sec. 2.6. Moreover, these models conserve baryon number and proton decays are forbidden. This is related to the fact that the LQ only couple to a quark-lepton pair, and not to a quark-quark pair. This then means that this LQ contributes to B_s^0 and B^0 mixing only at loop level, as shown in Fig. 2.9. The loop contains one of the new heavy fermions, meaning that the contribution to the mixing can be further suppressed with new fermions of a sufficiently high mass.

2.14 Purpose of the thesis

In summary, anomalies have appeared in several observables related to the decays of beauty hadrons. These anomalies appear in various observables in loop level $b \rightarrow s\ell^+\ell^-$ mediated decays, and can be resolved by changing a single effective coupling constant in the theory describing these decays. Some anomalies also appear in tree level decays, and NP models unifying quark and leptons are able to explain both the tree and loop level anomalies. The anomalies could be a sign of the presence of physics beyond the SM. However, the current limited precision of the experimental measurements, and the uncertainties associated with several theory predictions make it impossible to draw a definite conclusion.

A significant discrepancy measured in a theoretically clean observable with respect to its SM expectation would assuredly confirm the presence of NP. The current thesis is dedicated to the update of the R_K measurement using a dataset approximately twice as large as that used to perform the measurement shown in Eq. 2.40. This has the potential to bring the R_K discrepancy to a level of four standard deviations, assuming the central value remains the same as the previous measurement.

3. Experimental apparatus

The R_K measurement presented in this thesis is performed using B mesons produced in proton-proton collisions generated at the Large Hadron Collider and recorded with the LHCb detector. This section summarises the aspects of the detector that are relevant for the R_K measurement.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [54] was designed and constructed at the European Organisation for Nuclear Research (CERN). The purpose of the LHC is to accelerate and collide protons at the unprecedented centre-of-mass energy of 14 TeV, and it is also used to accelerate heavy ions. The LHC is situated in a 27 km long tunnel approximately 100 m below the ground on the Swiss/French border close to Geneva.

After being accelerated up to 450 GeV in a set of several pre-accelerators, protons are injected into the LHC in two beams circulating in opposite directions. To accelerate the protons, the LHC uses 16 radio frequency chambers. Moreover, 1232 dipole magnets are used to keep the protons on their circular trajectories, and 392 quadrupole magnets are used to squeeze and collimate the beams. All magnets are superconducting, cooled down to 1.9 K with liquid helium. The beams are not continuous but consist of so-called bunches made of approximately 10^{11} protons, the space between them set the rate of the collisions. The protons (or ions) are then made to collide at four points, where four main particle detectors record the products of the collisions: ALICE [55], dedicated to the study of ion collisions, the two general purpose detectors ATLAS [56] and CMS [49], and LHCb, dedicated to heavy flavour physics.

The data used for the analysis presented in this thesis were taken in 2011, 2012, 2015 and 2016. The 2011 and 2012 data are referred to as Run 1 and the 2015 and 2016 data as Run 2. At the time of writing, data were collected for two additional years, 2017 and 2018. A two year shutdown in 2019, 2020 will be used to upgrade the detectors. Collisions took place at a higher energy and a higher rate in Run 2 than in Run 1 [57, 58]. Tab. 3.1 shows the bunch spacing, the integrated luminosity recorded at LHCb $\int \mathcal{L} dt$, the centre-of-mass energy of the collisions $\sqrt{s}$ and the cross-section $\sigma_{b\bar{b}}$ for producing pairs of beauty quarks and antiquarks in the LHCb acceptance. Due to the enhanced $\sigma_{b\bar{b}}$ in Run 2 with respect to Run 1, the LHC produced, for a given luminosity, approximately twice as many B mesons

in Run 2 compared to Run 1.

Table 3.1: Centre-of-mass energy, $b\bar{b}$ production cross-section, bunch spacing and integrated luminosity recorded at LHCb for the four years of LHC data used in the analysis presented in this thesis.

year	Run 1		Run 2	
	2011	2012	2015	2016
$\sqrt{s}$ [TeV]	7	8	13	13
$\sigma_{b\bar{b}}$ [μb] [59]	~ 72	~ 82	~ 144	~ 144
bunch spacing [ns]	50	50	25	25
$\int \mathcal{L} dt$ [fb^{-1}]	1.1	2.1	0.3	1.6

3.2 Interaction of charged particles with matter

Before introducing the LHCb detector, this section first reviews the way in which particles going through a medium interact with the atoms in this medium. The focus is mainly on electrons, kaons and muons with a momentum of ranging between 1 and 100 GeV, as they are the most relevant for the R_K analysis.

The quantity of interest in this section is the mean energy loss per distance travelled in the medium, $\langle dE/dx \rangle$. It is shown in Fig. 3.1 for electrons and muons going through iron. The details of this plot are discussed in the sections below.

3.2.1 Energy loss by ionisation and excitation

When a charged particle goes through a medium, it loses energy interacting with electrons in that medium by exciting or ionising atoms. The energy loss corresponding to this effect is shown with the dashed lines in Fig. 3.1. The energy loss can be computed using the Bethe-Bloch formula [5], which considers the Rutherford scattering of the incoming particle against electrons in the effective potential of the medium.

The interaction of particles with electrons in a medium is the pillar of tracking, *i.e.* reconstructing the trajectories of particles through a detector. In a gas detector, a particle ionises the atoms of the gas, and in a silicon detector, a particle creates electron-hole pairs. The charges deposited can be detected and constitute a ‘hit’. Several hits deposited by a

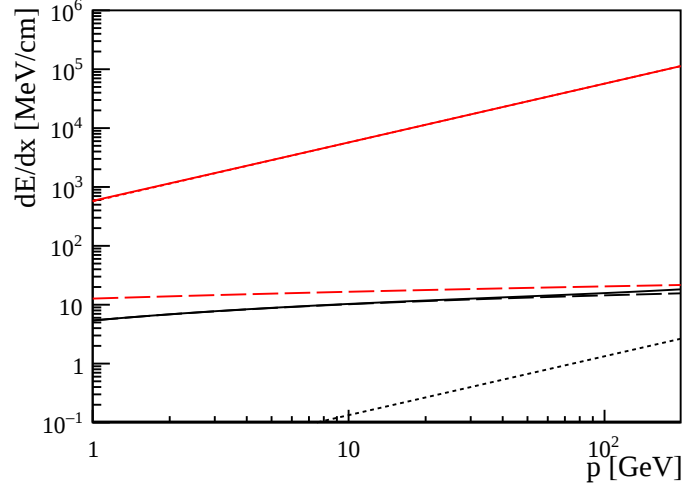


Figure 3.1: Energy lost per distance travelled in iron for muons (black) and electrons (red) as a function of momentum. The solid line corresponds to the total energy loss, the dotted line to ionisation and the dashed line to bremsstrahlung.

particle can be fitted together to reconstruct the trajectory of the particle, referred to as a ‘track’.

For muons going through the LHCb detector, interaction with electrons in the medium is the main source of energy loss, and remains very small. As can be seen in Fig. 3.1, in the momentum range of interest, $\langle dE/dx \rangle$ for muons is almost flat and corresponds to a region where the energy losses are minimal. Muons at LHCb are therefore referred to as Minimum Ionising Particles (MIP). This is however not the case for electrons, as discussed in the next section.

3.2.2 Energy loss by bremsstrahlung

Charged particles going through a medium also interact with nuclei emitting photons¹. The photons emitted are almost collinear to the incoming particle, which is slowed down by the emission. This process is called bremsstrahlung, which is German for ‘braking radiation’. For an electron going through matter, the average distance after which bremsstrahlung reduces the electron energy by a factor $1/e$ is referred to as a ‘radiation length’, X_0 . The radiation length depends on the nature of the medium and is proportional to the inverse squared mass of the incoming particle: $X_0 \sim 1/m^2$ [60]. Hence, as illustrated in Fig. 3.1,

¹This interaction also occurs with atomic electrons, but is more frequent with large nuclei that carry a larger electric charge.

it is much larger for electrons than for muons. In the following sections, X_0 refers to the electron radiation length.

3.2.3 Electromagnetic showers

High energy photons going through a medium mainly lose energy via pair production, *i.e.* by producing an electron-positron pair in the electromagnetic field of the atoms in the medium. The mean free path of a high energy photon in a medium before pair producing is $\frac{9}{7}X_0$ [5]. Hence, a high energy electron in a thick piece of material will deposit its entire energy in an electromagnetic shower: it emits bremsstrahlung photons that will in turn produce pairs of electrons and positrons, which will in turn emit bremsstrahlung photons and so on.

3.2.4 Hadronic interaction and showers

Unlike leptons, hadrons can interact with nuclei in a medium via the strong interaction. The distance a hadron can travel before undergoing an inelastic hadronic interaction is characterised by the nuclear interaction length X_h , which is typically much larger than the radiation length X_0 . In thick materials, highly energetic hadrons lose their energy emitting secondary particles (pions, electrons, photons, etc.) in hadronic showers, which are typically broader than electromagnetic showers. Showers provide a destructive way to measure the energy of a particle, as explained in Sec. 3.7.

3.2.5 Cherenkov radiation

When a charged particle moves through a dielectric medium, it can excite atoms and molecules which then relax to their ground state emitting photons. When the particle travels at a higher velocity than the speed of light in the medium, the photons emitted in the wake of the particle form a coherent front in the shape of a light cone. The light cone of this so-called Cherenkov radiation has an opening angle θ_c related to the velocity of the particle β and the refractive index n ,

$$\cos \theta_c = \frac{1}{n\beta}. \quad (3.1)$$

By measuring the opening angle of the Cherenkov light cone emitted by a particle and knowing the momentum of the particle, the mass of the particle, and hence its identity, can be inferred.

3.3 Layout of the LHCb detector

The LHCb detector [61] is mainly dedicated to the study of beauty hadrons. A schematic of the detector is shown in Fig. 3.2. Its most striking characteristic is that it is a forward detector, covering the pseudorapidity range $2 < \eta < 5$. This design is chosen because B mesons are produced very boosted in the proton collisions, and their decay products are collimated relatively close to the beam line. The various LHCb subdetectors, as well as the information they provide for the study of B decays, are described in detail in the following sections. Fig. 3.2 also defines the coordinate system used throughout this thesis: the z axis goes along the beam pipe, the y points upwards, and the x axis points to the left of a person standing at $z = 0$ and facing the detector. The transverse momentum of a particle p_T is then defined as the component of the momentum in the x, y plane. The angle between the momentum of a particle and the z axis is denoted θ , and is related to the pseudorapidity η with

$$\eta = -\log \left(\tan \left(\frac{\theta}{2} \right) \right). \quad (3.2)$$

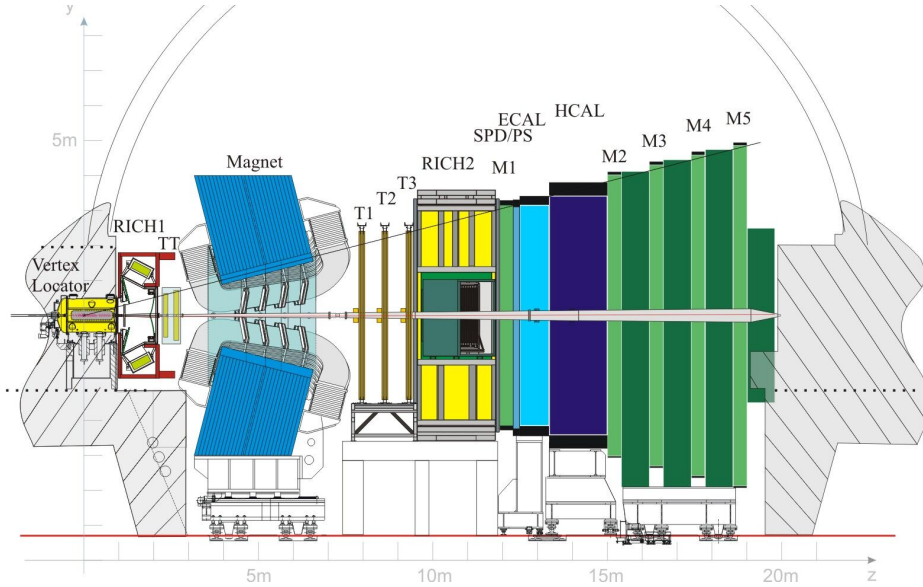


Figure 3.2: Schematic of the LHCb detector, taken from [61].

3.4 Tracking and vertexing

The subdetectors and algorithms used to reconstruct tracks in the LHCb detector are presented in this section.

3.4.1 Vertex locator

The VERtEX LOcator (VELO) [62], on the very left of Fig. 3.2, is the subdetector closest to the interaction point, and is designed to be able to resolve the decay vertices of beauty or charmed mesons from primary vertices (proton-proton collision). The VELO is made of two halves, one on the left and one on the right of the beam, and each half is composed of 21 modules. Each module contains two semi-circular silicon-strip sensors: one with radial strips to measure the polar angle of hits (ϕ sensors) and one with circular strips to measure the axial distance of hits (R sensors). These two kinds of sensors are illustrated in Fig. 3.3. The strip pitch ranges between $\sim 40\text{ }\mu\text{m}$ close to the beam to $\sim 100\text{ }\mu\text{m}$ furthest from the beam, achieving a maximal precision on a single hit of $\sim 4\text{ }\mu\text{m}$.

During the data-taking, the sensors are positioned 8 mm away from the beam line. During injection, when the beams are not completely stable, the sensors are retracted by $\sim 3\text{ cm}$.

The proton beams generate rapidly-varying electromagnetic fields that can be harmful to the VELO and its electronics. The VELO sensors are therefore separated from the beam by a corrugated aluminium foil. This is illustrated on the right of Fig. 3.3.

The precision the VELO achieves on the position of primary vertices depends largely on the number of tracks that compose the vertex: the resolution on the x (z) position is approximately $35\text{ }\mu\text{m}$ ($260\text{ }\mu\text{m}$) for a vertex with five tracks, and approximately $12\text{ }\mu\text{m}$ ($80\text{ }\mu\text{m}$) for a vertex with 30 tracks [63].

3.4.2 Tracking system

The LHCb tracking system is formed of tracking stations upstream and downstream of a large dipole magnet. The momentum of tracks is inferred from their bending in the magnetic field.

The first part of the tracking system is the Tracker Turicensis (TT) [64] placed just in front of the magnet. It consists of four layers of silicon strip detectors with a strip pitch of $183\text{ }\mu\text{m}$.

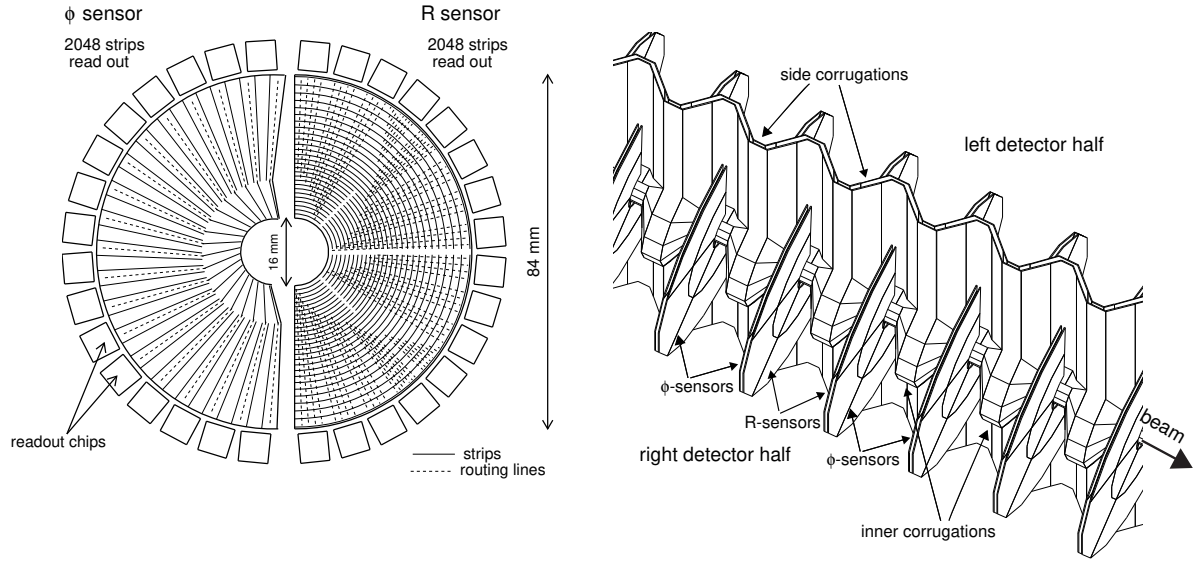


Figure 3.3: Schematic of the two kinds of VELO sensors (left). The ϕ sensors measure the polar angle and r sensors measure the radial distance from the beam pipe. Schematic illustrating how the sensors are placed around the beam pipe (right). The corrugated foil protecting the sensors is also visible in this diagram. Both figures are taken from Ref. [62].

Just downstream of the magnet is a set of three tracking stations, labeled T1, T2, T3 in Fig. 3.2, each containing four tracking modules placed in layers. The density of tracks is typically higher towards the beam line. Hence, the tracking modules are instrumented using a silicon tracker in the central region, and a gas-based detector in the outer region. The central region constitutes the inner tracker (IT), and the outer region the outer tracker (OT).

Each IT [65] module is instrumented in a similar way as the TT, using silicon strip detectors with a strip pitch of 198 μm . Each OT [66, 67] module contains two layers of drift-tubes, each tube consisting of a hollow cylinder with inner diameter of 4.9 mm filled with gas, and an anode wire in the centre. When particles pass through the tubes and ionise the gas, electrons drift towards the wire. The drift time is measured, and knowing the drift velocity, the position of the track is then inferred.

3.4.3 Magnet

The magnet [68] is a warm dipole, producing a magnetic field along the y axis. The maximal magnetic field at the centre of the magnet is approximately 1.1 T, and the integrated

magnetic field along the path of a particle going through the detector is approximately 4 Tm. Typically, the component p_x of a particle’s momentum receives a kick of approximately 1 GeV when going through the magnet.

3.4.4 Track reconstruction and performance

To fit tracks from hits deposited in the various tracking subdetectors, a Kalman filter algorithm [69,70] is used, which in essence is similar to a χ^2 minimisation. The fit considers all tracks as MIPs, in the sense that the fit shape takes into account energy loss due to ionisation, but not bremsstrahlung [71,72]. From the track fit and the bending of the tracks in the magnet, the momentum of the particles is determined.

Fake tracks that consist of fake hits or hits from different particles are referred to as ‘ghosts’. They are characterised by a larger χ^2 and fewer hits than a track coming from a single particle. Combining information from the track fit, a ghost probability, $\text{prob}_{\text{ghost}}$, is estimated for each track [73,74].

For MIPs, the tracking efficiency depends on the kinematics of the track, and is above 95% for tracks with $p_T > 2.5$ GeV. The momentum resolution is below 0.5% [75]. Due to bremsstrahlung, electrons give more smeared tracks and can be more frequently stopped before reaching the end of the tracking system. The electron tracking efficiency is approximately 15% lower than for MIPs, and the momentum resolution, which depends on the kinematics, is on the order of 10%.

3.5 Particle identification

This section details how particle identification (PID) is performed at LHCb, *i.e.* how a given particle species is associated with one track. Together with the momentum information provided by the tracking system, the PID allows the energy of a charged track to be computed using $E = \sqrt{p^2 + m^2}$.

For the present analysis, the three sub-detectors that provide the majority of the PID information are the RICH, the calorimeter system, and the muon system. They are described in more detail in the following sections. These sub-detectors provide a likelihood $L_\beta(t)$ of a track t belonging to a given particle species β . For the muon, electron or pion

hypothesis, these three likelihoods can then be combined into a global likelihood L^{tot} ,

$$L_{\mu}^{\text{tot}}(t) = L_{\mu}^{\text{RICH}} \cdot L_{\text{non } e}^{\text{calo}}(t) \cdot L_{\mu}^{\text{muon}}(t), \quad (3.3)$$

$$L_e^{\text{tot}}(t) = L_e^{\text{RICH}} \cdot L_e^{\text{calo}}(t) \cdot L_{\text{non } \mu}^{\text{muon}}(t), \quad (3.4)$$

$$L_K^{\text{tot}}(t) = L_K^{\text{RICH}} \cdot L_{\text{non } e}^{\text{calo}}(t) \cdot L_{\text{non } \mu}^{\text{muon}}(t), \quad (3.5)$$

where the superscripts indicate which sub-detector provides the likelihood information. The analysis uses the variable $\text{DLL}_{\beta}(t)$, which expresses how much more probable it is that t belongs to particle species β , rather than being a pion,

$$\text{DLL}_{\beta}(t) = \log L_{\beta}(t) - \log L_{\pi}(t). \quad (3.6)$$

Another way of combining PID information from different sub-detectors is to train a neural network to differentiate between particle species [76]. This defines the $\text{probNN}_{\beta}(t)$ variable, which is also used in this analysis. This variable is close to 1.0 if t is likely to be a track from a β , and close to 0 otherwise. The performance of the PID variables in reducing backgrounds with mis-identified tracks in this analysis is discussed in Sec. 6.7 and 7.3.

3.6 Cherenkov detectors

There are two Ring Imaging Cherenkov (RICH) [77] detectors at LHCb, upstream and downstream of the magnet, providing PID for charged particles. They are labeled RICH1 and RICH2 in Fig. 3.2. Both detectors are filled with a gas in which charged particles radiate a cone of Cherenkov light. The light is then reflected by a set of mirrors onto a plane of hybrid photomultipliers situated outside of the LHCb acceptance. Projected onto the photomultiplier plane, the Cherenkov cone forms a ring, the radius of which can be used as a measure of the Cherenkov angle.

The two Cherenkov detectors contain two different radiators to provide PID in different momentum ranges. Upstream of the magnet, RICH1 contains C_4F_{10} with $n = 1.0018$ (the refractive indices depend slightly on pressure and temperature) which provides PID in the momentum range $1 < p < 60 \text{ GeV}$. Downstream of the magnet, RICH2 contains CF_4 with $n = 1.0005$, which provides PID in the range $15 < p < 100 \text{ GeV}$.

The Cherenkov angle as a function of momentum for electrons, muons, kaons and pions is shown in Fig. 3.4, illustrating how the measurement of the angle, together with the momentum of the track provided, separate these four particle species. It is also visible

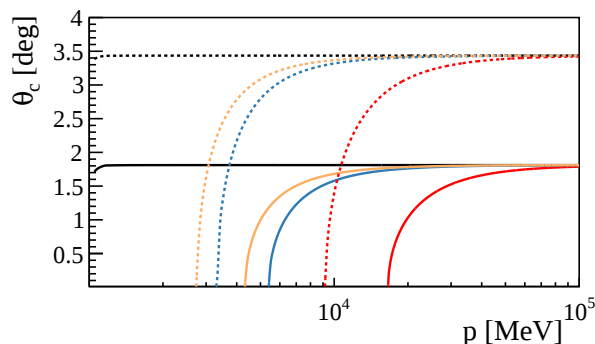


Figure 3.4: Cherenkov angle as a function of momentum for three different particle species: electron (black), pion (orange), muon (blue) and proton (red). The dotted lines correspond to the RICH1 subdetector, the solid lines to the RICH2 subdetector.

that electrons, being lighter, always saturate the RICH and produce rings of maximal radius. Although the RICH detectors provide the main input for the PID of kaons versus pions, electrons are identified mainly using the calorimeter system, as explained in the next section.

3.7 Calorimeter system

The calorimeter system helps to discriminate between photons, electrons and hadrons. It is made of three main parts: the scintillating pad (SPD) and pre-shower detector (PS), the electromagnetic calorimeter (ECAL) and the hadronic calorimeter (HCAL). Each of these subdetectors is presented in this section.

The SPD and PS are almost identical and both consist of scintillating pads. When charged particles go through a pad, they produce scintillation light, which is absorbed and guided by wavelength shifting fibres to photomultiplier tubes. The two detectors are separated by a 15 mm thick lead plate. The size of the SPD and PS pads is chosen to match that of the ECAL cells described below.

The ECAL is a sampling lead detector. It is made of 6016 cells, each of them consisting of plates of lead interlayered with scintillator plates. When an electromagnetic shower develops in the lead plates, secondary particles produce scintillation light, which is guided to the readout by wavelength shifting fibres. The energy deposited by the shower can be deduced from the intensity of the scintillation light collected. The ECAL cells come in three sizes: $4.04 \times 4.04 \text{ cm}^2$ in the inner region, $6.06 \times 6.06 \text{ cm}^2$ in the middle and

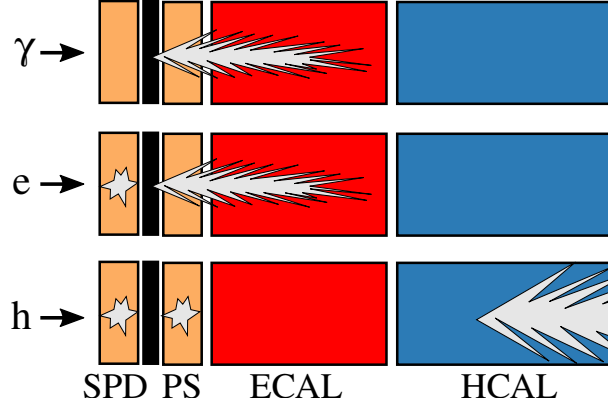


Figure 3.5: Schematic of the calorimeter system illustrating the energy deposit of a photon (γ), an electron (e), and a charged hadron (h).

$12.12 \times 12.12 \text{ cm}^2$ in the outer region.

The HCAL works in a similar fashion as the ECAL. It is made of 1488 cells consisting of iron plates and scintillating plates. The cells come in two transverse sizes: $13.13 \times 13.13 \text{ cm}^2$ in the inner region and $26.6 \times 26.6 \text{ cm}^2$ in the outer region.

The way in which the different components of the calorimeter system can discriminate between particles is illustrated in Fig. 3.5. The lead plate between the SPD/PS detector has a thickness corresponding to $2.5 X_0$ but only $0.08 X_h$. Hence, most of photons and electrons start showering, but not the hadrons. The ECAL thickness corresponds to $25 X_0$ and $0.8 X_h$, meaning that it can contain most electromagnetic showers entirely. Hadrons typically start showering in the HCAL, or late in the ECAL. The HCAL has a thickness corresponding to $5.6 X_h$, typically not enough to contain a full hadronic shower, but sufficient for the purpose of particle identification and trigger. The SPD helps discriminate between photons and electrons, as a hit there indicates the presence of a charged particle. The ECAL and HCAL are also very important for the trigger, as detailed in Sec. 3.10.

3.8 Muon system

The muon system [78, 79] provides a very powerful way to discriminate muons against hadrons, in particular pions. Pions are very problematic as they have a mass very close to that of muons and they constitute most of the tracks generated in proton-proton collisions.

The muon system consists of five stations, one upstream of the calorimeters (M1) and four downstream (M2-M5 in Fig. 3.2). The downstream stations are interleaved with

three 80 cm-thick iron shields. These three shields cover, together with the calorimeters, approximately $21 X_h$.

Each muon station is divided into 276 chambers designed to detect charged tracks. The chambers have an increasing size from the beam pipe. The 12 chambers closest to the beam pipe in M1, where the occupancy is highest, are gas electron multiplier (GEM) detectors [80]. The other 264 chambers are multi-wire proportional chambers (MWPC) [81].

The way in which the muon system provides particle identification information is based on the fact that only muon tracks are able to reach and leave hits in the muon stations, as hadrons and electrons are stopped before in the calorimeters or iron shields. In this analysis, two PID variables are computed from the information collected from the muon stations [82]. The first one is the binary `isMuon` variable. A track satisfies the `isMuon` requirements if it deposits hits in two (if $p < 6$ GeV), three (if $p < 10$ GeV) or four (if $p > 10$ GeV) stations among M2 to M5. The second variable is a likelihood, which is based on the pattern of hits recorded in the muon stations, around the extrapolated position of a track. The muon system is also used by the trigger, as explained in Sec. 3.10.

3.9 Bremsstrahlung recovery

When an electron loses momentum by emitting bremsstrahlung before reaching the magnet, it results in a momentum measurement that is below the real value. In the case of a $B^+ \rightarrow K^+ e^+ e^-$ decay, the momentum loss by the electrons will result in an underestimated B^+ mass. This effect can be partially compensated by recovering some of the bremsstrahlung energy deposited in the ECAL. In fact, as bremsstrahlung photons are collinear to the trajectory of the electron, their energy can be searched for in a region defined by extrapolating the upstream track segment, found in the VELO and TT, to the ECAL. The momentum of the recovered bremsstrahlung cluster is then added to that of the electron. Moreover, the likelihood of an ECAL cluster being compatible with being a bremsstrahlung associated with a given track contributes to the total DLL_e .

In both $B^+ \rightarrow K^+ e^+ e^-$ and $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ decays, approximately half of the electrons have at least one bremsstrahlung cluster associated with them. Electrons with a bremsstrahlung cluster associated have an improved momentum resolution and better PID performance compared to electrons without an associated bremsstrahlung cluster.

3.10 Trigger system

Approximately 35 kBytes of memory are necessary to store an LHCb event. The rate of visible collisions, approximately 10 – 20 MHz, is too high to allow every single event to be stored. Instead, the trigger system [83] is used to select the events that are most likely to contain a B meson, which are the only events stored. The trigger decision is based on the fact that tracks coming from a B meson decay typically have a relatively large transverse momentum and are detached from the primary vertex.

The trigger system is divided into two stages: the low level trigger (L0), implemented in electronic boards, and the high level trigger (HLT), which is software-based and runs on a computer farm. Both stages are described below.

3.10.1 Low level trigger

The L0 trigger reduces the data rate to approximately 1 MHz, which is a manageable input for the HLT. The L0 decisions are based on information provided by the calorimeter and muon systems only. The trigger algorithm first discards events with too many hits in the SPD detector. Such high occupancy events would otherwise take too long to fully reconstruct. The L0 algorithm then searches for objects with a high p_T : either a single hadron, electron, photon, or muon, or a system formed of two muons. In this analysis, only the decisions based on a single muon, electron or hadron are used (see Sec. 6.8). The trigger algorithms taking these decisions are respectively referred to as **L0Muon**, **L0Electron**, **L0Hadron**.

The **L0Muon** algorithm [84] detects aligned hits in all five muon stations. From the spatial coordinates of the track formed by these hits, the L0 computes an approximate estimate of the p_T of the track. The event is saved if the p_T is above a threshold, which is typically around 1.8 GeV, but the threshold value changes depending on the data-taking conditions.

The **L0Electron** algorithm [85] computes the transverse energy E_T in clusters of 2×2 ECAL cells. The E_T is computed as

$$E_T = E \cdot \sin \theta, \quad (3.7)$$

where E is the energy deposited in the four cells and θ is the polar angle of the line joining the centre of the four cells and the proton interaction point. If the energy cluster in the

ECAL is associated with hits in the PS and SPD, and if the E_T is above a threshold, the event is saved. The threshold is typically around $2.5 - 3.0$ GeV.

The L0Hadron algorithm works in a very similar fashion as L0Electron, requiring hadronic showers to have a transverse energy above a threshold. Energy deposits are searched for in clusters of 2×2 HCAL cells. The E_T corresponding to these HCAL clusters is computed including the energy found in the adjacent ECAL cells. The E_T threshold is typically between 3.5 and 4.0 GeV.

3.10.2 High level trigger

Events that satisfy the L0 requirements are then processed by the HLT. The HLT reduces the data rate from the 1 MHz L0 output rate to 5 kHz in Run 1 and 12.5 kHz in Run 2, which is low enough to store the events on disk. The HLT runs in two stages, HLT1 and HLT2. The 1 MHz L0 rate is first reduced to approximately 70 kHz by HLT1, and subsequently further reduced by HLT2.

Using information from the tracking system, the HLT1 performs a fast and partial event reconstruction, fitting only tracks that have a high impact parameter IP as measured in the VELO. If some tracks satisfy a set of requirements based on the p_T and the IP, the event is passed to the HLT2 algorithm. Details about the HLT1 selection used for the R_K analysis are presented in Sec. 6.8.

The HLT2 algorithm reconstructs all tracks in the event that satisfy $p_T > 0.3$ GeV. The selection which is then performed is either exclusive, *i.e.* dedicated to select one particular decay, or inclusive. The HLT2 selection used in this analysis, detailed in Sec. 6.8.2, is inclusive and is essentially based on looking for systems of two or three tracks the topology of which is compatible with coming from the decay of a heavy object. The way in which the HLT2 is implemented differs between Run 1 and Run 2. In the Run 1, the detector is not fully calibrated before the HLT2 selection is run, meaning that the variables used to perform the HLT2 selection differ slightly from the fully calibrated variables used for physics analyses. In Run 2, the detector is fully calibrated before running the HLT2 selection.

3.11 Material in the LHCb detector

As electrons interact in the material of the detector, the probability of detecting an electron track is very sensitive to the amount of material in the detector. The material upstream

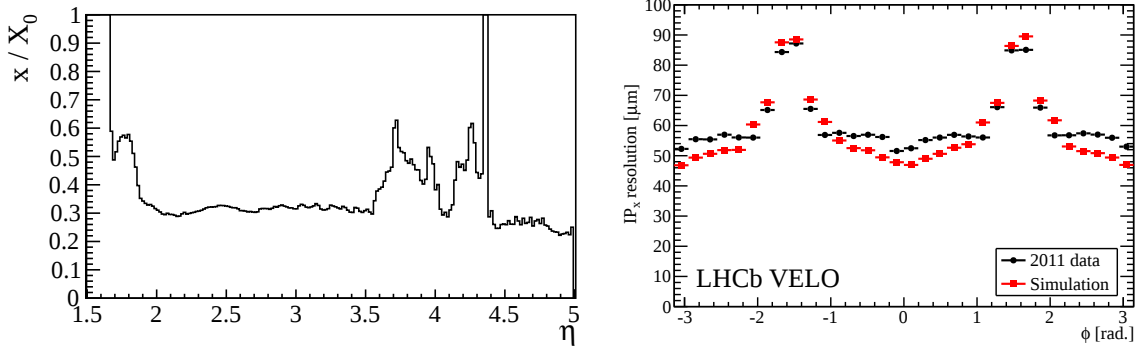


Figure 3.6: Amount of material, determined from simulation (left), traversed by a particle going from the proton-proton interaction point to the magnet, as a function of the rapidity. Resolution on the impact parameter as a function of the polar angle (right), for simulation and data collected in 2011, taken from Ref. [63].

of the magnet is of particular importance, as it worsens both the tracking efficiency and the momentum resolution and reduces the energy of electrons reaching the magnet. The latter also reduces the trigger efficiency.

Fig. 3.6 shows the amount of material traversed by a particle going from the proton interaction point to the magnet, as a function of the rapidity η , in units of X_0 . For $2 < \eta < 3.5$, the material distribution is approximately flat at a third of a radiation length. At high rapidity, closer to the beam pipe, the amount of the material is higher and varies quickly as a function of η . In particular, the peak at $\eta \approx 4.3$ is due to the beam pipe inside RICH1.

The material distribution also depends on the polar angle ϕ , and is higher around $\phi = \pm\pi/2$ rad where the corrugated foil in the VELO is situated. This impacts directly the IP resolution, as illustrated in Fig. 3.6 right. The IP resolution is much worse for tracks going through the foil. The comparison between data and simulation also illustrates that the amount of material in the VELO is imperfectly known. The uncertainty on the material budget upstream of the magnet is estimated to be 10% [86].

Part II

The R_K analysis

4. Analysis strategy

In this chapter, the strategy of the R_K measurement is presented, as well as the main challenges of the analysis.

4.1 $K^+e^+e^-$ and $K^+\mu^+\mu^-$ final states at LHCb

Throughout the analysis, two final states are considered which both have three tracks reconstructed and identified as $K^+\ell^+\ell^-$, where the lepton pair $\ell^+\ell^-$ is either a di-electron pair (e^+e^-) or a di-muon pair ($\mu^+\mu^-$). The 2D distributions $m(K^+e^+e^-), q^2$ and $m(K^+\mu^+\mu^-), q^2$ are shown in Fig. 4.1. The events shown are those from the entire dataset with the selection criteria and trigger requirements discussed in Sec. 6.3 and Sec. 6.8 applied. Various features are visible in these plots:

- A peaking structure centred at $m(K^+\ell^+\ell^-) \approx 5.28$ GeV and $q^2 \approx 9.6$ GeV², corresponding to the resonant B -decay $B^+ \rightarrow K^+J/\psi$, with the J/ψ subsequently decaying as $J/\psi \rightarrow \ell^+\ell^-$. The radiative tails from this decay appear as a diagonal elongated band;
- Another peaking structure centred at $m(K^+\ell^+\ell^-) \approx 5.28$ GeV and $q^2 \approx 13.6$ GeV², corresponding to the resonant B -decay $B^+ \rightarrow K^+\psi(2S)$, with $\psi(2S)$ decaying as

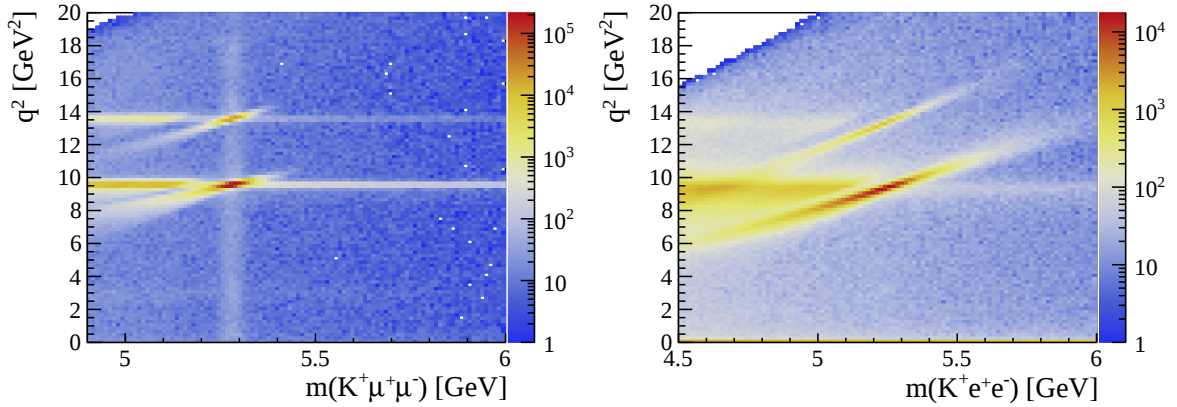


Figure 4.1: Two-dimensional distributions of $m(K^+\ell^+\ell^-), q^2$ from the muon (left) and electron (right) datasets. The entire Run 1 and Run 2 datasets are shown with the preselection and trigger requirements discussed in Chpt. 6 and Sec. 6.8 applied.

$\psi(2S) \rightarrow \ell^+ \ell^-$. Again, the radiative tail is visible;

- To the left of these peaks, elongated horizontal structures corresponding to partially-reconstructed decays with a missing track. A smaller horizontal structure on the right of the peak corresponds to a J/ψ associated with a random K^+ track, both not necessarily coming from a B -decay;
- Finally, the non-resonant rare decay mode $B^+ \rightarrow K^+ \ell^+ \ell^-$, corresponding to a vertical band at $m(K^+ \ell^+ \ell^-) \approx 5.28$ GeV and spanning the full q^2 -range.

Comparing the $K^+ e^+ e^-$ and $K^+ \mu^+ \mu^-$ data in Fig. 4.1, it is clearly visible that the mass and q^2 resolutions are much worse in the di-electron case than in the di-muon case, and that the di-muon data sample contains significantly more events. This is due to fundamental differences when reconstructing final states containing a di-electron pair or a di-muon pair at LHCb. As explained in Sec. 3, muons are MIPs that lose a negligible fraction of their momentum while travelling through the LHCb detector. This is not the case for an electron track from *e.g.* a $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ decay, which loses typically $\sim 20\%$ of its momentum in bremsstrahlung before reaching the calorimeters. The large radiative tail on the lower mass side of the $B^+ \rightarrow K^+ e^+ e^-$ mass peak is due to bremsstrahlung emitted before the magnet, which results in a lower momentum measured with the magnet spectrometer. This effect is partly corrected for by the bremsstrahlung recovery process discussed in Sec. 3.9. To avoid double counting the radiated energy, a given energy cluster can be added to one electron track only. Using the bremsstrahlung recovery, the momentum resolution for an electron track from $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ decay is approximately 10%. This is still very significantly worse than the momentum resolution for a muon track, which is typically 0.5%. The $m(\ell^+ \ell^-)$ distribution for simulated $B^+ \rightarrow K^+ e^+ e^-$ and $B^+ \rightarrow K^+ \mu^+ \mu^-$ events is shown in Fig. 4.2, illustrating the difference in resolution.

Decays with two electrons in the final state can therefore be classified into three bremsstrahlung categories: the 0γ category, where neither of the two electrons have recovered bremsstrahlung clusters added to them, 1γ where one of the two electron has recovered bremsstrahlung energy added, and 2γ where the two electrons both have recovered bremsstrahlung energy added. The mass and momentum resolution, as well as the particle identification efficiency associated with an event, is strongly dependent on the bremsstrahlung category to which an event belongs.

The electron triggers are significantly less efficient than the muon triggers, reducing the number of recorded $B^+ \rightarrow K^+ e^+ e^-$ events. For electrons, the momentum resolution

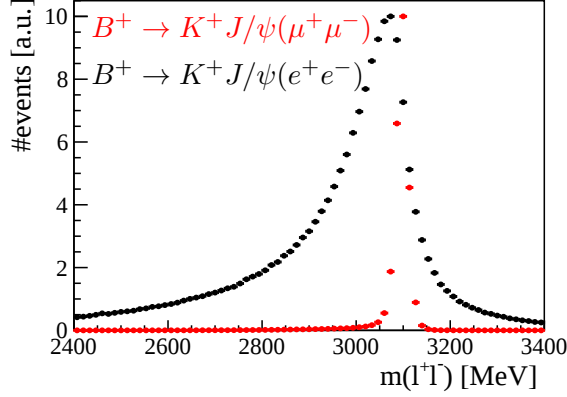


Figure 4.2: Distributions of $m(\ell^+\ell^-)$ for $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulated Run 1 events. In the case of $B^+ \rightarrow K^+ J/\psi(e^+e^-)$, the mass peak is much wider and shows extended tails to the left and right. The tail to the left is due to unrecovered bremsstrahlung. The tail to the right corresponds to recovered bremsstrahlung which, due to noise or the finite resolution of the calorimeter, has an over-estimated energy.

and tracking efficiency depend strongly on the amount of material in the detector, which must be well known to reproduce the efficiency in simulated events. Moreover, the trigger efficiency and bremsstrahlung recovery depend on the performance of the calorimeters, which is affected by *e.g.* ageing and the occupancy of the detector. Both ageing effects and the occupancy are poorly modelled in the simulation.

The differences in the detection of $K^+e^+e^-$ and $K^+\mu^+\mu^-$ final states have two consequences for the analysis. First, due to the smaller electron detection efficiency, the precision of the analysis is statistically limited by the yield of $B^+ \rightarrow K^+e^+e^-$ events. Second, the systematic uncertainties can be expected to be dominated by effects related to the detection of electrons. In fact, the muon detection efficiencies have been shown to be under good control in previous LHCb analyses, in particular the angular analysis of $B^0 \rightarrow K^{*0}\mu^+\mu^-$ [34]. The present analysis therefore focuses primarily on maximising the $B^+ \rightarrow K^+e^+e^-$ signal significance, and on demonstrating that the estimate of the $B^+ \rightarrow K^+e^+e^-$ selection efficiency is well controlled. In order to capture the effect of the trigger and calorimeter performance, as well as to control the electron tracking efficiencies and their dependence on the amount of material in the detector, control samples selected from the data are used to correct the simulation.

4.2 Normalisation and control modes

As shown in Sec. 2.6, the branching fractions of the decays $J/\psi \rightarrow e^+e^-$ and $J/\psi \rightarrow \mu^+\mu^-$ are measured to be equal with a sub-percent precision [5]. To cancel systematic effects, the resonant decays $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ and $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ are therefore used as control channels and R_K is measured in a double ratio

$$\begin{aligned} R_K &= \frac{\int_{1.1 \text{ GeV}^2}^{6.0 \text{ GeV}^2} \frac{d\mathcal{B}(B^+ \rightarrow K^+ \mu^+ \mu^-)}{dq^2} dq^2}{\int_{1.1 \text{ GeV}^2}^{6.0 \text{ GeV}^2} \frac{d\mathcal{B}(B^+ \rightarrow K^+ e^+ e^-)}{dq^2} dq^2} \cdot \frac{\mathcal{B}(B^+ \rightarrow K^+ J/\psi(e^+e^-))}{\mathcal{B}(B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-))} \\ &= \frac{N(B^+ \rightarrow K^+ \mu^+ \mu^-)}{\varepsilon(B^+ \rightarrow K^+ \mu^+ \mu^-)} \cdot \frac{\varepsilon(B^+ \rightarrow K^+ e^+ e^-)}{N(B^+ \rightarrow K^+ e^+ e^-)} \\ &\quad \cdot \frac{\varepsilon(B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-))}{N(B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-))} \cdot \frac{N(B^+ \rightarrow K^+ J/\psi(e^+e^-))}{\varepsilon(B^+ \rightarrow K^+ J/\psi(e^+e^-))}, \end{aligned}$$

where $N(X)$ indicates the estimate of the yield of decay mode X , which is obtained from a fit to the invariant mass $m(K^+\ell^+\ell^-)$ with a suitable cut on q^2 ; and $\varepsilon(X)$ is the efficiency for selecting decay mode X , as computed from simulation samples which are corrected for various effects using data calibration samples. Hereafter, the $B^+ \rightarrow K^+\ell^+\ell^-$ decay in the range $1.1 < q^2 < 6.0 \text{ GeV}^2$ is referred to as the “rare mode”, whereas the $B^+ \rightarrow K^+ J/\psi(\ell^+\ell^-)$ mode is referred to as the “normalisation mode”.

The selection applied to the normalisation modes is kept identical to that applied to the rare modes, except for the q^2 selection that differentiates between the rare and normalisation decays. In this way, many systematic uncertainties cancel in the ratio between these two modes. Indeed, the absolute size of *e.g.* tracking, particle identification or trigger efficiencies of one mode need not be known exactly, only the ratio of efficiencies between the rare mode and the corresponding control mode must be controlled, *i.e.* $\varepsilon(B^+ \rightarrow K^+ e^+ e^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(e^+e^-))$ and $\varepsilon(B^+ \rightarrow K^+ \mu^+ \mu^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-))$.

If the kinematic distributions of variables related to the rare mode were identical to those of the normalisation mode, the above efficiency ratios would be unity and the measurement would be free from any efficiency-related systematic uncertainties. Residual systematic uncertainties arise due to differences in the kinematic distributions of the rare modes and their corresponding normalisation modes. Distributions for various kinematic variables for simulated $B^+ \rightarrow K^+\ell^+\ell^-$ and $B^+ \rightarrow K^+ J/\psi$ events are shown in Fig. 4.3. Some variables show a near-perfect agreement between the rare and normalisation mode, such as the quality of the vertex fit $\chi_{\text{DV}}^2(B^+)$, the significance of the impact parameter $\chi_{\text{IP}}^2(B^+)$, the pseudorapidity of all tracks η , and the fraction of an electron track’s energy emitted

via bremsstrahlung before the magnet, denoted $p^{\text{brem}}(e)/p^{\text{tot}}(e)$. For other variables, the distributions between the rare decay and the normalisation differ, and the cancellation of the efficiency, $\varepsilon(B^+ \rightarrow K^+ e^+ e^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(e^+ e^-))$ and $\varepsilon(B^+ \rightarrow K^+ \mu^+ \mu^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-))$, will be imperfect *e.g.* the p_T of all tracks and the dilepton angle $\alpha_{\ell^+ \ell^-}$. If the detection efficiency is mis-modelled in a way that is dependent on such variables — for example, in the case where the trigger efficiency would be computed wrongly at low p_T — the measurement of R_K could be biased. Hence, to compute R_K correctly, it must be ensured that the dependency of all efficiencies on these variables is fully understood.

In order to demonstrate that the efficiencies are controlled, several cross-checks are performed. The first of these cross-checks is the measurement of the single ratio of branching fractions between the normalisation modes,

$$\begin{aligned} r_{J/\psi} &= \frac{\mathcal{B}(B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-))}{\mathcal{B}(B^+ \rightarrow K^+ J/\psi(e^+ e^-))} \\ &= \frac{N(B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-))}{\varepsilon(B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-))} \cdot \frac{\varepsilon(B^+ \rightarrow K^+ J/\psi(e^+ e^-))}{N(B^+ \rightarrow K^+ J/\psi(e^+ e^-))}. \end{aligned} \quad (4.1)$$

Given Eq. 2.33, this ratio is expected to be unity at the sub-percent level. As $r_{J/\psi}$ is a single ratio, the muon and electron efficiencies have to be controlled directly with respect to one another. This is therefore a stringent cross-check, as systematic effects do not cancel as they do in the double ratio of Eq. 4.1. As $N(B^+ \rightarrow K^+ J/\psi(e^+ e^-))$ and $N(B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-))$ are both relatively large, the statistical uncertainty is small and the total uncertainty on $r_{J/\psi}$ is dominated by such systematic effects.

As a further cross-check, $r_{J/\psi}$ is measured differentially, as a function of one (see Sec. 10.2) or several (see Sec. 10.3) kinematic variables. This checks if the dependency of the efficiencies on a given kinematic variable is properly understood and therefore cancels in the double ratio. This is particularly important for variables that have a different distribution in the normalisation and rare mode. For example, while the p_T distributions of the rare mode and the normalisation mode are different, good control of the $r_{J/\psi}$ single ratio as a function of p_T demonstrates that the efficiencies as a function of p_T are correctly understood. As detailed in Sec. 10.3, the rare decay can be described using four kinematic variables, whereas the normalisation mode, by virtue of being constrained to have $m(\ell^+ \ell^-) = m(J/\psi)$, is a function of only three kinematic variables. The differential $r_{J/\psi}$ cross-check hence accesses only projections of the full space in one, two or three dimensions.

As a third cross-check, the double ratio of branching fractions $R_{\psi(2S)}$ between the

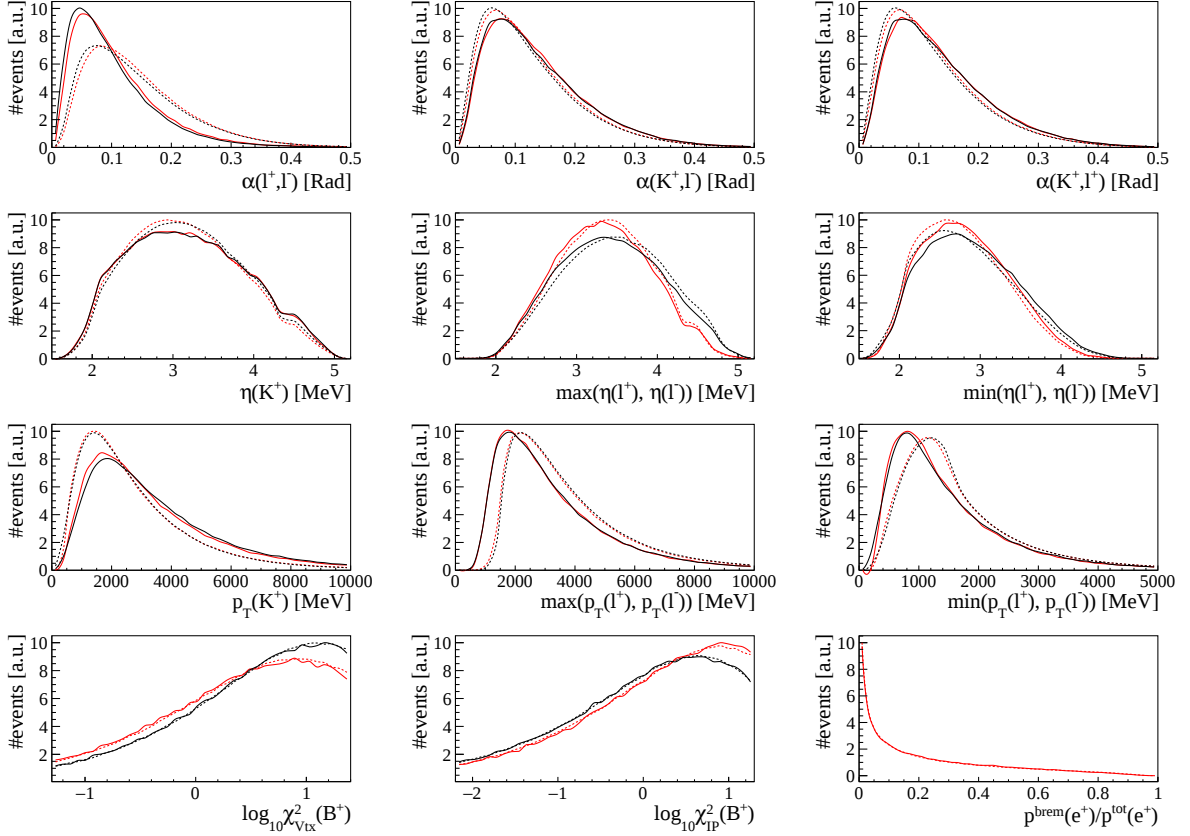


Figure 4.3: Distributions of kinematic variables for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode (dashed red line), $B^+ \rightarrow K^+ e^+e^-$ mode (solid red line), $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode (dashed black line) and $B^+ \rightarrow K^+ \mu^+\mu^-$ mode (solid black line). The first row of plots from the top shows the angle between the two leptons, or one lepton and the kaon. The second row shows the rapidity distributions, and the third row the transverse momentum distributions of all the final state particles. The bottom left plot shows the distribution for the quality of the B^+ vertex fit. The bottom centre plot shows the significance of the B^+ impact parameter. The bottom right plot shows the fraction of momentum radiated away by bremsstrahlung and is shown only for the electron modes. The distributions correspond to Run 2 simulated events, to which the preselection discussed in Chpt. 6 has been applied.

decays $B^+ \rightarrow K^+ \psi(2S)(\ell^+\ell^-)$ and $B^+ \rightarrow K^+ J/\psi(\ell^+\ell^-)$ is measured:

$$\begin{aligned}
 R_{\psi(2S)} &= \frac{\mathcal{B}(B^+ \rightarrow K^+ \psi(2S)(\mu^+\mu^-))}{\mathcal{B}(B^+ \rightarrow K^+ \psi(2S)(e^+e^-))} \cdot \frac{\mathcal{B}(B^+ \rightarrow K^+ J/\psi(e^+e^-))}{\mathcal{B}(B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-))} \\
 &= \frac{N(B^+ \rightarrow K^+ \psi(2S)(\mu^+\mu^-))}{\varepsilon(B^+ \rightarrow K^+ \psi(2S)(\mu^+\mu^-))} \cdot \frac{\varepsilon(B^+ \rightarrow K^+ \psi(2S)(e^+e^-))}{N(B^+ \rightarrow K^+ \psi(2S)(e^+e^-))} \\
 &\quad \cdot \frac{\varepsilon(B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-))}{N(B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-))} \cdot \frac{N(B^+ \rightarrow K^+ J/\psi(e^+e^-))}{\varepsilon(B^+ \rightarrow K^+ J/\psi(e^+e^-))}.
 \end{aligned} \tag{4.2}$$

This ratio of branching fractions $\mathcal{B}(\psi(2S) \rightarrow e^+e^-)/\mathcal{B}(\psi(2S) \rightarrow \mu^+\mu^-)$ is expected to be unity, as no NP is expected to impact the radiative decays $\psi(2S) \rightarrow e^+e^-$ or $\psi(2S) \rightarrow \mu^+\mu^-$. The precision on this ratio from previous measurements is of the order of $\sim 10\%$ [5]. As it is a double ratio, $R_{\psi(2S)}$ is expected to have reduced systematic uncertainties. Hence, testing that the measurement of $R_{\psi(2S)}$ is unity is a way to make sure that all systematic effects cancel well in the double ratio, with the caveat that in the $B^+ \rightarrow K^+\psi(2S)(\ell^+\ell^-)$ mode, the q^2 is larger than in the rare mode.

4.3 Outline

The remainder of this thesis is organised as follows. In Chpt. 5, all the data and simulation samples used in this analysis are introduced. In Chpt. 6, the full selection chain is presented. In Chpt. 7, the corrections applied to the simulation are reported, and the efficiency computation is detailed. In Chpt. 8, the fit strategy is detailed and the results of the fits to the control modes are reported. In Chpt. 9, the systematic uncertainties are reported. In Chpt. 10, the results of the cross-checks, integrated and differential $r_{J/\psi}$, $R_{\psi(2S)}$ are presented. In Chpt. 11, the fits to the rare mode are shown and the final result is presented. In Chpt. 12, conclusions are presented.

5. Data and simulation samples

All the samples used in the analysis are introduced in this chapter.

5.1 Data samples

The analysis presented in this thesis is based on combinations of three charged tracks: one track being identified as a kaon, and two tracks identified as oppositely charged electrons or muons, *i.e.* $K^+e^+e^-$ or $K^+\mu^+\mu^-$. The tracks are taken from collision data collected at LHCb in Run 1 and Run 2, corresponding to the conditions shown in Tab. 3.1. In each year, approximately half of the data is collected with the LHCb magnetic field pointing downwards, and the other half with the magnetic field pointing upwards.

In addition to the $K^+e^+e^-$ and $K^+\mu^+\mu^-$ samples, a sample of $K^+e^+\mu^-$ candidates is used as a proxy for background events selected from fragments of multiple decays - so-called combinatorial background. This sample is used to optimise the multivariate selection, as detailed in Sec. 6.9. The region $5100 < m(K^+e^+\mu^-) < 5350$ MeV was removed in order to avoid any possible NP signal from $B^+ \rightarrow K^+e^+\mu^-$ decays. Outside of this region, the sample is dominated by random combinations of tracks.

Other final states are used to measure the kaon and muon particle identification efficiencies, as detailed in Sec. 7.3. They correspond to the standard calibration samples listed in Ref. [87].

5.2 Simulation samples

To tune the signal selection and compute efficiencies, simulated events are used. All the simulated samples are generated using the LHCb Gauss software [88]. Gauss is based on various packages: PYTHIA 6 and 8 [89, 90] are used to generate the proton-proton interactions and the b hadronisation, EVTGEN is used to generate the B decays [91], PHOTOS is used to generate final state radiation [92] and GEANT4 to simulate the interaction of the daughter particles with the LHCb detector [93, 94].

The rare decay $B^+ \rightarrow K^+\mu^+\mu^-$ is generated using the transition form factors (see Sec. 2.9) described in [95]. The rare decay $B^+ \rightarrow K^+e^+e^-$ is generated using the transition form factors from [96]. The systematic uncertainty related to the form factors is described in Sec. 9.9.

6. Selection

This chapter describes the selection used to isolate signal events and reduce backgrounds. After the application of selection criteria, the yields are extracted by fitting the reconstructed mass $m(K^+\ell^+\ell^-)$. Two main kinds of backgrounds are expected to contribute: those that accumulate at some particular $m(K^+\ell^+\ell^-)$ mass, under or close to the signal peak (so-called ‘peaking’ backgrounds) and those that are widely distributed in $m(K^+\ell^+\ell^-)$ (non-peaking). In this analysis, the only non-peaking background corresponds to combinatorial events which originate from particles selected from multiple different decay modes. When combined, the reconstructed particles happen to fall into the mass region of interest. Such combinatorial backgrounds affect both the rare and normalisation modes, in both the electron and muon cases, and are reduced primarily by the use of the multivariate selection described in Sec. 6.9. The remaining backgrounds discussed below are all peaking backgrounds. In the limit of a perfect resolution, they would be all reconstructed at masses below the signal region at the nominal B^+ mass. Resolution effects result in some background events leaking into the signal region. Leakage also happens in the case where a particle is mis-identified. Where possible without excessive loss of signal efficiency, these partially-reconstructed backgrounds are reduced by the use of specific selection criteria.

In making yield estimates, the branching fractions are always taken from Ref. [5]. In particular, the branching fractions for the rare modes over the whole q^2 range are assumed to be:

$$\mathcal{B}(B^+ \rightarrow K^+ e^+ e^-) = (5.5 \pm 0.7) \times 10^{-7}, \quad (6.1)$$

$$\mathcal{B}(B^+ \rightarrow K^+ \mu^+ \mu^-) = (4.43 \pm 0.24) \times 10^{-7}. \quad (6.2)$$

In the next sections the different types of peaking backgrounds are first described and then the different steps in the selection (preselection, trigger and multivariate selections) are each detailed.

6.1 Types of backgrounds

In this section, the different types of backgrounds are described and the naming conventions are introduced. The selection requirements used to reduce these backgrounds, as well as the expected yields remaining after this selection is applied, are described in the following sections.

6.1.1 Charmed and strange partially-reconstructed backgrounds

One type of peaking background stems from partially-reconstructed b -hadron decays where one or several tracks are missing. The first type of partially-reconstructed background is a $b \rightarrow s\ell^+\ell^-$ decay, where the s quark hadronises into an excited strange resonance. For example, the decay $B^0 \rightarrow K^{*0}(K^+\pi^-)\ell^+\ell^-$, where the π^- is not reconstructed. This type of partially-reconstructed background is denoted ‘strange’. The second type of background is mediated by the hadronic transition $b \rightarrow c\bar{c}s$, where the $c\bar{c}$ system hadronises into an excited charmonium resonance, it is denoted ‘charmed’. For example, $B^+ \rightarrow \psi(2S)(J/\psi\pi^0)K^+$ where the π^0 is lost. Due to the poorer mass resolution of the electron modes, both charmed and strange partially-reconstructed backgrounds can leak closer to signal in the case of the electron channels.

6.1.2 Peaking backgrounds with a single mis-identified track

In both the rare and normalisation modes, there is only one fully-reconstructed background with a single mis-identified track: $B^+ \rightarrow J/\psi\pi_{[\leftrightarrow K^+]}^+$ or $B^+ \rightarrow \ell^+\ell^-\pi_{[\leftrightarrow K^+]}^+$, where the notation $\pi_{[\leftrightarrow K^+]}^+$ indicates that the π^+ is incorrectly identified as a K^+ . This background is expected to peak right under the signal, but is suppressed by a factor $|V_{cd}/V_{cs}|^2 \approx 0.05$ in the case of the normalisation modes and $|V_{td}/V_{ts}|^2 \approx 0.05$ in the case of the rare modes.

6.1.3 Peaking backgrounds with two mis-identified tracks

The first fully reconstructed double mis-ID background stems from $B^+ \rightarrow K^+\pi_{[\leftrightarrow \ell^+]}^+\pi_{[\leftrightarrow \ell^-]}^-$ decays with both pions mis-identified as leptons. The branching fraction of this decay is about 30 times larger than the rare decay [5]. Particle identification requirements are therefore necessary to suppress it. There is no equivalent background for the control modes, as $J/\psi \rightarrow \pi^+\pi^-$ and $\psi(2S) \rightarrow \pi^+\pi^-$ decays are forbidden.

A second kind of double mis-ID background is referred to as “swaps”. For example, a $B^+ \rightarrow K_{[\leftrightarrow \ell^+]}^+J/\psi(\ell_{[\leftrightarrow K^+]}^+\ell^-)$ transition where the K^+ is mis-identified as the ℓ^+ and vice-versa could end up in the $1.1 < q^2 < 6.0 \text{ GeV}^2$ q^2 -range and pollute the rare mode.

6.1.4 Cascade backgrounds

Partially reconstructed semi-leptonic V_{cb} decays, followed by a semi-leptonic V_{cs} decay, are referred to as “cascade decays”. They are decays of the type $H_b \rightarrow H_c(K^+\ell^-\bar{\nu}_\ell X)\ell^+\nu_\ell Y$, where H_b is a beauty hadron (B^+ , B^0 , B_s^0 or Λ_B), H_c an open charm hadron (D^0 , D^+ , D^* , etc.), and X , Y are missing particles. It is also possible that such decays can be selected with a mis-identified particle in the final state. In this case, the decay can be $H_b \rightarrow H_c(K^+\pi_{[\leftrightarrow\ell^-]}^- X)\ell^+\nu_\ell Y$ or $H_b \rightarrow H_c(K^+\ell^-\bar{\nu}_\ell X)\pi_{[\leftrightarrow\ell^+]}^+ Y$, where the pion is mis-identified as a lepton. As at least two neutrinos are missing, these decays peak at lower masses than the fully reconstructed signal but they can have branching fractions on the order of a few percent *i.e.* approximately four orders of magnitude larger than the rare decay signal modes.

Backgrounds with semileptonic V_{ub} decays have much lower branching fractions ($\mathcal{O}(10^{-4})$) and contain mis-identified particles. The dominant B^+ decays have two mis-identified particles in the final state. These are $B^+ \rightarrow \rho\ell^+\nu_\ell$, $B^+ \rightarrow \omega\ell^+\nu_\ell$, $B^+ \rightarrow \eta\ell^+\nu_\ell$, where ρ , ω and η decay to a final state with two pions, one being mis-identified as the ℓ^- and the other as the K^+ . There are also B_s^0 decays with one mis-identified particle in the final state. These are decays of the form $B_s^0 \rightarrow H_s^-\ell^+\nu_\ell$, where H_s^- is a strange resonance decaying to K^- and several pions, one of which must be mis-identified as ℓ^- .

6.1.5 Backgrounds with converted photons

Another type of background to be considered is any fully or partially-reconstructed decay containing a photon in the final state that converts into a di-electron pair in the material of the detector. This type of background is expected to lie at very low q^2 , but could potentially contaminate the signal q^2 region due to resolution effects.

6.2 The selection chain

A candidate $B^+ \rightarrow K^+\ell^+\ell^-$ decay with all its three tracks reconstructed must satisfy a chain of selection requirements which consists of the following steps:

1. Cut-based preselection requirements that belong to four categories:
 - (a) Decay quality requirements that reduce the sample to a manageable level;

- (b) Fiducial requirements to simplify the efficiency computation and vetoes to reduce cascade backgrounds;
 - (c) Particle identification (PID) requirements to reduce mis-ID backgrounds;
 - (d) Selection of the q^2 range that differentiates the rare and control modes from each other;
2. Trigger requirements;
 3. Multivariate selection to further reduce the combinatorial background;
 4. Finally, selection of a range in $m(K^+\ell^+\ell^-)$ mass in which a fit is made to determine the signal and background yields.

Events that satisfy the cut-based preselection are hereafter referred to as ‘preselected events’. Events that satisfy the whole selection chain, including the trigger, BDT and mass range, are hereafter referred to as ‘fully selected events’.

All the selection requirements applied to $K^+\mu^+\mu^-$ and $K^+e^+e^-$ final states are shown in Tab. 6.1, except the mass and q^2 range requirements which differ for the rare, $\psi(2S)$ and J/ψ modes. The following sections describe each of the selection steps in more detail.

6.3 Decay quality requirements

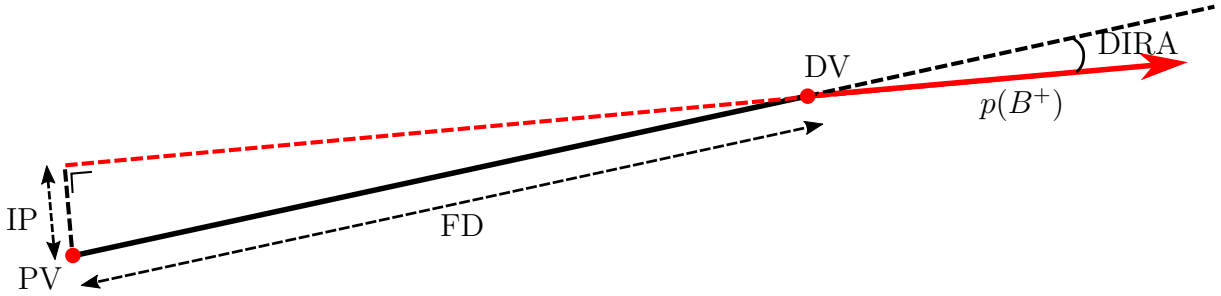


Figure 6.1: Schematic of a B^+ decay defining quantities used for the selection. The primary vertex (PV) and decay vertex (DV) are shown, as well as the distance between the two, or flight distance (FD). The directional angle (DIRA) and impact parameters (IP) are also represented.

Although the multivariate algorithm discussed in Sec. 6.9 reduces the combinatorial yield to a very low level, it is useful to apply some loose decay quality requirements at the

Table 6.1: Offline selection cuts applied to the electron modes (left) and muon modes (right).

Event quality		Event quality	
$\chi_{\text{FD}}^2(B^+)$	> 100	$\chi_{\text{FD}}^2(B^+)$	> 121
$\cos(\text{DIRA})$	< 0.995	$\cos(\text{DIRA})$	< 0.9999
$\chi_{\text{IP}}^2(B^+)$	< 25	$\chi_{\text{IP}}^2(B^+)$	< 16
$\chi_{\text{DV}}^2/\text{ndof}(B^+)$	< 9	$\chi_{\text{DV}}^2/\text{ndof}(B^+)$	< 8
$\chi_{\text{FD}}^2(ee)$	> 16	$\chi_{\text{FD}}^2(\mu\mu)$	> 9
$\chi_{\text{DV}}^2/\text{ndof}(ee)$	< 9	$\chi_{\text{DV}}^2/\text{ndof}(\mu\mu)$	< 12
$\chi_{\text{IP}}^2(e)$	> 9	$\chi_{\text{IP}}^2(\mu)$	> 9
$\chi_{\text{IP}}^2(K)$	> 9	$\chi_{\text{IP}}^2(K)$	> 6
$p_{\text{T}}(K)$	$> 400 \text{ MeV}$		
nSPDHits	$< 600 \text{ (Run 1)}$	nSPDHits	$< 600 \text{ (Run 1)}$
	$< 450 \text{ (Run 2)}$		$< 450 \text{ (Run 2)}$
Cascade vetoes		Cascade & mis-ID vetoes	
$m(K^+e^-)$	$> 1885 \text{ MeV}$	$m(K^+\mu^-)$	$> 1885 \text{ MeV}$
$m_{e\rightarrow\pi}^{\text{track}}(K^+e^-)$	$\notin m(D^0) \pm 40 \text{ MeV}$	$m_{\mu\rightarrow\pi}(K^+\mu^-)$	$> 1885 \text{ MeV}$
Fiducial cuts		$m_{K\rightarrow\mu}(K^+\mu^-)$	$\notin m(J/\psi) \pm 60 \text{ MeV}$
$p_{\text{T}}(e)$	$> 0.5 \text{ GeV}$	$m_{K\rightarrow\mu}(K^+\mu^-)$	$\notin m(\psi(2S)) \pm 60 \text{ MeV}$
$p(e)$	$> 3 \text{ GeV}$	Fiducial cuts	
$\text{hasRich}(K, e)$	$= \text{true}$	$\text{inMuonAcc}(K, \mu)$	$= \text{true}$
$\text{hasCalo}(e)$	$= \text{true}$	$\text{hasRich}(K, \mu)$	$= \text{true}$
$x_{\text{ECAL}}(e)$	$> 363.6 \text{ mm}$	$p_{\text{T}}(\mu)$	$> 0.8 \text{ GeV}$
$\text{or } y_{\text{ECAL}}(e)$	$> 282.6 \text{ mm}$	PID cuts	
PID cuts		$\text{probNN}_K(K)$	> 0.2
$\text{probNN}_K(K)$	> 0.2	$\text{isMuon}(K)$	$= \text{false}$
$\text{DLL}_e(K)$	< 0	$\text{DLL}_\mu(\mu)$	> -3
$\text{DLL}_e(e)$	> 3	$\text{isMuon}(\mu)$	$= \text{true}$
$\text{prob}_{\text{ghost}}(K, e)$	< 0.3	$\text{prob}_{\text{ghost}}(K, \mu)$	< 0.3

start of the selection in order to eliminate the bulk of the combinatorial background and reduce the sample size to a manageable level. Fig. 6.1 defines the quantities that are used to characterise the decay quality. The decay quality requirements are given at the top

of Tab. 6.1 and are as follows:

- The selected tracks are required to form a di-lepton and a B^+ vertex of good quality, by requiring the χ^2 of the decay vertex fit, χ_{DV}^2 , to be small;
- The decay must be well aligned, meaning that the angle between the momentum of the B^+ candidate and the vector linking the primary vertex to the decay vertex (DIRA) must be close to zero. The significance of the impact parameter χ_{IP}^2 of the B^+ with respect to the primary vertex is also required to be small.
- To avoid selecting random tracks coming from the primary vertex, all tracks are required to be significantly detached from it. This is done by requiring the χ_{IP}^2 of each track to be large, and by requiring the significance of the flight (*i.e.* the distance between the primary and secondary vertices) χ_{FD}^2 to be large;
- The number of hits in the SPD (Sec. 3.7), nSPDHits, is required to not be too large¹.

6.4 Definition of q^2 and mass ranges

The only two differences between the selection of the rare modes and of the control modes are the mass ranges and q^2 ranges. They are summarised in Tab. 6.2. For the $B^+ \rightarrow K^+ J/\psi(\ell^+\ell^-)$ ($B^+ \rightarrow K^+ \psi(2S)(\ell^+\ell^-)$) mode, the mass of the B^+ candidate is computed by constraining the di-lepton mass to $m(\ell^+\ell^-) = m(J/\psi)$ ($m(\ell^+\ell^-) = m(\psi(2S))$) and is denoted $m_{\text{DTF}}^{J/\psi}(K^+\ell^+\ell^-)$ ($m_{\text{DTF}}^{\psi(2S)}(K^+\ell^+\ell^-)$) [97]. The choice of q^2 ranges is explained below, and the choices of mass ranges are justified later in Sec. 6.10.

The lower bound of the q^2 range has been increased from 1 GeV^2 used in the Run 1 R_K analysis to the 1.1 GeV^2 that is used in other $b \rightarrow s\ell^+\ell^-$ analyses. This allows the contamination from $B^+ \rightarrow K^+ \phi(\rightarrow \ell^+\ell^-)$ to be minimised and is particularly important for the electron mode where, because of the bremsstrahlung recovery, events can migrate to higher q^2 . The reconstructed q^2 shape for simulated $\phi(\rightarrow e^+e^-)$ events is shown in Fig. 6.2. Based on the $B^+ \rightarrow K^+ \phi(\rightarrow \ell^+\ell^-)$ branching fraction measured in [98], and the branching fraction for $B^+ \rightarrow K^+ e^+e^-$ from [5], the $B^+ \rightarrow K^+ \phi(\rightarrow e^+e^-)$ yield in the range $1.1 < q^2 < 6.0 \text{ GeV}^2$ is expected to be 0.1% of the $B^+ \rightarrow K^+ e^+e^-$ yield and is neglected.

¹As the requirement on nSPDHits is also placed by the trigger, placing it at the reconstruction level has no effect on the total selection efficiency, but simplifies the calibration of the PID and trigger efficiencies (see Sec. 7.3 and Sec. 7.4).

Table 6.2: Signal and control samples and their corresponding q^2 and mass ranges. These are the only requirements that are different between the selection of the rare mode and that of the control mode.

	e^+e^- mode	$\mu^+\mu^-$ mode
signal mode	$1.1 < q^2 < 6.0 \text{ GeV}^2$	$1.1 < q^2 < 6.0 \text{ GeV}^2$
	$4.88 < m(K^+e^+e^-) < 6.20 \text{ GeV}$	$5.18 < m(K^+\mu^+\mu^-) < 5.60 \text{ GeV}$
J/ψ mode	$6.00 < q^2 < 12.96 \text{ GeV}^2$	$8.68 < q^2 < 10.09 \text{ GeV}^2$
	$4.88 < m_{\text{DTF}}^{J/\psi}(K^+e^+e^-) < 5.70 \text{ GeV}$	$5.18 < m_{\text{DTF}}^{J/\psi}(K^+\mu^+\mu^-) < 5.70 \text{ GeV}$
$\psi(2S)$ mode	$9.92 < q^2 < 16.40 \text{ GeV}^2$	$12.5 < q^2 < 14.2 \text{ GeV}^2$
	$4.88 < m_{\text{DTF}}^{\psi(2S)}(K^+e^+e^-) < 5.70 \text{ GeV}$	$5.18 < m_{\text{DTF}}^{\psi(2S)}(K^+\mu^+\mu^-) < 5.70 \text{ GeV}$

The lower q^2 bound used is sufficient to reduce any backgrounds with a converted photon to a completely negligible level. This is demonstrated using a sample of simulated $B^0 \rightarrow K^{*0}\gamma$ events, where some of the photons convert to a di-electron pair. The q^2 distribution for this decay is shown on the right of Fig. 6.2. Out of the 2×10^6 simulated events, none is reconstructed in the signal q^2 region.

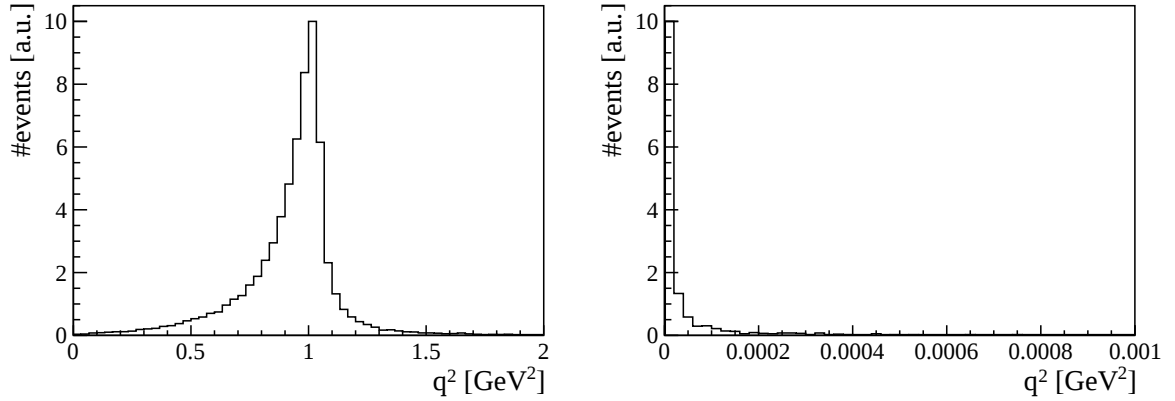


Figure 6.2: Reconstructed q^2 shapes for simulated $B^+ \rightarrow K^+ \phi (\rightarrow e^+ e^-)$ events (left) and simulated $B^0 \rightarrow K^{*0} \gamma$ events where the photon has converted into a di-electron pair (right).

6.5 Fiducial cuts

Fiducial cuts are applied to veto regions where the detector is less performant, or where some efficiencies are unknown. Requirements on p and p_T are applied to ensure that

the signal spans the same phase space as the calibration samples used to determine the efficiency of the PID requirements (see Sec. 7.3.2), which are produced independently of this analysis.

Similarly, tracks are required to be within the acceptance of the relevant sub-detectors to maximise the rejection of mis-ID backgrounds: muons are required to be within the muon chambers (using the variable `inMuonAcc`), electrons to have associated hits in the calorimeters (`hasCalo`), and all tracks must have associated hits in the RICH detectors (`hasRich`).

Some cells in the centre of the ECAL are not read out. Electrons which hit these cells are vetoed by requiring the x and y coordinates of electron candidates at the ECAL level, x_{ECAL} and y_{ECAL} , to satisfy the requirement $x_{ECAL} > 363.6$ mm or $y_{ECAL} > 282.6$ mm. This requirement rejects less than 1% of $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events. As the trigger and PID efficiencies are not corrected in fine bins of x_{ECAL} and y_{ECAL} , as detailed in Sec. 7.4, vetoing this region guarantees a better agreement between data and simulation.

Approximately 8% of $B^+ \rightarrow K^+ e^+ e^-$ events and 22% of $B^+ \rightarrow K^+ \mu^+ \mu^-$ events are removed by the fiducial cuts. The loss of $B^+ \rightarrow K^+ \mu^+ \mu^-$ events is mostly due to the $p_T(\mu) > 0.8$ GeV requirement. This requirement could be avoided by using an alternative PID calibration sample. This is however not done in the present analysis, as the precision on R_K is limited by the number of electron events.

6.6 Cascade background vetoes

Cascade backgrounds are suppressed by requiring $m(K^+ \ell^-) > m(D^0)$. Mis-ID cascades where the pion is mis-identified as a lepton, *i.e.* $H_b \rightarrow H_c(K^+ \pi_{[\leftrightarrow \ell^-]}^-) \ell^+ \nu_\ell Y$, are suppressed by a similar veto, switching the lepton mass hypothesis to that of the pion. In the muon case, the veto is simply $m_{\mu \rightarrow \pi}(K^+ \mu^-) > m(D^0)$. In the electron case, the D^0 mass is computed switching off the bremsstrahlung recovery to improve the mass resolution and the veto used is $|m_{e \rightarrow \pi}^{\text{track}}(K^+ e^-) - m(D^0)| > 40$ MeV. These cascade vetoes are illustrated in Fig. 6.3.

The cascade vetoes retain $\sim 97\%$ of preselected $B^+ \rightarrow K^+ \mu^+ \mu^-$ and $\sim 95\%$ of $B^+ \rightarrow K^+ e^+ e^-$ events, and reduce the cascade backgrounds to a very small level. The expected yields, N^{exp} , for simulated cascade events can be seen in Tab. 6.3. The yields are obtained after applying the full selection including the cascade vetoes as well as the mass window cut described in Sec. 6.10. It can be seen that a small contamination from

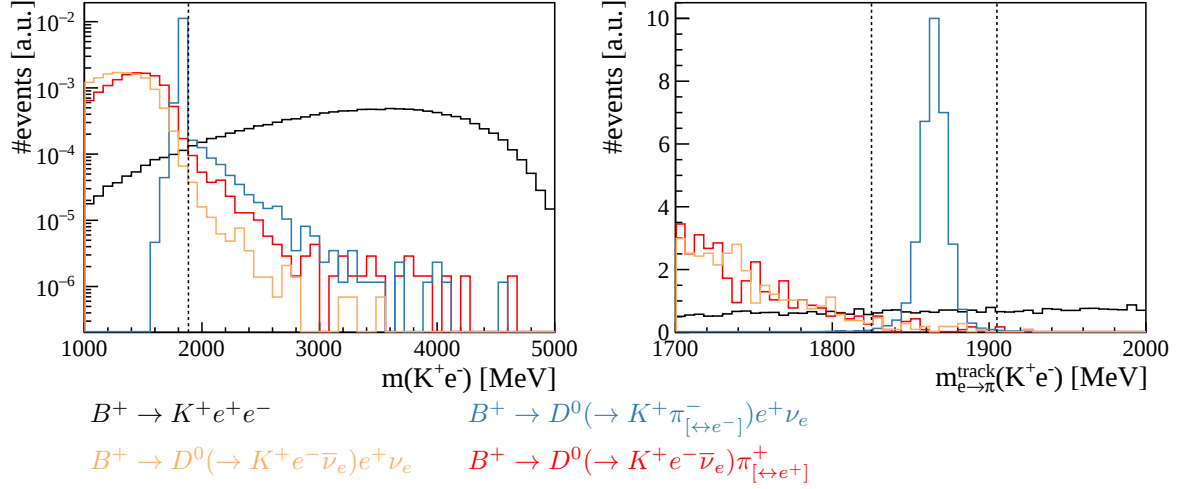


Figure 6.3: K^+e^- mass distributions from signal and cascade simulation samples. The decay to which each colour corresponds is displayed below the plots. All the distributions are normalised to unity. In the left-hand plots, the events on the left of the dashed vertical lines are vetoed. In the right-hand plots, the events between the two dashed vertical lines are vetoed. In the right-hand plot, the mass is computed from the track information only, whereas, in the left-hand plot, the bremsstrahlung correction to the momentum of the electron is taken into account, resulting in a tail to the right. All plots are taken from preselected Run 1 simulated events.

$B^+ \rightarrow D^0(\rightarrow K^+\ell^-)\ell^+\nu_\ell$ remains after applying the veto. This component is taken into account in the fit to the rare mode by including it in the partially-reconstructed background shape.

Table 6.3: Expected level of cascade and mis-ID backgrounds as a fraction of the relevant $B^+ \rightarrow K^+\ell^+\ell^-$ signal yield, when the full selection has been applied, including the q^2 cut $1.1 < q^2 < 6.0 \text{ GeV}^2$.

Background type	$\frac{N^{\text{exp}}(bkg)}{N^{\text{exp}}(B \rightarrow K\ell\ell)}$	$\frac{N^{\text{exp}}(bkg)}{N^{\text{exp}}(B \rightarrow K\mu\mu)}$
Cascades		
$B^+ \rightarrow D^0(\rightarrow K^+\ell^-\bar{\nu}_\ell)\ell^+\nu_\ell$	3.26%	0.00%
$B^+ \rightarrow D^0(\rightarrow K^+\pi_{[\leftrightarrow\ell^-]}^-\ell^+\nu_\ell)$	0.00%	0.00%
$B^+ \rightarrow D^0(\rightarrow K^+\ell^-\bar{\nu}_\ell)\pi_{[\leftrightarrow\ell^+]}^+$	0.00%	0.00%
Single and double mis-ID		
$B^+ \rightarrow \pi_{[\leftrightarrow K^+]}^+\ell^+\ell^-$	0.41%	0.41%
$B^+ \rightarrow K^+\pi_{[\leftrightarrow\ell^+]}^+\pi_{[\leftrightarrow\ell^-]}^-$	0.83%	0.60%
$B^+ \rightarrow K_{[\leftrightarrow\ell^+]}^+J/\psi(\ell_{[\leftrightarrow K^+]}^+\ell^-)$	0.04%	0.00%

6.7 Particle identification

To reduce mis-identified backgrounds and swaps, particle identification (PID) requirements are used. Backgrounds with fake tracks are reduced by making a loose requirement on $\text{prob}_{\text{ghost}}$. In order to identify the kaon, a $\text{probNN}_K(K) > 0.2$ requirement is applied in both muon and electron modes. To reduce backgrounds containing mis-identified leptons, in the muon mode a combination of isMuon and DLL_μ cuts are used; and in the electron mode a requirement on DLL_e is used. These requirements reduce the peaking mis-identified backgrounds to a negligible level, as can be seen in Tab. 6.3. In the fit to the rare modes, the peaking mis-identified backgrounds are therefore neglected. In the case of the normalisation modes, a contamination from mis-identified $B^+ \rightarrow \pi^+ J/\psi$ events remains at a level of 0.45% and is included in the fit described in Sec. 8.2. This component cannot be neglected due to the very high yield of the normalisation modes.

6.8 Trigger

To allow for better control of the selection efficiency, a specific set of trigger lines is used that are either triggered on some element of the signal, triggered on signal (TOS), or independently of the signal (TIS). The requirements that are placed at the L0, HLT1 and HLT2 trigger levels are described in the following sections.

6.8.1 L0 trigger requirements

The R_K ratio is extracted separately, using electrons events triggered by three exclusive L0 categories, which use a different L0 trigger line:

$e\text{TOS}$: the event is selected by the L0Electron algorithm based on one of the two electrons;

$h\text{TOS}!$: the event is selected by the L0Hadron algorithm based on the kaon. The events from the previous category are excluded and this is denoted with the ‘!’;

$\text{TIS}!$: the event is selected by the L0 algorithm based on any other track in the event, excluding events from the two previous categories.

Since the muon trigger is much more efficient than the calorimeter-based trigger, only the muon trigger is used to select $K^+ \mu^+ \mu^-$ events (μTOS category, requiring the L0Muon trigger line to be TOS on one of the two muons). To perform cross-checks, inclusive hadron

and TIS categories, which do not exclude e TOS or e TOS and h TOS triggers, are also used and are denoted h TOS and TIS respectively.

Table 6.4: Trigger thresholds for the electron, hadron and muon L0 triggers for each year of data-taking. These are approximate as they change slightly during each year. For the electron and hadron triggers, the fiducial requirements on E_T are also shown.

	2011	2012	2015	2016
$E_T^{L0}(e)$ threshold [GeV]	2.5	3.0	2.7	2.4
$E_T(e)$ fiducial [GeV]	3.0	3.0	2.7	2.7
$E_T^{L0}(K)$ threshold [GeV]	3.5	3.7	3.6	3.8
$E_T(K)$ fiducial [GeV]	3.5	3.5	3.5	3.5
$p_T(\mu)$ threshold [GeV]	1.5	1.8	2.8	1.9

In the L0 trigger algorithm, electrons or hadrons can trigger if their transverse energy E_T^{L0} , as measured by the calorimeter system, is above a threshold. Muons can trigger if their transverse momentum p_T^{L0} , as measured by the muon system, is above a threshold. The precision on E_T^{L0} and p_T^{L0} is much lower than on their counterparts E_T and p_T measured by the tracking system. As detailed in Sec. 7.4, to avoid having to perfectly calibrate the calorimeter resolution in the simulation, an additional fiducial requirement is applied on E_T that tightens the trigger requirement. The L0 trigger requirements and the fiducial requirements depend on the year of data-taking, and are all summarised in Tab. 6.4. The L0 trigger thresholds are approximate, as they are regularly slightly changed during the data-taking and fluctuate because of the aging of the detector.

6.8.2 HLT1 trigger requirements

If an event satisfies one of the L0 trigger algorithms, at least one of the three signal tracks is then required to satisfy a set of HLT1 requirements. These differ between Run 1 and Run 2.

In Run 1, one of the three signal tracks must satisfy $p_T > 1.6$ GeV (or $p_T > 1.0$ GeV if it is identified as a muon), $p > 3$ GeV and $\chi_{IP}^2 > 16$. These quantities are measured using the fast HLT1 reconstruction, which is much more precise than the L0 algorithm, see Sec. 3.10.

To trigger the HLT1 stage in Run 2, one of the three signal tracks must satisfy the following requirement

$$\log(\chi_{IP}^2) > \frac{1.0}{(p_T[\text{GeV}] - 1.0)^2} + \frac{b}{25} \cdot (25 - p_T[\text{GeV}]) + \log(7.4), \quad (6.3)$$

where the constant b is set to either of two values during the data-taking: 73% of the Run 2 luminosity is collected using $b = 1.1$, and 27% using $b = 2.3$. In the simulation, both values of b are used in the same proportions as in the data. The effect of this requirement on the $p_T(e), \log \chi_{IP}^2(e)$ two-dimensional distribution can be seen in Fig. 6.4.

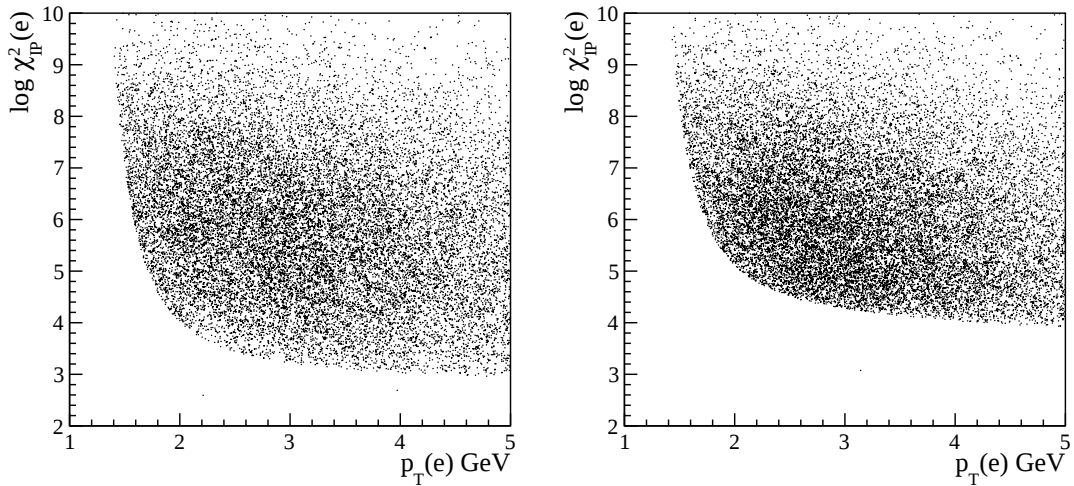


Figure 6.4: Distribution of $p_T(e), \log \chi_{IP}^2(e)$ for electrons selected from $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ candidates from the Run 2 data. The electrons are required to satisfy the HLT1 requirement (see Eq. 6.3). The left plot has $b = 1$ in Eq. 6.3, whereas the right plot corresponds to events selected with $b = 2.3$.

6.8.3 HLT2 trigger requirements

At the HLT2 level, the signal tracks are required to satisfy the topological trigger requirements. This trigger is based on a multivariate algorithm which is applied to either all three signal tracks, or a subset of two signal tracks, selecting groups of tracks that look like they come from the decay of a B hadron. One variable used as an input is the so-called corrected mass m_{corr} , which is an estimate of the mass of the mother particle, accounting for possible missing tracks in the decay:

$$m_{\text{corr}} = \sqrt{m^2 + p_{\text{T}}^{\text{miss}2}} + |p_{\text{T}}^{\text{miss}}|,$$

where p_T^{miss} is the missing momentum transverse to the direction of flight of the trigger candidate, a trigger candidate corresponding to the combination of two or three tracks. The other kinematic variables used as inputs are similar to those used in the multivariate algorithm trained specifically for this analysis and described in Sec. 6.9. More detail about the topological trigger can be found in Ref. [99].

6.9 Multivariate selection

To suppress combinatorial events, TMVA Boosted Decision Trees (BDTs) [100, 101] are used. The training of the BDTs is explained in the next section. The optimisation as well as the expected performance of the BDTs is explained in the following section.

6.9.1 Multivariate selection training

Four BDTs are trained separately for Run 1 and Run 2 samples, and for the $B^+ \rightarrow K^+ \mu^+ \mu^-$ and $B^+ \rightarrow K^+ e^+ e^-$ modes. They are trained using data events taken from the upper mass sideband ($m(K^+ \ell^+ \ell^-) > 5.4 \text{ GeV}$) as the background sample, and simulated $B^+ \rightarrow K^+ \ell^+ \ell^-$ events as the signal sample. The preselection requirements, the $1.1 < q^2 < 6.0 \text{ GeV}^2$ and trigger requirements are applied to both the signal and background samples before training the classifier. All correction weights discussed in Chpt. 7 are applied to the simulated sample. The training and testing is performed using the k-folding technique [102] with ten folds to enable the entire sample to be used to both train and test without inducing any significant statistical bias.

The variables that best discriminate between signal and combinatorial background are the p_T of all the particles, the quality of the vertex fitting and the significance of the displacement of the tracks from the primary vertex. Hence, the classifier uses the twelve discrimination variables which are listed in Tab. 6.5.

Variables related to the PID are not included in the training, as their distributions are badly modelled in the simulation. Correcting the PID distributions using data events from calibration samples in order to use them in the BDT is not envisaged here. This would require determining the dependency of the PID variables on the BDT training variables, and the size of the electron PID sample is too small to allow that to be done (see Sec. 7.3).

Two different boosting methods are tested: adaptive and gradient boosting [101, 103, 104]. For all trigger categories and for both the electron and muon modes, gradient boosting offers a better performance and is therefore the algorithm used. The ROC curves

Table 6.5: Variables used by the BDT classifier.

B^+	$p_T, \log \chi_{\text{IP}}^2, \chi_{\text{DV}}^2, \text{DIRA}, \chi_{\text{FD}}^2$
$\ell^+\ell^-$	$p_T, \log \chi_{\text{IP}}^2$
K^+	$p_T, \log \chi_{\text{IP}}^2$
ℓ	$\min, \max(p_T), \min, \max(\log \chi_{\text{IP}}^2)$

illustrating the performance of the two boosting methods are shown in Fig. 6.5 for Run 2 data. The area under the ROC curve is smaller for the muon mode than for its electron equivalent, but the muon mode has much less background before the application of the BDT.

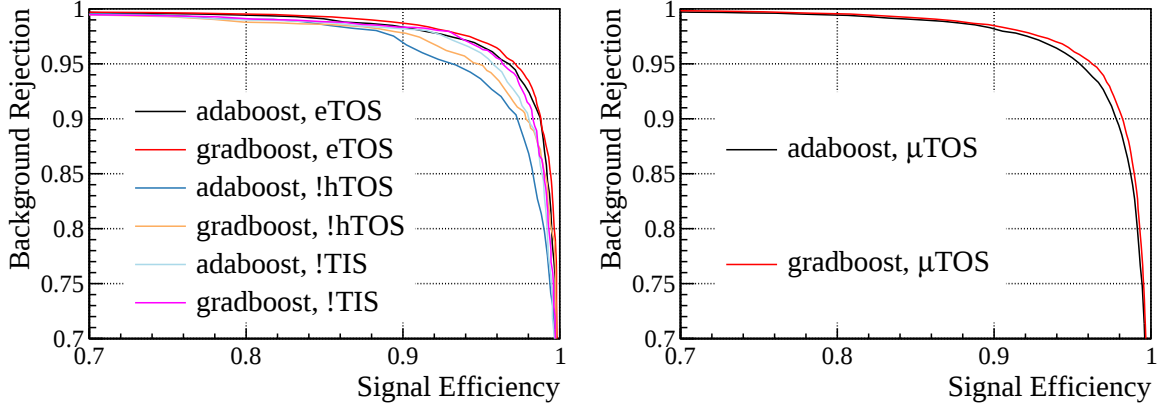


Figure 6.5: ROC curves for the rare modes in different trigger conditions for the Run 2 data, for the electron mode in the three exclusive trigger categories (left) and the muon mode (right). The curves corresponding to the BDT trained with the gradient boosting (gradboost), and adaptive boosting (adaboost) methods are shown. It is clear from these curves that the gradient boosting has the best performance.

6.9.2 Multivariate selection optimisation and performance

The selection requirement placed on the BDT output, hereafter referred to as the working point of the BDT, is found by optimising the expected significance $\mathcal{S}$,

$$\mathcal{S} = \frac{S}{\sqrt{S+B}}, \quad (6.4)$$

where S and B are the number of signal and background events expected in the signal region. The signal region is defined as the mass range $5210 < M(K^+\mu^+\mu^-) < 5350$ for the muon mode and $5000 < M(K^+e^+e^-) < 5380$ for the electron mode.

In the case of the muon mode, the expected number of background events in the signal region is extrapolated from the upper mass sideband using a fit with an exponential function. In the case of the electron mode, using such a fit is less precise, because the number of events in the sideband is small and the signal region larger. Instead, a data sample consisting of combinations of $K^+e^+\mu^-$ tracks, which provides a proxy for $K^+e^+e^-$ combinatorial background, is used. After the application of the PID selection and cascade vetoes, only a negligible amount of partially-reconstructed or mis-identified B hadron decays with $K^+e^+\mu^-$ in the final state are expected in the signal region. The $K^+e^+\mu^-$ sample is therefore dominated by combinatorial events. This sample is used to compute

$$N_{\text{comb}}^{\text{SR}} = N_{\text{comb}}^{\text{SB}} \cdot \frac{N_{K^+e^+\mu^-}^{\text{SR}}}{N_{K^+e^+\mu^-}^{\text{SB}}}, \quad (6.5)$$

where $N_{\text{comb}}^{\text{SR}}$ ($N_{\text{comb}}^{\text{SB}}$) is the number of combinatorial events in the signal region (sideband region). The ratio of yields $\frac{N_{K^+e^+\mu^-}^{\text{SR}}}{N_{K^+e^+\mu^-}^{\text{SB}}}$ is extracted from an exponential fit to $m(K^+e^+\mu^-)$. The fit is performed in the range $4880 < m(K^+e^+\mu^-) < 6250$ MeV for Run 1, and $4880 < m(K^+e^+\mu^-) < 6700$ MeV for Run 2, with the region $5100 < m < 5350$ MeV removed in order to avoid any signal from lepton flavour violating modes. This fit is illustrated in Fig. 6.6.

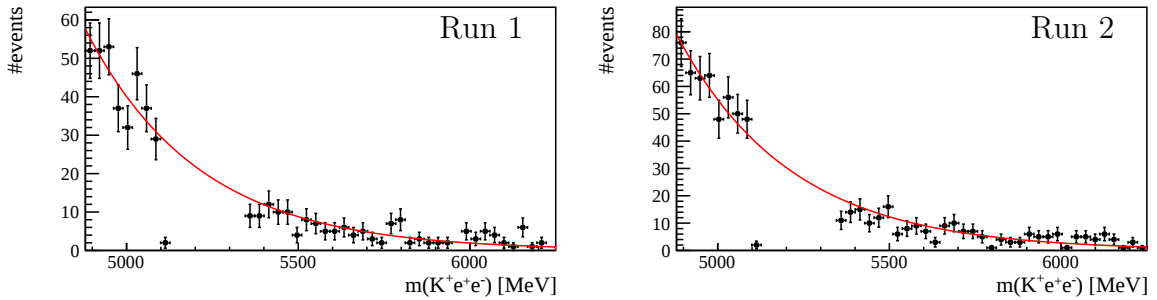


Figure 6.6: Illustration of the exponential fit to the $K^+e^+\mu^-$ sample to extrapolate the number of events from the sideband ($m > 5400$ MeV) to the signal region ($5000 < m(K^+e^+e^-) < 5380$ MeV). The data used is that from Run 1 (left) and Run 2 (right). The region $5100 < m < 5350$ MeV is blinded.

The expected number of signal events S is computed as

$$S = \varepsilon_{\text{rare}}^{\text{BDT}} \cdot \frac{\varepsilon_{\text{rare}}^{\text{pres, trig}}}{\varepsilon_{J/\psi}^{\text{pres, trig}}} \cdot \frac{\mathcal{B}(B^+ \rightarrow K^+ \ell^+ \ell^-)}{\mathcal{B}(B^+ \rightarrow K^+ J/\psi (\ell^+ \ell^-))} \cdot N^{\text{pres, trig}}(B^+ \rightarrow K^+ J/\psi), \quad (6.6)$$

where $\mathcal{B}$ indicates a branching fraction taken from Ref. [5], $N^{\text{pres, trig}}(B^+ \rightarrow K^+ J/\psi)$ is the number of control events satisfying the preselection and trigger requirements that are extracted from the fit described in Sec. 7.2, $\varepsilon_{J/\psi}^{\text{pres, trig}}$ and $\varepsilon_{\text{rare}}^{\text{pres, trig}}$ are the efficiencies of the preselection and trigger criteria on the control modes or rare modes, $\varepsilon_{\text{rare}}^{\text{BDT}}$ is the efficiency of the BDT cut on the rare mode. The efficiencies are computed using simulation samples that are fully corrected according to the procedure described in Chpt. 7.

To find the optimal working point of the BDT, the expected significance is fitted with a fourth order polynomial in order to smooth the curve and to mitigate statistical fluctuations. The expected significances $\mathcal{S}$ for the electron mode in the $e\text{TOS}$ category as well as for the muon mode in Run 2 are shown in Fig. 6.7

It is also checked that the BDT working point and electron DLL_e cut yield the optimal significance. The ROC curves made separately for each fold are checked to be compatible with each other, and the BDT is checked to not be overtrained.

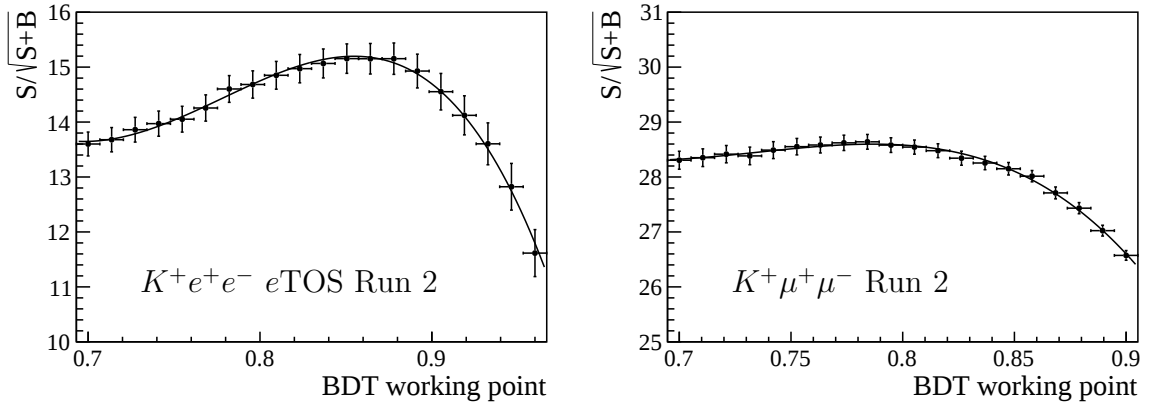


Figure 6.7: Expected significance as a function of the BDT working point for the $B^+ \rightarrow K^+ e^+ e^-$ decay in the $e\text{TOS}$ category (left), and $B^+ \rightarrow K^+ \mu^+ \mu^-$ decay (right). Both plots are for the Run 2 data only, and the working point for the Run 1 data is found using the exact same method.

The BDT optimal working points, as well as the expected signal and combinatorial yields in the signal window, are shown in Tab. 6.6. In the case of the electron mode, the combinatorial yield is estimated either by using the $K^+ e^+ \mu^-$ sample as explained above, or, as a cross-check, by using an exponential fit to the $K^+ e^+ e^-$ upper mass sideband

directly. The two methods are in good agreement with each other, but using the $K^+e^+\mu^-$ sample yields a better statistical precision.

In order to keep the selection of the rare and control modes as similar as possible, the same BDT working point, which is chosen on the basis of the assumed rare decay yields, is used for $B^+ \rightarrow K^+\ell^+\ell^-$, $B^+ \rightarrow K^+J/\psi(\ell^+\ell^-)$ and $B^+ \rightarrow K^+\psi(2S)(\ell^+\ell^-)$.

Table 6.6: Expected signal and combinatorial yields in the fit window for the electron and muon rare modes in different trigger and data-taking conditions, for the BDT working point (wp) optimising the expected significance $\mathcal{S}$. The uncertainty on the signal yields takes only into account the statistical uncertainty of the simulation samples. In the case of the electron mode, the combinatorial level is either estimated using the $K^+e^+\mu^-$ sample as a proxy (comb. 1), or extrapolating the number of events from the upper mass sideband to the signal region using an exponential fit to the $K^+e^+e^-$ mass (comb. 2).

	wp	signal	comb. 1	comb. 2
Run 1				
$B^+ \rightarrow K^+e^+e^-$ eTOS	0.89	206.7 ± 2.1	115 ± 23	210 ± 110
$B^+ \rightarrow K^+e^+e^-$ hTOS!	0.79	70.9 ± 1.4	171 ± 34	160 ± 60
$B^+ \rightarrow K^+e^+e^-$ TIS!	0.87	76.2 ± 1.3	48 ± 19	60 ± 60
$B^+ \rightarrow K^+\mu^+\mu^-$ μ TOS	0.81	1003 ± 6		244 ± 30
Run 2				
$B^+ \rightarrow K^+e^+e^-$ eTOS	0.86	345.6 ± 0.9	199 ± 29	170 ± 50
$B^+ \rightarrow K^+e^+e^-$ hTOS!	0.86	90.7 ± 1.8	97 ± 24	120 ± 60
$B^+ \rightarrow K^+e^+e^-$ TIS!	0.85	98.8 ± 1.6	86 ± 18	100 ± 40
$B^+ \rightarrow K^+\mu^+\mu^-$ μ TOS	0.79	886 ± 5		192 ± 26

6.10 Fit window

All the selection requirements described so far are used to suppress mis-identified, combinatorial and cascade backgrounds. However, the strange and charmed partially-reconstructed backgrounds, defined in Sec. 6.1, are not suppressed by any dedicated selection. The discrimination power comes from the fact that these backgrounds peak at a lower $K^+\ell^+\ell^-$ mass than the signal, and a fit to the $K^+\ell^+\ell^-$ mass is used to separate them.

In the muon case, the mass resolution enables placing a cut at $m(K^+\mu^+\mu^-) > 5.18$ GeV, which suppresses all the partially-reconstructed backgrounds, while keeping more than

99% of the signal. Partially reconstructed backgrounds are expected to peak below $m(B) - m(\pi) \approx 5.14$ GeV.

In the electron case, bremsstrahlung results in the signal mass peak having a long tail to the left and it is not possible to place a tight cut on $m(K^+e^+e^-)$ to suppress all the partially-reconstructed backgrounds while keeping a large proportion of the signal. Instead, it is necessary to include these backgrounds in the fit.

The partially-reconstructed backgrounds to the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ control mode are studied using an inclusive simulation sample containing $H_b \rightarrow J/\psi X$ decays, to which the preselection is applied. The beauty hadron H_b is either a B^+ , B^0 and B_s^0 ². The fractions of events in each possible decay are set using the fragmentation fractions from Ref. [105] and branching fractions from Ref. [105]. As explained in Sec. 8.1, the decays $B^+ \rightarrow J/\psi(\rightarrow e^+e^-\gamma)K^+$ and $B^+ \rightarrow J/\psi(\rightarrow e^+e^-\gamma)\pi^+$ are not included in this inclusive sample, and are taken into account in the fit as the signal and peaking mis-ID background components, respectively.

The J/ψ -constrained mass $m_{\text{DTF}}^{J/\psi}$ shape from the inclusive sample is shown in Fig. 6.8. It can be seen that at high mass, the most significant components belong to the strange partially-reconstructed background category; whereas at low mass, charmed components dominate. The strange components all have an equivalent in the rare mode (substituting the J/ψ for a dielectron e^+e^-) but the charmed components have no such equivalent. In Fig. 6.8, H_s^* refers to either $K^* \rightarrow K^+\pi$ or $\phi \rightarrow K^+K^-$, whereas H_s^{**} refers to heavier resonances decaying to more than two particles and H_c refers to a charm resonance heavier than the J/ψ .

The partially-reconstructed backgrounds to the $B^+ \rightarrow K^+e^+e^-$ rare mode are studied using four samples of preselected events from the partially-reconstructed simulated decays $B^0 \rightarrow K^{*0}(\rightarrow K^+\pi^-)e^+e^-$, $B^+ \rightarrow K_1^+(\rightarrow K^+\pi^+\pi^-)e^+e^-$, $B^+ \rightarrow K_2^+(\rightarrow K^+\pi^+\pi^-)e^+e^-$ and $B \rightarrow J/\psi X$. In addition, simulated $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ decays are also used which, due to bremsstrahlung, can leak into the signal q^2 region and peak at low $m(K^+e^+e^-)$ mass. The mass distributions for these five background samples are shown in Fig. 6.8. In the region with $m(K^+e^+e^-) \gtrsim 4.88$ GeV, the partially-reconstructed background is completely dominated by the $B \rightarrow K^*(\rightarrow K^+\pi)e^+e^-$ component. By contrast, the leakage from $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ is large in the region with $m(K^+e^+e^-) \lesssim 4.88$ GeV. The $B^+ \rightarrow K_{1,2}^+(\rightarrow K^+\pi^+\pi^-)e^+e^-$ component also starts to be significant below $m(K^+e^+e^-) \lesssim 4.88$ GeV, which is potentially problematic in the fit, as the $B^+ \rightarrow K_{1,2}^+e^+e^-$ branching

²Contributions from B_c^+ or Λ_b^0 are negligible

fraction is not known³. Given the negligibly small contribution, Fig. 6.8 also demonstrates that charmed partially-reconstructed backgrounds are irrelevant for the rare mode.

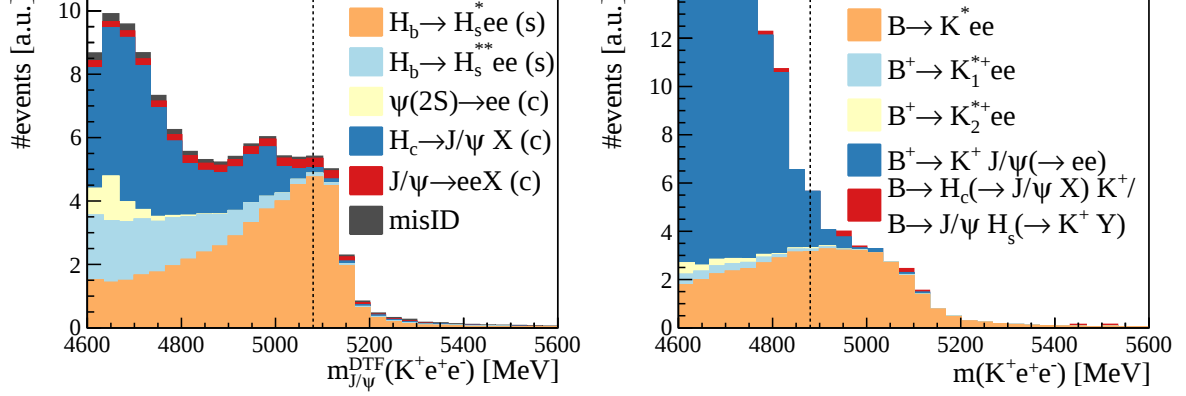


Figure 6.8: Constrained B mass distribution for partially-reconstructed backgrounds to the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode (left), to the rare $B^+ \rightarrow K^+ e^+e^-$ mode (right). In the left figure, the labels (s) and (c) stand for the strange and charmed partially-reconstructed background components. The preselection, including the cut $1.1 < q^2 < 6.0 \text{ GeV}^2$, is applied to all samples. The lower mass cut (i.e. lower limit of the fit window) used in the present analysis is shown with a dotted black line.

Considering Fig. 6.8, the lower mass cut is chosen to be $\sim 300 \text{ GeV}$ below the nominal B^+ mass, that is $m_{\text{DTF}}^{J/\psi}(K^+e^+e^-) > 4.88 \text{ GeV}$ for the normalisation mode $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ and $m(K^+e^+e^-) > 4.88 \text{ GeV}$ for the rare mode $B^+ \rightarrow K^+ e^+e^-$. In the case of the normalisation mode, this cut is more than 99% efficient on signal, and removes most of the charmed partially-reconstructed backgrounds. In the case of the rare mode, the cut $m(K^+e^+e^-) > 4.88 \text{ GeV}$ is 85% efficient on signal and eliminates most of the leakage from $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ and the background from $B^+ \rightarrow K_{1,2}^+(\rightarrow K^+\pi^+\pi^-)e^+e^-$ events. The mass windows used for all modes are summarised in Tab. 6.2.

6.11 Multiple candidates

After the whole selection, the fraction of events containing several candidates is at the sub-percent level for all modes. The fraction of events having multiple candidates for the

³For the purpose of the plot, the branching fraction has been set to

$$\mathcal{B}(B^+ \rightarrow K_{1,2}^+ e^+ e^-) = \mathcal{B}(B^0 \rightarrow K^{*0} e^+ e^-) \cdot \mathcal{B}(B^+ \rightarrow K_{1,2}^+ J/\psi) / \mathcal{B}(B^+ \rightarrow K^{*0} J/\psi)$$

$B^+ \rightarrow K^+ J/\psi$ mode in both simulation and data is reported in Tab. 6.7. When multiple candidates are found for the same event, one candidate chosen at random is kept.

Table 6.7: Fraction of events for which more than one $B^+ \rightarrow K^+ J/\psi$ candidates are reconstructed, when the full selection chain is applied.

	simulation				data	
	rare ee	rare $\mu\mu$	J/ψ ee	J/ψ $\mu\mu$	J/ψ ee	J/ψ $\mu\mu$
Run 1	0.30%	0.01%	0.20%	0.04%	0.10%	0.01%
Run 2	0.47%	0.03%	0.32%	0.03%	0.17%	0.03%

7. Efficiencies and corrections to the simulation

The total efficiency of the full selection summarised in Sec. 6.2 can be factorised into several terms:

$$\varepsilon^{\text{tot}} = \varepsilon^{\text{geom}} \cdot \varepsilon^{\text{rec,pres}} \cdot \varepsilon^{\text{PID}} \cdot \varepsilon^{\text{trig}} \cdot \varepsilon^{\text{BDT}} \quad (7.1)$$

where:

- $\varepsilon^{\text{geom}}$ is the efficiency of all tracks having a polar angle between 10 mrad and 400 mrad (an approximation of the LHCb acceptance);
- $\varepsilon^{\text{rec,pres}}$ is the efficiency of the tracking, vertex reconstruction and preselection, including the q^2 cut that is used to separate the signal and normalisation modes (see Tab. 6.2), but omitting the PID cuts;
- ε^{PID} is the efficiency of all the PID cuts;
- $\varepsilon^{\text{trig}}$ is the efficiency of the L0 and HLT trigger requirements;
- ε^{BDT} is the efficiency of the BDT selection.

Each efficiency is computed on events that have satisfied the previous requirement in the list, *e.g.* $\varepsilon^{\text{trig}}$ is computed with events that have satisfied the PID selection. The total efficiency ε^{tot} is computed using simulated samples which are corrected using control channels selected from the data. These corrections are applied using three types of weights which are computed on an event-by-event basis. Firstly, the PID distributions are mis-modelled in the simulation. Therefore, the PID requirements are not applied to simulated events and instead, events are assigned a weight $w^{\text{PID}} = \varepsilon_{K^+}^{\text{PID}} \cdot \varepsilon_{\ell^-}^{\text{PID}} \cdot \varepsilon_{\ell^+}^{\text{PID}}$, where $\varepsilon_{K^+}^{\text{PID}}$ and $\varepsilon_{\ell^\pm}^{\text{PID}}$ are PID efficiencies extracted from control samples. The independence of the electron weights is established in Sec. 7.3.3. Secondly, the trigger efficiencies are corrected using a weight w^{trig} corresponding to a data/simulation ratio of trigger efficiencies. Thirdly, after applying the trigger and PID corrections, the kinematics of simulated B^+ mesons are corrected to match those observed in the data, resulting in kinematic weights labelled w^{kin} . Two sets of kinematic weights are computed. The first is labelled $w_{\text{gen}}^{\text{kin}}$ and corrects the momentum and rapidity distributions of all B^+ generated in the simulation. These weights

directly impact on the reconstruction efficiency. The second is labelled $w_{\text{rec}}^{\text{kin}}$ and correct for the momentum and rapidity distributions, as well as variables related to the vertex quality. This set of weights is not applied to all simulated B mesons, but only to those that have been reconstructed and have satisfied the preselection requirements. Correcting vertex quality variables is important in order to be able to compute the efficiency of the BDT selection properly, but has a negligible impact on the efficiency of the loose preselection requirements.

The expression for the total efficiency is then,

$$\varepsilon^{\text{tot}} = \underbrace{\frac{\sum_{\text{rec}} w_{\text{gen}}^{\text{kin}}}{\sum_{\text{all}} w_{\text{gen}}^{\text{kin}}}}_{\varepsilon^{\text{geom}, \varepsilon^{\text{rec}, \text{pres}}}} \cdot \underbrace{\frac{\sum_{\text{sel}} w_{\text{rec}}^{\text{kin}} \cdot w^{\text{PID}} \cdot w^{\text{trig}}}{\sum_{\text{rec}} w_{\text{rec}}^{\text{kin}}}}_{\varepsilon^{\text{PID}, \varepsilon^{\text{trig}}, \varepsilon^{\text{BDT}}}}, \quad (7.2)$$

where $\sum_{\text{all}}$ indicates a sum over all simulated B decays, $\sum_{\text{rec}}$ indicates a sum over reconstructed events to which the preselection requirements — except the PID requirements — have been applied, $\sum_{\text{sel}}$ indicates a sum over fully selected events, to which the full selection chain — except the PID requirements — has been applied.

For the rare modes, the branching fraction of interest is defined in a range of the “true” di-lepton invariant mass, $1.1 < q_{\text{true}}^2 < 6.0 \text{ GeV}^2$ (the exact definition of q_{true}^2 is explained in detail in Sec. 7.1). To extract it, the efficiency-corrected yield of $B^+ \rightarrow K^+ \ell^+ \ell^-$ events in $1.1 < q_{\text{true}}^2 < 6.0 \text{ GeV}^2$, $\mathcal{N}_{\text{in}}(B^+ \rightarrow K^+ \ell^+ \ell^-)$, must be computed (*i.e.* $\mathcal{N}_{\text{in}}(B^+ \rightarrow K^+ \ell^+ \ell^-)$ is the total yield of $B^+ \rightarrow K^+ \ell^+ \ell^-$ events in $1.1 < q_{\text{true}}^2 < 6.0 \text{ GeV}^2$ before any detector effects). This efficiency-corrected yield is related to the yield of fully selected rare decay events in the range $1.1 < q^2 < 6.0 \text{ GeV}^2$, $N^{\text{sel}}(B^+ \rightarrow K^+ \ell^+ \ell^-)$, by

$$N^{\text{sel}}(B^+ \rightarrow K^+ \ell^+ \ell^-) = \mathcal{N}_{\text{all}}(B^+ \rightarrow K^+ \ell^+ \ell^-) \cdot \varepsilon^{\text{tot}}(B^+ \rightarrow K^+ \ell^+ \ell^-) \quad (7.3)$$

$$= \mathcal{N}_{\text{in}}(B^+ \rightarrow K^+ \ell^+ \ell^-) \cdot \frac{1}{f^{q^2}} \cdot \varepsilon^{\text{tot}}(B^+ \rightarrow K^+ \ell^+ \ell^-), \quad (7.4)$$

where $\mathcal{N}_{\text{all}}$ is the total number of events in the whole q^2 range, and f^{q^2} is the fraction of $B^+ \rightarrow K^+ \ell^+ \ell^-$ events in the range $1.1 < q_{\text{true}}^2 < 6.0 \text{ GeV}^2$. In practice, f^{q^2} is computed from simulated events as

$$f^{q^2} = \frac{\sum_{\text{gen}, 1.1 < q_{\text{true}}^2 < 6.0 \text{ GeV}^2} w_{\text{gen}}^{\text{kin}}}{\sum_{\text{gen}, \text{all } q^2} w_{\text{gen}}^{\text{kin}}}. \quad (7.5)$$

Note that while this ratio is dependent on the generated q^2 distribution, and hence the form factors assumed in the model, the selection efficiency has a component that, barring small resolution effects, cancels this dependence.

Hence, the ratio of branching fractions between the rare mode, in the range $1.1 < q_{\text{true}}^2 < 6.0 \text{ GeV}^2$, and the $B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-)$ control mode is given by

$$\frac{\mathcal{B}(B^+ \rightarrow K^+ \ell^+ \ell^-)_{\text{in}}}{\mathcal{B}(B^+ \rightarrow K^+ J/\psi)} = \frac{\mathcal{N}_{\text{in}}(B^+ \rightarrow K^+ \ell^+ \ell^-)}{\mathcal{N}(B^+ \rightarrow K^+ J/\psi)} \quad (7.6)$$

$$= \frac{\mathcal{N}_{\text{all}}(B^+ \rightarrow K^+ \ell^+ \ell^-)}{\mathcal{N}(B^+ \rightarrow K^+ J/\psi)} \cdot f^{q^2} \quad (7.7)$$

$$= \frac{N^{\text{sel}}(B^+ \rightarrow K^+ \ell^+ \ell^-)}{N^{\text{sel}}(B^+ \rightarrow K^+ J/\psi)} \cdot \frac{\varepsilon^{\text{tot}}(B^+ \rightarrow K^+ J/\psi)}{\varepsilon^{\text{tot}}(B^+ \rightarrow K^+ \ell^+ \ell^-)} \cdot f^{q^2} \quad (7.8)$$

where the subscript ‘in’ means in the range $1.1 < q_{\text{true}}^2 < 6.0 \text{ GeV}^2$ and ‘all’ means over the full q^2 range. The following sections give more detail on how each efficiency is computed and how each correction weight is determined.

7.1 Truth-matching, ghosts, and the definition of q_{true}^2

Once the simulation software has generated a proton-proton collision and propagated the final state particles into the detector, the full tracking and reconstruction software is run on the simulated event, just as it would be for the real data. For a simulated signal track, accessing the particle that generated it, and the value of its ‘true’ momentum prior to any detector effect is vital to correct the kinematic (Sec. 7.7) and q^2 resolution (Sec. 7.8). The algorithm that provides this information is referred to as the truth-matching algorithm.

For a small fraction of simulated signal events, the truth-matching algorithm fails and one or several tracks in these events end up being mis-classified as ghost tracks. The fraction of ghosts is higher for the electron mode than for the muon modes, as summarised in Tab. 7.1. It is then impossible to tell whether a simulated track classified as a ghost is a signal or background track. This background in the simulation of signal decays must be suppressed to avoid biasing the efficiencies and this is done in three ways. First, only events with one track classified as a ghost, and the two others correctly matched to a lepton or a kaon, are retained. Second, applying the $\text{prob}_{\text{ghost}}$ cut (see in Tab. 6.1) also helps reduce the fraction of combinatorial events. Third, applying the BDT requirement eliminates the majority of the remaining combinatorial events. This means that $\varepsilon^{\text{rec,pres}}$, ε^{PID} and $\varepsilon^{\text{trig}}$ cannot be computed taking signal events mis-classified as ghosts into account, only the total efficiency ε^{tot} , including the BDT cut, take such signal events into account.

Hence, events containing a track classified as a ghost are excluded in extracting all the correction weights, but are included in computing the final total efficiency.

Table 7.1: Relative fractions of simulated events classified as ghosts in fully selected $B^+ \rightarrow K^+ \ell^+ \ell^-$ and $B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-)$ events, after removing multiple candidates.

	e^+e^-		$\mu^+\mu^-$	
	rare	J/ψ	rare	J/ψ
Run 1	4.2%	3.7%	1.5%	1.6%
Run 2	5.1%	4.2%	2.0%	1.9%

The true q^2 , q_{true}^2 , is here defined not only prior to any detector effect, but also prior to FSR (see Sec. 2.11). As electrons are relatively light, they can lose a significant amount of energy via FSR. The simulation of FSR, which is handled by PHOTOS, has been tested and validated in Ref. [41]. The distributions of the reconstructed $q^2 = |p_{e^+} + p_{e^-}|^2$, as well as the true pre-FSR q_{true}^2 , and true post-FSR $q_{\text{true,FSR}}^2$ are shown in Fig. 7.1 for simulated $B^+ \rightarrow K^+ e^+ e^-$ and $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ events. The $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ distribution post-FSR has a large tail to the left due to FSR from the electrons, which is absent on the pre-FSR distribution.

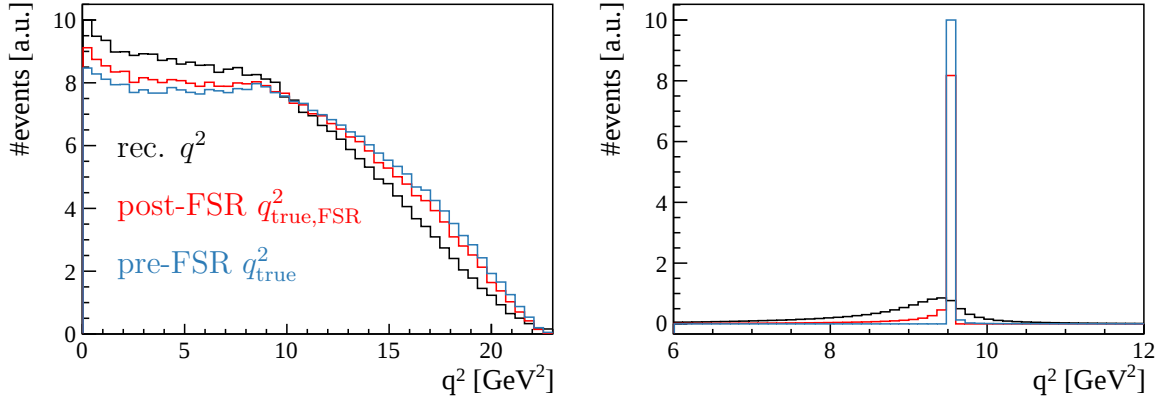


Figure 7.1: Distribution of the reconstructed q^2 (black), true post-FSR q^2 (red) and true pre-FSR q^2 (blue) for simulated $B^+ \rightarrow K^+ e^+ e^-$ events (left) and $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ events (right).

7.2 Fit to the preselected control modes

A fit to the control modes $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ and $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ is performed after applying the preselection and trigger requirements, but without applying the BDT cut. This is done, firstly, to obtain the yield of $B^+ \rightarrow K^+ J/\psi$ events, that is used to optimise the BDT (see Sec. 6.9), and, secondly, to obtain background-subtracted data samples that are used to extract the data/simulation correction weights. In order to statistically subtract the background, the *sPlot* method is used [106].

The fits to $m_{\text{DTF}}^{J/\psi}$ for preselected $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and *eTOS* $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ in the Run 2 data, as well as their pulls, are shown in Fig. 7.2. A description of the fit shapes can be found in Sec. 8.1.

7.3 Corrections to particle identification efficiencies

The performance of the RICH detector and calorimeters is not accurately simulated by the LHCb simulation software. As a result, the efficiencies of the PID requirements summarised in Sec. 6.7 cannot be predicted using the simulation samples. Instead, calibration samples from the data are used to compute the PID efficiencies as a function of kinematic variables. The way in which the calibration samples are selected and the PID efficiencies parametrised has been used in many previous LHCb analyses and is documented in Ref. [87]. The computation for the efficiencies of all particle species relevant to the present analysis are presented below.

7.3.1 Kaon, muon and pion (mis-)identification efficiencies

The PID efficiencies for kaons and muons, as well as the pion mis-identification efficiencies to a kaon (which are different in the muon and electron modes), are measured using data calibration samples of $D^0 \rightarrow K^+\pi^-$ decays and $J/\psi \rightarrow \mu^+\mu^-$ decays. These two-body decays are selected cleanly without the application of PID criteria, and a ‘tag & probe’ method is used to determine the efficiencies: no PID requirement is applied to the track used to measure the PID efficiency (probe), but a tight PID requirement is applied to the other track (tag). It is moreover insured that all requirements applied to the calibration sample that may affect the PID efficiency as measured on the probe are aligned to those used for the signal and control modes, see Tab. 6.1. Fiducial cuts are applied to ensure that the particles are within the acceptance of the relevant detector (RICH, calorimeters

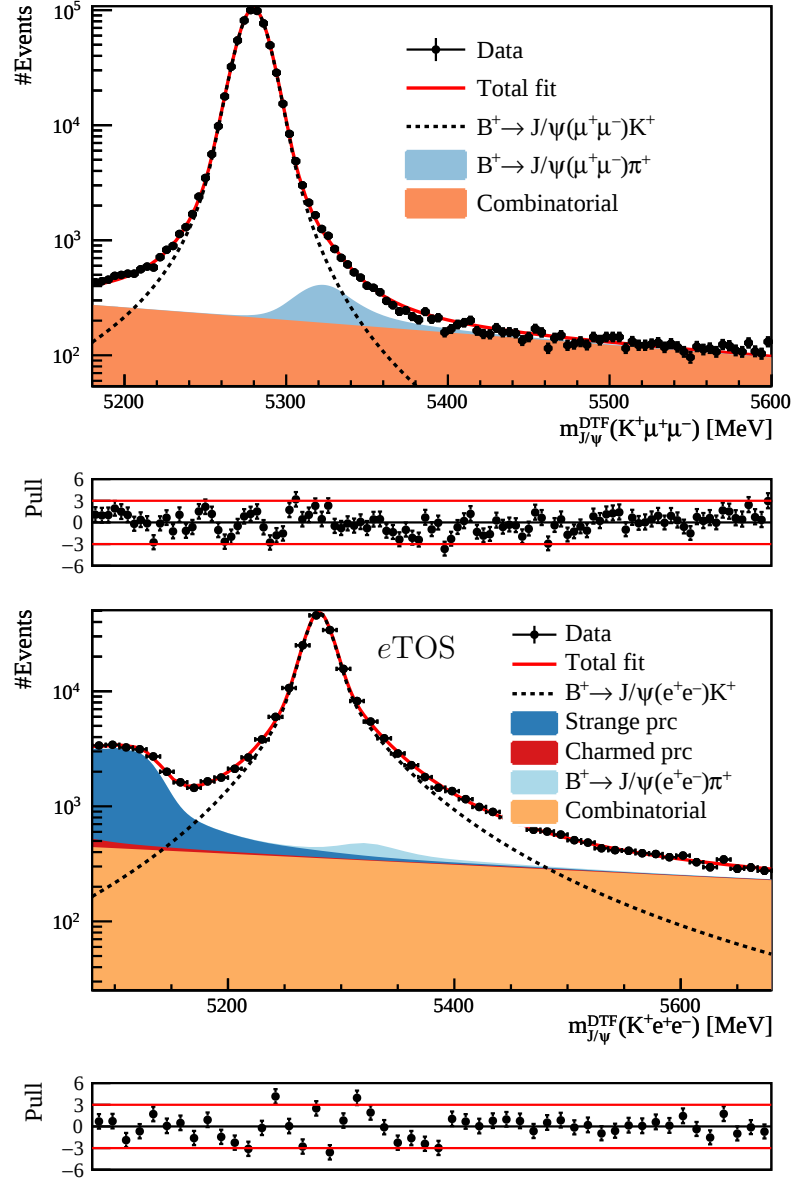


Figure 7.2: Fit to preselected $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ (top) and $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ (bottom) events in the $eTOS$ trigger category. Both fits correspond to the Run 2 data. The different components of the fit are indicated in the legend, where ‘prc’ stands for partially-reconstructed background. The fit quality, while not perfect, is adequate for the purpose of extracting the control channel yield with an error that is irrelevant compared to that of the $B^+ \rightarrow K^+ e^+ e^-$ yield, which is two orders of magnitude smaller.

or muon chambers), and the occupancy cut on nSPDHits is also applied to the calibration samples.

In each selected calibration sample, a fit to the reconstructed mass of the D^0 or J/ψ meson from which the probe originates is performed in order to subtract the background using the *sWeight* method [106]. In each kinematic bin, the efficiency ε^{PID} of a PID requirement is computed using this background-subtracted sample.

The reconstruction of the control modes, the fits and the extraction of the efficiencies are performed centrally for all LHCb analyses [107]. In the case of the muons and kaons from the muon modes, calibration histograms are produced separately for the isMuon cuts and for the DLL (or probNN) cut, in order to be able to study the isMuon efficiency separately from the other requirement. All the PID efficiencies are broken down into bins of year, magnet polarity, momentum and rapidity. Efficiencies are expected to depend on the occupancy. However, they are not binned in any occupancy proxy (like nSPDHits or nTracks) because these proxies are badly described by the simulation, and, being drawn from the same data samples, the occupancy of the calibration samples is very similar to that of the signal samples.

To find the optimal binning in momentum and rapidity, the following procedure is used, inspired by Ref. [108]:

1. An initial binning is chosen such that events are equally-distributed in ten rapidity bins and 100 momentum bins;
2. The 2D rapidity, momentum histograms are projected into two 1D histograms;
3. In each of the two projections, if the efficiency values in two adjacent bins are less than 2.5 standard deviations apart, the bins are merged together.

This method allows a compromise to be found between having enough bins to cover the efficiency variation, and having enough events in each bin to be able to compute the efficiency precisely.

As an example, the projections of the calibration histograms onto the momentum for muons and kaons are shown for the 2016 magnet down data taking conditions in Fig. 7.3. The rest of the kaon and muon PID histograms in other data-taking conditions are similar and are not reported in this thesis.

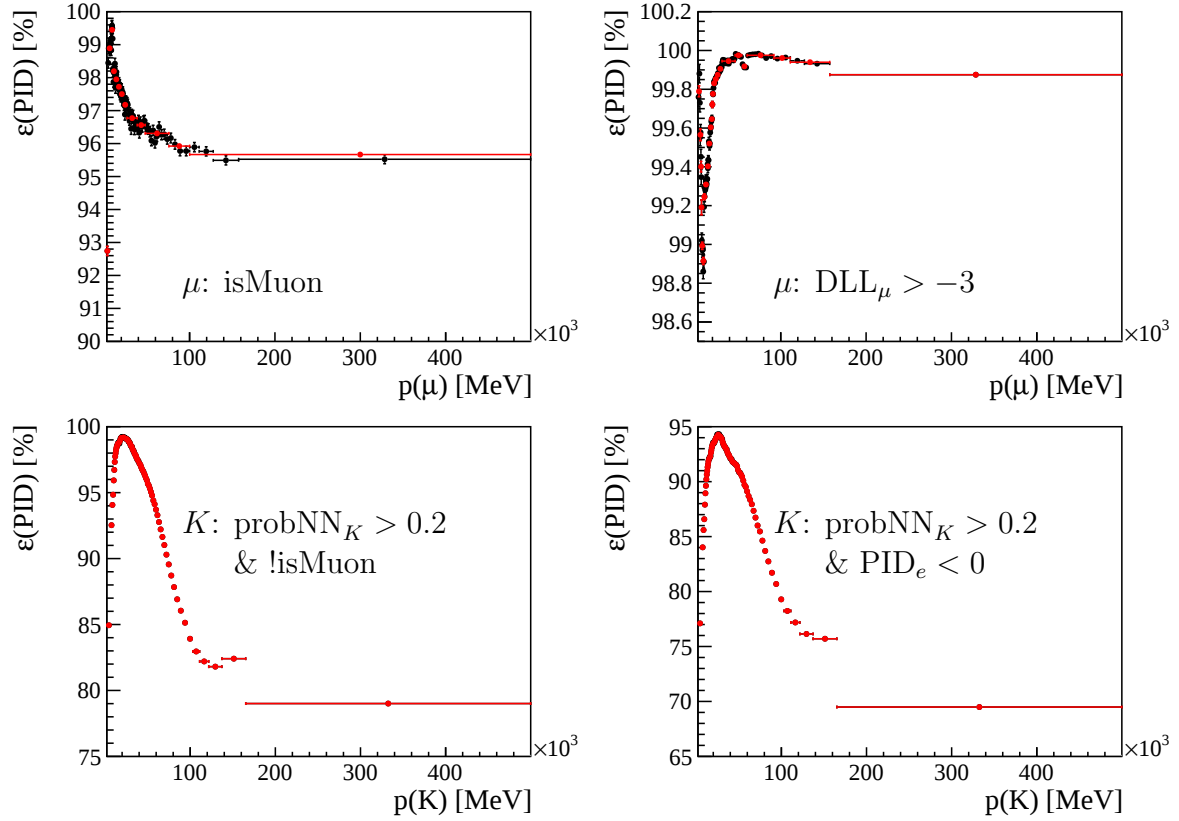


Figure 7.3: Efficiencies of the PID requirements applied to muons (top) and kaons (bottom) as a function of their momentum. The efficiencies are shown before (red) and after (black) the reweighting algorithm.

7.3.2 Electron identification efficiencies

The calibration of the electron PID efficiencies uses a tag & probe method on the decay of the normalisation mode $B^+ \rightarrow K^+ J/\psi(e^+e^-)$, but the sample is reconstructed using a different selection. In particular, the tag is required to satisfy $p_T > 1.5$ GeV and $DLL_e > 5$, whereas all PID requirements on the probe are removed.

The electron PID efficiency is binned in several quantities: the year of data-taking, momentum, rapidity, and two bins for whether the probe electron has a bremsstrahlung photon associated with it or not. The latter variable is denoted `hasBrem`. The electron PID efficiencies in the two magnet polarities are found to be compatible within their errors.

The *sWeight* method that was used to subtract the background in order to extract muon and kaon PID efficiencies cannot be used to precisely extract electron PID efficiencies, because the level of combinatorial background is higher, and the mass shape of the

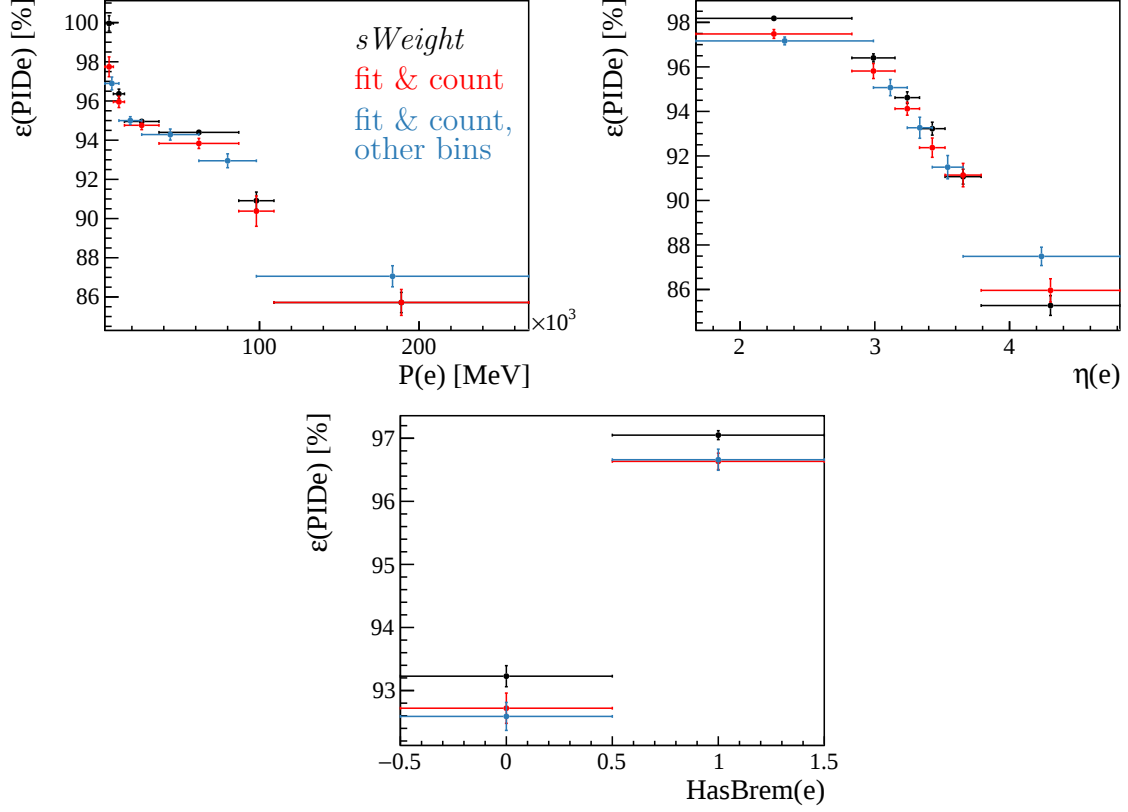


Figure 7.4: Efficiency of the PID requirement $DLL_e > 3$ measured from the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ calibration sample in the 2016 data. Efficiencies are shown as a function of momentum (top left), rapidity (top right) and whether the electron track has a bremsstrahlung cluster associated with it. The black dots correspond to the efficiencies computed with the *sWeight* method, and the red dots to the efficiencies computed with the fit & count method (nominal). The blue dots correspond to an alternative binning scheme used to compute systematic uncertainties (see Sec. 9.4).

combinatorial background is correlated with the PID variable. The *sWeight* method is therefore used only to find the optimal binning in p and η using the algorithm described in Sec. 7.3.1. In each kinematic bin, a fit to the J/ψ -constrained B mass is performed to extract the yields before and after applying the PID cut, and the ratio of these two yields gives the efficiency of the cut. This method is referred to as the fit & count method.

The electron PID efficiencies measured on the 2016 data sample, and projected onto p , η and hasBrem are shown in Fig. 7.4. The efficiencies computed with the *sWeight* method, as well as the default ones computed with the fit & count method are shown, and the difference illustrates the use of the *sWeight* method would result in a biased efficiency at the level of a few percent. The histograms made using the 2011, 2012 or 2015 data are

similar and are not reported in this thesis.

7.3.3 Combination of PID efficiencies and biases

The weight w^{PID} in Eq. 7.2 that is used to calculate the efficiency of the PID requirements is computed as the product of the PID efficiencies for each track, extracted from the PID calibration histograms,

$$w^{\text{PID}} = \varepsilon^{\text{PID}}(K^+) \cdot \varepsilon^{\text{PID}}(\ell^+) \cdot \varepsilon^{\text{PID}}(\ell^-). \quad (7.9)$$

As explained below Eq. 7.1, the PID efficiencies ε^{PID} are defined on preselected events, before applying the trigger requirements.

The way in which the PID efficiencies are computed could result in two potential biases:

1. Factorisation bias: The assumption that the PID efficiencies of the three tracks are independent might not hold perfectly. This is particularly the case for the electron tracks, as the DLL_e variable requires inputs from the calorimeters, which have a low granularity (*i.e.* two electron clusters, or their bremsstrahlung clusters, can be superimposed in the calorimeter and might correlate their DLL responses);
2. Trigger bias: the trigger requirements used in the calibration samples could also bias the PID response in two ways. For the muon mode, the `L0Muon` trigger might bias the efficiency of the `isMuon` requirement. For the electron mode, the `L0Electron` trigger might bias the efficiency of the requirement on DLL_e .

Concerning the muon PID efficiencies, the factorisation bias is absent and the trigger bias is reduced to a negligible level by requiring each event in the calibration sample to have been triggered independently of the muon probe. However, some biases impact the kaon and electron PID efficiencies. These biases are studied on $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulated events, as detailed below.

The efficiency of the PID cut $\text{probNN}_K > 0.2 \ \& \ \text{DLL}_e < 0$ as a function of momentum for simulated kaons in the decay $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ is shown in Fig. 7.5. In black, only the preselection is applied. In red, the electron PID requirements are applied. In blue, any trigger line is required to have fired. In orange, the event is required to have been triggered independently of the probe. The agreement between the red and black shows that the efficiency of the kaon PID requirement is independent of the electron PID requirement. There is a slight disagreement between black and blue, showing that there is a small trigger

bias. The kaon PID calibration sample uses a decay different from that used for Fig. 7.5: $D^0 \rightarrow K^+\pi^-$, so the exact size of the bias cannot be deduced from Fig. 7.5

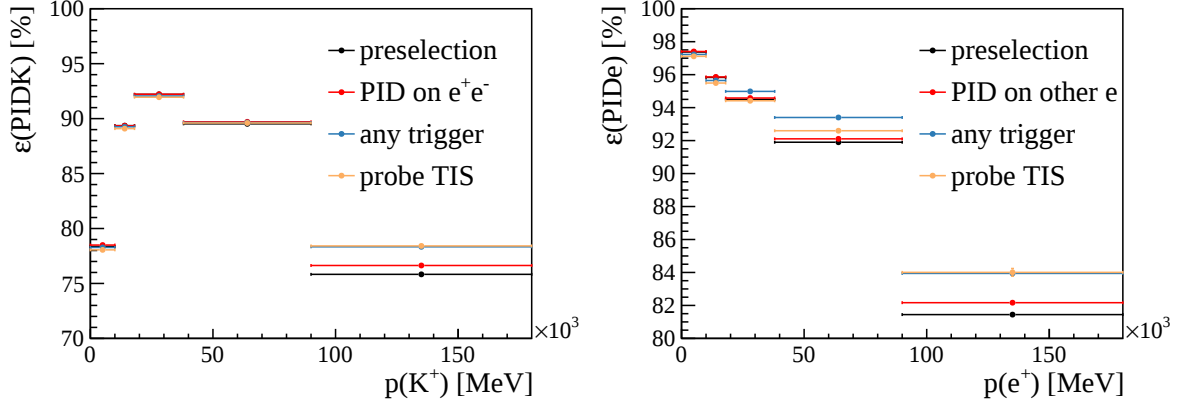


Figure 7.5: Efficiency of the kaon (left) and electron (right) PID cut as a function of momentum, from simulated $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the Run 2 data-taking conditions. The different colours highlight the trigger and factorisation biases as explained in the text.

The efficiency of the PID cut $DLL_e > 3$ applied on one of the electrons in simulated $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events is shown in Fig. 7.5. In black, only the preselection is applied. In red, the PID cut is applied to the other electron in the event. In blue, any trigger line is required to have fired. In orange, the probe electron is required to be TIS at the L0 and HLT trigger levels. The agreement between red and black shows that the two electron PID efficiencies factorise. There is, however, a small trigger bias. Unlike in the muon case, where this could be corrected with the TIS requirement on the probe, this is not the case with electrons, as there is still a disagreement between the orange and black.

To summarise, the PID efficiencies factorise well between the three particles in the decay. There is however a small trigger bias when computing the electron and kaon PID efficiencies. The systematic uncertainty associated with this bias is discussed in Sec. 9.4.

7.4 Calibration of L0 trigger efficiencies

The efficiencies of the four L0 trigger categories defined in Sec. 6.8 are known to be described imperfectly in the simulation. Therefore, they are calibrated from the data using $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ and $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ decays, to which the preselection requirements, including the PID cuts, are applied. The *sWeight* method is employed

to subtract the background and obtain statistically clean data samples, using the fits described in Sec. 7.2. After applying the PID cuts, the normalisation samples have a very low background level and, in contrast to the electron PID calibration histograms, the bias related to the $sWeight$ is negligible.

The weights w^{trig} applied to the simulation are defined as a data-over-simulation ratio of trigger efficiencies. This is different from the PID weights, which are defined as PID efficiencies measured from the data. This has to be the case as the HLT requirement is very dependent on the L0 response, and ignoring this correlation would introduce a significant bias (on the order of 10% on the total trigger efficiency).

As explained below Eq. 7.1, the trigger efficiency is defined on the sample of events that have passed the preselection and PID requirements. This sample is accessible in the simulation, but not fully in the real data: events that are recorded on tape must have satisfied the requirements of at least one trigger line at the three trigger levels. Hence, the efficiency of one particle from the B decay, referred to as the probe, satisfying the requirements of one L0 trigger line can only be measured on a sample of events that have been triggered by another particle. This other particle is referred to as the tag. It is then possible that the trigger requirement on the tag biases the trigger efficiency of the probe. Therefore, when measuring the efficiency of one trigger line, several different tags are used to assess the size of the bias.

7.4.1 L0 efficiency measurements on simulated events

The trigger efficiencies are measured separately in each year of data-taking. They are first studied using simulated $B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-)$ events. The trigger efficiencies of the L0Muon, L0Electron, L0Hadron lines, as well as that of the TIS trigger category, are shown in Fig. 7.6 for 2016 simulated events. For each of the four trigger categories, the ‘direct’ trigger efficiency is shown in black. This corresponds to the trigger efficiency measured on events that satisfy the preselection, including the PID requirements, without using any tag. The ‘direct’ efficiency is therefore accessible in the simulation but not in the data, and is used in the simulation to assess the size of the tag bias. The trigger efficiencies using various tags are shown with different colours.

In the case of the L0Muon, L0Electron, L0Hadron lines, all the tagged efficiencies are close to the direct ones, indicating that the tag bias is small. The systematic uncertainty the tag bias introduces is discussed in Sec. 9.5. The histograms shown in Fig. 7.6 are explained in more detail below.

μ TOS calibration histograms

The efficiency of a muon satisfying the **L0Muon** line requirement, $\varepsilon_{\text{sim}}^{\text{L0}\mu}(\mu)$, is measured as a function of $p_{\text{T}}(\mu)$ using $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ events. The efficiency, measured using events simulated with the 2016 data-taking conditions, is shown in Fig. 7.6a.

The efficiency is measured on events that have been triggered by three possible tags. The first tag is any track in the event, not originating from the signal B , which triggers any L0 line. The second tag is the other muon from the $J/\psi \rightarrow \mu^+ \mu^-$ decay that triggers the **L0Muon** line. The third tag is the kaon from the B decay which triggers the **L0Hadron** line. These three tags are respectively labelled ‘TIS tag’, ‘ μ tag’ and ‘ K tag’ in Fig. 7.6a. The TIS tag is used for the nominal trigger corrections.

In order to reduce the uncertainty due to the limited size of the trigger calibration samples, the trigger efficiency histograms are fitted with the sum of two error functions with four free parameters,

$$f \cdot \frac{1}{2} \cdot \left(1 + \operatorname{erf} \left(\frac{E_{\text{T}} - t}{\sqrt{2}\sigma_1} \right) \right) + (1 - f) \cdot \frac{1}{2} \cdot \left(1 + \operatorname{erf} \left(\frac{E_{\text{T}} - t}{\sqrt{2}\sigma_2} \right) \right), \quad (7.10)$$

where t is the threshold, σ_1 and σ_2 the resolution of each error function and f the fraction of each error function. Using a single error function with one resolution parameter only does not yield a satisfactory fit. This may be due to the fact that the resolution of the muon system depends on the transverse coordinate, *i.e.* better closer to the beam pipe where the GEM detectors are used. The fit function is shown as a solid line on top of all trigger calibration histograms. The data is not fitted below $p_{\text{T}}(\mu) < 800$ MeV, as all events in this region are removed by a preselection requirement (see Tab. 6.1).

e TOS calibration histograms

The efficiency of an electron e satisfying the **L0Electron** line requirement, $\varepsilon_{\text{sim}}^{\text{L0}e}(e)$, is parametrised as a function of two variables: the transverse energy and which of the three ECAL regions the electron deposits its energy in. As explained in Sec. 3.7, the size of the ECAL cells depends on the ECAL region (smallest cells in the inner part and largest cells in the outer part), and therefore the response of the **L0Electron** trigger is expected to depend on the ECAL region. The **L0Electron** efficiency, measured using simulated $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ events in the 2016 data-taking conditions, is shown in Fig. 7.6b. The efficiency is measured as a function of E_{T} and for the outer region in the ECAL. The equivalent plots for the two other ECAL regions are similar and are not shown in this thesis.

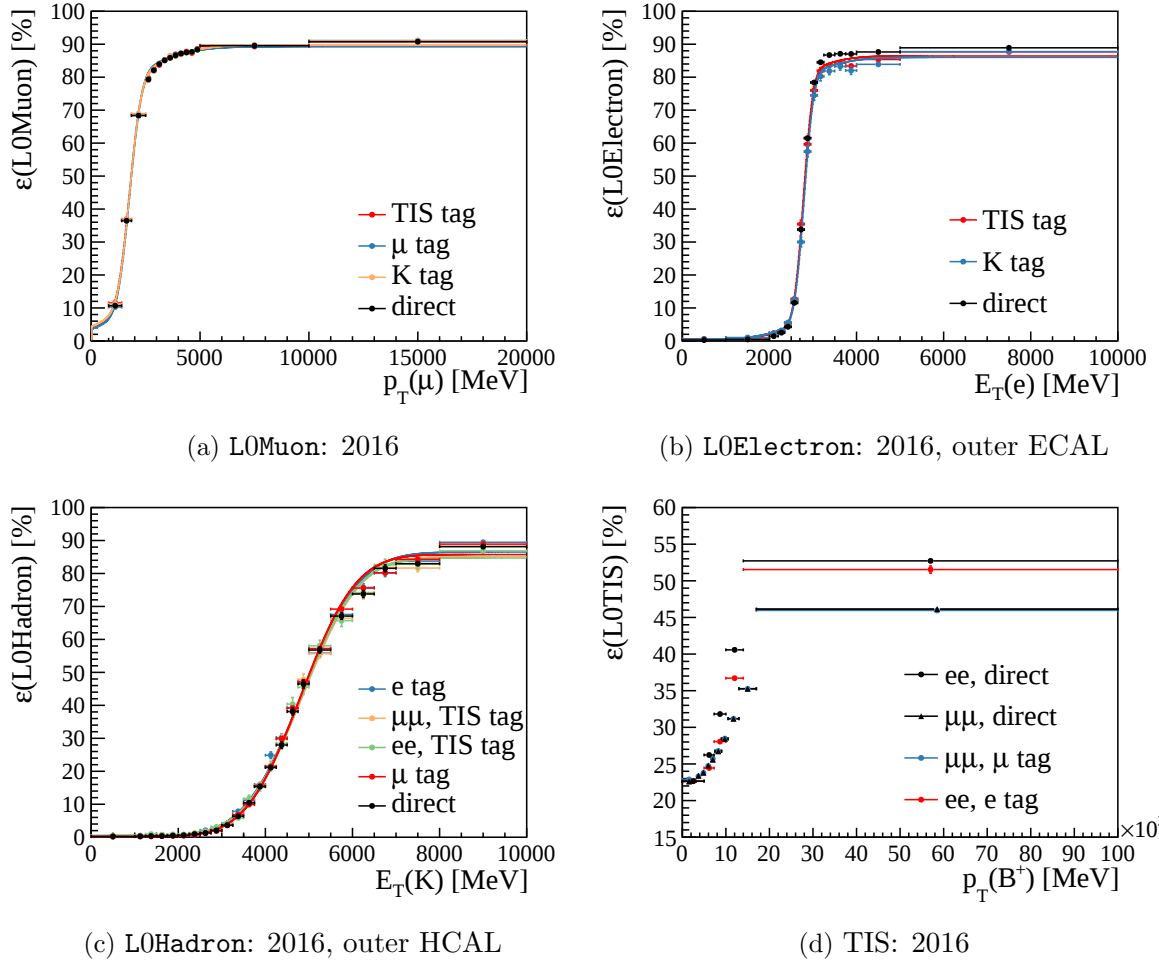


Figure 7.6: Efficiencies of the L0Muons trigger algorithm (top left), L0Electron (top right), L0Hadron (bottom left) and TIS (bottom right). They are respectively shown as a function of the transverse momentum of the muon, transverse energy of the electron, transverse energy of the kaon and transverse momentum of the B^+ in $B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-)$ simulated decays. All plots correspond to the 2016 data-taking conditions, and the L0Electron and L0Hadron plots show the efficiency measured in the outer region of the calorimeter only. The different colours correspond to the different tags discussed in the text. The efficiencies shown in red always correspond to the nominal trigger efficiency measurement, and those in black to the direct (no tag) measurement. The fits to the efficiency histograms are shown with coloured lines.

The efficiency is measured on events that have been triggered at the L0 by either of two tags. The first tag is a hadron or a muon independent of the signal B that triggers the L0Hadron or L0Muons lines. This is labelled as ‘TIS tag’ in Fig. 7.6b, and it is the tag used for the nominal trigger corrections. The second tag is a kaon in the B decay that triggers

the L0Hadron line. This is labelled as ‘ K tag’ in Fig. 7.6b, and has a poorer statistical precision than the previous category.

The trigger efficiencies are fitted with an error function with three free parameters:

$$\frac{1}{2} \cdot \left(1 + \operatorname{erf} \left(\frac{E_T - t}{\sqrt{2}\sigma} \right) \right) + a, \quad (7.11)$$

where t is the threshold, σ the resolution and a the efficiency at $E_T = 0$. The parameter a is non-zero because noise or hits in the ECAL from other tracks in the event can fire the trigger and be falsely matched to the candidate, independently of its E_T value. The fit function is again shown as a solid line on top of the trigger calibration histogram.

h TOS calibration histograms

In a similar way to the L0Electron efficiency, the efficiency of the kaon K satisfying the L0Hadron line requirement, $\varepsilon_{\text{sim}}^{\text{L0h}}(K)$ is measured on simulated data. Following the same arguments as for the L0Electron efficiency, the L0Hadron efficiency is parametrised as a function of the transverse energy of the kaon and which of the two HCAL regions the kaon deposits its energy in.

The L0Hadron efficiency measured using the kaon in $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ and $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ simulated events is shown in Fig. 7.6c for the 2016 data-taking conditions. The efficiency is measured on events where either of four tags have satisfied the L0 trigger requirements. The first one is using $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ events that have been triggered by one of the two electrons satisfying the requirements of the L0Electron line. The second one is using $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ events that have been triggered by one of the two muons satisfying the requirements of the L0Muon line. The third one is using $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ events that have been triggered by any particle not originating from the signal B that has satisfied the requirements of any L0 trigger line. The fourth and last one is using the same strategy as in the previous case but using $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ events. These four possible tags are respectively labelled as ‘ e tag’, ‘ μ tag’, ‘ ee TIS tag’ and ‘ $\mu\mu$ TIS tag’ in Fig. 7.6c. The μ tag is used as the nominal tag for the hadron trigger corrections, because it has the highest yield and hence the best precision.

The data and simulation histograms are fitted with the same function that was used for the L0Electron efficiency, Eq. 7.11. The fit function is again shown as a solid line on top of all the trigger calibration histograms.

TIS calibration histograms

The efficiency of an event satisfying the TIS requirement, $\varepsilon_{\text{sim}}^{\text{LOTIS}}$, is measured as a function of the transverse momentum of the B , $p_T(B)$, as it is expected to be correlated with the transverse momentum of the system resulting from the hadronisation of the opposite b -quark, which directly impacts on the TIS efficiency.

Approximately 55% of $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ and 60% of $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ TIS events are triggered by a hadron, and the TIS efficiency therefore depends on whether the kaon from the signal decay has passed the L0Hadron line requirement or not. As one of the trigger categories used to extract R_K is the exclusive TIS! category, the TIS calibration histograms are extracted only for events where the kaon did not fire the hadron trigger. The TIS efficiency measured on $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events from the data collected in 2016 is shown in Fig. 7.6d.

The TIS efficiency measured on selected $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ and $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events without a tag are labelled ‘ ee direct’ and ‘ $\mu\mu$ direct’ in Fig. 7.6d and are incompatible with each other. The efficiency is also measured using two different tags: First, the efficiency is measured on $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events that have been triggered in the L0 trigger algorithm by one of the two electrons satisfying the L0Electron requirements. This is labelled as ‘ ee , e tag’ in Fig. 7.6d. Second, the efficiency is measured as for the ‘ ee , e tag’, but using $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events and the L0Muon trigger. This is labelled as ‘ $\mu\mu$, μ tag’ in Fig. 7.6d.

The e tag is used as the nominal tag for the TIS trigger corrections in the electron mode. However, it is clear from Fig. 7.6d that the tag bias in the electron mode is large, and that the muon mode cannot be used to correct the TIS efficiencies in the electron mode. The difference in the TIS efficiency between the electron and muon modes as well as the tag bias are due to ‘fake TIS’ events, as discussed in more detail in Sec. 9.5. The systematic uncertainty associated to the tag bias is also discussed in Sec. 9.5.

7.4.2 L0 efficiency measurements on data and data/simulation ratio

For $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ data, the trigger efficiencies ($\varepsilon_{\text{data}}^{\text{L0}\mu}$, $\varepsilon_{\text{data}}^{\text{L0e}}$, $\varepsilon_{\text{data}}^{\text{L0h}}$ and $\varepsilon_{\text{data}}^{\text{LOTIS}}$) are measured, parametrised and fitted in the exact same way as for simulated events. The ratio of trigger efficiencies between data and simulation for the year 2016 are shown in Fig. 7.7. The efficiency ratios for the other years of data-taking

are similar and not reported in this thesis. In the case of the L0Muon, L0Electron and L0Hadron trigger lines, it is clearly visible that at low E_T or p_T , the discrepancy between data and simulation is large, and the difference between the efficiencies measured with various tags is also large. This justifies the use of the trigger fiducial requirement, shown with the dashed black line, which removes events that would be associated with a large systematic uncertainty.

7.4.3 Trigger efficiency combination and factorisation bias

Using the L0Electron efficiencies measured on simulation and data, the e TOS correction weights are given by

$$w^{e\text{TOS}} = \frac{1 - (1 - \varepsilon_{\text{data}}^{\text{L0e}}(e^+)) \cdot (1 - \varepsilon_{\text{data}}^{\text{L0e}}(e^-))}{1 - (1 - \varepsilon_{\text{sim}}^{\text{L0e}}(e^+)) \cdot (1 - \varepsilon_{\text{sim}}^{\text{L0e}}(e^-))}, \quad (7.12)$$

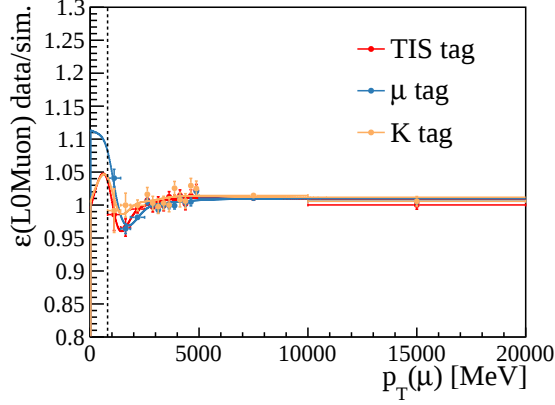
and similarly for the μ TOS correction weights.

This formula is only exact, however, if the trigger efficiencies from the two leptons are independent of each other. This has been checked to hold in the muon case, as can be seen in Fig. 7.8: the trigger efficiency of one muon does not depend on whether the other muon in the decay has passed the L0Muon requirement or not. However, in the electron case, the trigger efficiencies of the two leptons factorise less well, as can be seen in Fig. 7.8. If one electron is required to satisfy the L0Electron requirements, then the probability of the other electron also satisfying the L0Electron requirements goes down. However, this bias is much reduced if the distance between the two electrons in the calorimeter is required to be above 1 m. This is related to the fact that each front-end board in the ECAL can only save the highest- E_T candidate. Hence, if the two signal electrons end up too close in the ECAL, only one is saved as having satisfied the trigger requirements. The systematic uncertainty related to the imperfect factorisation of the trigger efficiencies of the two electrons is discussed in Sec. 9.5.

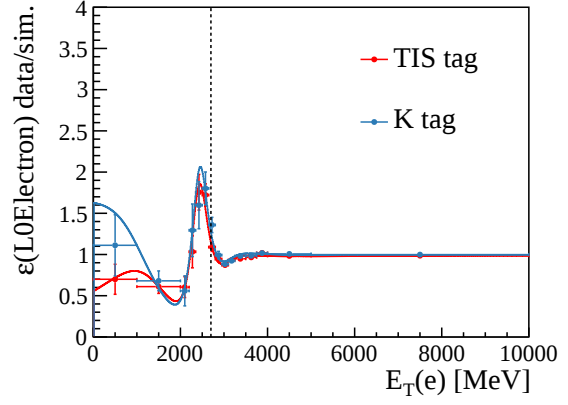
The weight applied to simulated events to correct the trigger L0 efficiency in the h TOS! category is computed as

$$w^{h\text{TOS!}} = \frac{\varepsilon_{\text{data}}^{\text{L0h}}(K^+)}{\varepsilon_{\text{sim}}^{\text{L0h}}(K^+)} \cdot \frac{(1 - \varepsilon_{\text{data}}^{\text{L0e}}(e^+)) \cdot (1 - \varepsilon_{\text{data}}^{\text{L0e}}(e^-))}{(1 - \varepsilon_{\text{sim}}^{\text{L0e}}(e^+)) \cdot (1 - \varepsilon_{\text{sim}}^{\text{L0e}}(e^-))}.$$

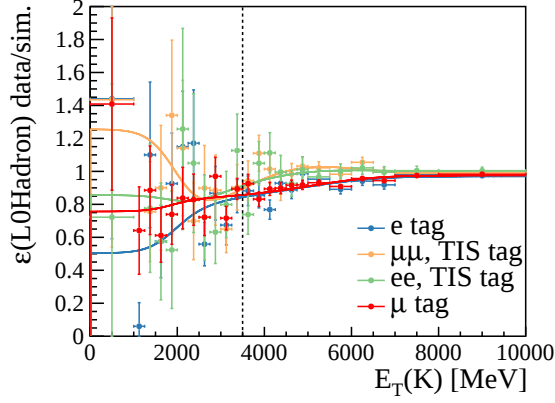
Simulated $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events indicate that the efficiency of the L0Hadron trigger line does not depend on whether one of the electrons has passed the L0Electron requirement, as can be seen in Fig. 7.9.



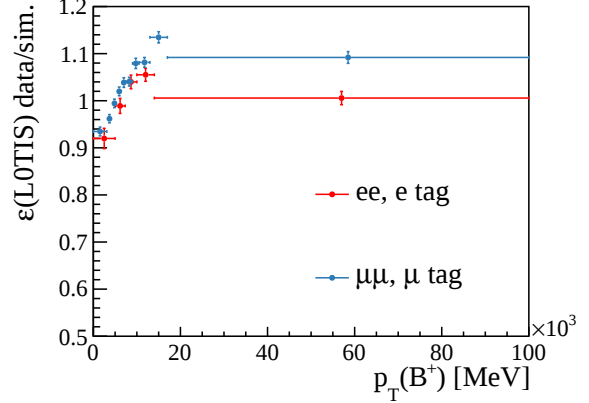
(a) L0Muon: 2016



(b) L0Electron: 2016, outer ECAL



(c) L0Hadron: 2016, outer HCAL



(d) TIS: 2016

Figure 7.7: Data over simulation ratio of efficiencies of the L0Muon trigger algorithm (top left), L0Electron (top right), L0Hadron (bottom left) and TIS (bottom right). They are shown as a function of the transverse momentum of the muon, transverse energy of the electron, transverse energy of the kaon and transverse momentum of the B^+ in $B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-)$ simulated decays. All plots correspond to the 2016 data-taking conditions, and the L0Electron and L0Hadron plots show the efficiency measured in the outer region of the calorimeter only. The different colours correspond to the different tags discussed in the text. The efficiencies shown in red always correspond to the nominal trigger efficiency measurement. The coloured lines on top of the histograms correspond to the ratio of the functions used to fit the data and simulation efficiency histograms. The vertical dashed lines correspond to the fiducial requirements shown in Tab. 6.4.

The weight applied to simulated events in order to correct the trigger L0 efficiency in

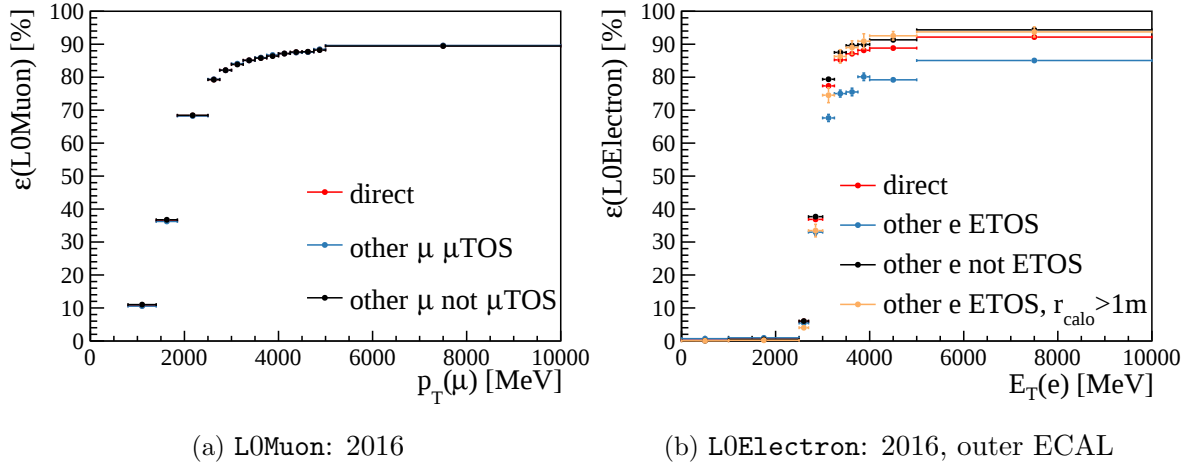


Figure 7.8: Efficiency of the L0Muon trigger as a function of the transverse momentum of a muon probe (left) and efficiency of the L0Electron trigger as a function of the transverse energy of an electron probe (right) for $B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-)$ simulated events, corresponding to the 2016 data-taking conditions. The efficiency is shown in the case where no trigger requirement is placed on the other lepton in the B^+ decay (red), when the other lepton is required to have fired the L0 trigger (blue), and when the other lepton is required not to have fired the L0 trigger (black). In the L0Electron case, the efficiency is also shown when the other electron has fired but is more than 1 m away from the electron probe (orange).

the TIS! category is computed as:

$$w^{\text{TIS!}} = \frac{\epsilon_{\text{data}}^{\text{L0TIS}}(B^+)}{\epsilon_{\text{sim}}^{\text{L0TIS}}(B^+)} \cdot \frac{(1 - \epsilon_{\text{data}}^{\text{L0e}}(e^+)) \cdot (1 - \epsilon_{\text{data}}^{\text{L0e}}(e^-))}{(1 - \epsilon_{\text{sim}}^{\text{L0e}}(e^+)) \cdot (1 - \epsilon_{\text{sim}}^{\text{L0e}}(e^-))} \cdot \frac{1 - \epsilon_{\text{data}}^{\text{L0h}}(K^+)}{1 - \epsilon_{\text{sim}}^{\text{L0h}}(K^+)}.$$

As the trigger calibration histograms are computed after requiring the kaon to not have fired the hadron trigger, the above formula correctly takes into account the bias that the hadron TOS requirement induces on the TIS efficiency. However, this formula also relies on the fact that the TIS efficiency is independent of whether one of the electrons has fired the trigger. As can be expected from the tag bias already discussed this is not exactly the case, as can be seen in Fig. 7.9. The systematic uncertainty related to this is discussed in Sec. 9.5.

7.4.4 Hidden variables in L0 efficiency parametrisation

All the trigger correction weights have been parametrised as a function of one or two variables. It could however be the case that the trigger efficiencies depend on one or several extra variables. If the distributions of one of these hypothetical extra variables differ

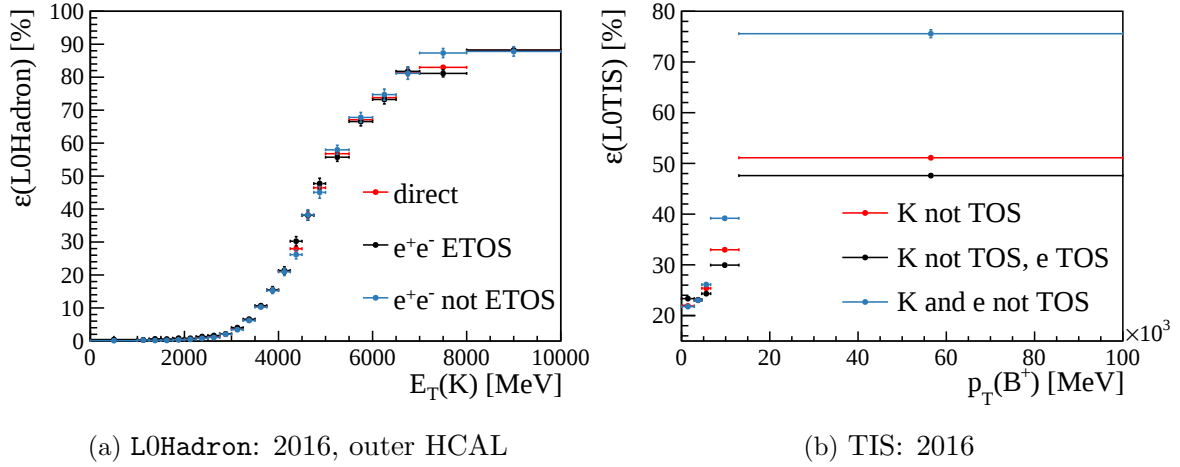


Figure 7.9: Efficiency of the L0Hadron trigger algorithm as a function of the transverse energy of the kaon (left), and efficiency of the TIS trigger as a function of the transverse momentum of the B^+ , for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulated events corresponding to the 2016 data-taking conditions. In the L0Hadron case, the efficiency corresponds to that of kaons detected in the outer region of the calorimeter. The efficiency is shown in the case where no requirement is placed on the electrons (red), when one of the electrons fired the L0Electron trigger (black) and when neither of the two electrons fired the L0Electron trigger (blue). In the TIS case, the kaon is always required to not have fired the L0Hadron trigger.

between the control modes and the rare modes, the ratio of trigger efficiencies $\epsilon^{\text{trig}}(B^+ \rightarrow K^+ \ell^+ \ell^-) / \epsilon^{\text{trig}}(B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-))$ will be biased. The size of the bias is extracted by computing the L0 trigger efficiency histograms using simulated $B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-)$, $B^+ \rightarrow K^+ \psi(2S)(\ell^+ \ell^-)$ and $B^+ \rightarrow K^+ \ell^+ \ell^-$ in the range $1.1 < q^2 < 6.0 \text{ GeV}^2$. For all categories, the trigger histograms computed with the three channels are not significantly different, showing that the number of variables in the parametrisation is sufficient and that there is no significant bias in these cases. For each L0 trigger line, one efficiency histogram is shown in Fig. 7.10 in the Run 2 data-taking conditions and for the three different channels.

7.5 Calibration of HLT efficiencies

The efficiency of the HLT1 and HLT2 requirements is measured on $B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-)$ data and simulated events as a function of the transverse momentum of the B^+ candidate. The HLT efficiencies are computed using preselected events where a particle independent of the B^+ decay has triggered both the HLT1 and HLT2 trigger algorithms. In the muon

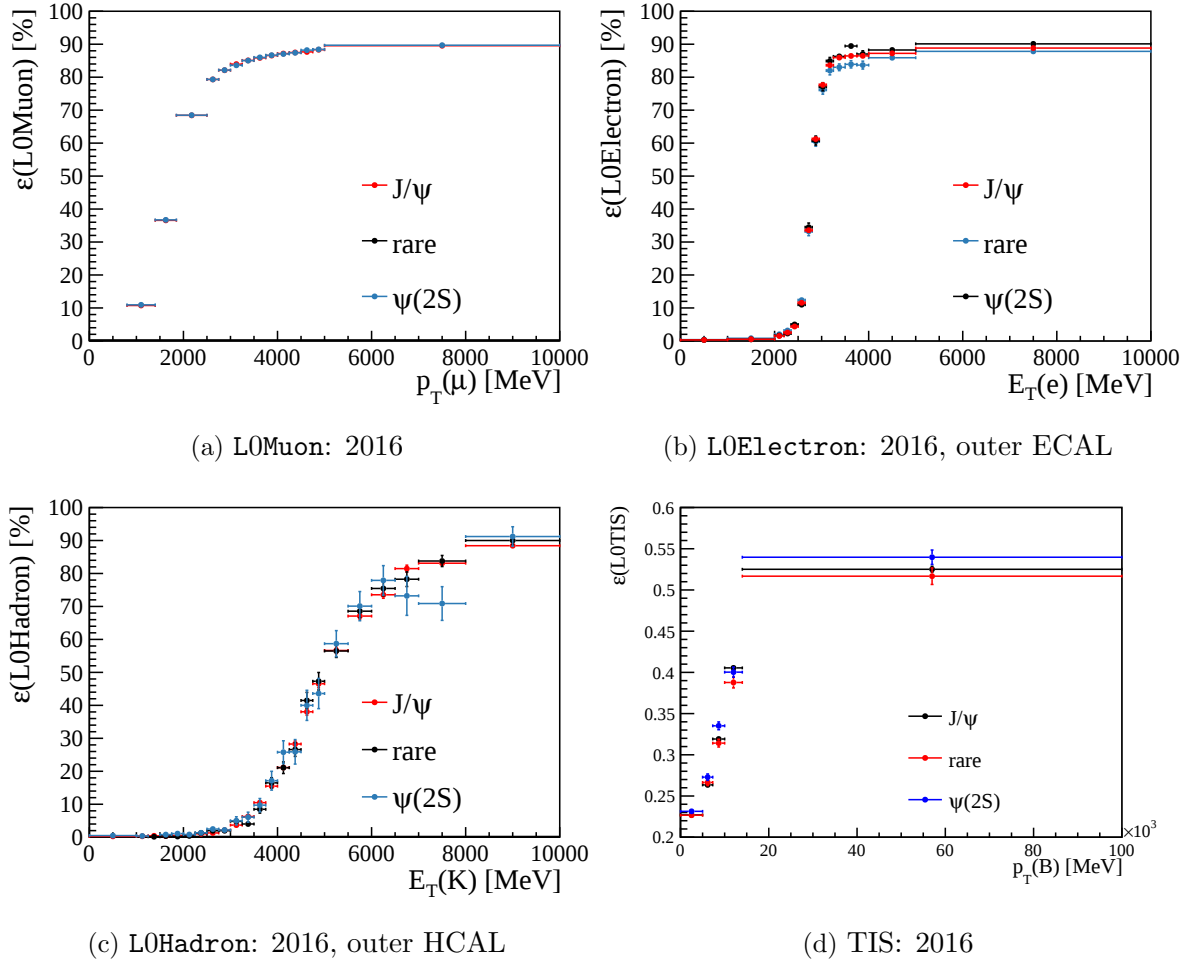


Figure 7.10: Efficiency of the L0Muon (top left), L0Electron (top right), L0Hadron (bottom left) and TIS requirements for preselected $B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-)$, $B^+ \rightarrow K^+ \psi(2S)(\ell^+ \ell^-)$ and $B^+ \rightarrow K^+ \ell^+ \ell^-$ simulated events in the 2016 data-taking conditions. The efficiencies computed in the three channels are statistically compatible in the three TOS cases but not in the TIS case.

case, the events are required to pass the L0Muon requirements. In the electron case, HLT efficiencies are extracted in three different cases depending on whether the events pass the L0Electron, L0Hadron or TIS requirements.

To extract the HLT efficiencies in the data, the *sWeight* method cannot be used because the level of background is too high for events selected with an HLT1 and HLT2 TIS requirement. Instead, the HLT efficiencies are computed by fitting the B^+ mass before and after applying the HLT cut, using the fit model described in Sec. 8.2.

The ratio of HLT efficiencies between data and simulation, for the years of data-taking

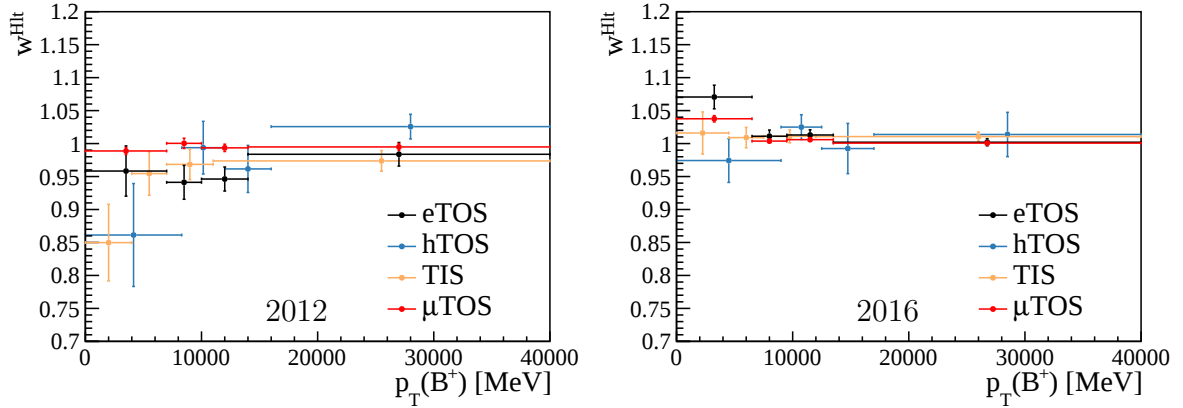


Figure 7.11: Data over simulation ratios of HLT trigger efficiencies as a function of the transverse momentum of the B^+ candidate, corresponding to 2012 (left) and 2016 (right) data-taking conditions. The black, blue and orange points correspond to the ratios measured on $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events that pass the preselection requirements, that are TIS in the HLT1 and HLT2 triggers, and that belong to either e TOS (black) h TOS (blue) or TIS (orange) L0 categories. The red points correspond to ratios measured on $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events that pass the preselection requirements, that are TIS in the HLT1 and HLT2 triggers, and that belong to the μ TOS L0 category.

2012 and 2016, is shown in Fig. 7.11 for $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ μ TOS events and for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ e TOS, h TOS and TIS events.

As expected, because of the Run 2 HLT fiducial cuts discussed in Sec. 6.8, the agreement between data and simulation for the HLT efficiencies is better in Run 2 than in Run 1. The ratio of efficiencies shown in Fig. 7.11 are therefore used as a nominal correction in the Run 1 efficiency computation, but not in Run 2.

7.6 Kinematic corrections

The B^+ momentum and rapidity spectra are generated imperfectly in the simulation. Moreover, the reconstructed χ_{DV}^2 and $\chi_{IP}^2(B^+)$ distributions are also badly modelled in the simulation. The simulation is therefore corrected using kinematic weights, which correspond to data over simulation ratios of kinematic distributions stored in histograms. Two sets of kinematic weights are used in the efficiency computation:

1. A first set that corrects the two-dimensional $(p_T(B^+); \eta(B^+))$ distribution to match that in the data, in order to compute the geometric and reconstruction efficiencies, ϵ^{geom} and $\epsilon^{\text{rec,pres}}$;

2. A second set that corrects the two-dimensional $(p_T(B^+); \eta(B^+))$ distribution, followed by the one-dimensional χ_{DV}^2 and $\chi_{\text{IP}}^2(B^+)$ distributions. This is done iteratively in four loops to handle the small correlations between the three distributions. These weights are used in the computation of the rest of the efficiencies, $\varepsilon^{\text{presel}}$, ε^{PID} , $\varepsilon^{\text{trig}}$ and ε^{BDT} .

The samples used to extract the correction weights are the data and simulated samples of preselected $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ events in the μTOS category. This particular sample is used because it has the largest number of events, and the largest signal-to-background ratio. Moreover, it is the mode for which the PID and trigger efficiencies are best under control, as the former does not show any trigger bias (see Sec. 7.3.3) and the latter does not suffer from any significant tag or factorisation bias (see Sec. 7.4.1 and 7.4.3). In addition, the momentum resolution is much better for muons than for electrons. Kinematic weights computed from the $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ mode are only used for cross-checks and to compute systematics, as is described in Sec. 9.3.

The PID and L0 corrections are used to correct the simulation sample, and the *sWeight* method is used to subtract the background in the data sample. The two-dimensional and one-dimensional histograms contain respectively 20×10 and 150 iso-populated bins.

The kinematic weights are computed separately for each year of data-taking and, in Run 1, separately for simulated events generated with PYTHIA 6 and PYTHIA 8, as the B^+ kinematics differ slightly in the two versions of the software (as stated in Sec. 5.2, in Run 2, all simulated samples are generated with PYTHIA 8). When the weights are applied on an electron mode, the true $(p_T(B^+); \eta(B^+))$ is used, as the reconstructed $(p_T(B^+); \eta(B^+))$ has a poor resolution due to bremsstrahlung.

7.6.1 Correction to the occupancy distribution

The occupancy of the detector is known to be poorly described in the simulation. The resolution, tracking or trigger efficiencies are all mildly dependent on the occupancy. Some occupancy proxies, such as the number of hits recorded in the SPD detector, `nSPDHits`, could be considered as quantities that should be corrected in addition to those already mentioned. However, as can be seen in Fig. 7.12, correcting the `nSPDHits` distribution in simulated $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ events to obtain a good agreement with the data results in poorer agreement for some other occupancy proxies. In particular, the occupancy of the ECAL, both described in terms of number of reconstructed photons and in terms of the sum of the transverse momenta of the reconstructed photons, shows a much better

agreement before correcting the nSPDHits distribution. This is particularly important because the L0Electron efficiency and the bremsstrahlung recovery (and therefore the electron momentum and q^2 resolution) depend strongly on the occupancy of the ECAL. The efficiency of the cut on nSPDHits, *i.e.* nSPDHits < 450 in Run 2 and < 600 in Run 1, cancels perfectly in the efficiency ratio $\varepsilon(B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-))/\varepsilon(B^+ \rightarrow K^+ \ell^+ \ell^-)$ and therefore does not present any issue in the present analysis.

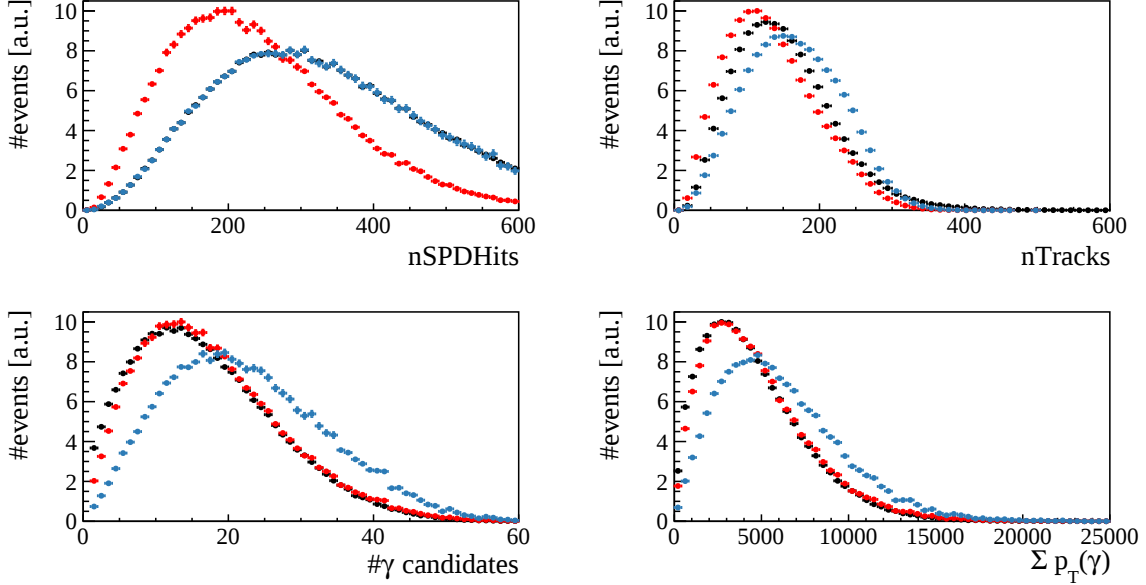


Figure 7.12: Distributions for various occupancy proxies for $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ events. The black histograms correspond to background-subtracted data, red to uncorrected simulation and blue simulation in which the nSPDHits distribution is corrected to match that in the data (as it is clearly visible in the top left plot). The top distributions correspond to the occupancy of the SPD detector (left) and the number of tracks reconstructed in the event (right). The bottom plots correspond to the number of photons reconstructed in the ECAL (left) and the sum of the transverse momenta of the photons reconstructed in the ECAL (right).

The reason why different occupancy proxies show different levels of agreement between simulation and data is explained by a number of factors. Firstly, different sub-detectors can be more or less sensitive to various badly modelled effects, such as back-splash (*i.e.* a portion of a shower in the calorimeter that leaves hits in the SPD detector) or secondary interactions (*i.e.* low energy particles coming from the interaction of higher energy particles with the material of the detector). In particular, the SPD detector is expected to be more sensitive than the ECAL to low energy, badly modelled effects. Secondly, the energy or

rapidity distributions of the particles emerging from the primary proton-proton interaction is also expected to be badly modelled, impacting the various occupancy proxies in different ways.

In summary, it is impossible to satisfactorily correct the occupancy of the detector by simply reweighting a single proxy. In particular, reweighting nSPDHits would lead to a strong disagreement between the ECAL occupancy in the simulation and in data in a way that would impact on the L0Electron efficiency. Therefore, in this analysis, the nominal efficiencies are computed without correcting any efficiency proxy. The systematic uncertainty related to the occupancy is discussed in Sec. 9.6.

7.7 Kinematic distributions in simulation and data

The distributions of several kinematic variables for preselected $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ events in the μ TOS category are shown in Fig. 7.13. Similar distributions are shown for preselected $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ events in the e TOS category in Fig. 7.14. All distributions correspond to the Run 2 data-taking conditions. For each mode and each kinematic variable, three distributions are shown: the first corresponds to the background-subtracted data, the second to the simulation with the trigger and PID correction weights but not the kinematic correction weights, and the third corresponds to the simulation with all the weights, including the kinematic corrections.

In the case of the muon mode, the agreement between the data and the simulation with corrected kinematics is perfect for all the distributions. In the case of the electron mode, the agreement between the data and the simulation is much improved by applying the kinematic weights. However, some discrepancies persist in distributions related to angles, such as the lepton rapidity or the angle between the two electrons. These are further discussed in Sec. 9.7. Some discrepancies are also visible in variables related to the alignment between the B^+ momentum and the B^+ line of flight, such as the $\chi_{\text{IP}}^2(K^+ e^+ e^-)$ or the DIRA. However, these discrepancies can be at least partly attributed to the fact that these variables are correlated to the mass fit variable $m_{\text{DTF}}^{J/\psi}$. The background subtraction, which relies on the *sWeight* method, is therefore imperfect. For these variables, the agreement between data and simulation is instead probed with the differential $r_{J/\psi}$ test discussed in Sec. 10.2.

A supplementary set of kinematic distributions, in the Run 1 and Run 2 conditions, is shown in Appendix B. The observations made for the Run 2 data remain valid for the

Run 1 data.

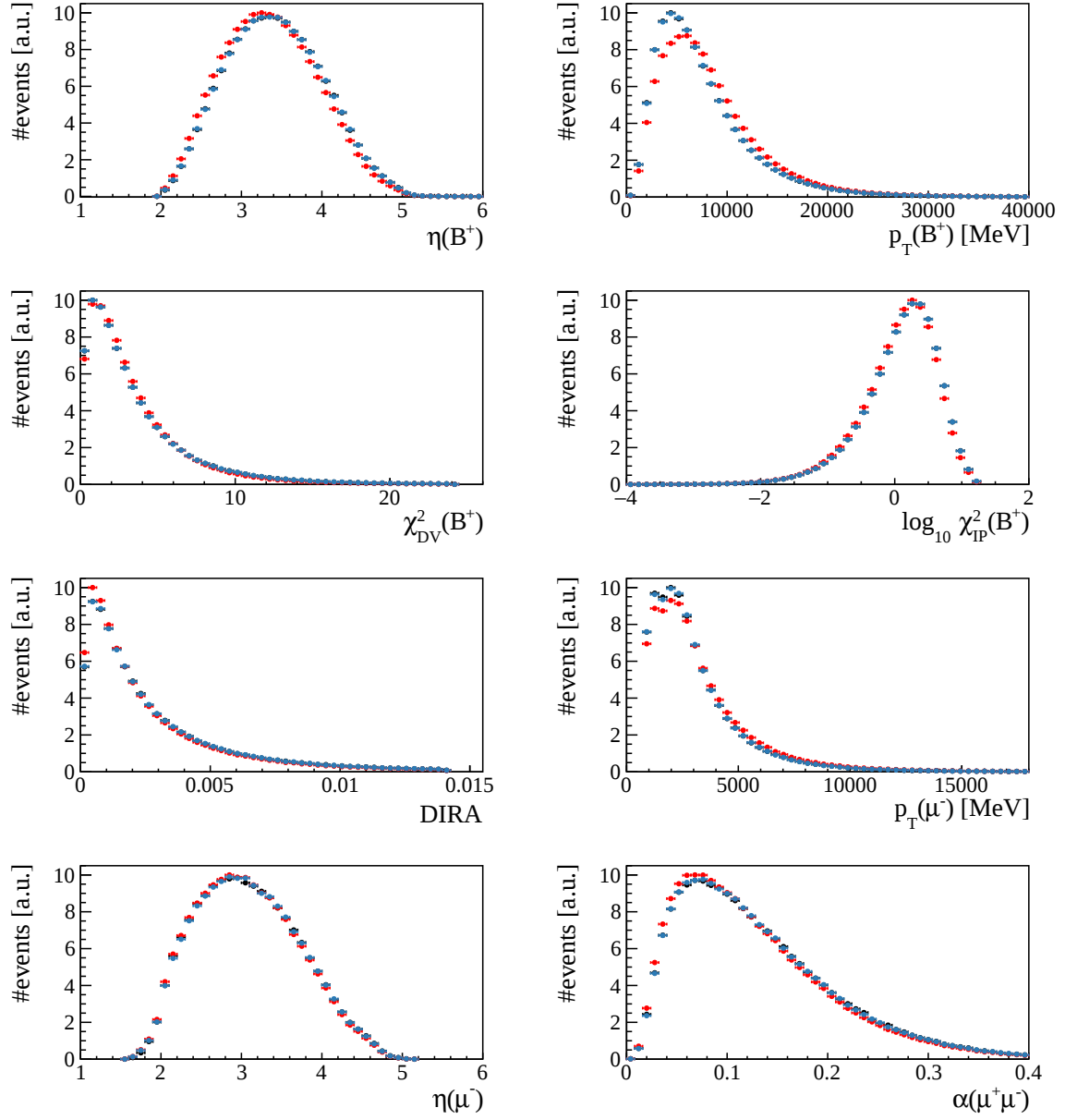


Figure 7.13: Kinematic distributions for preselected $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ events in the μ TOS category and in the Run 2 data-taking conditions. Each plot shows the background-subtracted data (black), simulation (red) and simulation with kinematic correction weights (blue).

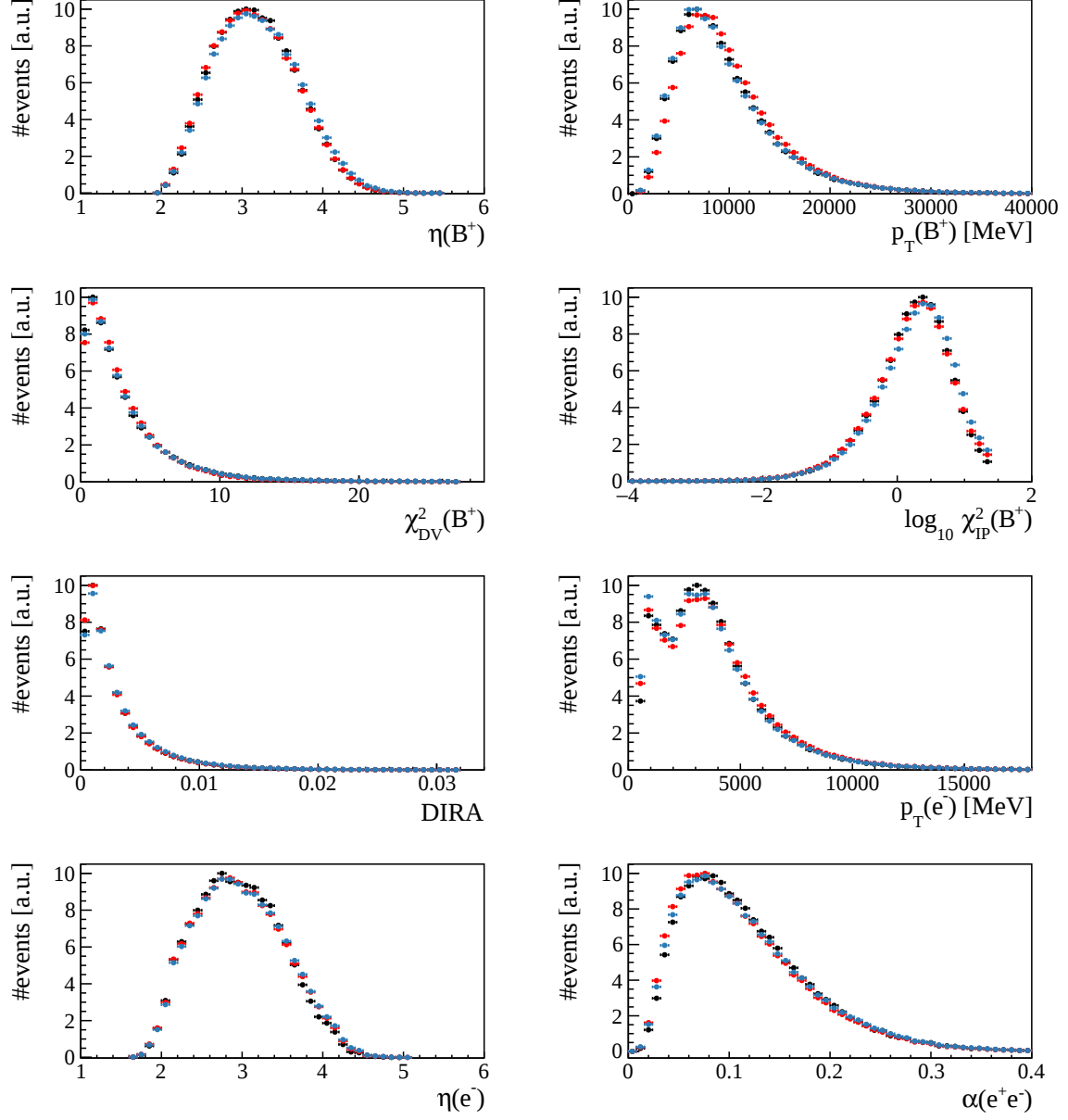


Figure 7.14: Kinematic distributions for preselected $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the ϵ TOS category and in the Run 2 data-taking conditions. Each plot shows the background-subtracted data (black), simulation (red) and simulation with kinematic correction weights (blue).

7.8 q^2 and B mass resolution in the electron modes

In the simulation, the dielectron invariant mass has a $\sim 10\%$ better resolution than in the data. For the rare mode, the fraction of events migrating in and out of the true q^2 range $1.1 < q_{\text{true}}^2 < 6.0 \text{ GeV}^2$ is then underestimated in the simulation sample and this impacts on the B mass lineshape, meaning that the efficiency of the mass window cut is overestimated when computed from simulated events. This difference in resolution is expected for two main reasons. First, differences in the material budget of the detector between data and simulation impact on the amount of bremsstrahlung emitted, which in turn impacts on the momentum resolution. Second, the poor description of the occupancy and of the calorimeter response impacts on the bremsstrahlung recovery, which influences the mass shape.

In order to study the resolution, the $m(e^+e^-)$ distribution is fitted in the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ control mode both in data and in simulation, in each of the trigger and bremsstrahlung categories separately. The full selection is applied to the samples used in the fit, adding a cut on the constrained B mass, $5.20 < m_{\text{DTF}}^{J/\psi}(K^+e^+e^-) < 5.68 \text{ GeV}$, to reduce the partially-reconstructed backgrounds in the data. In the simulation, $m(e^+e^-)$ is fitted using a sum of two Gaussian functions. One of the Gaussian functions has a power-law tail to the left, and the other to the right. The same model is used for the $m(K^+e^+e^-)$ fit described more precisely in Chpt. 8. The shape obtained from the simulation is then used to fit the data, allowing some parameters to float. The width and mean are allowed to float using a width scale factor s_σ and a mean shift parameter $\Delta\mu$, which relates the width and mean of the distribution observed in the data to that of the simulation,

$$\mu_{\text{data}} = \mu_{\text{sim}} + \Delta\mu, \quad \sigma_{\text{data}} = s_\sigma \cdot \sigma_{\text{sim}}. \quad (7.13)$$

where μ_{sim} (σ_{sim}) represents the mean (width) of the distribution obtained from the fit to the simulation and μ_{data} (σ_{data}) the mean (width) of the distribution obtained in the fit to the real data. In the bremsstrahlung categories 1γ and 2γ , the upper power law tail is also allowed to float using one extra free parameter. In the data, the residual combinatorial component is modelled using an exponential function. The fits in the Run 2 data-taking conditions and in the $e\text{TOS}$ trigger category, in all three bremsstrahlung categories combined, are shown in Fig. 7.15.

In each bremsstrahlung category, the averages of s_σ , $\Delta\mu$ and μ_{sim} among the three trigger categories, denoted $\overline{s_\sigma}$, $\overline{\Delta\mu}$ and $\overline{\mu_{\text{sim}}}$, are computed. In all the simulation samples,

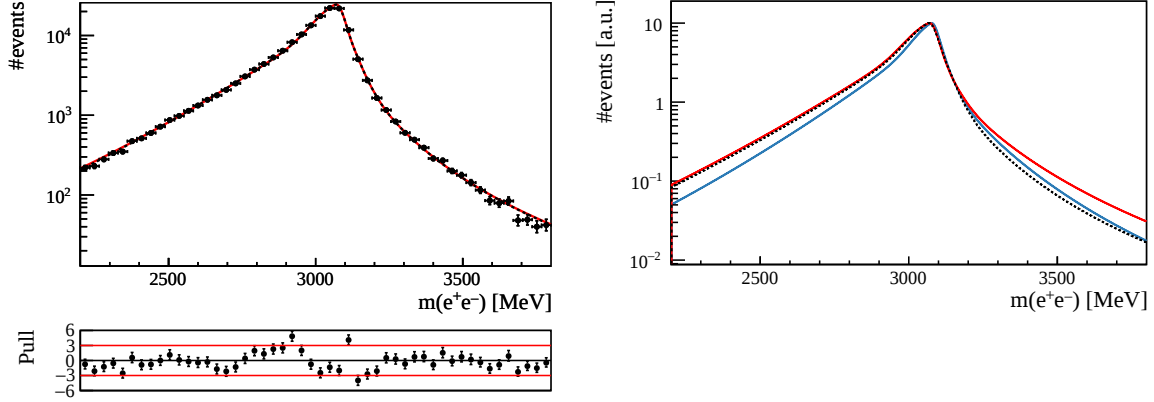


Figure 7.15: Fit to the mass $m(e^+e^-)$ for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ data events (left). The data points are shown in black, the total fit function is shown with a red line and the signal component with a dotted black line. The combinatorial background component is too low to be visible. The fit is performed in all bremsstrahlung categories separately, but the figure shows the three categories combined. Fit functions to the $m(e^+e^-)$ distributions (right) in data (dotted black, same as the dotted black shape on the left plot), simulation with no smearing (blue) and simulation after smearing (red). Both plots are made using events in the e TOS category and Run 2 data-taking conditions.

$m(e^+e^-)$ — and therefore q^2 — is smeared to obtain m^{smeared} , in such a way that if m follows a Gaussian distribution of width σ and mean μ , m^{smeared} will follow a Gaussian distribution of width $\overline{s_\sigma} \cdot \sigma$ and mean $\mu + \overline{\Delta\mu}$. The width scale factors vary mildly across bremsstrahlung categories and are all close to $\overline{s_\sigma} \approx 1.1$. The shift factor $\overline{\Delta\mu}$ varies between -10 and 10 MeV depending on the bremsstrahlung category. Fig. 7.15 right shows the fitted shapes for $m_{\text{sim}}(e^+e^-)$, $m^{\text{smeared}}(e^+e^-)$ and $\mu^{\text{data}}(e^+e^-)$ on $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events, showing the improved agreement between simulation and data after applying the smearing. As expected, some small discrepancy remains in the description of the right tail, as the smearing does not impact the tail parameters. The systematic related to this discrepancy is computed in Sec. 9.8. The mass $m(K^+e^+e^-)$, the resolution of which is dominated by the electron momentum resolution, is also smeared according to the exact same method and with the same smearing factors.

It is expected that smearing the q^2 distribution has an impact on the fraction of rare events that migrate in and out of the range $1.1 < q^2 < 6.0$ GeV². These fractions, before and after applying the smearing, are shown in Tab. 7.2. It can be seen that although the fraction of events migrating inside the q^2 range is not negligible, the impact of applying the smearing to the simulation is limited, changing the number of events migrating in by only $\sim 1\%$ of the number of events in $1.1 < q^2 < 6.0$ GeV².

Table 7.2: Number of $B^+ \rightarrow K^+ e^+ e^-$ events migrating in and out of the range $1.1 < q^2 < 6.0 \text{ GeV}^2$, expressed as a percentage of the total number of events in the true range $1.1 < q_{\text{true}}^2 < 6.0 \text{ GeV}^2$. These fractions are estimated using simulated events in both Run 1 and Run 2 data-taking conditions, with and without applying the smearing. The first two columns show the fraction of events from the true range $1.1 < q_{\text{true}}^2 < 6.0 \text{ GeV}^2$, that migrate inside the reconstructed range $1.1 < q^2 < 6.0 \text{ GeV}^2$. The third column show the fraction of events inside the true range that stay in the reconstructed range, and the last two columns show the fraction of events that migrate from inside the true range to outside of the reconstructed range.

	up→in [%]	down→in [%]	in→in [%]	in→up [%]	in→down [%]
Run 1					
No smearing	8.00 ± 0.23	0.34 ± 0.05	96.94 ± 0.14	1.31 ± 0.10	1.75 ± 0.11
smearing	7.90 ± 0.23	0.43 ± 0.06	96.83 ± 0.15	1.53 ± 0.10	1.65 ± 0.11
Run 2					
No smearing	7.77 ± 0.18	0.48 ± 0.05	96.86 ± 0.12	1.40 ± 0.08	1.74 ± 0.09
smearing	8.78 ± 0.20	0.36 ± 0.04	96.50 ± 0.13	1.39 ± 0.08	2.11 ± 0.10

7.9 Summary and list of efficiencies

The total efficiencies related to the $B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-)$ and $B^+ \rightarrow K^+ \ell^+ \ell^-$ modes, as well as the fractions f^{q^2} , computed taking into account all corrections as well as the ghosts, are listed in Tab. 7.3. The individual efficiencies defined in Eq. 7.1 are listed in Appendix C. The efficiency difference between the J/ψ mode and the rare mode mostly stems from the fact that the selection for the rare mode contains the q^2 cut $1.1 < q^2 < 6.0 \text{ GeV}^2$. The fractions of $B^+ \rightarrow K^+ \ell^+ \ell^-$ events which, at generator level, are within the true q^2 range $1.1 < q_{\text{true}}^2 < 6.0 \text{ GeV}^2$ are:

$$f^{q^2}(B^+ \rightarrow K^+ e^+ e^-) = 29.72 \pm 0.05 \%, \quad (7.14)$$

$$f^{q^2}(B^+ \rightarrow K^+ \mu^+ \mu^-) = 26.75 \pm 0.04 \%. \quad (7.15)$$

A more complete list of efficiencies, factorised down according to Eq. 7.1, is shown in Appendix C.

Table 7.3: Total efficiencies for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$, $B^+ \rightarrow K^+ e^+e^-$, $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow K^+ \mu^+\mu^-$ decays, as well as efficiency ratios $\varepsilon(B^+ \rightarrow K^+ \ell^+ \ell^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-))$, for the Run 1 and Run 2 data-taking conditions. All efficiencies are displayed in percent. All corrections are taken into account, and the events classified as having ghosts are also accounted for.

	$J/\psi \ \varepsilon \ [\%]$	rare $\varepsilon \ [\%]$	ratio $[\%]$
Run 1			
electron modes			
e TOS	2.4785 ± 0.0096	0.683 ± 0.005	27.55 ± 0.24
h TOS!	0.4087 ± 0.0033	0.2304 ± 0.0028	56.4 ± 0.80
TIS!	0.728 ± 0.005	0.2247 ± 0.0029	30.9 ± 0.40
muon modes			
μ TOS	14.046 ± 0.027	3.261 ± 0.0016	23.22 ± 0.12
Run 2			
electron modes			
e TOS	4.324 ± 0.013	1.230 ± 0.008	28.44 ± 0.21
h TOS!	0.4721 ± 0.0033	0.315 ± 0.004	66.7 ± 0.9
TIS!	1.124 ± 0.006	0.381 ± 0.005	33.9 ± 0.4
muon modes			
μ TOS	14.424 ± 0.028	3.348 ± 0.015	23.21 ± 0.11

8. Yields and fits

As Eq. 4.1 shows, the measurement of R_K requires two kinds of inputs: efficiencies ε , described in Chpt. 7, and yields N that are extracted using fits to the data. The present chapter details how these fits are performed. First, the fit shapes and the fit strategy are introduced in Sec. 8.1. Then, the results of the fit to the resonant modes $B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-)$ and $B^+ \rightarrow K^+ \psi(2S)(\ell^+ \ell^-)$ are shown in Sec. 8.2. The fits to the rare modes are shown together with the final result in Chpt. 11, as the rare mode data are kept blind during the development of the analysis. Only the fit strategy for the rare mode is discussed in this chapter, in Sec. 8.3. The strategy used to combine yields and efficiencies from the different modes and the two runs into a single value for R_K is explained in Sec. 8.4.

8.1 Fit method and shapes

All the fits are performed using an extended maximum likelihood as implemented in the ROOFIT package [109], where the optimisation of the likelihood is based on the MINUIT algorithm [110]. The line shape f^{rt} used to fit the mass distribution in the data can be generically expressed as:

$$f^{rt}(m) = N_{\text{sig}}^{rt} S^{rt}(m) + \sum_i N_{\text{bkg},i}^{rt} B_i^{rt}(m) \quad (8.1)$$

where S^{rt} and B_i^{rt} are probability density functions (PDF) describing the signal and background distributions (the index i runs over all different backgrounds), N_{sig}^{rt} and $N_{\text{bkg},i}^{rt}$ are their associated yields and m is either $m(K^+ \ell^+ \ell^-)$, $m_{\text{DTF}}^{J/\psi}(K^+ \ell^+ \ell^-)$ or $m_{\text{DTF}}^{\psi(2S)}(K^+ \ell^+ \ell^-)$ depending on the fit. The index $r = 1, 2$ indicates whether the fit refers to the Run 1 or Run 2 conditions. The index t runs over the different trigger categories. For the muon modes, only one trigger category is defined (μTOS), in contrast to the three trigger categories defined for those modes with electrons ($e\text{TOS}$, $h\text{TOS!}$ and TIS!). The fit functions f^{rt} , S^{rt} and B_i^{rt} depend on various fit parameters that are omitted in Eq. 8.1.

Additionally for electron channels, the shape of the signal distribution, S^{rt} , depends strongly on the bremsstrahlung category (0γ , 1γ , or 2γ , see Sec. 4.1). Therefore, the signal PDF S^{rt} is expressed as a sum of three shapes, one for each bremsstrahlung category:

$$S^{rt}(m) = r_{0\gamma}^{rt} S_{0\gamma}^{rt}(m) + r_{1\gamma}^{rt} S_{1\gamma}^{rt}(m) + (1 - r_{0\gamma}^{rt} - r_{1\gamma}^{rt}) S_{2\gamma}^{rt}(m) \quad (8.2)$$

In the fit to the data, the values of $r_{0\gamma}^{rt}$ and $r_{1\gamma}^{rt}$, the fractions of events in the 0γ and 1γ categories, are constrained to the values observed in simulation, with an uncertainty of 1% coming from the limited statistics of the simulation sample.

In the following, the shapes used for S^{rt} and B_i^{rt} are described in more detail. Systematic variations of the models assumed are considered in Sec. 9.10.

8.1.1 Background shapes

In all fits to the data, the combinatorial background is fitted with an exponential shape with one free parameter λ ,

$$B_{\text{comb.}}^{rt}(m) \propto e^{-\lambda m}. \quad (8.3)$$

The fits to the electron modes must also take into account partially-reconstructed backgrounds. These are not fitted using an analytical shape, but using an adaptive kernel estimation [111] obtained from the mass distribution of simulated partially-reconstructed events. A kernel function is a sum of many Gaussian functions, one for each simulated event, the width of which is adapted to the local density of simulated events. The shape obtained from simulation is then used to fit the data with a floating yield.

8.1.2 Signal shapes

A natural choice to fit the signal peak in mass is a Gaussian function, the width of which describes the resolution of the detector. This is however insufficient for two reasons. First, radiative energy losses create a tail towards lower masses, particularly prominent in the case of the electron mode, that a Gaussian shape does not take into account. A tail also appears towards higher masses when using the constrained mass m_{DTF} . Such a tail towards higher masses also appears in the electron mode, when the bremsstrahlung recovery algorithm overestimates the amount of energy radiated. Second, the mass resolution is not constant but can vary from one event to another. A minimal solution to this latter problem is to use a sum of several Gaussian shapes. To model the radiative tails, a power law is used. Such a shape with a Gaussian core and a power law tail is referred to as a Crystal Ball function (CB) [112].

For the majority of the fits, a sum of two CB functions with opposite tails has sufficient freedom to model the signal mass peak well. This shape is hereafter referred to as a double CB. Such a shape has eight free parameters. Four describe the core of the function: the mean μ in common for the two Gaussian cores, two widths σ_1 and σ_2 , and the parameter

c_g corresponding to the fraction of area covered by the first Gaussian. The two power-law tails each have two free parameters: the exponent n , and the parameter α , defined such that $\mu + \alpha \sigma$ corresponds to the point at which the distribution switches from the Gaussian core to the power-law tail.

In the case of the m_{DTF} fit to the muon control modes, the double CB shape is insufficient to describe the mass peak accurately. Instead, an Hypatia function [113] is used. The core of this distribution corresponds to an infinite sum of Gaussians, the widths of which follow a hyperbolic distribution. In practice, the Hypatia function has nine free parameters. Each of the two tails are described by the two parameters α and n . The core is described by five parameters, including the mean μ and the root-mean-square σ of the distribution.

In each fit, all the shape parameters of the double CB or Hypatia function are fixed using a fit to the mass peak in simulated signal events. All parameters remain fixed when fitting the data, except for the mean μ and width σ which are allowed to float using scale and shift parameters s_σ and $\Delta\mu$ in the same way as in Eq. 7.13. The parameters s_σ and $\Delta\mu$ are used to take into account the fact that the momentum calibration and the detector resolutions are imperfectly described in the simulation.

8.2 Fit to resonant modes

In order to obtain the yield of the resonant modes $B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-)$ and $B^+ \rightarrow K^+ \psi(2S)(\ell^+ \ell^-)$, a fit is performed to the masses $m_{\text{DTF}}^{J/\psi}(K^+ \ell^+ \ell^-)$ and $m_{\text{DTF}}^{\psi(2S)}(K^+ \ell^+ \ell^-)$.

8.2.1 Fit to $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$

The fit to $m_{\text{DTF}}^{J/\psi}(K^+ \mu^+ \mu^-)$ using data $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ events surviving the full selection is shown in Fig. 8.1 for the Run 2 conditions. The values of the free-floating and constrained parameters are also shown in the figure.

There are two background contributions that contaminate this fit. The first is the combinatorial background, of yield N_{comb} , fitted with an exponential shape. The second is the $B^+ \rightarrow \pi^+ J/\psi(\mu^+ \mu^-)$ decay, which is described with a double CB. The parameters of this shape are fixed from a fit to simulated $B^+ \rightarrow \pi^+ J/\psi(\mu^+ \mu^-)$ events, except the mean and width, which are allowed to float using the same corrective parameters $\Delta\mu$ and s_σ as for the signal. The ratio of yields of the $B^+ \rightarrow \pi^+ J/\psi(\mu^+ \mu^-)$ component to

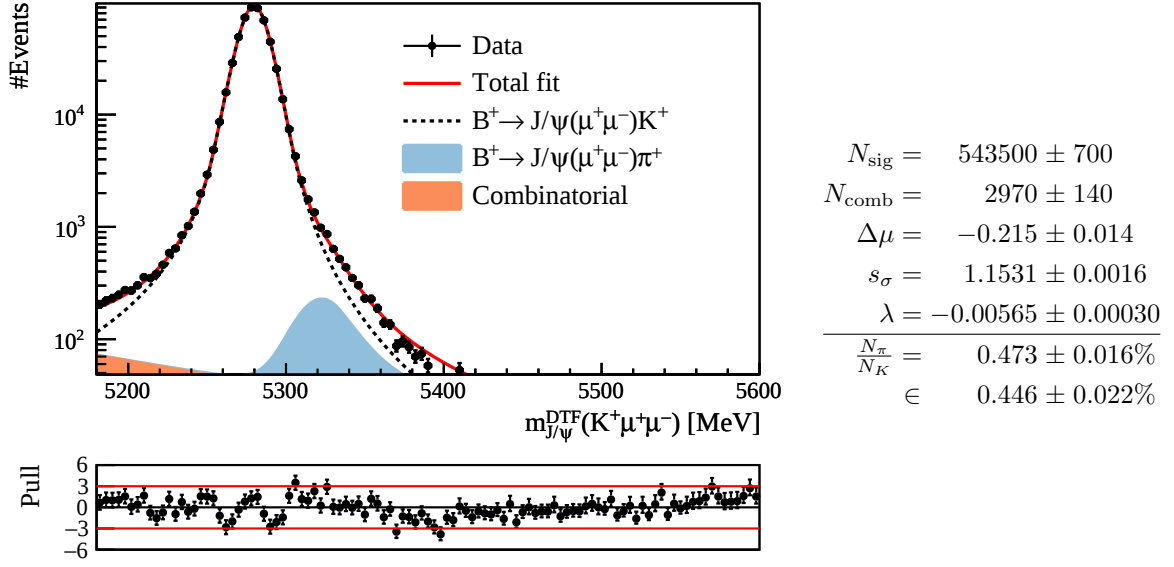


Figure 8.1: Fit to $m_{\text{DTF}}^{J/\psi}$ for $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ events in the Run 2 data, and the pull between the fit function and the data points. The different components taken into account in the fit, shown with different colours, are described in the text. The free-floating parameters are displayed in the table to the right of the figure. Constrained parameters are shown below the line at the bottom of the table, where the notation $\mu \pm \sigma$ indicates the mean (μ) and standard deviation (σ) of the constraint.

the $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ signal is Gaussian-constrained to the product of the ratio of branching fractions [5] and the ratio of efficiencies,

$$\frac{N_\pi}{N_K} = \frac{\mathcal{B}(B^+ \rightarrow \pi^+ J/\psi(\mu^+ \mu^-))}{\mathcal{B}(B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-))} \cdot \frac{\varepsilon_{J/\psi K}}{\varepsilon_{J/\psi \pi}}. \quad (8.4)$$

The width of this constraint is set to the error derived from the branching fraction measurements, which is 10%. This is much larger than the error expected on the ratio of efficiencies.

8.2.2 Fit to $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$

In the case of the fit to $m_{\text{DTF}}^{J/\psi}(K^+ e^+ e^-)$ for $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ events, the signal shape is described as the sum of three distributions, each modelling the mass shape of one of the three bremsstrahlung categories. Each of these three shapes is described by a double CB. Mis-identified $B^+ \rightarrow \pi^+ J/\psi(e^+ e^-)$ events also contribute in this fit range. In the absence of a $B^+ \rightarrow \pi^+ J/\psi(e^+ e^-)$ simulation sample, simulated $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ events are

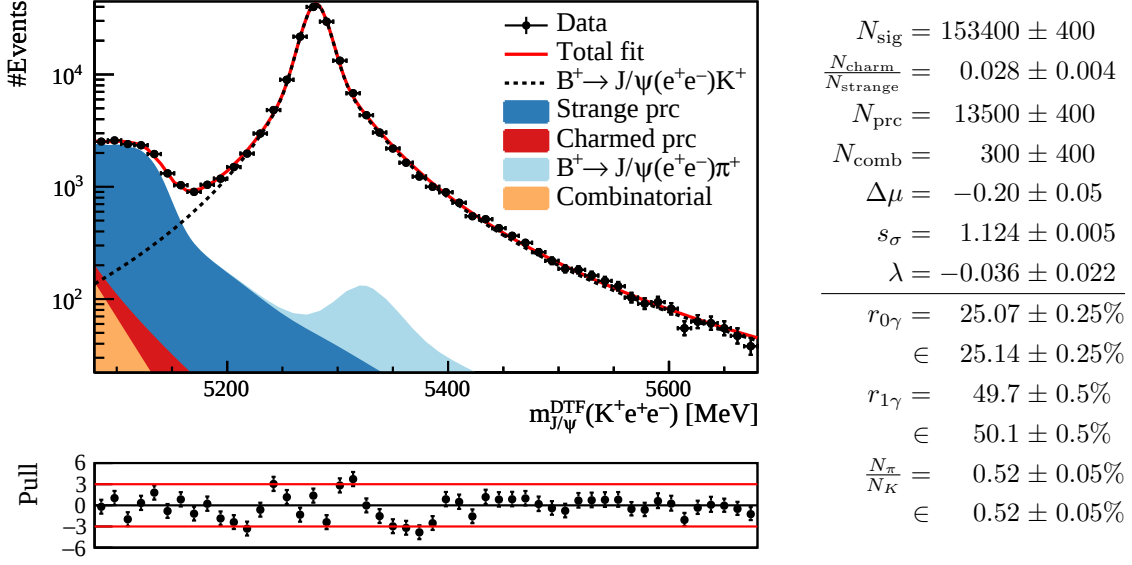


Figure 8.2: Fit to $m_{\text{DTF}}^{J/\psi}$ for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the Run 2 data and selected in the $e\text{TOS}$ trigger category, and the pull between the fit function and the data points. The different components of the fit are indicated in the legend, where ‘prc’ stands for partially-reconstructed background. The free-floating parameters are displayed in the table to the right of the figure. Constrained parameters are shown below the line at the bottom of the table, where the notation $\in \mu \pm \sigma$ indicates the mean (μ) and standard deviation (σ) of the constraint.

used to extract the shape of the background from $B^+ \rightarrow \pi^+ J/\psi(e^+e^-)$ decays with the pion mis-identified as a kaon. The method employed estimates the shift in mass induced by a $\pi \rightarrow K$ mis-identification for each candidate. The mass $m_{\text{DTF}}^{J/\psi}(\pi_{[\leftrightarrow K^+]}^+ e^+ e^-)$ is estimated using

$$m_{\text{DTF}}^{J/\psi}(\pi_{[\leftrightarrow K^+]}^+ e^+ e^-) \approx \left((m_{\text{DTF}}^{J/\psi}(K^+ e^+ e^-))^2 + \frac{E_B}{E_K} \cdot (m_K^2 - m_\pi^2) \right)^{\frac{1}{2}} \quad (8.5)$$

for each $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ event, where E_B and E_K are the energy of the B^+ and K^+ candidates, and m_K and m_π are the mass of the kaon and pion, respectively. More detail on this method is given in Appendix D. The shape of the distribution obtained is fitted using a double CB. The yield of misidentified $B^+ \rightarrow \pi^+ J/\psi(e^+e^-)$ events, denoted N_π is Gaussian-constrained to the product of the ratio of branching fractions [5] and efficiencies. The means and widths of the PDFs describing the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ and $B^+ \rightarrow \pi^+ J/\psi(e^+e^-)$ components are left free to float but common scale and width parameters, $\Delta\mu$ and s_σ , are used.

Due to the worse resolution for electrons, charmed and strange partially-reconstructed

backgrounds (see Sec. 6.10) are relevant in this sample. These background components are modelled by a kernel function, which is made using the $H_b \rightarrow J/\psi X$ simulation sample. The total partially-reconstructed background yield, N_{prc} , and the ratio of strange to charmed component, $N_{\text{charm}}/N_{\text{strange}}$ are free-floating. Finally, the combinatorial shape, with yield N_{comb} , is again described with an exponential function.

The fit to the electron resonant mode $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ in the Run 2 data is shown in Fig. 8.2. The fitted values of all free-floating and constrained parameters are displayed in the same figure.

8.2.3 Fit to $B^+ \rightarrow K^+ \psi(2S)(\mu^+ \mu^-)$

The model used to fit candidates selected in the q^2 region around the $\psi(2S)$, shown in Fig. 8.3 for the Run 2 conditions, is very similar to that used to fit $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$. The fit models only differ in two respects. First, a double CB is used to fit the signal in place of the Hypatia shape. Second, the $B^+ \rightarrow \pi^+ \psi(2S)(\mu^+ \mu^-)$, which has a very small yield, is not included in this fit.

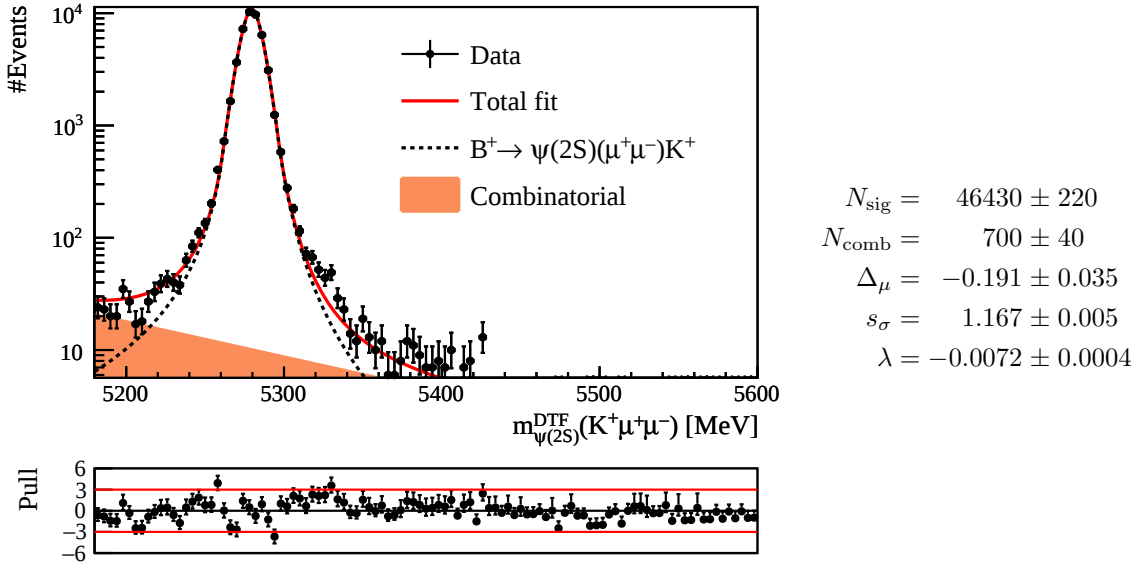
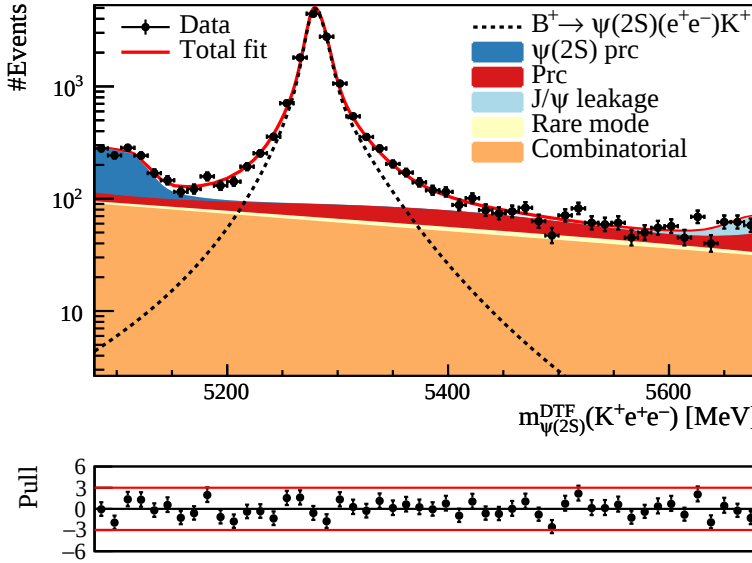


Figure 8.3: Fit to $m_{\text{DTF}}^{\psi(2S)}$ for $B^+ \rightarrow K^+ \psi(2S)(\mu^+ \mu^-)$ events in the Run 2 data, and the pull between the fit function and the data points. The different components of the fit are indicated in the legend, where ‘prc’ stands for partially-reconstructed background. The free-floating parameters are displayed in the table to the right of the figure.

8.2.4 Fit to $B^+ \rightarrow K^+\psi(2S)(e^+e^-)$

The fit to the $B^+ \rightarrow K^+\psi(2S)(e^+e^-)$ mode is shown in Fig. 8.4 for the Run 2 conditions. The values of all free-floating and constrained parameters are shown to the right of the figure. The signal peak is fitted with the same analytical shape used for the $B^+ \rightarrow K^+J/\psi(e^+e^-)$ signal peak, and the combinatorial is again fitted with an exponential function. Due to the lower number of events in this fit compared to the fit for $B^+ \rightarrow K^+J/\psi(e^+e^-)$, the $\Delta\mu$, s_σ parameters are fixed to the values obtained in the fit to $B^+ \rightarrow K^+J/\psi(e^+e^-)$. The parameters $r_{0\gamma}$ and $r_{1\gamma}$ are also fixed to their values from the simulation.

Several peaking backgrounds contribute in the fitted mass range. The first is from partially-reconstructed events of the type $B \rightarrow K^+\psi(2S)(e^+e^-)X$. This is fitted with a kernel function extracted from simulated $B^0 \rightarrow K^{*0}\psi(2S)(e^+e^-)$ events, which are expected to dominate this background. In the fit to the data, the yield of the $B \rightarrow K^+\psi(2S)(e^+e^-)X$ component, $N_{\psi(2S)\text{ prc}}$, is left free-floating.



$N_{\text{sig}} =$	12410 ± 120
$N_{\text{prc}} =$	840 ± 330
$N_{\psi(2S)\text{ prc}} =$	710 ± 50
$N_{\text{comb}} =$	2790 ± 330
$\lambda =$	-0.00178 ± 0.00022
$\frac{N_{\text{rare}}}{N_{\text{sig}}} =$	$1.74 \pm 0.23\%$
$\in$	$1.73 \pm 0.23\%$

Figure 8.4: Fit to $m_{\text{DTF}}^{\psi(2S)}$ for $B^+ \rightarrow K^+\psi(2S)(e^+e^-)$ events in the Run 2 data, and the pull between the fit function and the data points. The different components of the fit are indicated in the legend, where ‘prc’ stands for partially-reconstructed background. The free-floating parameters are displayed in the table to the right of the figure. Constrained parameters are shown below the line at the bottom of the table, where the notation $\in \mu \pm \sigma$ indicates the mean (μ) and standard deviation (σ) of the constraint.

A second peaking background comes from partially-reconstructed events of the type $B \rightarrow H_c(J/\psi(e^+e^-)X)K^+Y$, that migrate into the $\psi(2S)$ q^2 region due to bremsstrahlung

recovery. This background is actually dominated by $B^+ \rightarrow \psi(2S)K^+, \psi(2S) \rightarrow J/\psi X$ decays. This background is fitted with a kernel function extracted from the inclusive $H_b \rightarrow J/\psi X$ simulation sample. The yield of this component, N_{prc} , is also free-floating in the fit to the data.

A third peaking background stems from $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events that leak into the $\psi(2S)$ q^2 range. Such events are fitted using a kernel function extracted from $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulated events reconstructed in the $\psi(2S)$ q^2 range. The yield of this component $N_{J/\psi \text{ leak}}$ is fixed using the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ yield obtained from the fit shown in Fig. 8.2, corrected with the ratio of efficiencies for reconstructing $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the J/ψ q^2 range, $\varepsilon(J/\psi \text{ range})$, and in the $\psi(2S)$ q^2 range, $\varepsilon(\psi(2S) \text{ range})$,

$$N_{J/\psi \text{ leak}} = N_{\text{sig}}(B^+ \rightarrow K^+ J/\psi(e^+e^-)) \cdot \frac{\varepsilon(\psi(2S) \text{ range})}{\varepsilon(J/\psi \text{ range})}. \quad (8.6)$$

A fourth broadly peaking background stems from rare decay events in the $\psi(2S)$ q^2 range. The shape for this component is a kernel function extracted from $B^+ \rightarrow K^+ e^+ e^-$ simulated events reconstructed in the $\psi(2S)$ q^2 range. The ratio of this yield to the signal yield, $\frac{N_{\text{rare}}}{N_{\text{sig}}}$, is Gaussian-constrained to the product of the ratio of branching fractions [5] and the ratio of efficiencies. The standard deviation for this constraint is set to the uncertainty on $\mathcal{B}(B^+ \rightarrow K^+ e^+ e^-)$, which is the dominant uncertainty.

8.3 Fit to rare modes

To extract the yields of the rare modes, fits to the unconstrained masses $m(K^+ e^+ e^-)$ and $m(K^+ \mu^+ \mu^-)$ are performed. In order to fix the parameters $\Delta\mu$ and s_σ in this case, a fit to the control modes $B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-)$ is repeated, this time using the unconstrained mass $m(K^+ \ell^+ \ell^-)$ as for the rare mode. The results of the fits to the rare mode are presented in Chpt. 11.

8.3.1 Fit to $B^+ \rightarrow K^+ \mu^+ \mu^-$

The mass fit to $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$, shown in Fig. 8.5, is made in the same way as the constrained mass fit presented in Sec. 8.2.1, except that the signal peak is modelled with a double CB instead of a Hypatia function. The double CB indeed suffices for the purpose of the unconstrained mass fit. The fit quality is not perfect but suffices to set s_σ and $\Delta\mu$ for the rare mode, which is approximately 200 times less abundant.

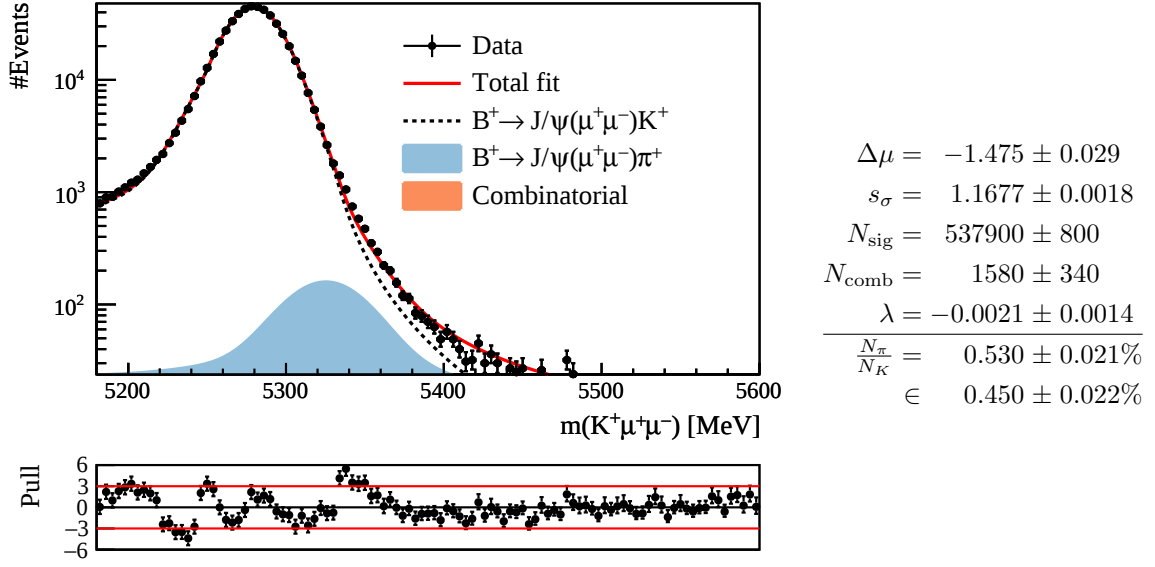


Figure 8.5: Fit to $m(K^+\mu^+\mu^-)$ for $B^+ \rightarrow K^+J/\psi(\mu^+\mu^-)$ events in the Run 2 data, and the pull between the fit function and the data points. The different components of the fit are indicated in the legend. The free-floating parameters are displayed in the table to the right of the figure. Constrained parameters are shown below the line at the bottom of the table, where the notation $\in \mu \pm \sigma$ indicates the mean (μ) and standard deviation (σ) of the constraint.

For the rare mode, the signal peak is also modelled with a double CB, the parameters of which are fixed from a fit to $B^+ \rightarrow K^+\mu^+\mu^-$ simulated events. For the fit to the data, the parameters $\Delta\mu$ and s_σ are fixed to the values obtained with the resonant mode. The only background component is the combinatorial, fitted with an exponential function.

8.3.2 Fit to $B^+ \rightarrow K^+e^+e^-$

The mass fit to $B^+ \rightarrow K^+J/\psi(e^+e^-)$ is shown in Fig. 8.6. The signal is modelled with a sum of three double CB's and, by contrast to the constrained mass fit described in Sec. 8.2.2, the exponent n^R of the right power-law tail is allowed to float with a scaling parameter s_n :

$$n_{\text{data}}^R = s_n \cdot n_{\text{sim}}^R, \quad (8.7)$$

where n_{sim} is the value of the exponent found by performing the fit to simulated events, and n_{data} is the value of the exponent that best fits the data. The disagreement between n_{data}^R and n_{sim}^R is due to the fact that the right power-law tail is due to noise in the ECAL that is interpreted as bremsstrahlung, and the amount of noise depends strongly on the occupancy

of the detector, which is known to be badly modelled in the simulation. In order to improve the precision on s_n , $\Delta\mu$ and s_σ , the partially-reconstructed background is removed by requiring $m_{\text{DTF}}^{J/\psi}(K^+e^+e^-) > 5.2$ GeV in addition to the rest of the selection. Hence, the only backgrounds that are present in this fit are the combinatorial and $B^+ \rightarrow \pi^+ J/\psi(e^+e^-)$ mis-ID, which are modelled in the same way as in the constrained mass fit.

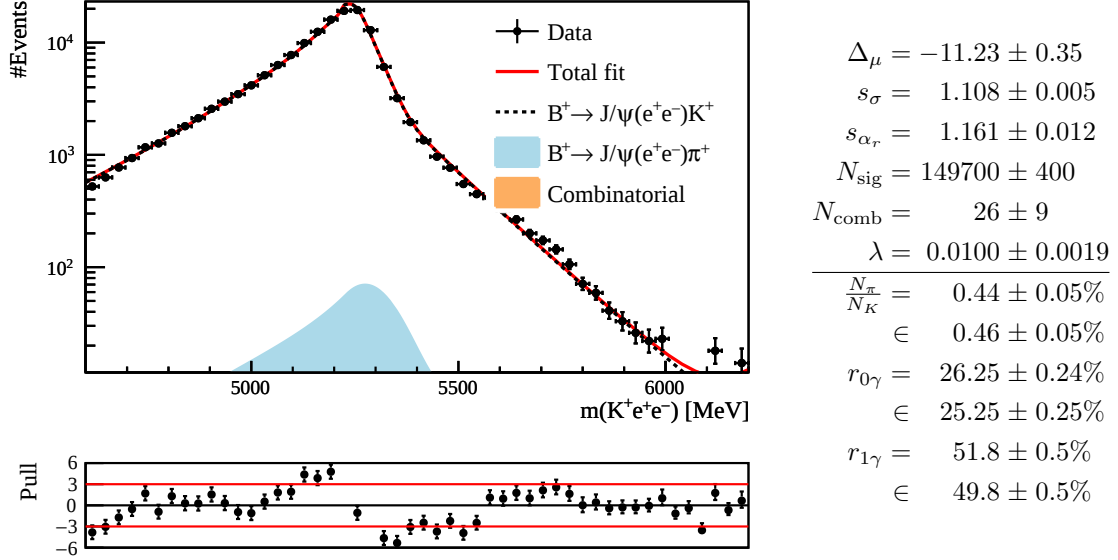


Figure 8.6: Fit to $m(K^+e^+e^-)$ for $B^+ \rightarrow K^+J/\psi(e^+e^-)$ events in the Run 2 data, and the pull between the fit function and the data points. The different components of the fit are indicated in the legend, where ‘prc’ stands for partially-reconstructed background. The free-floating parameters are displayed in the table to the right of the figure. Constrained parameters are shown below the line at the bottom of the table, where the notation $\in \mu \pm \sigma$ indicates the mean (μ) and standard deviation (σ) of the constraint.

For the rare mode, the signal peak is modelled with a sum of three double CB’s, the parameters of which are fixed from a fit to $B^+ \rightarrow K^+e^+e^-$ simulated events. For the fit to the data, the parameters $\Delta\mu$, s_σ and s_n are fixed to the values obtained from the fit to the resonant mode. There are three kinds of backgrounds in this fit. The first background is the partially-reconstructed background, fitted with a kernel function obtained from the $m(K^+e^+e^-)$ distribution in simulated $B^0 \rightarrow K^{*0}e^+e^-$ events. The second background is the combinatorial background, which is modelled with an exponential function. The third background is leakage of $B^+ \rightarrow K^+J/\psi(e^+e^-)$ events, fitted with a kernel function obtained using $B^+ \rightarrow K^+J/\psi(e^+e^-)$ events. The yield of this component $N_{J/\psi \text{ leak}}$ is Gaussian-constrained using the $B^+ \rightarrow K^+J/\psi(e^+e^-)$ yield obtained from the fit shown

in Fig. 8.2, corrected with the ratio of efficiencies for reconstructing $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the J/ψ q^2 range, $\varepsilon(J/\psi \text{ range})$, and in the rare q^2 range $1.1 < q^2 < 6.0 \text{ GeV}^2$,

$$N_{J/\psi \text{ leak}} = N_{\text{sig}}(B^+ \rightarrow K^+ J/\psi(e^+e^-)) \cdot \frac{\varepsilon(\text{rare range})}{\varepsilon(J/\psi \text{ range})}. \quad (8.8)$$

8.4 Extraction of R_K from the fits to the rare modes

The data corresponding to the rare modes are split into eight datasets. There are two $B^+ \rightarrow K^+ \mu^+ \mu^-$ datasets, one for each run. Additionally, there are six $B^+ \rightarrow K^+ e^+ e^-$ datasets, corresponding to each trigger category in each run. Six separate measurements of R_K could in principle be performed, denoted R_K^{rt} with the same rt indices as introduced in Eq. 8.1:

$$R_K^{rt} = \frac{N_{K\mu\mu}^r}{N_{Kee}^{rt}} \cdot c_K^{rt}, \quad (8.9)$$

with

$$c_K^{rt} \equiv \frac{N_{J/\psi ee}^{rt}}{N_{J/\psi\mu\mu}^r} \cdot \frac{\varepsilon_{Kee}^{rt}}{\varepsilon_{K\mu\mu}^r} \cdot \frac{\varepsilon_{J/\psi\mu\mu}^r}{\varepsilon_{J/\psi ee}^{rt}}, \quad (8.10)$$

where $N^{r(t)}$ is the number of events from each B^+ decay observed in the sample corresponding to a specific run and trigger category. The symbol $\varepsilon^{r(t)}$ denotes the total efficiency for each decay to be selected as part of one of these samples. These efficiencies are discussed in Chpt. 7. Swapping the rare mode yields and efficiencies for that of the $B^+ \rightarrow K^+ \psi(2S)(\ell^+ \ell^-)$ mode, six ratios $R_{\psi(2S)}^{rt}$ can be extracted, defining six quantities $c_{\psi(2S)}^{rt}$.

The six R_K^{rt} values must then be combined into a single measurement. The combination method must take into account the fact that the uncertainty on each R_K^{rt} cannot be assumed to be Gaussian distributed, as the yields N^{rt} are small. Moreover, the six measurements of R_K^{rt} are correlated for two reasons. First, they share the same values for the yields and efficiencies of the muon mode. Second, various sources of bias on the efficiencies are correlated, yielding correlated uncertainties between the six values of c_K^{rt} . The variances and covariances on c_K^{rt} are contained in a 6×6 matrix denoted $\mathbf{v}_c$.

In order to take into account all correlations when extracting a single R_K value from the whole data, the eight different datasets are actually not fitted separately. Instead, a combined measurement of R_K is extracted by performing one single simultaneous fit to all rare modes. This means that the yields $N_{K\mu\mu}^r$ and N_{Kee}^{rt} are found by optimising a single

combined likelihood. The log-likelihoods obtained from the invariant mass fits for the electron and muon rare modes, respectively $\log L_e^{rt}$ and $\log L_\mu^r$, read

$$\log L_e^{rt} = \sum_i \log f_e^{rt}(m_{rt}^i | N_{Kee}^{rt}), \quad (8.11)$$

$$\log L_\mu^r = \sum_i \log f_\mu^r(m_r^i | N_{K\mu\mu}^r), \quad (8.12)$$

where the sum runs over all the events in each $r(t)$ sample, $m_{r(t)}^i$ is the mass associated to one event, and $f^{e,\mu}$ are the PDFs used to model them. The background yields or shape parameters are omitted in the above expression. To combine these likelihoods, $f_e^{rt}(m_{rt}^i | N_{Kee}^{rt})$ is first re-parametrised to make it function of R_K , defining h_e^{rt} such that:

$$h_e^{rt}(m_{rt}^i | N_{K\mu\mu}^r, R_K, c_K^{rt}) \equiv f_e^{rt}\left(m_{rt}^i \left| \frac{N_{K\mu\mu}^r \cdot c_K^{rt}}{R_K} \right.\right). \quad (8.13)$$

The combined likelihood L^{tot} then reads

$$\begin{aligned} \log L^{\text{tot}} = & \sum_{rt} \sum_i \log h_e^{rt}(m_{rt}^i | N_{K\mu\mu}^r, R_K, \tilde{c}_K^{rt}) + \sum_r \sum_i \log f_\mu^r(m_i^{rt} | N_{K\mu\mu}^r) \\ & + \log G(\tilde{c}_K^{rt}, c_K^{rt}, \mathbf{v}_c), \end{aligned} \quad (8.14)$$

where G is a six-dimensional Gaussian function that constrains the values of $\tilde{c}_K^{rt}$, floating in the fit, to their expected values c_K^{rt} , within their estimated uncertainties described by the covariance matrix $\mathbf{v}_c$. The expected values of c_K^{rt} are determined using Eq. 8.10.

A supplementary Gaussian constraint is added to $\log L^{\text{tot}}$, which relates the amount of partially-reconstructed backgrounds in the $h\text{TOS!}$ ($t = 1$) and TIS! ($t = 2$) trigger categories to the $e\text{TOS}$ category ($t = 0$),

$$\frac{N_{\text{prc}}^{rt}/N_{\text{sig}}^{rt}}{N_{\text{prc}}^{r0}/N_{\text{sig}}^{r0}} = \frac{\varepsilon_{\text{prc}}^{rt}/\varepsilon_{\text{sig}}^{rt}}{\varepsilon_{\text{prc}}^{r0}/\varepsilon_{\text{sig}}^{r0}} = r_{\text{prc}}^{rt}, \quad t = 1, 2, \quad (8.15)$$

where $\varepsilon_{\text{prc}}^{rt}$ is the total selection efficiency for a partially-reconstructed decay, and $\varepsilon_{\text{sig}}^{rt}$ the total selection efficiency for a signal $B^+ \rightarrow K^+ e^+ e^-$ decay. The ratios r_{prc}^{rt} are determined from simulated events, using $B^0 \rightarrow K^{*0} e^+ e^-$ as a proxy for partially-reconstructed decays.

The values of the parameters c_K^{rt} , $c_{\psi(2S)}^{rt}$, and r_{prc}^{rt} are shown in Tab. 8.1. The uncertainties on these parameters are discussed in Chpt. 9.

Table 8.1: Values of the parameters c_K^{rt} and $c_{\psi(2S)}^{rt}$ used to extract R_K and $R_{\psi(2S)}$ (top table), and constraint on the partially-reconstructed background fractions r_{prc}^{rt} (bottom tables). All parameters are shown for both runs and all trigger categories.

	Run 1			Run 2		
	$e\text{TOS}$	$h\text{TOS!}$	TIS!	$e\text{TOS}$	$h\text{TOS!}$	TIS!
c_K^{rt}	0.1455	0.0501	0.0540	0.2809	0.0710	0.0806
$c_{\psi(2S)}^{rt}$	0.1357	0.0069	0.0362	0.2503	0.0083	0.0509

	Run 1		Run 2	
	$\frac{h\text{TOS!}}{e\text{TOS}}$	$\frac{\text{TIS!}}{e\text{TOS}}$	$\frac{h\text{TOS!}}{e\text{TOS}}$	$\frac{\text{TIS!}}{e\text{TOS}}$
r_{prc}^{rt}	0.91	1.07	0.72	0.93

8.5 Statistical and systematic uncertainties on R_K , $R_{\psi(2S)}$ and $r_{J/\psi}$

The most likely value for R_K , $\hat{R}_K$, is that which minimises the likelihood in Eq. 8.14. A 68% confidence interval on R_K is constructed from the likelihood using Wilks' theorem [114]: L^{tot} is minimised for a range of R_K values around the minimum, and the 68% confidence interval contains the values of R_K for which $\log L^{\text{tot}}(R_K) - \log L^{\text{tot}}(\hat{R}_K) < \frac{1}{2}$. The confidence interval takes into account the statistical fluctuations of the rare modes, together with the correlated uncertainties on c_K^{rt} . There are however two additional systematic uncertainties that are not taken into account by this method. The first one is the uncertainty σ_{shape} on $\frac{N_{K\mu\mu}^r}{N_{Kee}^{rt}}$ related to the fit shapes used to describe the signal and background components of the rare mode. The second one is the fit bias, which occurs if the minimum of the likelihood is systematically biased from the true value of R_K . The total uncertainty related to the fit shapes and the fit bias, σ_{fit} , is added after the likelihood minimisation by smearing $L^{\text{tot}}(R_K)$ with a Gaussian function of width σ_{fit} .

By convention, the uncertainty on R_K arising from the statistical fluctuations affecting $\frac{N_{K\mu\mu}^r}{N_{Kee}^{rt}}$ is referred to as ‘statistical uncertainty’. All the uncertainties on c_K^{rt} (represented by the covariance matrix $\mathbf{v}_c$), as well as σ_{fit} related to the fit shape and fit bias that affect $\frac{N_{K\mu\mu}^r}{N_{Kee}^{rt}}$, are referred to as ‘systematic uncertainties’. The systematic uncertainties include the statistical uncertainties on the yields $N^r(B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-))$, $N^{rt}(B^+ \rightarrow$

$K^+ J/\psi(e^+e^-)$), because these are correlated to the systematic uncertainties, given that the majority of the calibration histograms are extracted from the J/ψ decay modes.

Like R_K , the single ratio $r_{J/\psi}$ is also extracted in several trigger and run conditions yielding six values for $r_{J/\psi}$ denoted $r_{J/\psi}^i$. A covariance matrix $\mathbf{v}^{J/\psi}$ describing the uncertainty on the six $r_{J/\psi}^i$ values is constructed. The combined ratio $r_{J/\psi}$ is then computed as the weighted average of the six $r_{J/\psi}^i$ results,

$$r_{J/\psi} = \frac{\sum_{ij} (\mathbf{v}^{J/\psi})^{-1}_{ij} r_{J/\psi}^i}{\sum_{ij} (\mathbf{v}^{J/\psi})^{-1}_{ij}}. \quad (8.16)$$

Note that the combination is not done with a simultaneous fit, as in the case of R_K , because the data that is fitted to extract the $B^+ \rightarrow K^+ J/\psi$ yields are correlated with the data used to extract the corrections to the efficiencies.

The $R_{\psi(2S)}$ double ratio could be combined from its six independent values $R_{\psi(2S)}^i$ in the exact same way as for R_K . However, in the interest of simplicity, a simultaneous fit to the six $B^+ \rightarrow K^+ \psi(2S)(e^+e^-)$ and two $B^+ \rightarrow K^+ \psi(2S)(\mu^+\mu^-)$ peaks is not performed. Instead, the eight fits are performed independently and a covariance matrix $\mathbf{v}_{\text{stat}}^{\psi(2S)}$ is constructed, which describes the statistical uncertainties on the $B^+ \rightarrow K^+ \psi(2S)(e^+e^-)$ and $B^+ \rightarrow K^+ \psi(2S)(\mu^+\mu^-)$ yields in the different data-taking conditions. The combined $R_{\psi(2S)}$ result is then extracted using

$$R_{\psi(2S)} = \frac{\sum_{ij} (\mathbf{v}_{\text{tot}}^{\psi(2S)})^{-1}_{ij} R_{\psi(2S)}^i}{\sum_{ij} (\mathbf{v}_{\text{tot}}^{\psi(2S)})^{-1}_{ij}}, \quad (8.17)$$

where $\mathbf{v}_{\text{tot}}^{\psi(2S)} = \mathbf{v}_{\text{stat}}^{\psi(2S)} + \mathbf{v}_c^{\psi(2S)}$, $\mathbf{v}_c^{\psi(2S)}$ describes the systematic uncertainties on $c_{\psi(2S)}^t$ and is equivalent to $\mathbf{v}_c$ for R_K . The systematic uncertainty related to the fit shape is not taken into account in the $R_{\psi(2S)}$ ratio.

8.6 Fit validation and expected statistical precision

Before unblinding the data, the expected signal and background yields are estimated using the appropriate product of efficiencies and branching fractions. Moreover, the $m(K^+\ell^+\ell^-)$ shapes of the different components are known from simulated events. In the case of the combinatorial background, the expected yield and shape are estimated using the upper mass sideband, in the same way used to optimise the BDT in Sec. 6.9. From these estimates, the total number of events expected in the $m(K^+\mu^+\mu^-)$ and $m(K^+e^+e^-)$ fit window is

computed assuming a value R_K^{gen} for R_K , and their total distribution is determined. The expected distribution is used to generate 5000 pseudo-experiments which are each fitted to determine R_K . Each fit yields an estimate $\hat{R}_K$ for R_K , and an estimate of the statistical uncertainty denoted $\sigma(R_K)$.

Tab. 8.2 shows the mean of $\hat{R}_K$ in three different scenarios where R_K^{gen} is either set to 0.745, 1.0 or 1.25. The fit bias expected is given by $\langle \hat{R}_K \rangle - R_K^{\text{gen}}$, where the mean is computed over all the pseudo-experiments. The fit bias is at the level of 2 – 5% of the statistical uncertainty, depending on the value of R_K^{gen} . The exact bias therefore needs to be determined after unblinding, when the signal and background yields are known. The distribution of the pull of $\hat{R}_K$ with respect to R_K^{gen} indicates that the statistical uncertainty estimated by the fit has the correct coverage. This demonstrates the robustness of the fit method.

Table 8.2: Expected sensitivity on the determination of R_K , obtained from a fit to randomly generated pseudo-datasets with different generated hypotheses for the true value of R_K , R_K^{gen} . The relative error and the pull mean are shown in the third and last column. The results of the previous LHCb analysis [42] are given for comparison.

R_K^{gen}	$\langle \hat{R}_K \rangle \pm \langle \sigma(R_K) \rangle$	$\left\langle \frac{\sigma(R_K)}{\hat{R}_K} \right\rangle$	$\left\langle \frac{\hat{R}_K - R_K^{\text{gen}}}{\sigma(R_K)} \right\rangle$
1.000	1.001 ± 0.070	7.0%	−4.0%
0.745	0.747 ± 0.046	6.2%	−5.2%
1.250	1.255 ± 0.098	7.8%	−4.6%
Ref. [42]	$0.745^{+0.090}_{-0.074}$	11.0%	—

9. Systematic uncertainties

The systematic uncertainties and their correlations are presented in this chapter. The following paragraphs introduce the methods and define the quantities common to all systematic uncertainty calculations.

To illustrate how systematic uncertainties are computed, consider an efficiency computed in the two run conditions. The true value of the efficiency is ε_1 for Run 1 and ε_2 for Run 2, and these efficiencies are estimated in the nominal way to be $\bar{\varepsilon}_1$ for Run 1 and $\bar{\varepsilon}_2$ for Run 2. They are also computed in n alternative ways, taking n alternative values ε_1^i and ε_2^i , $1 \leq i \leq n$. The variances of ε_1 and ε_2 are denoted v_1 and v_2 and are computed using

$$v_1 = \frac{1}{n} \sum_{i=1}^n (\varepsilon_1^i - \bar{\varepsilon}_1)^2, \quad v_2 = \frac{1}{n} \sum_{i=1}^n (\varepsilon_2^i - \bar{\varepsilon}_2)^2. \quad (9.1)$$

The systematic uncertainties on ε_1 and ε_2 are denoted σ_1 and σ_2 and are defined as the square root of these variances.

In some situations, it is known whether a source of uncertainty induces a bias on ε_1 and a bias on ε_2 in a fully correlated way, or in an uncorrelated way. If the level of correlation is not known, the covariance, $\text{cov}_{1,2}$, between ε_1 and ε_2 is computed using

$$\text{cov}_{1,2} = \frac{1}{n} \sum_{i=1}^n (\varepsilon_1^i - \bar{\varepsilon}_1) \cdot (\varepsilon_2^i - \bar{\varepsilon}_2), \quad (9.2)$$

and the correlation factor, $\text{corr}_{1,2}$, between ε_1 and ε_2 is computed using

$$\text{corr}_{1,2} = \frac{\text{cov}_{1,2}}{\sigma_1 \sigma_2}. \quad (9.3)$$

The above quantities can be organised into a covariance matrix $\mathbf{v}$ or a fractional error matrix $\boldsymbol{\sigma}$, which is here defined as a matrix containing relative uncertainties on the diagonal and correlations otherwise. In the case of ε_1 and ε_2 , $\mathbf{v}$ and $\boldsymbol{\sigma}$ are

$$\mathbf{v} = \begin{pmatrix} v_1 & \text{cov}_{1,2} \\ \text{cov}_{1,2} & v_2 \end{pmatrix}, \quad \boldsymbol{\sigma} = \begin{pmatrix} \sigma_1/\bar{\varepsilon}_1 & \text{corr}_{1,2} \\ \text{corr}_{1,2} & \sigma_2/\bar{\varepsilon}_2 \end{pmatrix}. \quad (9.4)$$

When there exist several sources of biases, the combined covariance matrix is obtained by adding the individual covariance matrices from each source.

In this analysis, the covariance matrices of interest are the 6×6 matrices $\mathbf{v}_c$, $\mathbf{v}_c^{\psi(2S)}$ and $\mathbf{v}^{J/\psi}$ introduced in Sec. 8.4 and 8.5, which affect the three ratios R_K , $R_{\psi(2S)}$ and $r_{J/\psi}$. To ease their readability, they are written as 6×6 fractional error matrices, $\boldsymbol{\sigma}_c$, $\boldsymbol{\sigma}_c^{\psi(2S)}$ and $\boldsymbol{\sigma}^{J/\psi}$. These matrices are described in the next section.

9.1 Total systematic uncertainties

The fractional error matrices σ_c , $\sigma_c^{\psi(2S)}$, and $\sigma^{J/\psi}$ on c_K^{rt} , $c_{\psi(2S)}^{rt}$ and $r_{J/\psi}$ are shown in Tab. 9.1. For R_K , σ_c represents all the systematic uncertainties, except that related to the rare mode fit shape and fit bias. For $R_{\psi(2S)}$ and $r_{J/\psi}$, $\sigma_c^{\psi(2S)}$ and $\sigma^{J/\psi}$ correspond to the total systematic uncertainty.

Table 9.1: Fractional error matrices σ_c , $\sigma_c^{\psi(2S)}$, and $\sigma^{J/\psi}$, showing the fractional uncertainties and their correlation factors (in percent). They are obtained by propagating the systematic uncertainty due to all the different sources of bias to c_K^{rt} , $c_{\psi(2S)}^{rt}$ and $r_{J/\psi}$.

$$\begin{aligned}
 \sigma_c &= \begin{pmatrix} & \text{Run 1} & & & \text{Run 2} & & \\ & e\text{TOS} & h\text{TOS!} & \text{TIS!} & e\text{TOS} & h\text{TOS!} & \text{TIS!} \\ 1.63 & 14.34 & 56.55 & 30.96 & -4.03 & 41.33 & \\ & 3.88 & 20.03 & -6.13 & 9.68 & 28.83 & \\ & & 3.03 & 12.07 & -6.26 & 36.86 & \\ & & & 1.35 & 16.38 & 17.48 & \\ & & & & 8.56 & 28.12 & \\ & & & & & 2.37 & \end{pmatrix} \\
 \sigma_c^{\psi(2S)} &= \begin{pmatrix} 2.17 & 40.15 & 10.28 & 8.00 & 10.20 & 10.16 & \\ & 3.89 & 28.72 & -5.38 & 37.69 & 20.03 & \\ & & 2.15 & 2.45 & 14.14 & 12.74 & \\ & & & 1.55 & 17.48 & 46.81 & \\ & & & & 3.60 & 16.58 & \\ & & & & & 2.14 & \end{pmatrix} \\
 \sigma^{J/\psi} &= \begin{pmatrix} 6.03 & -20.14 & -14.92 & 59.69 & 43.55 & 41.02 & \\ & 7.01 & 36.13 & 29.44 & 26.50 & 4.27 & \\ & & 6.95 & 16.28 & -11.31 & 44.32 & \\ & & & 6.08 & 26.17 & 14.70 & \\ & & & & 8.63 & 60.91 & \\ & & & & & 8.37 & \end{pmatrix}
 \end{aligned}$$

The various sources of systematic uncertainties are listed in Tab. 9.2, indicating the

Table 9.2: List of systematic uncertainties on $r_{J/\psi}$, R_K , and $R_{\psi(2S)}$ expressed in percent. Some sources of systematic uncertainty originate from several different contributions, which are indicated in an indented sub-list.

		$\varepsilon(r_{J/\psi})$ [%]	$\varepsilon(R_K)$ [%]	$\varepsilon(R_{\psi(2S)})$ [%]
1.	Calibration samples size	1.81	0.63	0.63
2.	Kinematic reweighting	2.33	0.51	0.31
3.	PID calibration	0.32	0.17	0.09
3a.	Muon and kaon PID	0.08	0.01	0.01
3b.	Electron PID	0.32	0.09	0.17
4.	Trigger calibration	1.69	0.83	0.35
4a.	e TOS	1.14	0.83	0.33
4b.	h TOS	1.50	0.08	0.07
4c.	TIS	0.94	0.04	0.06
4d.	μ TOS	0.25	0.05	0.06
5.	Occupancy proxy	2.15	0.43	0.55
6.	Tracking efficiency	0.24	0.07	0.41
7.	q^2 and mass resolution	0.28	0.27	0.57
7a.	Right tail	0.12	0.05	0.32
7b.	Trigger bias	0.26	0.11	0.12
7c.	Momentum dependence	—	0.24	0.46
8.	Decay model	—	0.36	—
9.	Fit shape	—	1.66	—
9a.	Signal shape	—	1.34	—
9b.	Part. reco. shape.	—	0.98	—
Total		3.62	2.13	1.18

contribution of each source to the total relative uncertainty. The total uncertainty is not exactly equal to the sum in quadrature of all individual contributions, because the naïve addition in quadrature does not take into account the correlations between the different trigger categories and data-taking conditions. Tab. 9.2 also illustrates the power of the double ratio, every uncertainty from each source being smaller on R_K and $R_{\psi(2S)}$ than on $r_{J/\psi}$. The following sections explain how each systematic uncertainty is computed.

9.2 Size of the calibration and simulation samples

The statistical uncertainty related to the finite size of the normalisation channels and the calibration histograms is classified as a systematic uncertainty and is computed in this section. Most of the calibration histograms, for example the trigger efficiency histograms or electron PID calibration histograms discussed in Sec. 7.4 and in Sec. 7.3.2, are computed from the normalisation modes $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ and $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ themselves. The normalisation modes and the calibration histograms are therefore statistically correlated. In order to take the correlations into account when computing the statistical uncertainty, a bootstrapping method is used [115]. To generate a bootstrapped version of the analysis, each data or simulated event is given a weight that is distributed according to a Poisson distribution of unit mean. All the calibration histograms are re-extracted, all efficiencies are re-computed and all the fits to the control modes are re-made using these weights. This procedure is repeated 100 times, hence generating 100 versions of $r_{J/\psi}$, c_K^{rt} , and $c_{\psi(2S)}^{rt}$. The 100 bootstrapped versions of $r_{J/\psi}$, c_K^{rt} , and $c_{\psi(2S)}^{rt}$ are used to extract correlation matrices according to Eq. 9.4.

The uncertainties on the yields $N(B^+ \rightarrow K^+ J/\psi(\ell^+\ell^-))$ computed from the bootstrapping method are compatible with those estimated from the fit performed to obtain this yield. In principle, there is the possibility that the shift and scale factors $\Delta\mu$, s_σ , extracted in Sec. 8.3, are correlated to $N(B^+ \rightarrow K^+ J/\psi(\ell^+\ell^-))$. Consequently, $\Delta\mu$ and s_σ would also be correlated to c_K^{rt} , and the correlation should then be taken into account in the fit to the rare mode. The bootstrapping method is used to extract the correlation factors between c_K^{rt} and $\Delta\mu$, s_σ . The correlations are at a level of 5% and are therefore neglected in the fit to the rare mode. The absence of correlation between the yield and shape factors can be understood from the fact that, after the full selection chain is applied, the control samples are very clean, and the signal yield is almost equal to the total data sample size.

The finite size of the calibration samples results in an uncertainty shown in row 1 of Tab. 9.2. The uncertainty is smaller on the double ratios R_K and $R_{\psi(2S)}$ than on the single ratio $r_{J/\psi}$. This is because the statistical uncertainty of the calibration histograms mostly cancel in the ratio of efficiencies between the rare and J/ψ modes, or between the $\psi(2S)$ and J/ψ modes.

9.3 Kinematic reweighting scheme

As detailed in Sec. 7.7, the distributions for $p_T(B^+)$, $\eta(B^+)$, χ_{DV}^2 and $\chi_{IP}^2(B^+)$ are imperfect in the simulation and are corrected in all simulation samples using $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ events from the data in the μ TOS category. However, this method is expected to generate some biases. For example, if the LOMuon correction as a function of $p_T(\mu)$ is imperfect, this will bias the simulated $p_T(B^+)$ which will in turn bias the kinematic weights. Another possible bias stems from the fact that the discrepancy between data and simulation in the χ_{DV}^2 and $\chi_{IP}^2(B^+)$ distributions can be influenced by bremsstrahlung in the VELO for the electron mode, but not in the muon mode, so that using the muon mode to correct the electron mode might yield inaccuracies.

To study this potential bias, kinematic distributions in the $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ mode and in the $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ mode are studied using various sets of kinematic weights. For the muon mode, two sets of weights are compared:

1. μ TOS weights: these are the nominal weights computed from the $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ sample in the μ TOS trigger category;
2. μ TIS weights: these weights are computed using the $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ sample in the inclusive TIS trigger category.

For the electron mode, the two above sets of weights are used, as well as three others:

1. e TOS weights: these weights are computed from the $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ sample in the e TOS trigger category;
2. e TIS weights: these weights are computed from the $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ sample in the inclusive TIS trigger category;
3. mixed weights: the correction for $p_T(B^+)$, $\eta(B^+)$ and $\chi_{IP}^2(B^+)$ is taken from the $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ sample in the μ TOS trigger category, but the correction for χ_{DV}^2 is taken from the $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ sample in the e TOS trigger category.

Apart from the samples used to compute them, the procedure to subtract the background and compute the weights is exactly the same as that detailed in Sec. 7.7.

Efficiency-corrected yield plots are extracted using the various sets of weights and are shown in Fig. 9.1 for the Run 2 conditions, as the Run 1 equivalent plots show the same trends. These plots correspond to a data over simulation ratio of kinematic distributions,

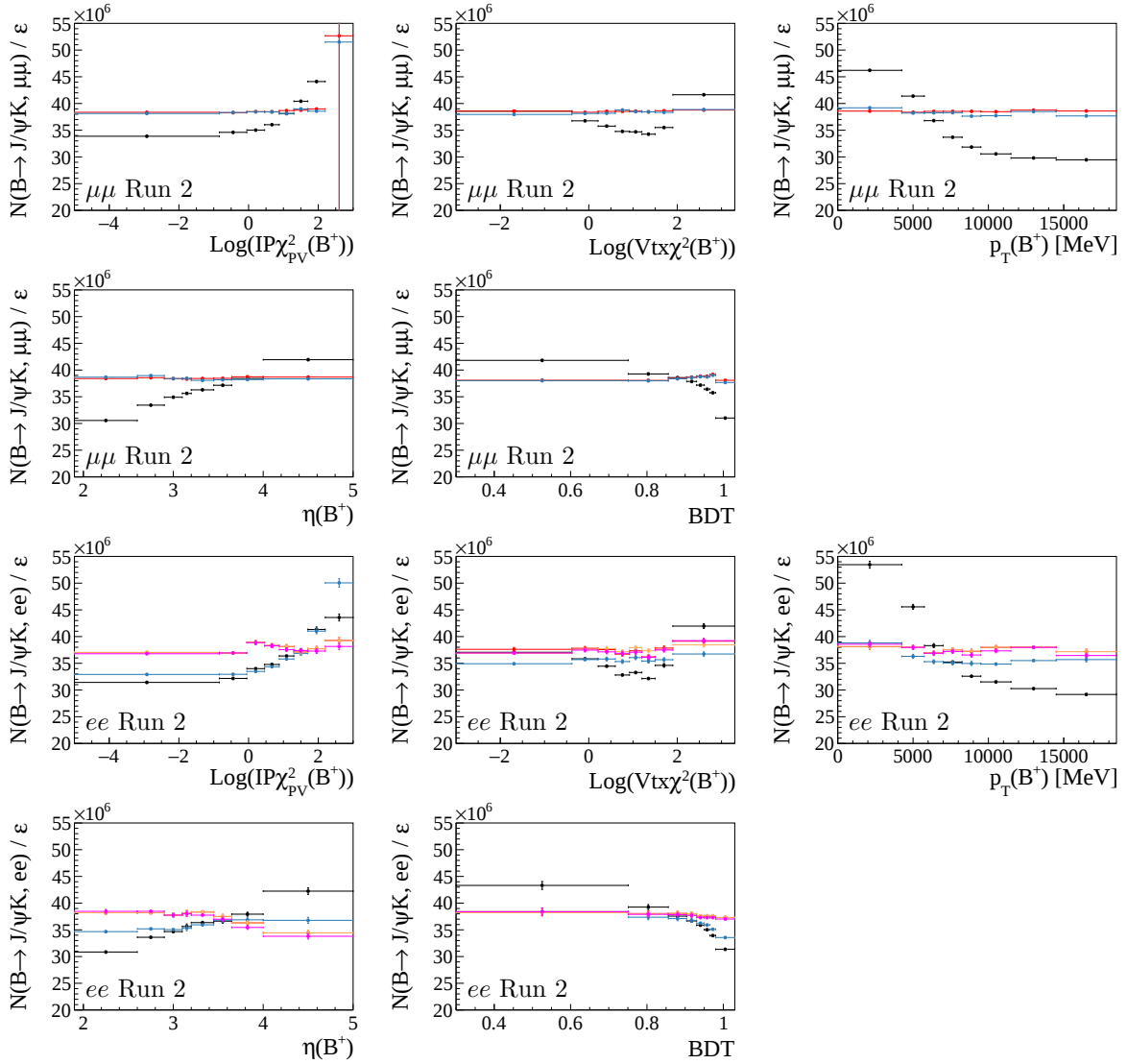


Figure 9.1: Efficiency-corrected yield as a function of several kinematic variables for the $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ mode (top five plots) and the $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ mode (bottom five plots). To compute the efficiencies, various kinematic weights are used and correspond to the different colours: no kinematic weights at all (black), the nominal μ TOS weights (red), μ TIS weights (orange). For the $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ mode, the distribution computed with the e TOS weights (blue) and the mixed weights (magenta) are also shown.

and the procedure used to extract them is detailed in Sec. 10.2. The following observations can be made:

- In both the electron and muon sample, using the μ TOS weights or μ TIS weights makes very little difference;

- In the electron mode, using the e TOS weights improves the data versus simulation agreement with regards to the χ_{DV}^2 distributions, but worsens the $\chi_{\text{IP}}^2(B^+)$ distribution. It also flattens the $\eta(B^+)$ efficiency-corrected yield plot;
- In the electron mode, using the mixed weights yields a very good agreement between data and simulation regarding the χ_{DV}^2 and $\chi_{\text{IP}}^2(B^+)$ distributions, without impacting the $\eta(B^+)$ and $p_{\text{T}}(B^+)$ distributions.

As mentioned above, the agreement between data and simulation in the χ_{DV}^2 distribution can be expected to improve when using weights from the electron mode directly. The $\chi_{\text{IP}}^2(B^+)$ distribution, however, is difficult to correct from the electron mode because this variable is very correlated to the B^+ mass, meaning that subtracting the background tends to bias this distribution.

The efficiency ratios relevant for $r_{J/\psi}$, R_K and $R_{\psi(2S)}$ are computed using the different kinematic weights and are used to extract covariance matrices according to Eq. 9.4. The contribution of the kinematic reweighting uncertainty to the total systematic uncertainty on $r_{J/\psi}$, R_K or $R_{\psi(2S)}$ is reported in row 2 of Tab. 9.2.

9.4 PID calibration

There are three sources of systematic uncertainty related to the PID efficiency calibration:

- The first source relates to the choice of binning scheme for the muon and kaon PID calibration histograms. To estimate an uncertainty, alternative binning schemes are used, which are determined by using the rebinning algorithm described in Sec. 7.3 using different thresholds: either 1.5, 2.0, 3.0 and 3.5 standard deviations, instead of the nominal 2.5 standard deviations.
- The second source is related to the choice of binning scheme for the electron PID histograms. In this case, due to the limited statistics in the electron PID histograms, only one alternative binning is used to assess the systematic uncertainty. This alternative binning is shown in blue (to compare with the nominal binning, shown in red) in Fig. 7.4.
- The last source is related to the trigger bias from the electron and kaon PID efficiencies. A systematic uncertainty is assessed by computing the difference in efficiencies obtained by using the nominal electron PID histogram (where the probe

has potentially triggered any line), or by using a PID histogram extracted by requiring the probe to be TIS with respect to any trigger line. The nominal and alternative histograms are shown in blue and orange in Fig. 7.5.

Covariance matrices corresponding to these three sources are computed and then added up to obtain the total uncertainty related to the PID calibration, which is very small. The contribution of the PID calibration uncertainty to the total systematic uncertainty on $r_{J/\psi}$, R_K or $R_{\psi(2S)}$ is reported in row 3 of Tab. 9.2.

9.5 Trigger calibration

The systematic uncertainties related to the L0 calibration histograms are shown in row 4 of Tab. 9.2. The different trigger categories are impacted by different systematic uncertainties that are discussed in the following sections.

9.5.1 Tag bias impacting the muon, electron and hadron trigger calibration

As discussed in Sec. 7.4, the tag used to extract the L0Muon, L0Electron and L0Hadron histograms introduces a bias on the trigger efficiencies. To evaluate the uncertainty related to this bias, the relevant efficiency ratios for R_K , $r_{J/\psi}$ and $R_{\psi(2S)}$ are computed with different trigger histograms using different tags. These histograms are shown in Fig. 7.6, with the nominal histogram shown in red and the alternative ones used to evaluate the systematic uncertainty in different colours.

For the h TOS! and μ TOS trigger categories, the tag bias is the only source of systematic uncertainty considered. This systematic uncertainty is shown in Tab. 9.2.

9.5.2 Parametrisation of the electron trigger efficiencies

As discussed in Sec. 7.4.3, the L0Electron efficiencies of the two electrons are correlated, and this is ignored in the nominal parametrisation of the electron trigger efficiencies. Therefore, an alternative way to parameterise the total electron trigger efficiency is used: the trigger efficiency is measured in bins of the maximal electron E_T and the distance between the two electrons in the ECAL. The trigger histograms are all extracted using the nominal TIS tag and are shown in Fig. 9.2 for the 2016 data-taking conditions. The histograms show the same behaviour in each of the data-taking conditions.

Covariance matrices are computed according to Eq. 9.4 for both the tag bias and the parametrisation bias, and are combined to get the total fractional error matrices due to the `L0Electron` histograms. The uncertainty related to the `L0Electron` efficiencies is dominated by the tag bias and is shown in Tab. 9.2. The uncertainty on the single ratio $r_{J/\psi}$ is large, at the level of a few percent, but cancels well in the double ratios $R_{\psi(2S)}$ and R_K .

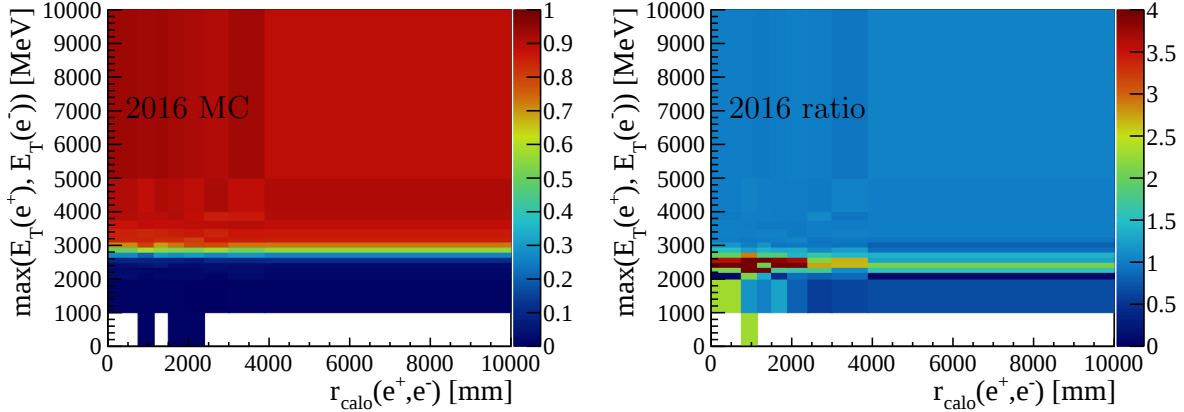


Figure 9.2: Efficiency of either of the two electrons in $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ decays triggering the `L0Electron` line in simulation (left) and the ratio between data and simulation (right). Both figures correspond to the 2016 data-taking conditions. All efficiencies are computed using the nominal TIS tag.

9.5.3 Biases caused by the TIS calibration histograms

As detailed in Sec. 7.4.1, the TIS efficiency depends on whether the signal electrons have triggered or not, hence using an electron tag introduces a large bias on the TIS efficiencies. However, all other possible tags (muon or kaon) introduce an even larger bias. In order to evaluate a systematic uncertainty related to the tag bias on the TIS efficiencies, the origin of this bias is studied. The most likely cause of this bias is the presence of ‘fake TIS’ events in the TIS sample, *i.e.* events triggered due to the presence of the signal electrons but falsely classified as TIS. An example would be those events triggered by the bremsstrahlung photons produced by the signal electrons. Such events can be found when analysing the content of the TIS! category in the $B^+ \rightarrow K^+ e^+ e^-$ simulation. The number of these fake TIS events depends on the kinematics of the signal electrons, and on whether the electrons fired the `L0Electron` trigger or not. Hence, the fake TIS efficiency can be expected to be a function of the p_T of the electrons, rather than a function of the B p_T .

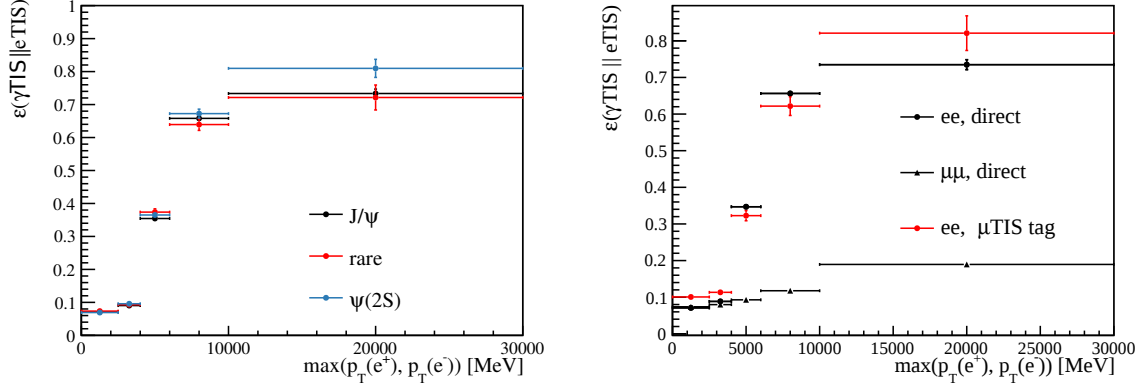


Figure 9.3: Efficiency $\varepsilon(\mu\text{TIS} \parallel h\text{TIS})$ as a function of the maximum electron p_T computed using different simulation samples (left) and requiring different tags (right). Both plots correspond to the 2016 data-taking conditions.

To estimate the impact of fake TIS events, an alternative strategy to correct the TIS efficiency is used. The TIS category is further split into two subcategories. The first corresponds to the situation where a photon or electron in the event has triggered, $\gamma\text{TIS} \parallel e\text{TIS}$, and is expected to contain the majority of the fake TIS events. The second are events triggered by a muon or a hadron, $\mu\text{TIS} \parallel h\text{TIS}$, expected to contain few fake TIS events. The global TIS efficiency is then calculated as

$$\varepsilon(\text{TIS}) = 1 - (1 - \varepsilon(\gamma\text{TIS} \parallel e\text{TIS})) \cdot (1 - \varepsilon(\mu\text{TIS} \parallel h\text{TIS})). \quad (9.5)$$

The efficiencies $\varepsilon(\gamma\text{TIS} \parallel e\text{TIS})$ and $\varepsilon(\mu\text{TIS} \parallel h\text{TIS})$ are measured as detailed below.

Figure 9.3 (left) shows the $\varepsilon(\gamma\text{TIS} \parallel e\text{TIS})$ efficiency as a function of the maximum electron p_T . The left plot shows that the efficiencies computed using $B^+ \rightarrow K^+ J/\psi(e^+e^-)$, $B^+ \rightarrow K^+ \psi(2S)(e^+e^-)$ or $B^+ \rightarrow K^+ e^+e^-$ simulated events are compatible. This demonstrates that this parametrisation allows corrections for $\varepsilon(\gamma\text{TIS} \parallel e\text{TIS})$ to be extracted using the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode, which can then be used for the $B^+ \rightarrow K^+ \psi(2S)(e^+e^-)$ and $B^+ \rightarrow K^+ e^+e^-$ modes. The right plot shows that the efficiencies can be measured using the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events which have been triggered by a muon independent of the signal, *i.e.* a μ TIS tag. This plot also shows a large difference between the efficiencies of the muon and electron modes, owing to the presence of fake TIS events in the electron mode, but not in the muon mode.

As shown in Fig. 9.4, $\varepsilon(\mu\text{TIS} \parallel h\text{TIS})$ in bins of $p_T(B^+)$ obtained for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$, $B^+ \rightarrow K^+ \psi(2S)(e^+e^-)$ and $B^+ \rightarrow K^+ e^+e^-$ simulated events are in good agreement, and $p_T(B^+)$ is therefore the variable used to parametrise this efficiency. In

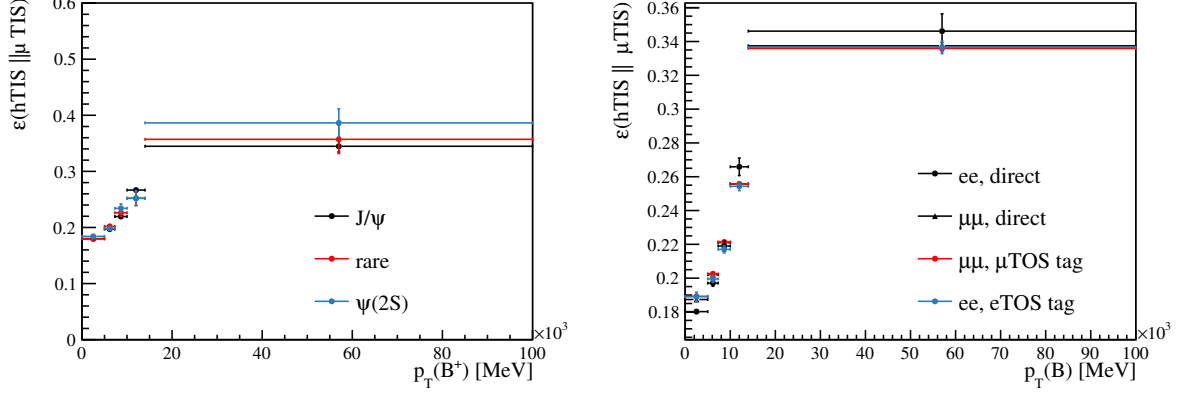


Figure 9.4: Efficiency $\varepsilon(\mu\text{TIS} \parallel h\text{TIS})$ as a function of the $p_T(B^+)$ computed using different simulation samples (left) and requiring different tags (right). Both plots correspond to the 2016 data-taking conditions.

addition, Fig. 9.4 also shows that the efficiency can be measured on $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events that are triggered by a signal electron ($e\text{TOS}$ tag), without introducing a significant bias.

Combining $\varepsilon(\gamma\text{TIS} \parallel e\text{TIS})$ and $\varepsilon(\mu\text{TIS} \parallel h\text{TIS})$ efficiencies, an alternative value for the TIS! efficiency is computed. The difference between this value and the nominal calculation is assigned as a systematic uncertainty and is shown in Tab. 9.2. The impact of the TIS parametrisation on the double ratios $R_{\psi(2S)}$ and R_K is negligible.

9.6 Occupancy proxy

As shown in Fig. 7.12, the occupancy of the detector is poorly described in the simulation. To study how this can impact on the total efficiency, the distributions of three occupancy proxies in the simulation are separately reweighted to match the data. This is done using the same algorithm as described in Sec. 7.7, adding one occupancy proxy into the list of variables that are reweighted. The three occupancy proxies considered are:

1. nSPDHits: the number of hits in the scintillating pad detector;
2. nTracks: the number of tracks found in the detector;
3. nPVs: the number of primary vertices found in the VELO.

To compute the TIS! efficiency, the TIS calibration histograms described in Sec. 7.4.1, which are very sensitive to the occupancy, are also re-computed after making the occupancy

correction. Using these three alternative ways to compute the total efficiencies and Eq. 9.4, systematic uncertainties related to the occupancy are computed for $r_{J/\psi}$, R_K and $R_{\psi(2S)}$ and are shown in row 5 of Tab. 9.2.

9.7 Material budget and tracking efficiency

The rapidity distributions for simulation and data in the Run 2 data-taking conditions are shown in Fig. 9.5. A discrepancy is visible at high rapidity, for all trigger categories. The same discrepancy is observed in the Run 1 conditions. The most likely cause of this discrepancy is a mismatch between the data and simulation material budget as a function of the rapidity: as the electron interaction cross-section with the material of the detector is relatively large, the tracking efficiency of electrons is very dependent on the material budget. Moreover, as discussed in Sec. 3.11, the description of the material in the simulation is known to not match that in the data exactly.

A data-driven way of measuring the electron tracking efficiency has been developed and is described in Ref. [116]. The measured efficiencies are used to correct the simulation, and the efficiency difference using these corrections or not is assigned as a systematic uncertainty on $r_{J/\psi}$. However, at the time of writing, the tracking efficiency measurement in Ref. [116] is still preliminary and has only been performed using the Run 2 data. Therefore, another method is used to assess a systematic on the double ratios R_K and $R_{\psi(2S)}$.

Weights are computed that correspond to data over simulation ratios of two-dimensional $\eta(e^+)$, $\eta(e^-)$ distributions, computed in the e TOS category and in all bremsstrahlung categories separately. The weights are shown in Fig. 9.6. The rapidity distributions for simulated events after applying the correction weights are shown in Fig. 9.5 in orange. The data versus simulation agreement is improved in all categories, except in the Run 1 TIS! category, where the rapidity correction seems to over-correct slightly. This is due to the fact that part of the discrepancy as a function of η is due to the imperfect parametrisation of the L0Electron efficiency. The impact of the rapidity correction weights on a whole range of kinematic variables is discussed further in Sec. 10.2.

The rapidity correction weights can be interpreted as a data over simulation ratio of electron tracking efficiencies, times an unknown normalisation constant k :

$$w(\eta(e^+), \eta(e^-)) = k \cdot \frac{\varepsilon_{\text{data}}^{\text{track}}(\eta(e^+)) \cdot \varepsilon_{\text{data}}^{\text{track}}(\eta(e^-))}{\varepsilon_{\text{sim}}^{\text{track}}(\eta(e^+)) \cdot \varepsilon_{\text{sim}}^{\text{track}}(\eta(e^-))}. \quad (9.6)$$

The unknown normalisation constant k prevents the use of these weights to compute the

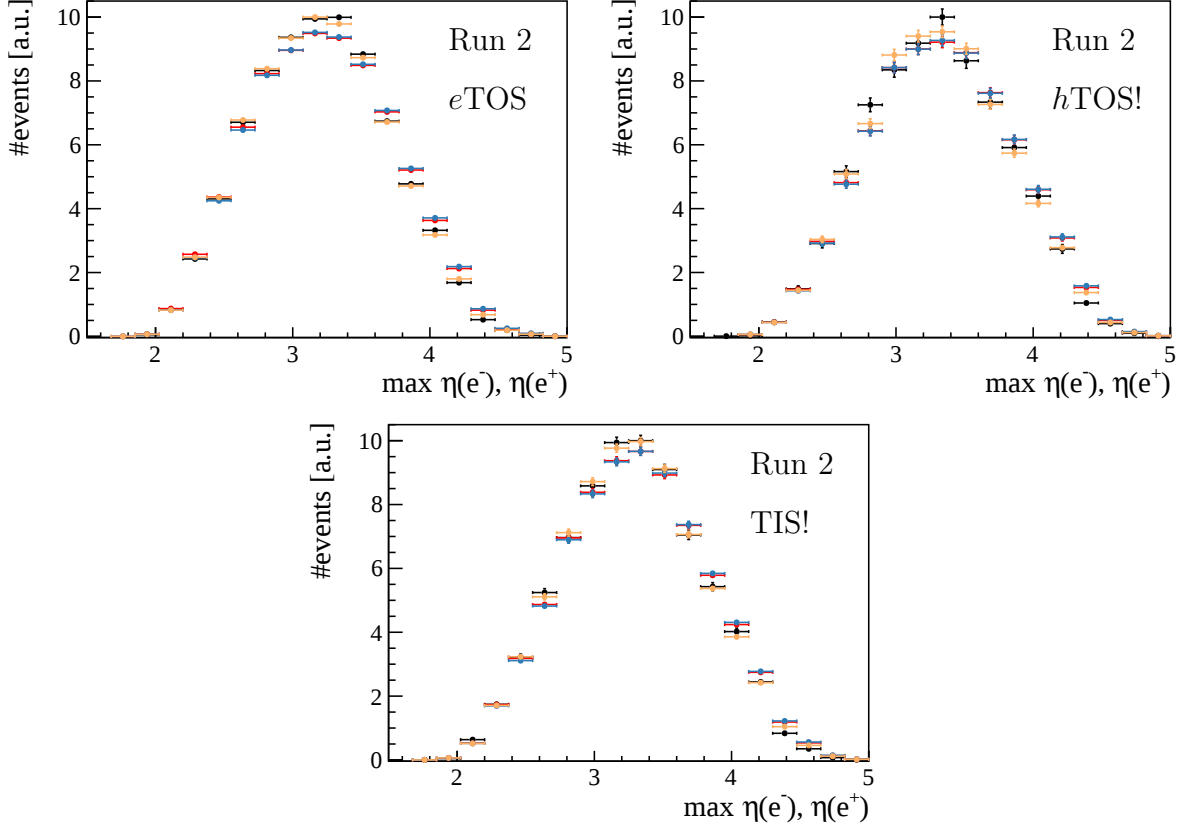


Figure 9.5: Distribution for the maximum electron rapidity for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in all three exclusive categories and in the Run 2 conditions. The black points correspond to background-subtracted data, the red to simulation without kinematic corrections, the blue to simulation with the nominal kinematic correction and the orange to the simulation with the extra rapidity correction.

impact of the rapidity correction weights on the total efficiency $\varepsilon^{\text{tot}}(B^+ \rightarrow K^+ J/\psi(e^+e^-))$, and therefore on $r_{J/\psi}$. However, in the ratio

$$\frac{\varepsilon^{\text{tot}}(B^+ \rightarrow K^+ e^+ e^-)}{\varepsilon^{\text{tot}}(B^+ \rightarrow K^+ J/\psi(e^+ e^-))},$$

the normalisation constant cancels out. Therefore, the systematic uncertainty on R_K related to the mis-modelling of the electron tracking efficiency is calculated as the difference between the ratio of total efficiencies computed with the rapidity correction weights, and without them. The correlation between all trigger categories and data-taking conditions is assumed to be 100%. The fractional uncertainties on R_K and $R_{\psi(2S)}$, shown in Tab. 9.2, are at the sub-percent level, which can be a priori surprising considering the large rapidity corrections. However, the rapidity corrections cancel almost perfectly in the double ratio,

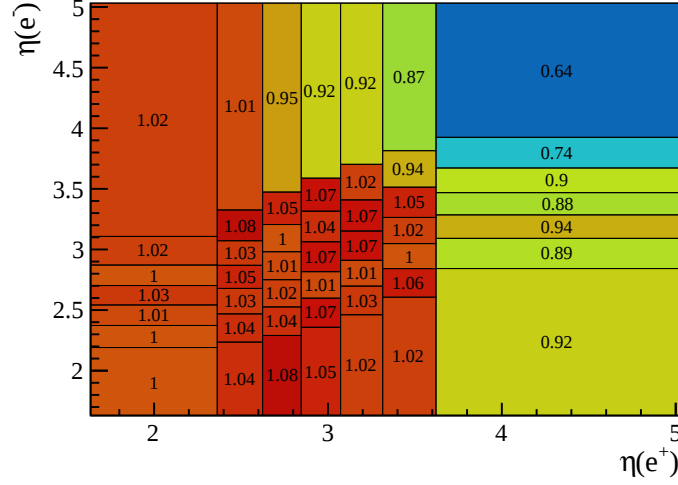


Figure 9.6: Data over simulation ratios of the two-dimensional electron rapidity distributions for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the Run 2 conditions.

due to the fact that the rapidity distributions are very similar in the rare mode, J/ψ mode and $\psi(2S)$ mode (see Fig. 4.3). The uncertainty related to the tracking efficiency is reported in row 6 of Tab. 9.2.

9.8 Electron q^2 and mass resolution

The electron mode q^2 and $m(K^+e^+e^-)$ smearing described in Sec. 7.8 is approximate. Systematic uncertainties related to various effects are described below. These various effects contribute together to a total systematic uncertainty shown in row 7 of Tab. 9.2. This systematic uncertainty does not cancel in the R_K or $R_{\psi(2S)}$ double ratios. It is however one of the sub-dominant uncertainties.

9.8.1 Right tail of the mass and q^2 distributions

As Fig. 7.15 shows, the description of the q^2 and $m(K^+e^+e^-)$ line shape is imperfect even after smearing, due to a residual discrepancy in the upper mass tails, which is related to the bremsstrahlung recovery. To assign a systematic for this residual discrepancy, an alternative method to correct the line shape for simulated events is used, relying on the fact that the size of the upper tail is dependent on the occupancy. The nSPDHits distribution is broken into three bins of uniform width and one $m(e^+e^-)$ shape is extracted

from the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulation in each of these three bins, and is denoted f_{SPD}^i , $1 \leq i \leq 3$. Each f_{SPD}^i is normalised to the fraction of simulated events in the i th nSPDHits bin. The J/ψ peak in $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ data is then fitted using the signal model,

$$N \cdot [(1 + a) \cdot f_{\text{SPD}}^1(m; s_\sigma, \Delta\mu) + f_{\text{SPD}}^2(m; s_\sigma, \Delta\mu) + (1 - a) \cdot f_{\text{SPD}}^3(m; s_\sigma, \Delta\mu)], \quad (9.7)$$

where N denotes the signal yield, s_σ and $\Delta\mu$ are the sigma scale and mean shift parameters, and a is a free parameter which allows the fraction of yields between the three nSPDHits bins to float in the data.

The fit to the Run 2 data obtained by using Eq. 9.7 as the signal shape and an exponential for the background is shown in Fig. 9.7. The signal shape describes the data accurately, including the upper mass tail. The fit favours a value of a very close to 1, which is the maximal value for this parameter, and is equivalent to eliminating all simulated events in the highest occupancy bin.

Hence, $m(K^+e^+e^-)$ and $m(e^+e^-)$ are corrected in the simulation using the result of the fit shown in Fig. 9.7. As for the nominal case, the distributions are smeared using s_σ and $\Delta\mu$ (see Eq. 7.13) but the nSPDHits distribution is also reweighted: the events in the lowest occupancy bin are reweighted by a factor two and the events in the highest occupancy bin are removed. The efficiencies and efficiency ratios relevant for $r_{J/\psi}$, R_K and $R_{\psi(2S)}$ are computed using this alternative smearing method, and the difference with the nominal is assigned as a systematic uncertainty, assuming a 100% correlation between all trigger categories and runs. This systematic uncertainty is shown in Tab. 9.2 and is labelled “right tail”.

9.8.2 Trigger bias on the mass and q^2 resolution

Another source of systematic uncertainty related to the q^2 resolution stems from the fact that the smearing and scaling factors depend mildly on which trigger category is used to perform the fit. The smearing and scaling factors averaged over all three trigger categories are used to smear the simulation in all three trigger categories. The uncertainty on these factors related to the trigger bias is computed as the *RMS* between the factors computed in the three trigger categories. To propagate the uncertainty to the efficiencies or efficiency ratios, the efficiencies are computed using smearing factors modified by plus or minus one standard deviation. This is used to extract a systematic uncertainty labelled “trigger bias” in Tab. 9.2. Again, correlations between all runs and trigger categories are assumed to be 100%.

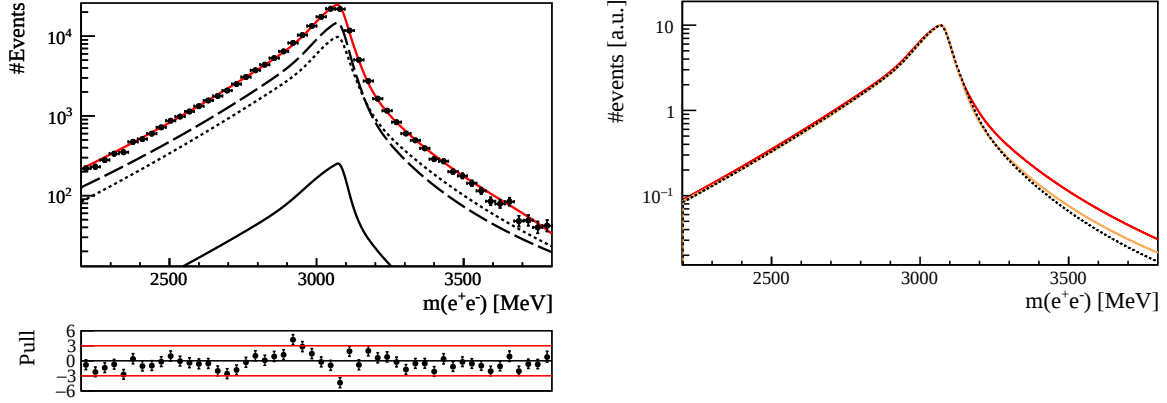


Figure 9.7: Fit to the mass $m(e^+e^-)$ for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ data events (left). The data points are shown in black, the total fit function is shown with a red line. The signal is broken into three components: the solid black line corresponds to the low occupancy bin, the dashed black line to the central occupancy bin, and the dotted black line to the high occupancy bin. The combinatorial background component is too low to be visible. The fit is performed in all bremsstrahlung categories separately, but the figure shows the three categories combined. Fit functions to the $m(e^+e^-)$ distributions (right) in data (dotted black, same as the signal component on the left plot), simulation with the nominal smearing (red) and simulation with the alternative smearing and occupancy reweighting (orange). Both plots are made using events in the $eTOS$ category and Run 2 data-taking conditions.

9.8.3 Momentum dependency of the mass and q^2 resolution

It is possible that the smearing factors depend on q^2 , and that the simulation should be smeared with a different factor at low q^2 than at $q^2 = m(J/\psi)^2$. For the electron mode, fitting the $\psi(2S)$ peak or the ϕ peak is very challenging due to the complicated backgrounds.

Negelecting the mass of both electrons, the di-electron invariant mass can be written as

$$m(e^+e^-) = 2\sqrt{p_{e^+}p_{e^-}} \sin\left(\frac{1}{2}\alpha_{\ell^+\ell^-}\right). \quad (9.8)$$

Simulated events indicate that the resolution on $\alpha_{\ell^+\ell^-}$ is $\sim 0.5\%$, whereas the resolution on p is $\sim 10\%$. Therefore, the resolution on $m(e^+e^-)$ is completely dominated by the momentum resolution itself, which is dependent on the momentum. To assign a systematic on the q^2 -dependence of the smearing factor, the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ channel in both data and simulation is used to measure the smearing constants in 2D bins of $p_{\max} = \max(p_{e^+}, p_{e^-})$, $p_{\min} = \min(p_{e^+}, p_{e^-})$, by performing a fit to the data and simulation in each of these bins. A projection of the data over simulation ratios of the smearing factors

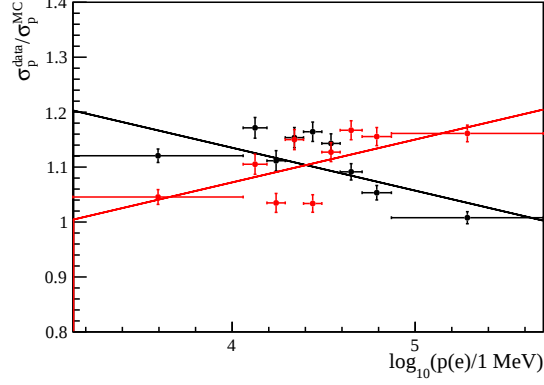


Figure 9.8: Data-over-simulation ratio of $m(e^+e^-)$ resolution for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the Run 2 data-taking conditions. Electrons with (red) and without (black) a bremsstrahlung cluster added are considered. A fit using a linear function is made to help visualise the trend of the momentum dependency.

is shown in Fig. 9.8, and a momentum dependence is visible. The uncertainty on R_K and $R_{\psi(2S)}$ related to the momentum dependence of the q^2 resolution is then taken as the difference between the ratio of efficiencies computed using the nominal momentum-independent smearing constants, and these momentum-dependent smearing constants. This systematic is reported in row 7c of Tab. 9.2. The correlations between all trigger categories and runs are assumed to be 100%.

9.9 Decay model

The uncertainties on the Wilson coefficients, as well as hadronic uncertainties impact on the q^2 distribution in simulated $B^+ \rightarrow K^+ \ell^+ \ell^-$ decays, and hence impact the selection efficiencies. These uncertainties are propagated to an uncertainty on R_K using the theoretical predictions from the flavio software package [117]. The predicted q^2 distributions are shown in Fig. 9.9. The q^2 distribution of the signal is reweighted 100 times to sample the theory uncertainties and the ratio of efficiencies $\varepsilon(B^+ \rightarrow K^+ e^+ e^-)/\varepsilon(B^+ \rightarrow K^+ \mu^+ \mu^-)$ is re-computed each time. This allows a systematic uncertainty on R_K to be extracted, shown in row 8 of Tab. 9.2. The correlations between all runs and triggers are 100%.

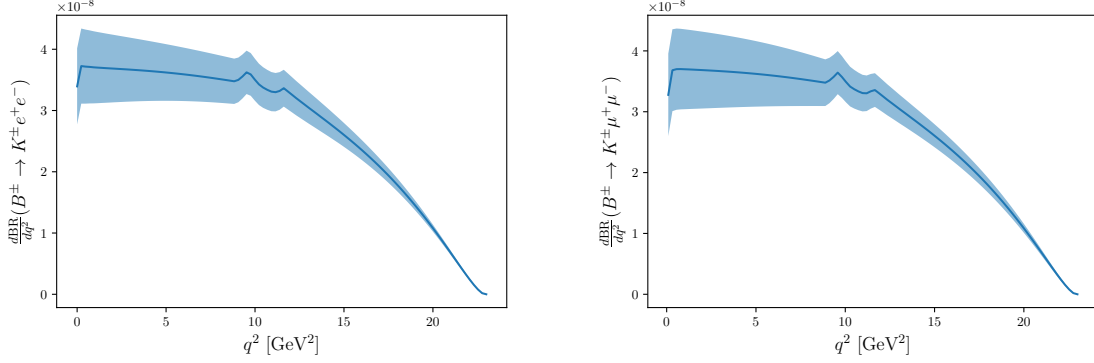


Figure 9.9: Differential branching fractions predictions from the flavio software for $B^+ \rightarrow K^+ e^+ e^-$ (left) and $B^+ \rightarrow K^+ \mu^+ \mu^-$ (right). The blue band shows the theoretical uncertainty that results from an extrapolation of the lattice calculation at high q^2 . The predictions in the region of the narrow charmonium resonances are not meaningful.

9.10 Fit shape

In order to evaluate the systematic uncertainty on R_K arising from the choice of fit model, pseudo-experiments are generated using an alternative model and fitted twice, once with the nominal model and once with the alternative model, to obtain two measurements of R_K . The mean of the difference between the two measurements is assigned as a systematic uncertainty. Alternative shapes are used for both the signal and partially-reconstructed background components. The alternative signal distribution is parameterised as a sum of three Gaussian functions for the 1γ and 2γ categories. In the 0γ category a single Crystal Ball distribution is used. The alternative partially-reconstructed background shape is obtained by reevaluating the kernel estimate after including the contribution from $B^+ \rightarrow K_{1,2}^+ (\rightarrow K^+ \pi^+ \pi^-) e^+ e^-$ decays estimated in Sec. 6.10. The systematic uncertainty related to the signal shape is shown in row 9a of Tab. 9.2, and the one related to the partially-reconstructed background shape in row 9b. By contrast to all the other systematic uncertainties in Tab. 9.2, this fit shape uncertainty affects the rare mode yields, rather than impacting on the efficiencies.

9.11 Partially reconstructed background constraint

In the combined fit to extract R_K , the amount of partially-reconstructed background across the three trigger categories is constrained using the r_{prc}^{rt} ratios defined in Eq. 8.15 and shown in Tab. 8.1. As the r_{prc}^{rt} factors are double ratios of efficiencies related to the electron modes, the systematic uncertainties discussed so far have a negligible effect on them. The dominant systematic uncertainty on r_{prc}^{rt} instead comes from the partially-reconstructed background model, as the TIS! and hTOS! efficiencies are expected to depend on the missing particle (*e.g.* a π^- in $B^0 \rightarrow K^{*0}(\rightarrow K^+\pi^-)e^+e^-$ or a π^0 in $B^+ \rightarrow K^{*+}(\rightarrow K^+\pi^0)e^+e^-$). Alternative values for the r_{prc}^{rt} factors are hence computed using a sample of simulated $B^+ \rightarrow K^{*+}(\rightarrow K^+\pi^0)e^+e^-$ decays with the π^0 not reconstructed. The difference between these values and the nominal ones, obtained from $B \rightarrow K^{*0}(\rightarrow K^+\pi^-)e^+e^-$ simulated decays with the π^- not reconstructed, is assigned as a systematic uncertainty. The fractional error matrix σ_{prc} for the r_{prc}^{rt} factors is shown in Tab. 9.3.

Table 9.3: Fractional error matrix on the r_{prc}^{rt} factors showing the fractional uncertainties and their correlation factors due to differences in efficiency between $B^0 \rightarrow K^{*0}e^+e^-$ and $B^+ \rightarrow K^{*+}e^+e^-$ events. All entries are shown in percent.

$$\sigma_{\text{prc}} = \begin{pmatrix} & \begin{matrix} \text{Run 1} \\ \frac{h\text{TOS!}}{e\text{TOS}} & \frac{\text{TIS!}}{e\text{TOS}} \end{matrix} & \begin{matrix} \text{Run 2} \\ \frac{h\text{TOS!}}{e\text{TOS}} & \frac{\text{TIS!}}{e\text{TOS}} \end{matrix} \\ \begin{matrix} \frac{h\text{TOS!}}{e\text{TOS}} \\ \frac{\text{TIS!}}{e\text{TOS}} \end{matrix} & \begin{matrix} 14.5 & 0.0 \\ & 1.4 \end{matrix} & \begin{matrix} 100 & 100 \\ 0.0 & 100 \\ 14.5 & 0.0 \\ & 1.4 \end{matrix} \end{pmatrix} \quad (9.9)$$

10. Cross-checks

Various cross-checks used to ensure the robustness of the analysis are presented in this chapter.

10.1 Integrated measurement of $r_{J/\psi}$ and $R_{\psi(2S)}$

The integrated $r_{J/\psi}$ single ratio and $R_{\psi(2S)}$ double ratio, which are defined in Eq. 4.1 and in Eq. 4.2, are determined using the fits described in Sec. 8.1 and the efficiencies computed with the nominal trigger, PID and kinematic corrections discussed in Sec. 7. The ratios are computed in Run 1 and Run 2 separately, using the three exclusive trigger categories for the electron mode and the μ TOS category for the muon mode. The ratios are displayed in Tab. 10.1. Statistical and systematic uncertainties are quoted separately for $R_{\psi(2S)}$, but are not separable in $r_{J/\psi}$, as explained in Chpt. 9. The $r_{J/\psi}$ and $R_{\psi(2S)}$ results in the different trigger categories and data-taking conditions are combined using Eq. 8.16 and Eq. 8.17. The combined results are also shown in Tab. 10.1. All individual and combined values are compatible with unity within their uncertainties.

10.2 Differential measurement of $r_{J/\psi}$

To test the data-simulation agreement of one-dimensional projections of kinematic distributions in $B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-)$ decays, the differential $r_{J/\psi}$ test is performed. The test proceeds in the following steps:

1. Using simulated $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ events that satisfy the preselection requirement (without the BDT applied), each kinematic variable is split into eight uniformly populated bins. This gives similar precision in each bin, allowing any trend to be spotted more easily;
2. For both data $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ and $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ events, fits to $m_{\text{DTF}}^{J/\psi}$ are performed to extract the yields of events in each of the eight bins. Two histograms storing the yields are obtained, one for the electron mode and one for the muon mode. They are denoted $h_{N,e}^i$ and $h_{N,\mu}^i$ with $1 \leq i \leq 8$;

Table 10.1: Integrated values of $r_{J/\psi}$ and $R_{\psi(2S)}$ computed for Run 1 and Run 2 data and for the combination of Run 1 and Run 2 data. The muon modes are measured in the μ TOS category and the electron modes are measured in the trigger category shown in the first column. The values of $R_{\psi(2S)}$ are shown with their statistical and systematic uncertainties separated.

	$r_{J/\psi}$	$R_{\psi(2S)}$
Run 1		
e TOS	1.063 ± 0.056	$0.984 \pm 0.014 \pm 0.021$
h TOS!	1.008 ± 0.079	$0.915 \pm 0.052 \pm 0.036$
TIS!	1.015 ± 0.082	$1.048 \pm 0.029 \pm 0.021$
Run 2		
e TOS	1.052 ± 0.074	$0.986 \pm 0.011 \pm 0.016$
h TOS!	1.053 ± 0.094	$0.889 \pm 0.049 \pm 0.029$
TIS!	1.112 ± 0.102	$0.971 \pm 0.023 \pm 0.021$
Combination		
	1.024 ± 0.037	$0.987 \pm 0.008 \pm 0.012$

- Using simulated $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ and $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events, two efficiency histograms are produced. Each of the eight bins of these histograms represents the probability of an event being reconstructed in that particular bin, denoted $h_{\varepsilon,e}^i$ and $h_{\varepsilon,\mu}^i$;
- The entry in each bin of the yield histograms is divided by the efficiency in each corresponding bin in the efficiency histograms to obtain two efficiency-corrected yield histograms $h_{\mathcal{Y},e}^i = h_{N,e}^i/h_{\varepsilon,e}^i$ and $h_{\mathcal{Y},\mu}^i = h_{N,\mu}^i/h_{\varepsilon,\mu}^i$. Each bin of the efficiency-corrected yield histograms represents the total number of $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ or $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ decay that were generated in the detector, before any reconstruction effect. Hence, if the data and simulated distributions of the kinematic variable of interest agree perfectly, the efficiency-corrected yield histograms will have uniformly populated bins;
- Finally, the muon efficiency-corrected yield histogram is divided by the electron efficiency-corrected yield histogram to obtain the $r_{J/\psi}$ histogram $h_{r_{J/\psi}}^i = h_{\mathcal{Y},\mu}^i/h_{\mathcal{Y},e}^i$, each bin of which should be compatible with unity, assuming a perfect agreement between data and simulation.

The simulation, particularly in the case of the electron mode, is however imperfect and some $r_{J/\psi}$ histograms are expected to show some level of variation across the bins. To quantify these variations, the $r_{J/\psi}$ flatness parameter d_f is computed. This parameter can be interpreted as follows: if the deviations from flatness across bins of the efficiency-corrected yield histogram are assumed to be genuine variations of the efficiency from a perfectly flat distribution then, given the distribution of events expected in the rare decay, what would the resulting deviation in the double ratio R_K be? More explicitly, d_f is computed as follows:

$$d_f = \left(\frac{\sum_{i=1}^8 h_{\varepsilon,\mu}^{\text{rare},i} \cdot h_{\mathcal{Y},\mu}^i}{\sum_{i=1}^8 h_{\varepsilon,\mu}^i \cdot h_{\mathcal{Y},\mu}^i} \cdot \frac{\sum_{i=1}^8 h_{\varepsilon,\mu}^i}{\sum_{i=1}^8 h_{\varepsilon,\mu}^{\text{rare},i}} \right) / \left(\frac{\sum_{i=1}^8 h_{\varepsilon,e}^{\text{rare},i} \cdot h_{\mathcal{Y},e}^i}{\sum_{i=1}^8 h_{\varepsilon,e}^i \cdot h_{\mathcal{Y},e}^i} \cdot \frac{\sum_{i=1}^8 h_{\varepsilon,e}^i}{\sum_{i=1}^8 h_{\varepsilon,e}^{\text{rare},i}} \right) - 1. \quad (10.1)$$

For each $r_{J/\psi}$ histogram, d_f must be comparable to the size of the total systematic uncertainty expected on R_K . A larger d_f would indicate an extra source of systematic uncertainty that does not cancel in the ratio $\varepsilon(B^+ \rightarrow K^+ \ell^+ \ell^-) / \varepsilon(B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-))$.

The $r_{J/\psi}$ histograms are shown in Fig. 10.1 (bottom of each row), for the Run 2 conditions, and for the electron mode selected in the e TOS category. Distributions are shown for the transverse momentum of the B^+ and leptons, as well as for variables related to the angles in the decay, *i.e.* rapidity, angle between the two leptons, and angle between the kaon and lepton. The angle θ_ℓ , corresponding to the angle between the two leptons in the B^+ rest frame is also shown. Distributions are also shown that relate to the alignment of the tracks, *i.e.* χ_{DV}^2 and $\chi_{\text{IP}}^2(B^+)$. In each bin, the $r_{J/\psi}$ value is shown with a red thick error bar corresponding to the statistical uncertainty of the fit. A thin black error bar indicates the total statistical and systematic uncertainty. The systematic uncertainty shown in Fig. 10.1 comes from the tracking efficiency corrections (see Sec. 9.7), the L0Electron calibration (see Sec. 9.5) and the kinematic reweighting (see Sec. 9.3). These are the three effects that most significantly impact the shape of the differential $r_{J/\psi}$ histograms in the e TOS category.

Some $r_{J/\psi}$ distributions show some trend: the high rapidity bins or the lowest dilepton angle bin are displaced from 1.0. However, the systematic uncertainty is large in these bins, illustrating the fact that the mis-modelled effects in the simulation (*e.g.* the material at high rapidity) are covered by the systematic uncertainties. All flatness parameters are at the permille level. This again illustrates the robustness of the ratio $\varepsilon(B^+ \rightarrow K^+ \ell^+ \ell^-) / \varepsilon(B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-))$.

Differential $r_{J/\psi}$ distributions in the Run 2 conditions with the electron mode in the $hTOS!$ and $TIS!$ category are shown in Appendix E. In these cases again, the d_f factors are all smaller than the expected systematic uncertainty on R_K in the corresponding trigger category. The $r_{J/\psi}$ distributions in the Run 1 show the same characteristics as the Run 2 ones shown here, and are therefore not reported in this thesis.

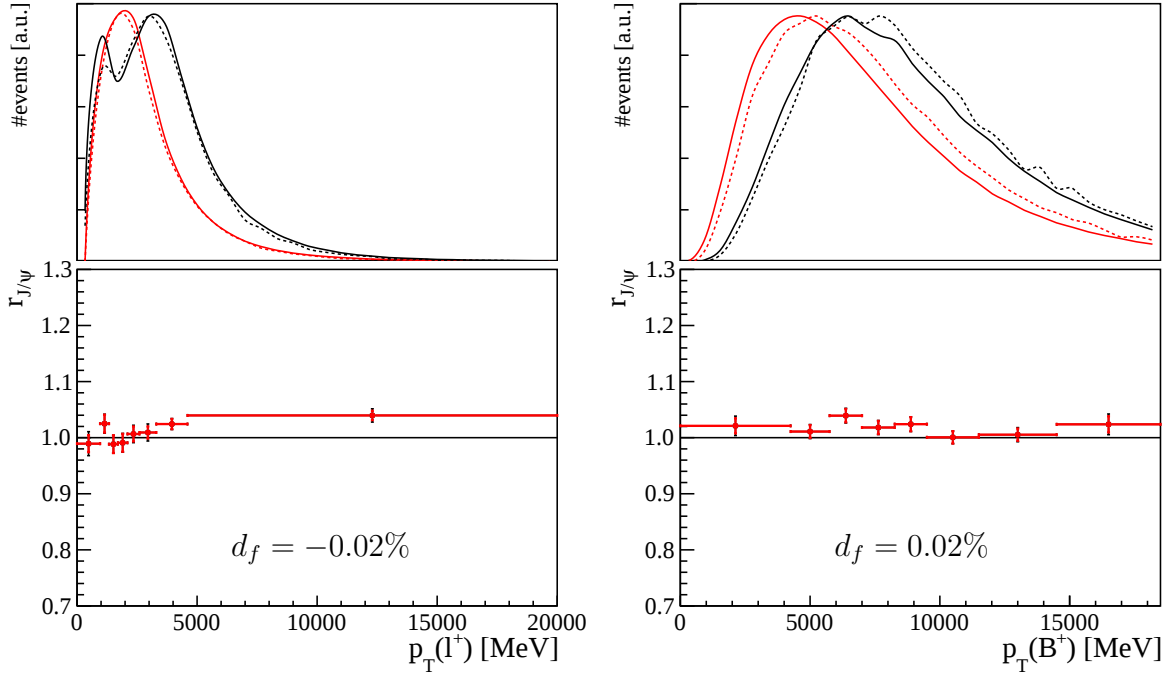


Figure 10.1: Binned $r_{J/\psi}$ values (bottom two plots) for the preselected $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the $eTOS$ category and the preselected $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the μTOS category, in the Run 2 data-taking conditions. The red thick error bars correspond to the statistical uncertainty and the black thin error bars to the systematic uncertainty. A large proportion of the bins have a statistically-dominated uncertainty, in which case the black error bar is hidden under the red. Kinematic distributions (top two plots) for simulated $B^+ \rightarrow K^+ \mu^+\mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+e^-$ (dotted black) and $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ (solid black) events are shown. Figure continues onto the next page. The value of the flatness parameter d_f is superimposed on each plot.

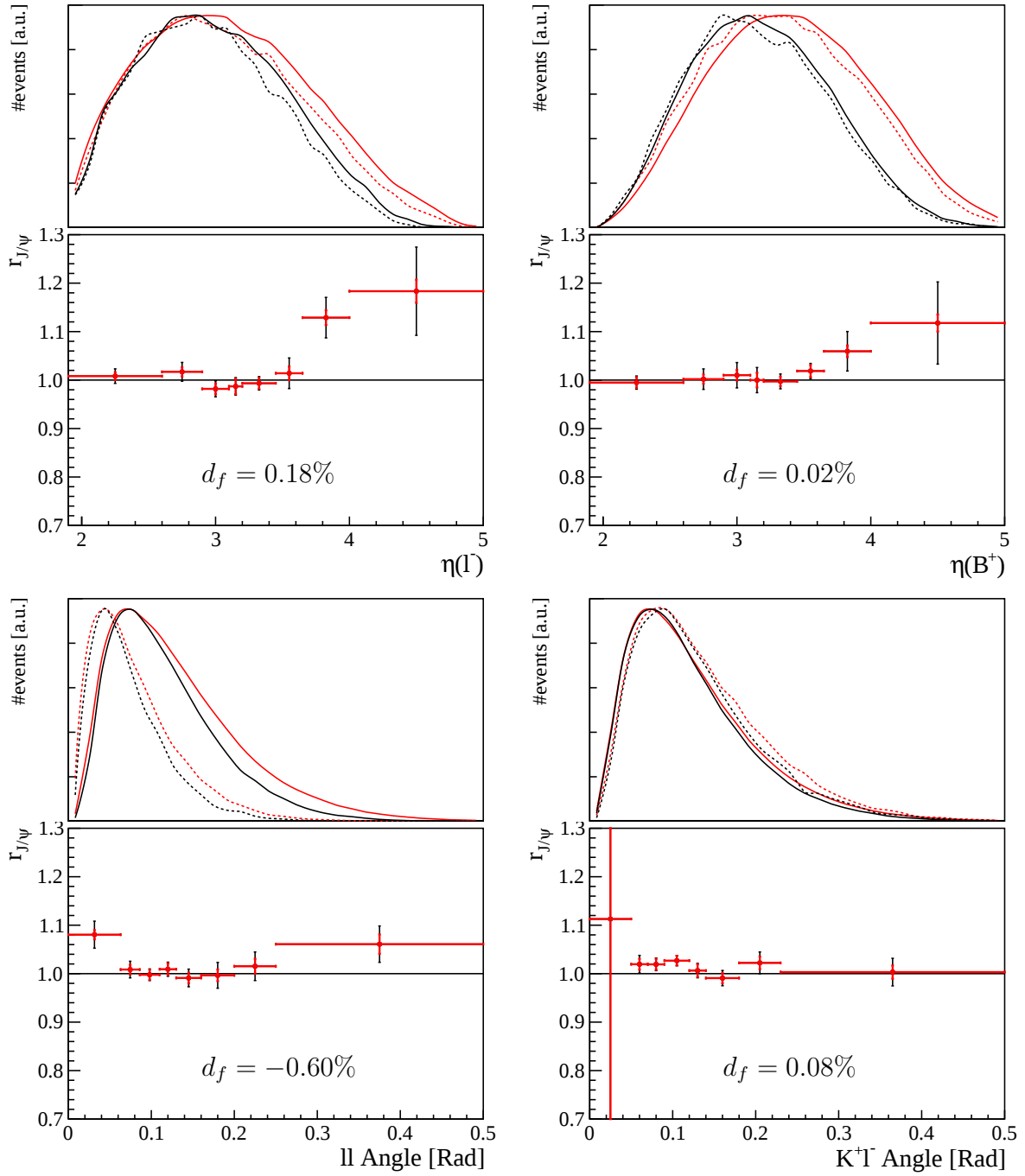


Figure 10.1: Continued from the previous page.

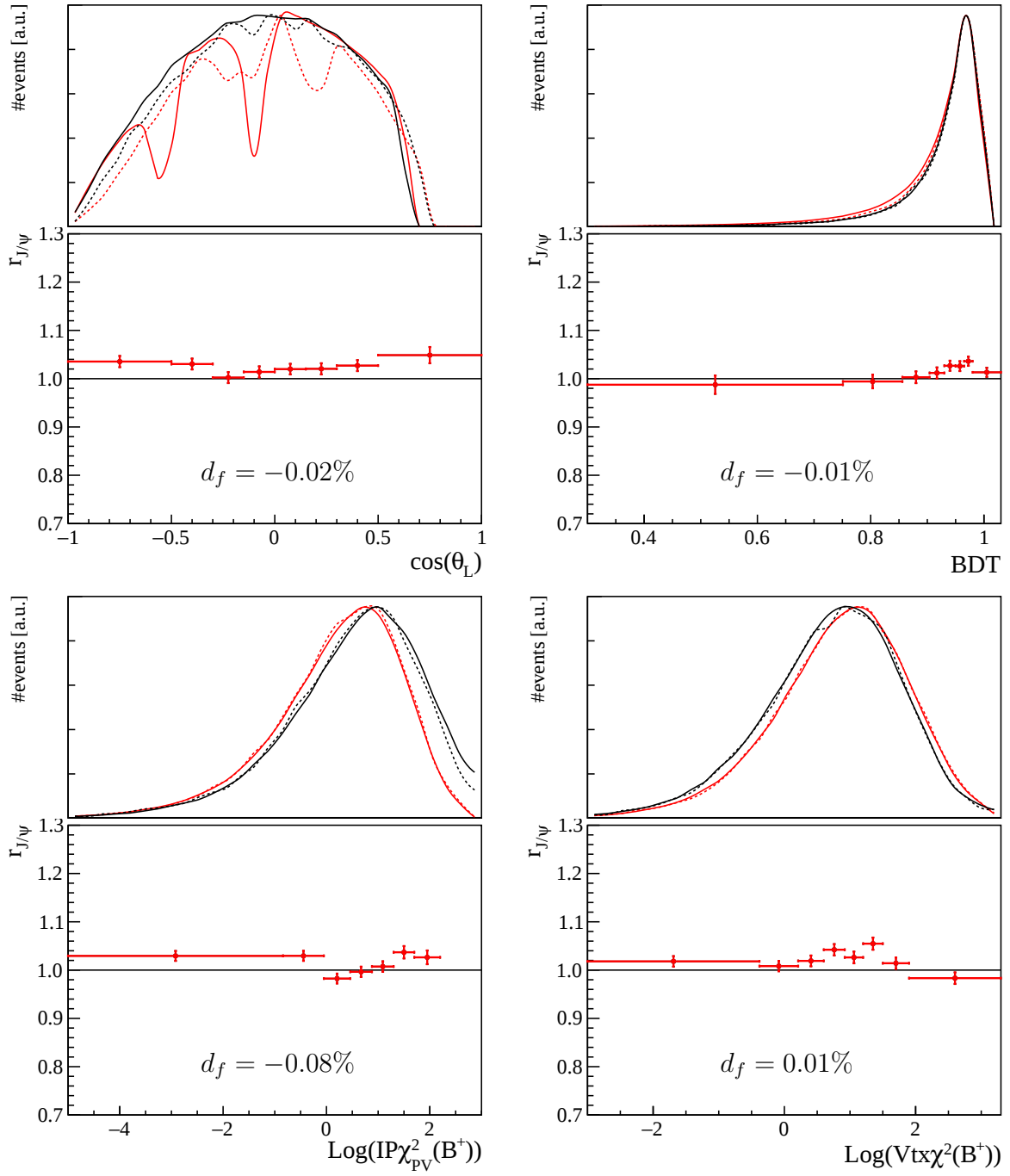


Figure 10.1: Continued from the previous page.

10.3 $R_{J/\psi}^{\text{bin}}$ test

While the q^2 distributions of the rare decay and the resonant control channels are clearly different, the detector performance (and hence the selection efficiency) depend on variables defined in the frame of the detector, such as the final state particles momenta and angles. As

$$q^2 = 4 p_{\ell^+} p_{\ell^-} \sin^2 \left(\frac{1}{2} \alpha_{\ell^+ \ell^-} \right), \quad (10.2)$$

where $\alpha_{\ell^+ \ell^-}$ is the angle between the two leptons, the efficiency can be examined as a function of some but not all of the variables p_{ℓ^+} , p_{ℓ^-} and $\alpha_{\ell^+ \ell^-}$. There is a significant overlap in the distributions of these variables between the rare and resonant decays, even if the decays do not overlap in q^2 . In the previous section, $r_{J/\psi}$ was examined as a function of a single variable (and with the parameter d_f , integrated over that variable, taking into account the spectrum of the rare decay). In this section, an attempt is made to look at $r_{J/\psi}$ in those regions of the parameter space where the control channel has kinematics closer to that of the rare decay. Taking again the example of q^2 , while it is not possible to look at $r_{J/\psi}$ in a bin of p_{ℓ^+} , p_{ℓ^-} and $\alpha_{\ell^+ \ell^-}$, it is possible to bin in two of these three variables. As detailed below, the decay kinematics can in fact be parameterised as a function of four variables and $r_{J/\psi}$ can then be examined in 2D projections of this 4D space. This test still has several shortcomings. In the 2D projection chosen, the rare decay and resonant decay will overlap only partially. The distribution in the two variables that are not examined are integrated over and, in general, the distribution in these variables will also differ in the rare decay and the resonant decay.

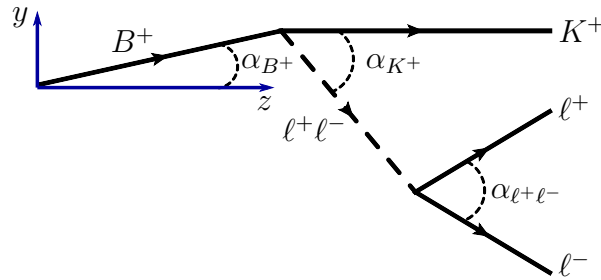


Figure 10.2: Illustration of the variables used to parametrise the $B^+ \rightarrow K^+ \ell^+ \ell^-$ decay.

To define the relevant regions, it must be first established how a $B^+ \rightarrow K^+ \ell^+ \ell^-$ decay can be parametrised in terms of variables defined in the frame of the detector. This is illustrated in Fig. 10.2. The B^+ momentum can be parametrised with three polar

coordinates: the magnitude of the momentum p_{B^+} , and an azimuthal and polar angle α_{B^+} and ϕ_{B^+} , all defined in the frame of the detector. The B^+ then decays into a kaon and a dilepton system. This is a two-body decay which can be fully parametrised with q^2 , the angle between the kaon and the dilepton system α_{K^+} , and the azimuthal angle of the dilepton system $\phi_{\ell^+\ell^-}$. Finally, the dilepton system ‘decays’ into two leptons, and this decay can be parametrised in terms of the angle between the two leptons $\alpha_{\ell^+\ell^-}$, and the azimuthal coordinate of one of the two leptons ϕ_ℓ . In summary, the $B^+ \rightarrow K^+\ell^+\ell^-$ decay can be parametrised with eight variables:

$$p_{B^+}, \alpha_{B^+}, \phi_{B^+}, \alpha_{\ell^+\ell^-}, \phi_{\ell^+\ell^-}, q^2, \alpha_{\ell^+\ell^-}, \phi_\ell. \quad (10.3)$$

Using the relation $q^2 = 2 \cdot p_{\ell^+} \cdot p_{\ell^-} \cdot (1 - \cos \alpha_{\ell^+\ell^-})$, q^2 and p_{B^+} in the above list of variables can be exchanged for the higher and lower lepton momenta, $\max p_\ell$ and $\min p_\ell$. The list of degrees of freedom therefore becomes

$$\alpha_{K^+}, \max p_\ell, \min p_\ell, \alpha_{\ell^+\ell^-}, (\phi_\ell, \phi_{\ell^+\ell^-}, \alpha_{B^+}, \phi_{B^+}). \quad (10.4)$$

The last four terms have been put in parentheses because the detector is assumed to be azimuthally symmetric, and the distribution of α_{B^+} is not related to the internal structure of the decay. Hence, the decay $B^+ \rightarrow K^+\ell^+\ell^-$ can be considered to depend upon four variables. The $B^+ \rightarrow K^+J/\psi(\ell^+\ell^-)$ decay has an extra constraint coming from $q^2 = m(J/\psi)^2$, which reduces this set to three variables.

The $R_{J/\psi}^{\text{bin}}$ test then proceeds as follows:

1. Two variables are picked among $\alpha_{K^+}, \max p_\ell, \min p_\ell, \alpha_{\ell^+\ell^-}$;
2. Using reconstructed $B^+ \rightarrow K^+J/\psi(e^+e^-)$ simulated events in the range $1.1 < q^2 < 6.0 \text{ GeV}^2$, two-dimensional bins are chosen so as to span the region occupied by the rare decay. As in the 1D $r_{J/\psi}$ test, the bins are chosen such that they contain an approximately equal number of simulated $B^+ \rightarrow K^+J/\psi(e^+e^-)$ events;
3. In each bin, the quantity $R_{J/\psi}^{\text{bin}}$ is defined as the following double ratio:

$$R_{J/\psi}^{\text{bin}} = \frac{\frac{N(B^+ \rightarrow K^+J/\psi(\mu^+\mu^-))|_{\in \text{bin}}}{\varepsilon(B^+ \rightarrow K^+J/\psi(\mu^+\mu^-))|_{\in \text{bin}}} \bigg/ \frac{N(B^+ \rightarrow K^+J/\psi(\mu^+\mu^-))}{\varepsilon(B^+ \rightarrow K^+J/\psi(\mu^+\mu^-))}}{\frac{N(B^+ \rightarrow K^+J/\psi(e^+e^-))|_{\in \text{bin}}}{\varepsilon(B^+ \rightarrow K^+J/\psi(e^+e^-))|_{\in \text{bin}}} \bigg/ \frac{N(B^+ \rightarrow K^+J/\psi(e^+e^-))}{\varepsilon(B^+ \rightarrow K^+J/\psi(e^+e^-))}}, \quad (10.5)$$

where the subscript $\in \text{bin}$ indicates that the event yield is measured in a given bin, and that the efficiency corresponds to the efficiency of an event being reconstructed in that bin.

As above, a significant deviation from $R_{J/\psi}^{\text{bin}}$ from unity in any of the bins would indicate a source of systematic uncertainty that may not cancel in the R_K double ratio. A flatness parameter that quantifies the potential effect on R_K is again constructed based on the deviations observed, taking into account the distribution of events in the rare decay.

Results of the $R_{J/\psi}^{\text{bin}}$ test in two dimensions are shown in Fig. 10.3. This figure is made using $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events selected in the e TOS trigger category, and $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ μ TOS events in the Run 2 data-taking conditions. Statistical uncertainties are shown with thick red error bars, and total statistical and systematic uncertainties with thin black error bars. The systematic uncertainties are computed using the same method as for the one-dimensional $r_{J/\psi}$ test in Sec. 10.2. As for the differential $r_{J/\psi}$ test, the flatness parameters are at most at the percent level, again giving confidence that the efficiency ratio $\varepsilon(B^+ \rightarrow K^+ \ell^+ \ell^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-))$ is understood over the whole phase space. The same conclusions are drawn in the Run 1 data-taking conditions.

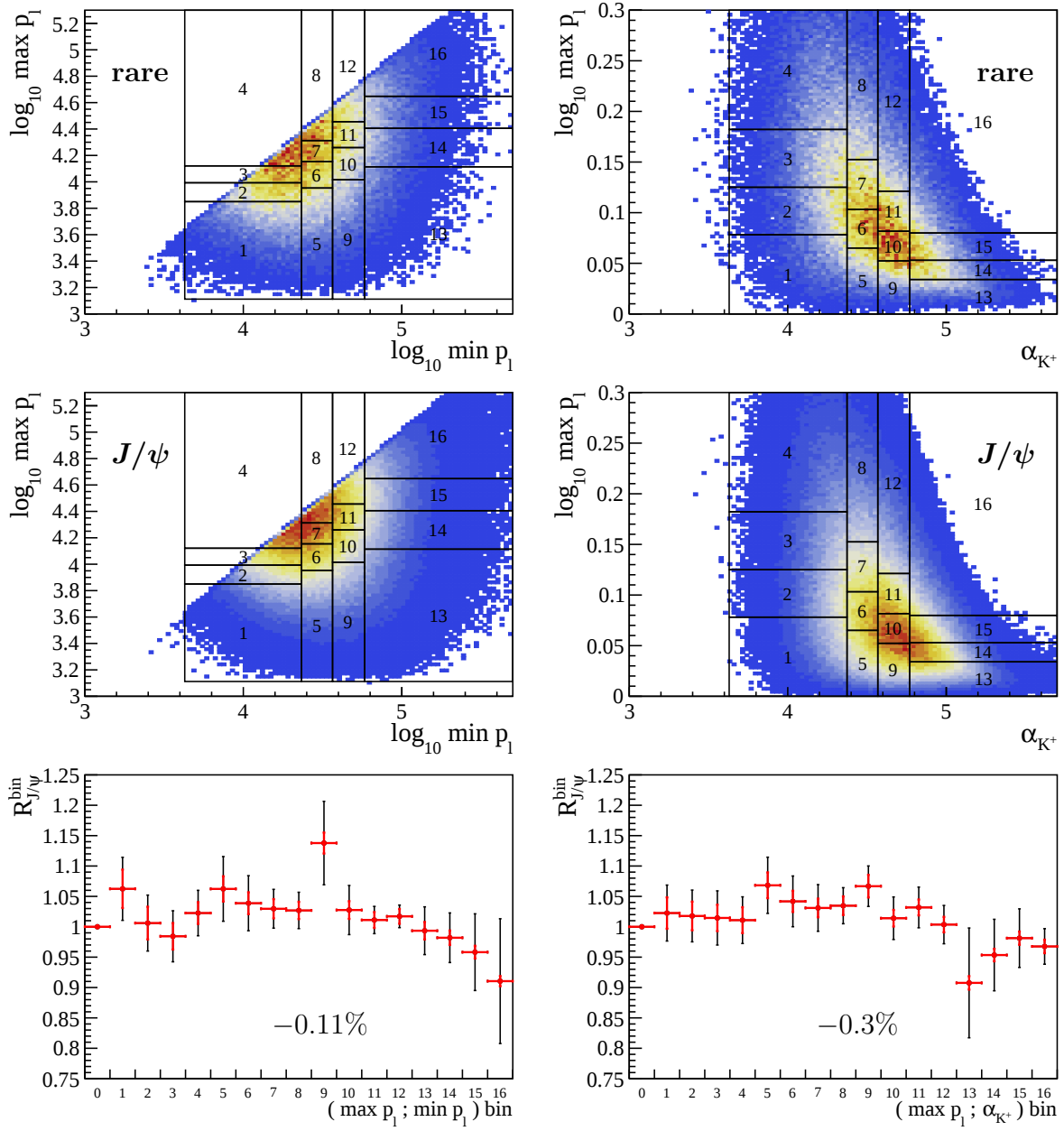


Figure 10.3: Result of the $R_{J/\psi}^{\text{bin}}$ test in two-dimensional bins, where the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode is in the e TOS category and the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode is in the μ TOS category, both in the Run 2 data-taking conditions. The $B^+ \rightarrow K^+ e^+e^-$ event distributions are shown at the top and the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ distribution in the centre, illustrating the binning scheme. The value of $R_{J/\psi}^{\text{bin}}$ in each bin is shown at the bottom, where the red thick error bars corresponds to the statistical uncertainty and the black thin error bars to the systematic uncertainty. The 0th bin corresponds to the $R_{J/\psi}^{\text{bin}}$ value computed over all phase-space, so it is by definition always equal to 1. Figure continues onto the next page.

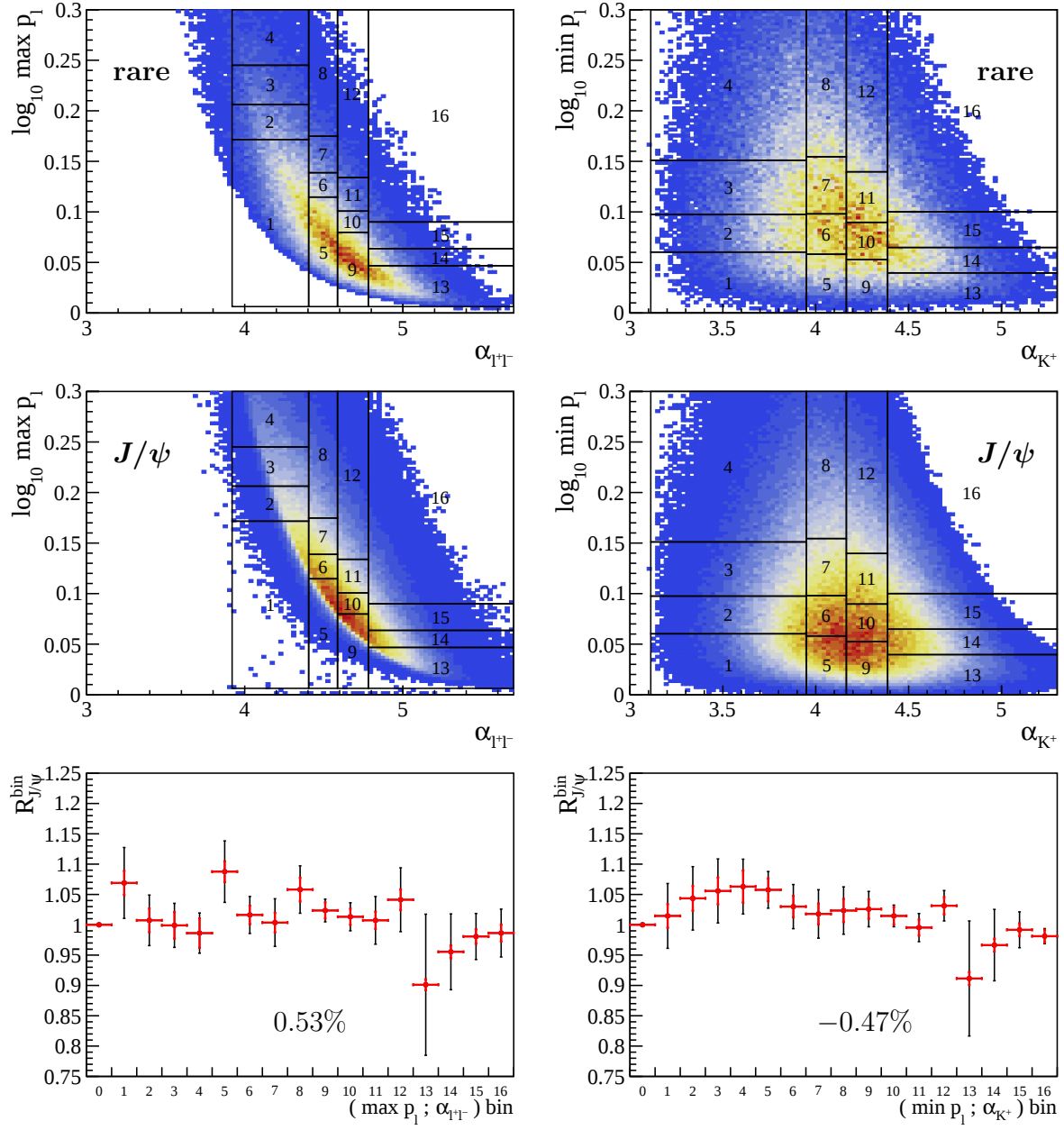


Figure 10.3: Continued from the previous page.

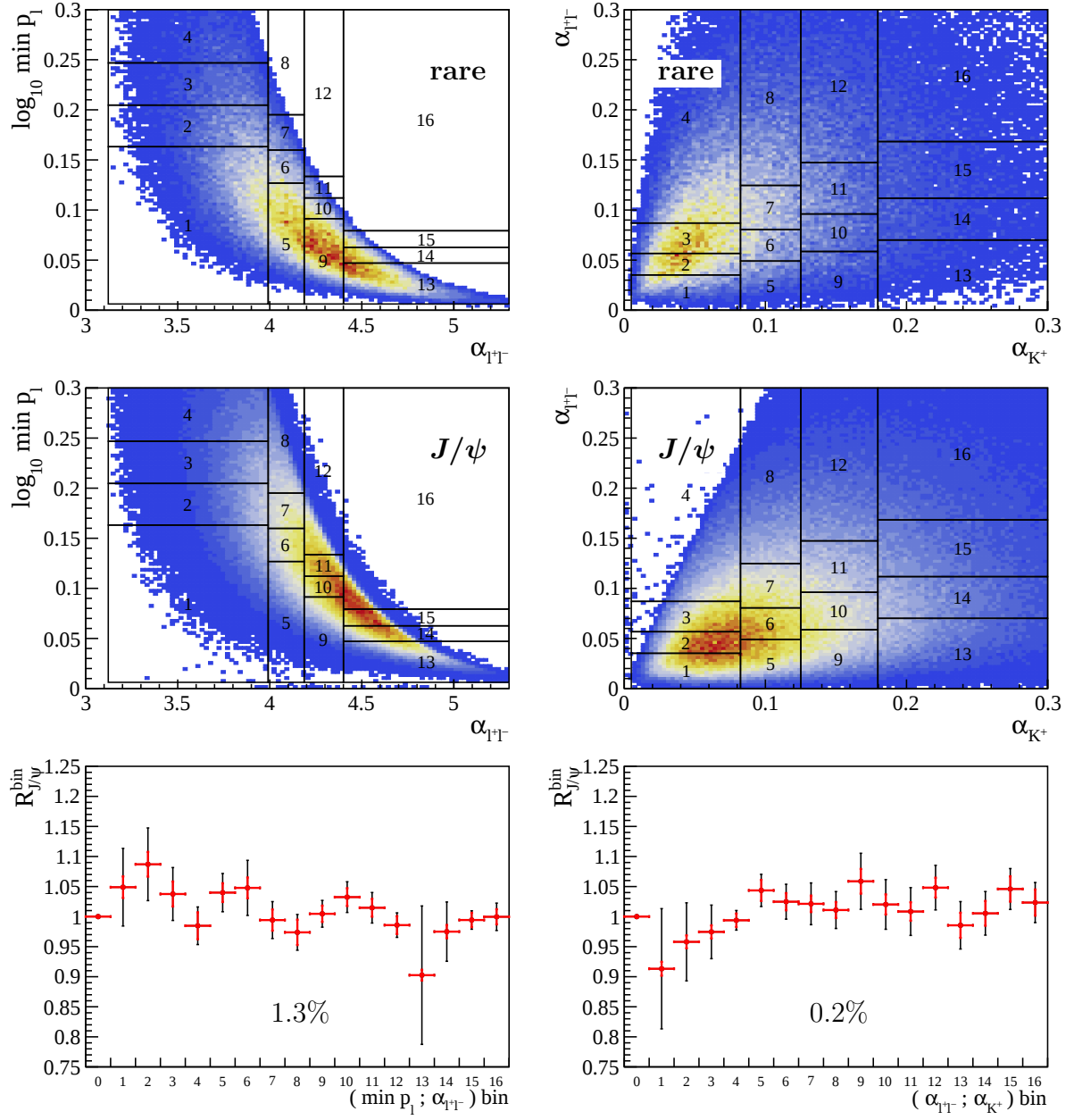


Figure 10.3: Continued from the previous page.

10.4 Branching fraction of the $B^+ \rightarrow K^+ \mu^+ \mu^-$ decay

Before proceeding to the unblinding of the data containing $B^+ \rightarrow K^+ e^+ e^-$ candidates used to measure R_K , the $B^+ \rightarrow K^+ \mu^+ \mu^-$ data are unblinded in order to measure the branching fraction of that decay and check the compatibility with the LHCb measurement using Run 1 data from Ref. [31].

The results for the mean differential branching fraction in the range $1.1 < q^2 < 6.0 \text{ GeV}^2$, measured in the Run 1 and Run 2 data-taking conditions separately, are

$$\begin{aligned} \text{Run 1: } \frac{dB}{dq^2} &= (24.4 \pm 0.9) \times 10^{-9} \text{ GeV}^{-2}, \\ \text{Run 2: } \frac{dB}{dq^2} &= (24.9 \pm 0.9) \times 10^{-9} \text{ GeV}^{-2}, \\ \text{Run 1 from Ref. [31]: } \frac{dB}{dq^2} &= (24.2 \pm 0.7) \times 10^{-9} \text{ GeV}^{-2}. \end{aligned}$$

The results are quoted with their statistical uncertainties only. The measurement in Ref. [31] uses a selection optimised for muonic B decays, which yields a higher number of $B^+ \rightarrow K^+ \mu^+ \mu^-$ candidates than in the analysis presented in this thesis. This explains the difference in statistical precision between the measurement performed in this thesis and the measurement in Ref. [31]. The differential branching fraction is also measured in five bins of q^2 and is shown in Fig. 10.4. The agreement between Run 1, Run 2 and the result from Ref. [31] is excellent.

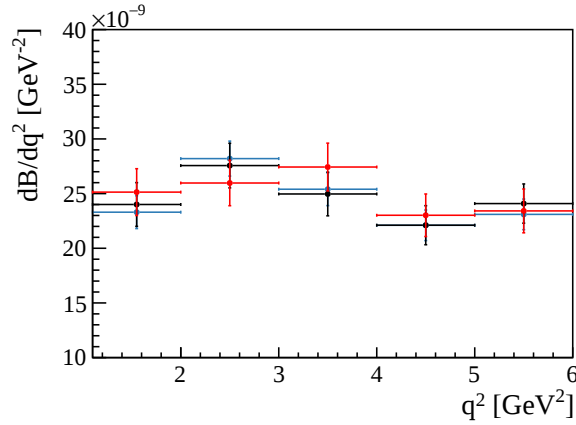


Figure 10.4: Differential branching fraction of the $B^+ \rightarrow K^+ \mu^+ \mu^-$ decay, measured using the data collected in the Run 1 (black) and Run 2 (red) data taking conditions. The result is compared with the measurement from Ref. [31] (blue). All uncertainties shown are statistical only.

10.5 Contamination from backgrounds with two mis-identified tracks

As explained in Sec. 6.7, contamination from fully reconstructed $B^+ \rightarrow K^+\pi^+\pi^-$ events, where both pions are mis-identified as electrons is expected to be reduced to a negligible level by the PID selection. To demonstrate that this is the case, the $B^+ \rightarrow K^+e^+e^-$ sample selected in data is reconstructed ignoring the bremsstrahlung correction and changing the mass hypothesis of each electron to that of the pion. Any contribution from $B^+ \rightarrow K^+\pi^+\pi^-$ decays would appear as a narrow peak at $m(B^+)$ in the $m(K^+e^+_{\rightarrow\pi^+}e^-_{\rightarrow\pi^-})$ mass distribution. As shown in Fig. 10.5, there is no trace of $B^+ \rightarrow K^+\pi^+\pi^-$ mis-identified events in the final $B^+ \rightarrow K^+e^+e^-$ data sample.

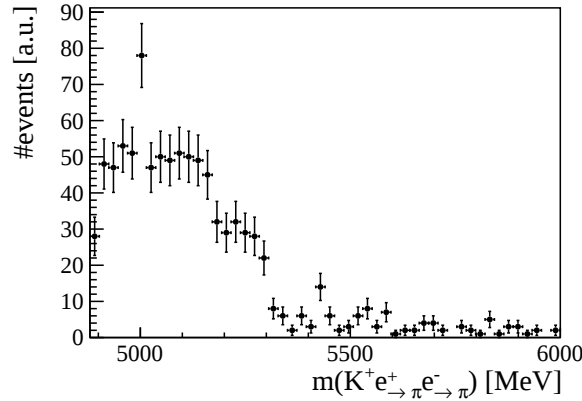


Figure 10.5: Distribution of the $m(K^+e^+_{\rightarrow\pi^+}e^-_{\rightarrow\pi^-})$ for $B^+ \rightarrow K^+e^+e^-$ selected candidates in the combined Run 1 and Run 2 sample.

The second kind of double mis-ID background is referred to as $B^+ \rightarrow K^+J/\psi(e^+e^-)$ “swaps”, where a $B^+ \rightarrow K^+_{[\leftrightarrow\ell^+]}J/\psi(\ell^+_{[\leftrightarrow K^+]}e^-)$ transition where the K^+ is mis-identified as the ℓ^+ and vice-versa. As in the previous case, this background is expected to be negligible after the PID selection. To demonstrate this, $B^+ \rightarrow K^+e^+e^-$ candidates in data are reconstructed under the exchanged mass hypothesis $K \leftrightarrow e$, and their invariant mass $m_{\text{DTF}}^{J/\psi}(K^+_{\rightarrow e^+}e^+_{\rightarrow K^+}e^-)$ is computed, where the mass $m(K^+e^-)$ is constrained to that of the J/ψ . Any contribution from $B^+ \rightarrow K^+_{[\leftrightarrow\ell^+]}J/\psi(\ell^+_{[\leftrightarrow K^+]}e^-)$ decays would appear as a sharp peak at the B^+ mass. As can be seen in Fig. 10.6, no such contribution is found in the final $B^+ \rightarrow K^+e^+e^-$ data sample.

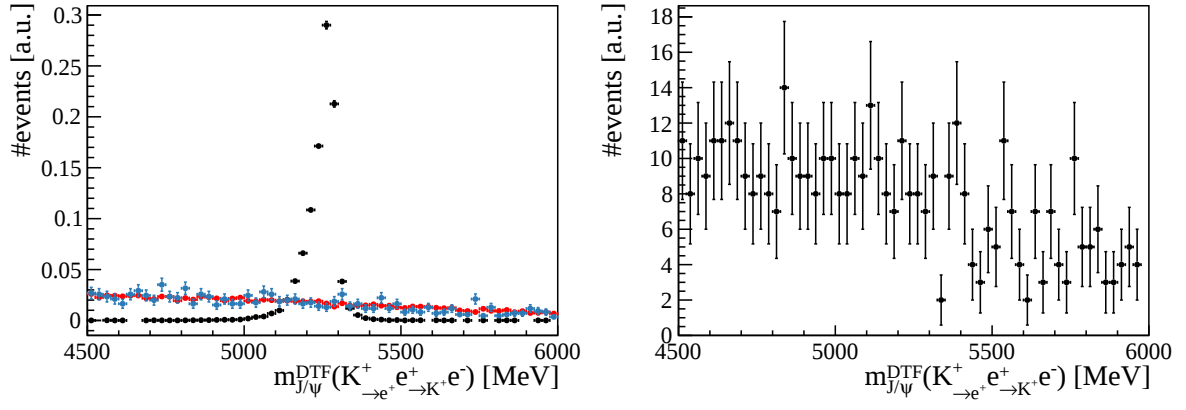


Figure 10.6: Distribution of the $m_{\text{DTF}}^{J/\psi}(K^+_{\rightarrow e^+} e^+_{\rightarrow K^+} e^-)$ invariant mass for $B^+ \rightarrow K^+_{[\leftrightarrow \ell^+]} J/\psi(\ell^+_{[\leftrightarrow K^+]} \ell^-)$ (black), $B^+ \rightarrow K^+ e^+ e^-$ and $B^0 \rightarrow K^{*0} e^+ e^-$ simulated events (left). The same distribution for $B^+ \rightarrow K^+ e^+ e^-$ selected candidates in the combined Run 1 and Run 2 sample (right).

11. Result

The fits to the rare modes are performed using the signal and background shapes described in Sec. 8.3. The fits to the $B^+ \rightarrow K^+ e^+ e^-$ and $B^+ \rightarrow K^+ \mu^+ \mu^-$ modes in the different trigger categories and data-taking conditions are performed simultaneously, yielding a single value for R_K , following the procedure described in Sec. 8.4. The coefficients c_K^{rt} , which allow the $B^+ \rightarrow K^+ e^+ e^-$ and $B^+ \rightarrow K^+ \mu^+ \mu^-$ yields to be converted to one value for R_K , are constrained to the central values shown in Tab. 8.1 using a multidimensional Gaussian constraint with widths and correlations shown in Tab. 9.1. The fits to $B^+ \rightarrow K^+ \mu^+ \mu^-$ in both the Run 1 and Run 2 data are shown in Fig. 11.2. The fits to $B^+ \rightarrow K^+ e^+ e^-$ in the Run 1 and Run 2 conditions are shown in Fig. 11.3-11.5.

In Fig. 11.3-11.5, it can be observed that the partially reconstructed background yield is higher in the Run 2 data than in the Run 1 data. On the contrary, the combinatorial background yield is lower in Run 2 than in Run 1. However, it was checked that the $B^+ \rightarrow K^+ e^+ e^-$ yield, and hence the value of R_K , does not change if the fraction of partially reconstructed background to signal is constrained to be the same between Run 1 and Run 2. Studies of the fit using pseudo-datasets show that the yields of these two background components are anti-correlated: the sum of the two yields is determined precisely in the fit, but the individual yields of each of the two background components have a large uncertainty. If, in a given fit to a pseudo-dataset, the partially reconstructed background yield is under-estimated, the combinatorial yield is usually over-estimated.

As explained in Sec. 8.5, the uncertainty on R_K established from the fit is corrected to include the fit bias and the uncertainty on the fit shape. The fit bias is established by repeating the study described in Sec. 8.6 using as input the measured yields for the partially-reconstructed and combinatorial backgrounds measured in the data. The final bias is estimated to be $(3.5 \pm 1.0)\% \times \sigma_{\text{stat}}$. The central value of R_K is shifted by this value, and the sum in quadrature of half of the size of the bias and its own uncertainty is considered as an additional systematic uncertainty on R_K . Finally, the systematic associated with the fit model computed in Sec. 9.10, is also combined with the fit uncertainty by convoluting the profiled likelihood with a Gaussian of suitable width. Fig. 11.1 shows the original likelihood profile obtained from the fit to the data, and the effect of the two additional systematic uncertainties.

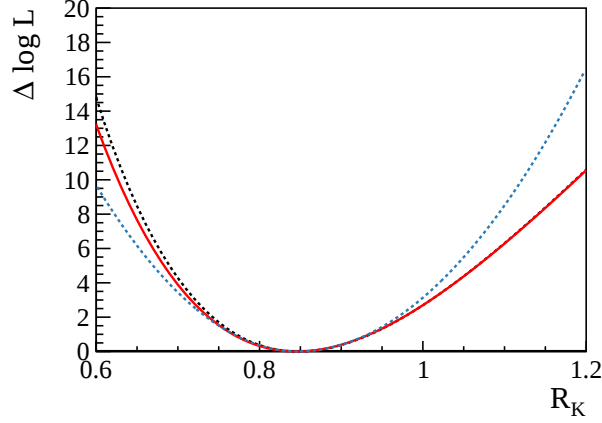


Figure 11.1: Profiled likelihood obtained from the fit to the data (dotted black), after the correction of the fit bias and with the additional systematic uncertainty related to the fit shape R_K (solid red). The blue dotted line depicts the expected shape of the likelihood profile if the uncertainties were Gaussian.

The final result for R_K is

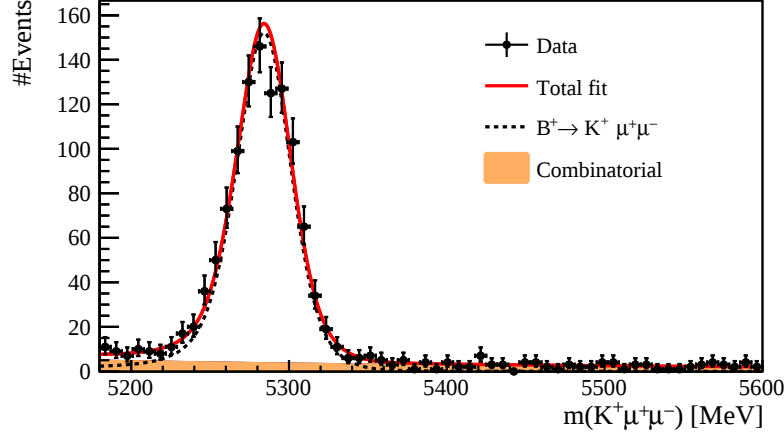
$$R_K = 0.846^{+0.060}_{-0.054} \text{ (stat.) }^{+0.016}_{-0.014} \text{ (syst.)}, \quad (11.1)$$

where the first uncertainty is statistical and the second is systematic. The likelihood profiled as a function of R_K is used to establish the level of compatibility of the result with the SM prediction. Taking into account a 1% variation of the SM prediction due to final state radiation [41], the compatibility of the measurement with the SM prediction is 2.5 standard deviations.

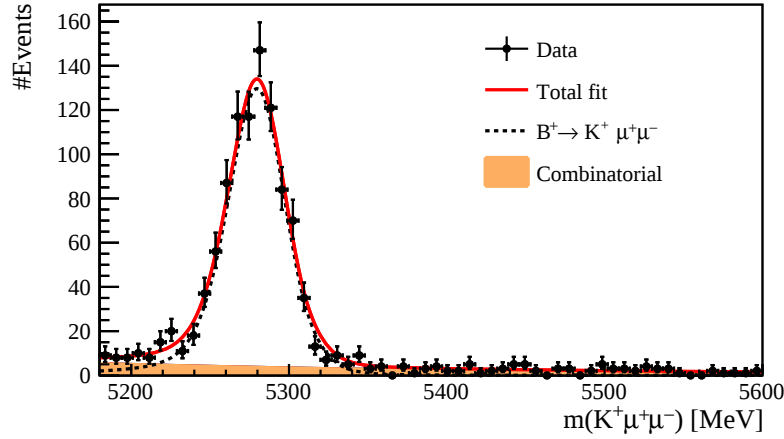
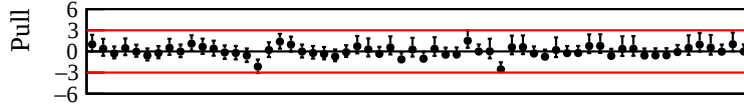
The fit is repeated to measure the value of R_K in Run 1 and Run 2 separately, in the three trigger categories separately, or separately in the six configurations defined by the trigger category and run. The results are shown in Tab. 11.1. The measurement of R_K in the Run 1 data is perfectly compatible with the previous LHCb R_K measurement. The Run 1 and Run 2 R_K values are compatible within 1.9 standard deviations.

Table 11.1: Values for R_K measured in the three trigger categories and in Run 1 (R1) and Run 2 (R2) separately (top table), in the three trigger categories separately (centre table), and in Run 1 and Run 2 separately (bottom table). The uncertainties shown are the combined systematic and statistical uncertainties from the fit.

	$eTOS$ R1	$eTOS$ R2	$hTOS!$ R1	$hTOS!$ R2	$TIS!$ R1	$TIS!$ R2
$R_K =$	$0.79^{+0.11}_{-0.10}$	$0.90^{+0.09}_{-0.08}$	$0.56^{+0.19}_{-0.12}$	$1.28^{+0.65}_{-0.35}$	$0.60^{+0.13}_{-0.10}$	$0.97^{+0.32}_{-0.21}$
	$eTOS$		$hTOS!$	$TIS!$		
$R_K =$	$0.87^{+0.07}_{-0.06}$	$0.87^{+0.24}_{-0.17}$	$0.75^{+0.13}_{-0.09}$			
	R1		R2			
$R_K =$	$0.71^{+0.08}_{-0.07}$	$0.93^{+0.09}_{-0.08}$				



$$\begin{aligned}
 N_{\text{sig}} &= 1044_{-35}^{+35} \\
 N_{\text{comb}} &= 205_{-20}^{+22} \\
 \lambda &= -2.34_{-0.76}^{+0.80} \times 10^{-3}
 \end{aligned}$$



$$\begin{aligned}
 N_{\text{sig}} &= 899_{-32}^{+33} \\
 N_{\text{comb}} &= 193_{-20}^{+21} \\
 \lambda &= -3.32_{-0.78}^{+0.81} \times 10^{-3}
 \end{aligned}$$

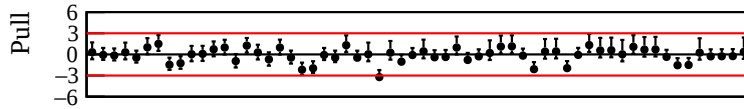


Figure 11.2: Fit to $m(K^+\mu^+\mu^-)$ for $B^+ \rightarrow K^+\mu^+\mu^-$ events in the Run 1 (top) and Run 2 (bottom) data, and the pull between the fit function and the data points. The free-floating parameters are displayed in the table to the right of the figure. Constrained parameters are shown at the bottom of the table, where the notation $\mu \pm \sigma$ indicates the mean (μ) and standard deviation (σ) of their Gaussian constraint.

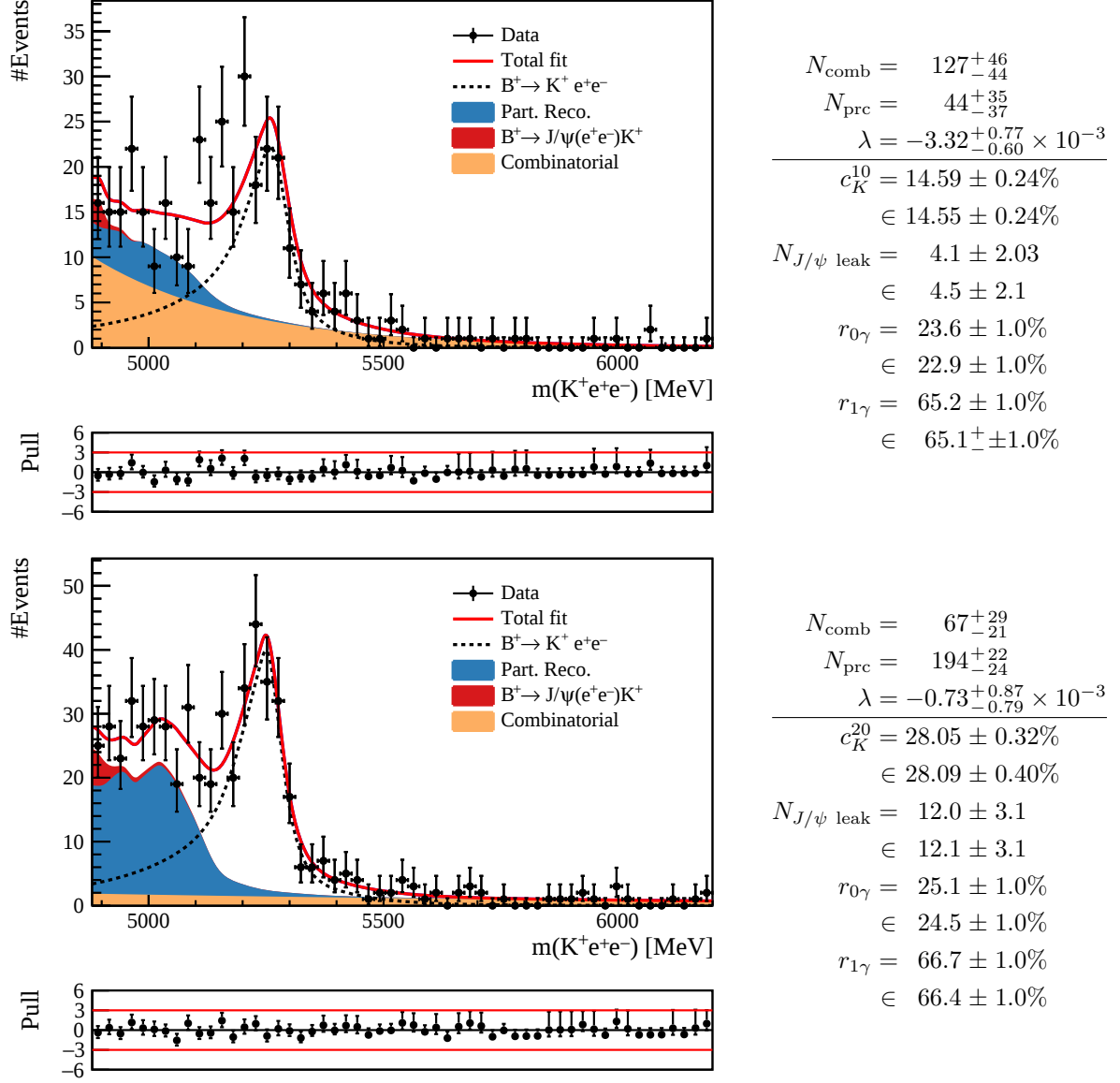


Figure 11.3: Fit to $m(K^+e^+e^-)$ for $B^+ \rightarrow K^+e^+e^-$ events in the e TOS trigger category and in the Run 1 (top) and Run 2 (bottom) data, and the pull between the fit function and the data points. The free-floating parameters are displayed in the table to the right of the figure. Constrained parameters are shown at the bottom of the table, where the notation $\in \mu \pm \sigma$ indicates the mean (μ) and standard deviation (σ) of their Gaussian constraint.

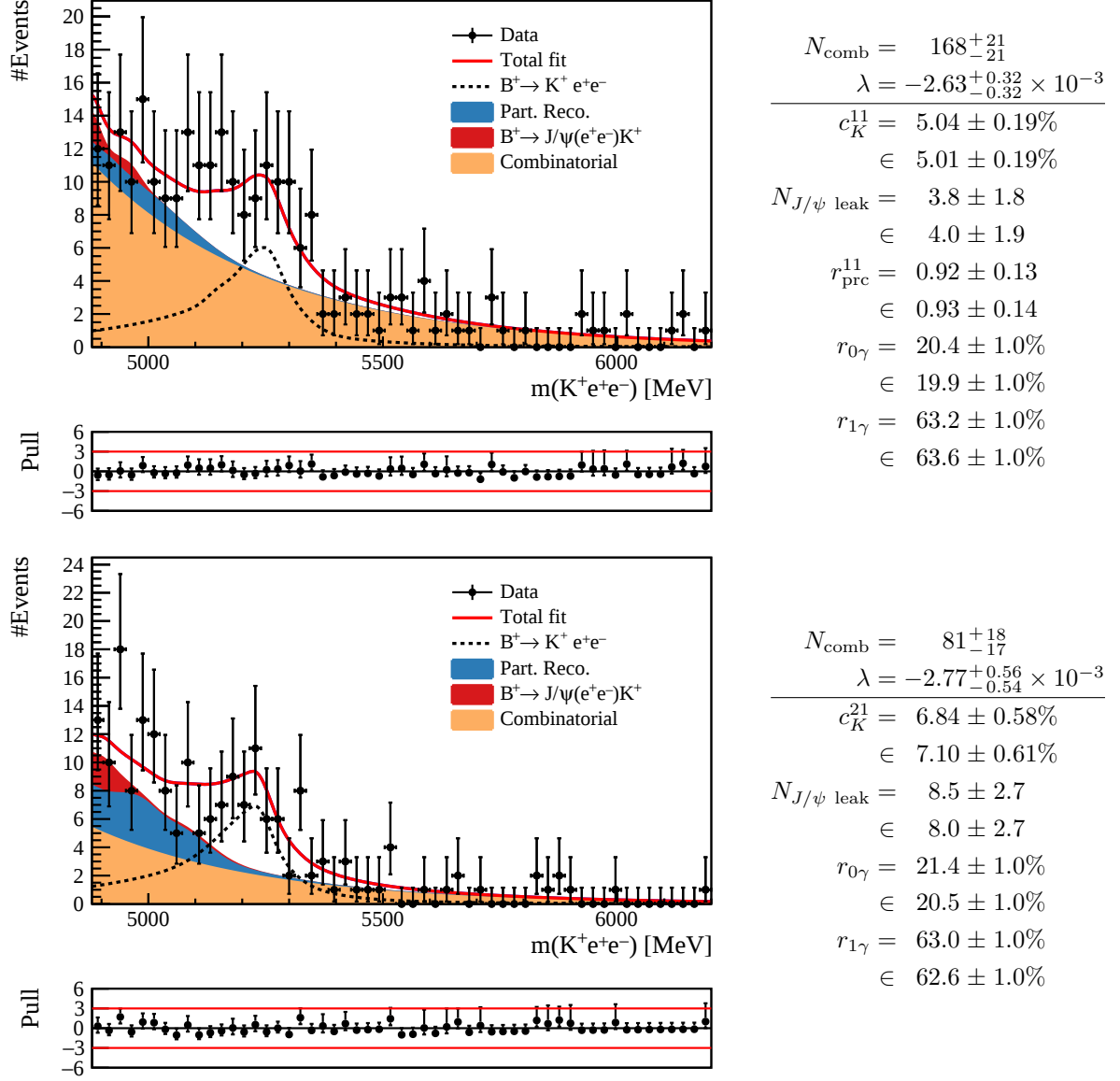


Figure 11.4: Fit to $m(K^+e^+e^-)$ for $B^+ \rightarrow K^+e^+e^-$ events in the $hTOS!$ trigger category and in the Run 1 (top) and Run 2 (bottom) data, and the pull between the fit function and the data points. The free-floating parameters are displayed in the table to the right of the figure. Constrained parameters are shown at the bottom of the table, where the notation $\in \mu \pm \sigma$ indicates the mean (μ) and standard deviation (σ) of their Gaussian constraint.

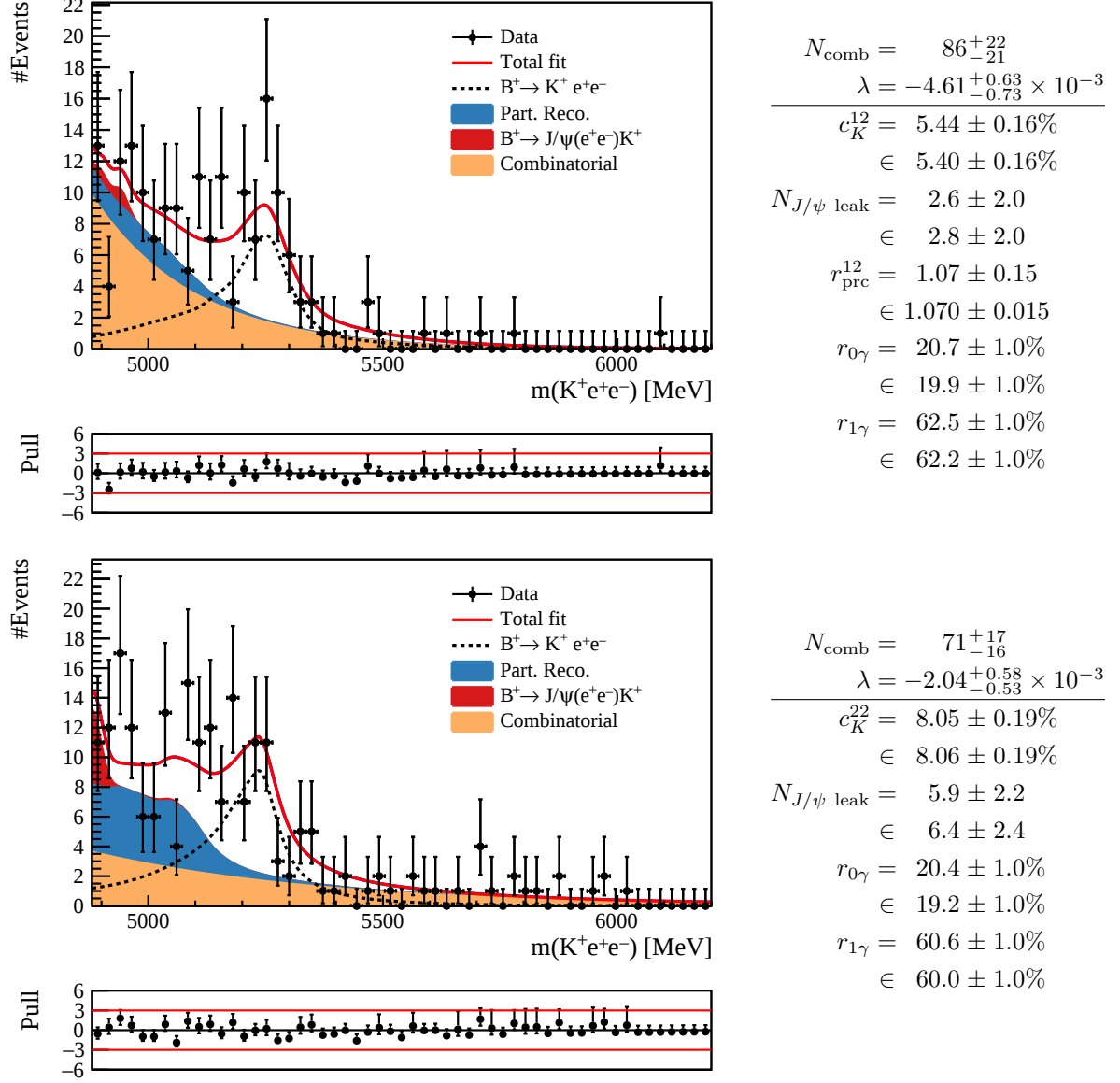


Figure 11.5: Fit to $m(K^+e^+e^-)$ for $B^+ \rightarrow K^+e^+e^-$ events in the TIS! trigger category and in the Run 1 (top) and Run 2 (bottom) data, and the pull between the fit function and the data points. The free-floating parameters are displayed in the table to the right of the figure. Constrained parameters are shown at the bottom of the table, where the notation $\in \mu \pm \sigma$ indicates the mean (μ) and standard deviation (σ) of their Gaussian constraint.

12. Conclusions and implications

This thesis presents a test of lepton flavour universality consisting of a measurement of the ratio of branching fractions of $B^+ \rightarrow K^+ \mu^+ \mu^-$ and $B^+ \rightarrow K^+ e^+ e^-$ decays, R_K . The result is

$$R_K = 0.846^{+0.060}_{-0.054} (\text{stat.})^{+0.016}_{-0.014} (\text{syst.}), \quad (12.1)$$

where the first uncertainty is statistical and the second is systematic. Compared to the previous R_K measurement [42], the relative uncertainty on R_K has been reduced by more than 40%. Moreover, the new calibration techniques and cross-checks the analysis employs give confidence in the fact that the detector's electron detection efficiencies are understood. However, the central value of the result is such that the discrepancy with respect to the SM expectation remains essentially the same, at the level of 2.5 standard deviations. The breaking of lepton flavour universality in beauty decays is not confirmed, nor excluded.

The 2017 and 2018 LHCb data consist of an additional $\sim 3.7 \text{ fb}^{-1}$ of proton collision data at a centre-of-mass energy of 13 TeV. This sample will contain a number of beauty mesons that is approximately double that used in this analysis. In the future, these new data will be used to further investigate the beauty anomalies on many different fronts.

Future update of the R_K and R_{K^*} results: The obvious next step is an update of the R_K and R_{K^*} analyses using the whole data available, hoping that the increase of precision will this time confirm or rule out the breaking of LFU in these decays. It will also be interesting to see whether the two R_{K^*} results extracted with the Run 1 data and Run 2 data show the same mild tension as the Run 1 and Run 2 R_K results, which may point to an issue of unknown origin with the Run 1 data.

Supplementary checks regarding electron detection: The extra data available will also allow new cross-checks of the electron detection efficiencies to be performed. A possible test is the measurement of the ratio of branching fractions between $D_s \rightarrow \phi(\rightarrow e^+ e^-) \pi^+$ and $D_s \rightarrow \phi(\rightarrow \mu^+ \mu^-) \pi^+$. This gives access to a di-lepton system at low q^2 , which would complement the checks made using the $B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-)$ control mode where the di-lepton system has a high q^2 .

Tests of LFU with other decays or observables: Lepton universality can also be tested in other electroweak penguin transitions, such as $\Lambda_b \rightarrow p^+ K^- \ell^+ \ell^-$, $B_s^0 \rightarrow \phi \ell^+ \ell^-$ or

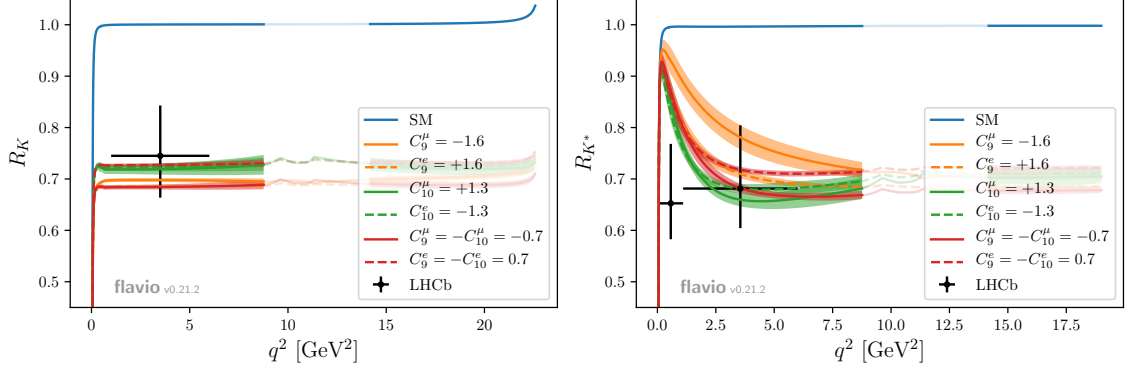


Figure 12.1: Theoretical prediction for R_K (left) and R_{K^*} (right) for different values of the Wilson coefficients and over the whole q^2 range. Taken from Ref. [37].

$B^+ \rightarrow \pi^+ \ell^+ \ell^-$. In case the LFU anomalies persist in R_K and R_{K^*} , measuring LFU ratios in other decays can help characterise the structure of any NP contribution. In addition to investigating other decays, other observables can also be measured. For example, LFU can be tested in the high q^2 region. This is a useful test, because a NP contribution in either C_9^μ or C_{10}^μ would induce the same branching fraction shift at low and high q^2 , as illustrated in Fig. 12.1. Additionally, LFU can be tested in angular observables, such as those defined in Ref. [118].

Searches for LFV: Searches for lepton flavour violating $b \rightarrow s \ell \ell'$ decays are particularly motivated if the presence of LFU breaking is confirmed. A leptoquark that can explain the B anomalies could mediate the decay $B^+ \rightarrow K^+ \mu^\pm \tau^\mp$ with a branching fraction at the level of 10^{-6} [51, 53]. Semi-tauonic decays are very challenging in the LHCb environment, but the data now available will make it possible to probe branching fractions down to this level.

Further investigations of $b \rightarrow s \mu^+ \mu^-$ transitions : Even if future analyses rule out LFU breaking in R_K and R_{K^*} , the anomalies in the $b \rightarrow s \mu^+ \mu^-$ branching fractions and angular observables remain to be explained. Improvements of the theory computations are awaited. On the experimental side, new analysis techniques can be employed to control some theory uncertainties (such as those related to the long distance contribution) from the data. This has already been done in Ref. [98] for $B^+ \rightarrow K^+ \mu^+ \mu^-$ decays, and will be applied to $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decays in the future.

LFU in tree level beauty decays: The R_D and R_{D^*} anomalies also remain to be explained. The new LHCb data will of course be useful to improve the precision on these observables. New observables can also be measured to further improve the understanding of these anomalies. Recently, lepton universality in tree level charmed beauty decays has been studied using LHCb data [119], and a measurement of the τ polarisation in $B \rightarrow D^* \tau^- \bar{\nu}_\tau$ decays has been performed using Belle data [120]. Such observables will help characterise the structure of any anomalous LFU breaking contribution in these decays.

In summary, research can be undertaken in many different directions in order to precisely understand the origin of the B anomalies. On top of the LHCb data collected in 2017 and 2018, the Belle II B factory will start collecting data in March 2019. Using this new detector will not only allow the LHCb results to be checked in another environment, but will also give access to observables that are difficult to measure at LHCb, such as those in beauty decays into final states with a τ . Later in 2021, the upgraded LHCb detector will collect data at an increased instantaneous luminosity, allowing further enhancement of the statistical precision on many observables. Even if the analysis presented in this thesis does not provide a definite answer on whether or not NP contribute to the decays of beauty mesons, it seems very likely that the answer will come within the next decade.

Bibliography

- [1] M. D. Schwartz, *Quantum Field Theory and the Standard Model*, Cambridge University Press, 2014.
- [2] M. Thomson, *Modern particle physics*, Cambridge University Press, New York, 2013.
- [3] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*, Addison-Wesley, Reading, USA, 1995.
- [4] Y. Grossman and P. Tanedo, *Just a Taste: Lectures on Flavor Physics*, in *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: Anticipating the Next Discoveries in Particle Physics (TASI 2016): Boulder, CO, USA, June 6-July 1, 2016*, pp. 109–295, 2018. [arXiv:1711.03624](#). doi: 10.1142/9789813233348_0004.
- [5] Particle Data Group, C. Patrignani *et al.*, *Review of particle physics*, Chin. Phys. **C40** (2016) 100001.
- [6] F. L. Wilson, *Fermi’s theory of beta decay*, American Journal of Physics **36** (1968) 1150.
- [7] C. S. Wu *et al.*, *Experimental Test of Parity Conservation in Beta Decay*, Phys. Rev. **105** (1957) 1413.
- [8] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579.
- [9] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264.
- [10] A. Salam and J. C. Ward, *Electromagnetic and weak interactions*, Phys. Lett. **13** (1964) 168.
- [11] Super-Kamiokande, Y. Fukuda *et al.*, *Evidence for oscillation of atmospheric neutrinos*, Phys. Rev. Lett. **81** (1998) 1562, [arXiv:hep-ex/9807003](#).
- [12] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508.
- [13] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321.

- [14] CMS collaboration, S. Chatrchyan *et al.*, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. **B716** (2012) 30, [arXiv:1207.7235](#).
- [15] ATLAS collaboration, 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. **B716** (2012) 1, [arXiv:1207.7214](#).
- [16] Gfitter Group, M. Baak *et al.*, *The global electroweak fit at NNLO and prospects for the LHC and ILC*, Eur. Phys. J. **C74** (2014) 3046, [arXiv:1407.3792](#).
- [17] T. Aoyama, M. Hayakawa, T. Kinoshita, and M. Nio, *Revised value of the eighth-order QED contribution to the anomalous magnetic moment of the electron*, Phys. Rev. **D77** (2008) 053012, [arXiv:0712.2607](#).
- [18] L. Calibbi and G. Signorelli, *Charged Lepton Flavour Violation: An Experimental and Theoretical Introduction*, Riv. Nuovo Cim. **41** (2018), no. 2 71, [arXiv:1709.00294](#).
- [19] ALEPH, DELPHI, L3, OPAL, SLD, LEP collaborations 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](#).
- [20] ALEPH, DELPHI, L3, OPAL, LEP collaborations, 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](#).
- [21] PiENU collaboration, A. Aguilar-Arevalo *et al.*, *Improved Measurement of the $\pi \rightarrow e\nu$ Branching Ratio*, Phys. Rev. Lett. **115** (2015) 071801, [arXiv:1506.05845](#).
- [22] NA62 collaboration, C. Lazzeroni *et al.*, *Precision Measurement of the Ratio of the Charged Kaon Leptonic Decay Rates*, Phys. Lett. **B719** (2013) 326, [arXiv:1212.4012](#).
- [23] KEDR collaboration, 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. **B731** (2014) 227, [arXiv:1311.5005](#).
- [24] T. Blake, T. Gershon, and G. Hiller, *Rare b hadron decays at the LHC*, Ann. Rev. Nucl. Part. Sci. **65** (2015) 113, [arXiv:1501.03309](#).

- [25] G. Buchalla, *Heavy quark theory*, in *Heavy flavor physics: Theory and experimental results in heavy quark physics and CP violation. Proceedings, 55th Scottish Universities Summer School in Physics, SUSSP 2001, St. Andrews, UK, August 7-23, 2001*, pp. 57–104, 2002. [arXiv:hep-ph/0202092](#).
- [26] A. Khodjamirian, T. Mannel, A. A. Pivovarov, and Y.-M. Wang, *Charm-loop effect in $B \rightarrow K^{(*)}\ell^+\ell^-$ and $B \rightarrow K^*\gamma$* , JHEP **09** (2010) 089, [arXiv:1006.4945](#).
- [27] S. Descotes-Genon, L. Hofer, J. Matias, and J. Virto, *On the impact of power corrections in the prediction of $B \rightarrow K^*\mu^+\mu^-$ observables*, JHEP **12** (2014) 125, [arXiv:1407.8526](#).
- [28] S. Jger and J. Martin Camalich, *Reassessing the discovery potential of the $B \rightarrow K^*\ell^+\ell^-$ decays in the large-recoil region: SM challenges and BSM opportunities*, Phys. Rev. **D93** (2016) 014028, [arXiv:1412.3183](#).
- [29] M. Della Morte, J. Heitger, H. Simma, and R. Sommer, *Non-perturbative Heavy Quark Effective Theory: An application to semi-leptonic B-decays*, Nucl. Part. Phys. Proc. **261-262** (2015) 368, [arXiv:1501.03328](#).
- [30] D. van Dyk, *The Decays $\bar{B} \rightarrow \bar{K}^{(*)}\ell^+\ell^-$ at Low Recoil and their Constraints on New Physics*, PhD thesis, Dortmund U., 2012.
- [31] LHCb collaboration, R. Aaij *et al.*, *Differential branching fractions and isospin asymmetries of $B \rightarrow K^{(*)}\mu^+\mu^-$ decays*, JHEP **06** (2014) 133, [arXiv:1403.8044](#).
- [32] LHCb collaboration, R. Aaij *et al.*, *Measurements of the S-wave fraction in $B^0 \rightarrow K^+\pi^-\mu^+\mu^-$ decays and the $B^0 \rightarrow K^*(892)^0\mu^+\mu^-$ differential branching fraction*, JHEP **11** (2016) 047, [arXiv:1606.04731](#), [Erratum: JHEP04,142(2017)].
- [33] LHCb collaboration, 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](#).
- [34] LHCb collaboration, *Angular analysis of the $B^0 \rightarrow K^{*0}\mu^+\mu^-$ decay using 3 fb^{-1} of integrated luminosity*, JHEP **02** (2016) 104.
- [35] S. Descotes-Genon, J. Matias, M. Ramon, and J. Virto, *Implications from clean observables for the binned analysis of $B \rightarrow K^*\mu^+\mu^-$ at large recoil*, JHEP **01** (2013) 048, [arXiv:1207.2753](#).

- [36] T. Hurth, F. Mahmoudi, and S. Neshatpour, *On the anomalies in the latest LHCb data*, Nucl. Phys. **B909** (2016) 737, [arXiv:1603.00865](#).
- [37] W. Altmannshofer, P. Stangl, and D. M. Straub, *Interpreting Hints for Lepton Flavor Universality Violation*, Phys. Rev. **D96** (2017) 055008, [arXiv:1704.05435](#).
- [38] L.-S. Geng *et al.*, *Towards the discovery of new physics with lepton-universality ratios of $b \rightarrow s\ell\ell$ decays*, Phys. Rev. **D96** (2017) 093006, [arXiv:1704.05446](#).
- [39] G. D’Amico *et al.*, *Flavour anomalies after the R_{K^*} measurement*, JHEP **09** (2017) 010, [arXiv:1704.05438](#).
- [40] B. Capdevila *et al.*, *Patterns of New Physics in $b \rightarrow s\ell^+\ell^-$ transitions in the light of recent data*, JHEP **01** (2018) 093, [arXiv:1704.05340](#).
- [41] M. Bordone, G. Isidori, and A. Pattori, *On the Standard Model predictions for R_K and R_{K^*}* , Eur. Phys. J. **C76** (2016) 440, [arXiv:1605.07633](#).
- [42] LHCb collaboration, *Test of lepton universality using $B^+ \rightarrow K^+\ell^+\ell^-$ decays*, Phys. Rev. Lett. **113** (2014) 151601.
- [43] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality with $B^0 \rightarrow K^{*0}\ell^+\ell^-$ decays*, JHEP **08** (2017) 055, [arXiv:1705.05802](#).
- [44] HFLAV, *Average of $R(D)$ and $R(D^*)$ for FPCP 2017*, <http://www.slac.stanford.edu/xorg/hflav/semi/fpcp17/RDRDs.html>. Accessed: February 2019.
- [45] BaBar collaboration, B. Aubert *et al.*, *The BaBar detector*, Nucl. Instrum. Meth. **A479** (2002) 1, [arXiv:hep-ex/0105044](#).
- [46] G. R. Farrar and M. E. Shaposhnikov, *Baryon asymmetry of the universe in the minimal Standard Model*, Phys. Rev. Lett. **70** (1993) 2833, [arXiv:hep-ph/9305274](#), [Erratum: Phys. Rev. Lett.71,210(1993)].
- [47] S. Aoki *et al.*, *Review of lattice results concerning low-energy particle physics*, Eur. Phys. J. **C77** (2017) 112, [arXiv:1607.00299](#).
- [48] CMS collaboration, *Search for singly produced third-generation leptoquarks decaying to a τ lepton and a b quark in proton-proton collisions at $\sqrt{s} = 13$ TeV*, Tech. Rep. CMS-PAS-EXO-17-029, CERN, Geneva, 2018. <https://cds.cern.ch/record/2308274>.

- [49] CMS collaboration, S. Chatrchyan *et al.*, *The CMS Experiment at the CERN LHC*, JINST **3** (2008) S08004.
- [50] J. C. Pati and A. Salam, *Lepton Number as the Fourth Color*, Phys. Rev. **D10** (1974) 275, [Erratum: Phys. Rev.D11,703(1975)].
- [51] L. Di Luzio, A. Greljo, and M. Nardecchia, *Gauge leptoquark as the origin of B-physics anomalies*, Phys. Rev. **D96** (2017) 115011, [arXiv:1708.08450](#).
- [52] L. Calibbi, A. Crivellin, and T. Li, *Model of vector leptoquarks in view of the B-physics anomalies*, Phys. Rev. **D98** (2018) 115002, [arXiv:1709.00692](#).
- [53] L. Di Luzio *et al.*, *Maximal Flavour Violation: a Cabibbo mechanism for leptoquarks*, JHEP **11** (2018) 081, [arXiv:1808.00942](#).
- [54] O. S. Brning *et al.*, *LHC Design Report*, CERN Yellow Reports: Monographs, CERN, Geneva, 2004.
- [55] ALICE collaboration, K. Aamodt *et al.*, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.
- [56] ATLAS collaboration, G. Aad *et al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.
- [57] M. Lamont, *Status of the LHC*, J. Phys. Conf. Ser. **455** (2013) 012001.
- [58] R. Steerenberg and J. Wenninger, *Operational challenges and performance of the LHC during Run II*, in *Proceedings, 61st ICFA Advanced Beam Dynamics Workshop on High-Intensity and High-Brightness Hadron Beams (HB2018): Daejeon, Korea, June 17-22, 2018*, p. THA2WD01, 2018. doi: 10.18429/JACoW-HB2018-THA2WD01.
- [59] LHCb collaboration, R. Aaij *et al.*, *Measurement of the b-quark production cross-section in 7 and 13 TeV pp collisions*, Phys. Rev. Lett. **118** (2017) 052002, [arXiv:1612.05140](#), [Erratum: Phys. Rev. Lett.119,no.16,169901(2017)].
- [60] Y.-S. Tsai, *Pair Production and Bremsstrahlung of Charged Leptons*, Rev. Mod. Phys. **46** (1974) 815, [Erratum: Rev. Mod. Phys.49,521(1977)].
- [61] LHCb collaboration, A. A. Alves, Jr. *et al.*, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.

- [62] LHCb collaboration, *LHCb VELO (VERtex LOcator): Technical Design Report*, CERN-LHCC-2001-011. LHCb-TDR-005.
- [63] R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) P09007, [arXiv:1405.7808](#).
- [64] LHCb collaboration, *LHCb reoptimized detector design and performance: Technical Design Report*, CERN-LHCC-2003-030. LHCb-TDR-009.
- [65] LHCb collaboration, *LHCb inner tracker: Technical Design Report*, CERN-LHCC-2002-029. LHCb-TDR-008.
- [66] LHCb collaboration, *LHCb outer tracker: Technical Design Report*, CERN-LHCC-2001-024. LHCb-TDR-006.
- [67] R. Arink *et al.*, *Performance of the LHCb Outer Tracker*, JINST **9** (2014) P01002, [arXiv:1311.3893](#).
- [68] LHCb collaboration, *LHCb magnet: Technical Design Report*, CERN-LHCC-2000-007. LHCb-TDR-001.
- [69] R. E. Kalman, *A New Approach to Linear Filtering and Prediction Problems*, Transactions of the ASME Journal of Basic Engineering (1960), no. 82 (Series D) 35.
- [70] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Meth. **A262** (1987) 444.
- [71] J. A. N. van Tilburg, *Track Simulation and Reconstruction in LHCb*, PhD thesis, Vrije U., Amsterdam, 2005.
- [72] M. Needham, *Electron Reconstruction Studies*, Tech. Rep. LHCb-2001-102, CERN, Geneva, Aug, 2001. <https://cds.cern.ch/record/691700>.
- [73] M. Needham, *Identification of Ghost Tracks using a Likelihood Method*, Tech. Rep. LHCb-2008-026. CERN-LHCb-2008-026. LPHE-2008-004, CERN, Geneva, May, 2008. <https://cds.cern.ch/record/1107564>.
- [74] A. Perieanu, *Identification of Ghost Tracks using Neural Networks*, Tech. Rep. LHCb-2007-158. CERN-LHCb-2007-158, CERN, Geneva, Jan, 2008. <https://cds.cern.ch/record/1079949>.

- [75] M. Needham, *Performance of the LHCb Track Reconstruction Software*, Tech. Rep. LHCb-2007-144. CERN-LHCb-2007-144, CERN, Geneva, Jan, 2008. <https://cds.cern.ch/record/1080556>.
- [76] D. Derkach *et al.*, *Machine-learning-based global particle-identification algorithms at the lhcb experiment*, Journal of Physics: Conference Series **1085** (2018), no. 4 042038.
- [77] LHCb collaboration, *LHCb RICH: Technical Design Report*, CERN-LHCC-2000-037. LHCb-TDR-003.
- [78] LHCb collaboration, *LHCb muon system: Technical Design Report*, CERN-LHCC-2001-010. LHCb-TDR-004.
- [79] A. A. Alves, Jr. *et al.*, *Performance of the LHCb muon system*, JINST **8** (2013) P02022, [arXiv:1211.1346](https://arxiv.org/abs/1211.1346).
- [80] G. Bencivenni *et al.*, *A triple-GEM detector with pad readout for the inner region of the first LHCb muon station*, Tech. Rep. LHCb-2001-051, CERN, Geneva, Jul, 2001. <https://cds.cern.ch/record/691604>.
- [81] A. Kashchuk *et al.*, *Asymmetric and double-cathode-pad wire chambers for the LHCb muon system*, Nucl. Instrum. Meth. **A555** (2005) 48.
- [82] F. Archilli *et al.*, *Performance of the Muon Identification at LHCb*, JINST **8** (2013) P10020, [arXiv:1306.0249](https://arxiv.org/abs/1306.0249).
- [83] T. Head, *The LHCb trigger system*, JINST **9** (2014) C09015.
- [84] E. Aslanides *et al.*, *The Level-0 muon trigger for the LHCb experiment*, Nuclear Instruments and Methods in Physics Research A **579** (2007) 989, [arXiv:0705.0310](https://arxiv.org/abs/0705.0310).
- [85] A. Martin Sanchez, P. Robbe, and M.-H. Schune, *Performances of the LHCb L0 Calorimeter Trigger*, Tech. Rep. LHCb-PUB-2011-026. CERN-LHCb-PUB-2011-026, CERN, Geneva, Jun, 2012. <http://cds.cern.ch/record/1407893>.
- [86] LHCb collaboration, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, JINST **10** (2015) P02007, [arXiv:1408.1251](https://arxiv.org/abs/1408.1251).
- [87] *PIDCalib Packages*, <https://twiki.cern.ch/twiki/bin/view/LHCb/PIDCalibPackage>. Accessed: November 2018.

- [88] M. Clemencic *et al.*, *The LHCb simulation application, Gauss: Design, evolution and experience*, J. Phys. Conf. Ser. **331** (2011) 032023.
- [89] T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026, [arXiv:hep-ph/0603175](#).
- [90] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, [arXiv:0710.3820](#).
- [91] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152.
- [92] 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](#).
- [93] Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.
- [94] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270.
- [95] A. Ali, P. Ball, L. T. Handoko, and G. Hiller, *A comparative study of the decays $B \rightarrow (K, K^*)\ell^+\ell^-$ in standard model and supersymmetric theories*, Phys. Rev. **D61** (2000) 074024, [arXiv:hep-ph/9910221](#).
- [96] D. Melikhov and B. Stech, *Weak form-factors for heavy meson decays: An Update*, Phys. Rev. **D62** (2000) 014006, [arXiv:hep-ph/0001113](#).
- [97] W. D. Hulsbergen, *Decay chain fitting with a Kalman filter*, Nucl. Instrum. Meth. **A552** (2005) 566, [arXiv:physics/0503191](#).
- [98] LHCb collaboration, R. Aaij *et al.*, *Measurement of the phase difference between short- and long-distance amplitudes in the $B^+ \rightarrow K^+\mu^+\mu^-$ decay*, Eur. Phys. J. **C77** (2017) 161, [arXiv:1612.06764](#).
- [99] M. Williams *et al.*, *The HLT2 Topological Lines*, Tech. Rep. LHCb-PUB-2011-002. CERN-LHCb-PUB-2011-002, CERN, Geneva, Jan, 2011. <http://cds.cern.ch/record/1323557>.
- [100] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*, Wadsworth international group, Belmont, California, USA, 1984.

- [101] A. Hoecker *et al.*, *TMVA - Toolkit for Multivariate Data Analysis*, ArXiv Physics e-prints (2007) [arXiv:physics/0703039](#).
- [102] A. Blum, A. Kalai, and J. Langford, *Beating the Hold-Out: Bounds for K-fold and Progressive Cross-Validation*, in *Proceedings of the Twelfth Annual Conference on Computational Learning Theory, COLT 1999, Santa Cruz, CA, USA, July 7-9, 1999*, pp. 203–208, 1999. doi: 10.1145/307400.307439.
- [103] P. J. Huber, in *Robust Estimation of a Location Parameter*, S. Kotz and N. L. Johnson, eds., pp. 492–518, Springer New York, New York, NY, 1992. doi: 10.1007/978-1-4612-4380-9_35.
- [104] R. E. Schapire and Y. Freund, *A decision-theoretic generalization of on-line learning and an application to boosting*, J. Comput. Syst. Sci. **55** (1997) 119.
- [105] LHCb collaboration, R. Aaij *et al.*, *Measurement of the fragmentation fraction ratio f_s/f_d and its dependence on B meson kinematics*, JHEP **04** (2013) 001, [arXiv:1301.5286](#).
- [106] M. Pivk and F. R. Le Diberder, *sPlot: A statistical tool to unfold data distributions*, Nucl. Instrum. Meth. **A555** (2005) 356, [arXiv:physics/0402083](#).
- [107] R. Aaij *et al.*, *Selection and processing of calibration samples to measure the particle identification performance of the LHCb experiment in Run 2*, [arXiv:1803.00824](#).
- [108] C. Burr *et al.*, *Prompt charm production in pp collisions at $\sqrt{s} = 13$ TeV in 2015 data*, tech. rep., Apr, 2015. <https://cds.cern.ch/record/2012800>.
- [109] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, eConf **C0303241** (2003) MOLT007, [arXiv:physics/0306116](#), [,186(2003)].
- [110] F. a. James, *MINUIT: Function Minimization and Error Analysis Reference Manual*, 1998. CERN Program Library Long Writeups, <https://cds.cern.ch/record/2296388>.
- [111] K. S. Cranmer, *Kernel estimation in high-energy physics*, Comput. Phys. Commun. **136** (2001) 198, [arXiv:hep-ex/0011057](#).

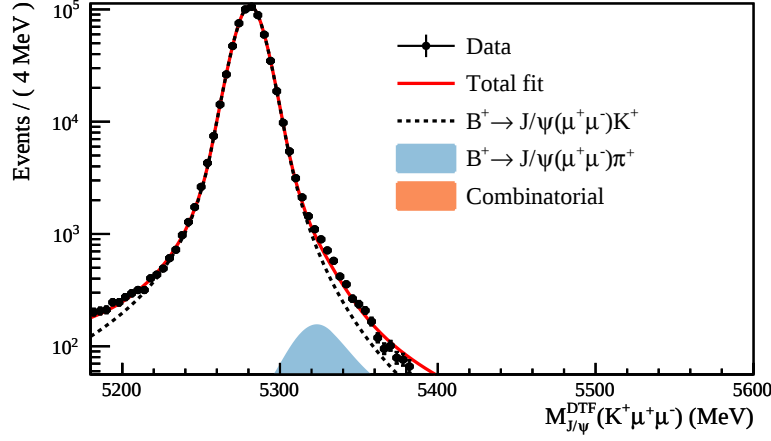
- [112] T. Skwarnicki, *A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances*, PhD thesis, Institute of Nuclear Physics, Krakow, 1986, DESY-F31-86-02.
- [113] D. Martinez Santos and F. Dupertuis, *Mass distributions marginalized over per-event errors*, Nucl. Instrum. Meth. **A764** (2014) 150, [arXiv:1312.5000](#).
- [114] S. S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, Ann. Math. Statist. **9** (1938) 60.
- [115] B. Efron, in *Bootstrap Methods: Another Look at the Jackknife*, S. Kotz and N. L. Johnson, eds., pp. 569–593, Springer New York, New York, NY, 1992. doi: 10.1007/978-1-4612-4380-9_41.
- [116] S. Beranek *et al.*, *Measurement of the electron detection and reconstruction efficiency*, <https://twiki.cern.ch/twiki/pub/LHCb/Velo2longElectronTrackingEfficiency/LHCb-ElEff-ANA-v1r2.pdf>.
- [117] D. M. Straub, *flavio: a Python package for flavour and precision phenomenology in the Standard Model and beyond*, [arXiv:1810.08132](#).
- [118] B. Capdevila, S. Descotes-Genon, J. Matias, and J. Virto, *Assessing lepton-flavour non-universality from $B \rightarrow K^* \ell \ell$ angular analyses*, JHEP **10** (2016) 075, [arXiv:1605.03156](#).
- [119] LHCb collaboration, 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](#).
- [120] Belle collaboration, S. Hirose *et al.*, *Measurement of the τ lepton polarization and $R(D^*)$ in the decay $\bar{B} \rightarrow D^* \tau^- \bar{\nu}_\tau$ with one-prong hadronic τ decays at Belle*, Phys. Rev. **D97** (2018) 012004, [arXiv:1709.00129](#).

Part III

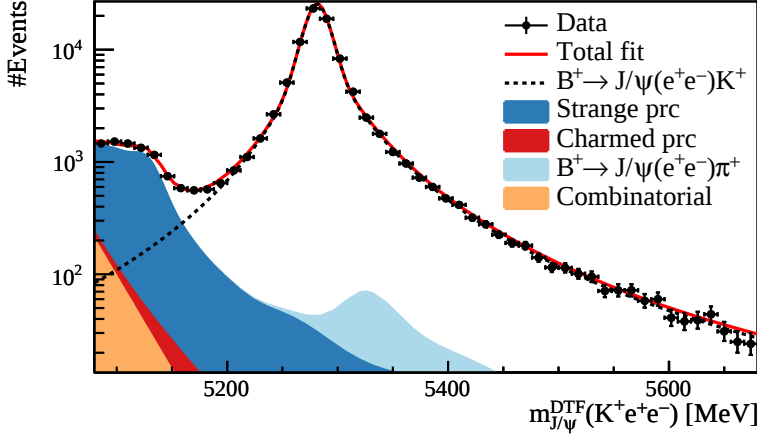
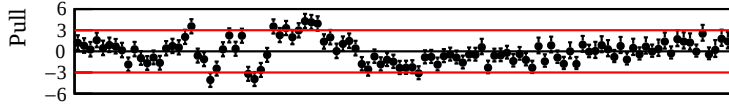
Appendix

A. Fit to the J/ψ mode in all data-taking conditions

The fits to m_{DTF} that were not shown in Sec. 8.2 are shown here instead. The signal yields extracted from these fits are used as an input to R_K . The fits to $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ events in the Run 1 data and $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ in the Run 1 data and selected in the $e\text{TOS}$ trigger category are shown in Fig. A.1. The fits to $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ events in the Run 1 and Run 2 data and selected in the $h\text{TOS!}$ trigger category are shown in Fig. A.2. The fits to $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ events in the Run 1 and Run 2 data and selected in the TIS! trigger category are shown in Fig. A.3.



$N_{\text{sig}} =$	618300 ± 800
$N_{\text{comb}} =$	2280 ± 140
$\Delta\mu =$	1.219 ± 0.013
$s_\sigma =$	1.1600 ± 0.0014
$\lambda =$	-0.0057 ± 0.0004
$\frac{N_\pi}{N_K} =$	$0.288 \pm 0.011\%$
$\in$	$0.246 \pm 0.012\%$



$N_{\text{sig}} =$	90220 ± 310
$\frac{N_{\text{charm}}}{N_{\text{strange}}} =$	0.027 ± 0.004
$N_{\text{prc}} =$	7680 ± 280
$N_{\text{comb}} =$	470 ± 250
$\Delta\mu =$	1.32 ± 0.06
$s_\sigma =$	1.112 ± 0.006
$\lambda =$	-0.038 ± 0.011
$r_{0\gamma} =$	$23.97 \pm 0.24\%$
$\in$	$24.11 \pm 0.24\%$
$r_{1\gamma} =$	$0.492 \pm 0.5\%$
$\in$	$50.0 \pm 0.5\%$
$\frac{N_\pi}{N_K} =$	0.47 ± 0.05
$\in$	$0.52 \pm 0.05\%$

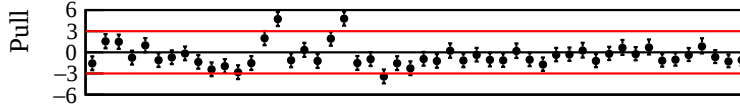


Figure A.1: Fit to $m_{\text{DTF}}^{J/\psi}$ for $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ events (top) and $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ events selected in the $e\text{TOS}$ trigger category (bottom). Both fits correspond to the Run 1 data-taking conditions. The pull between the fit function and the data points is also shown. The different components of the fit are indicated in the legend, where ‘prc’ stands for partially-reconstructed background. The free-floating parameters are displayed in the table to the right of the figure. Constrained parameters are shown at the bottom of the table, where the notation $\in \mu \pm \sigma$ indicates the mean (μ) and standard deviation (σ) of their Gaussian constraint.

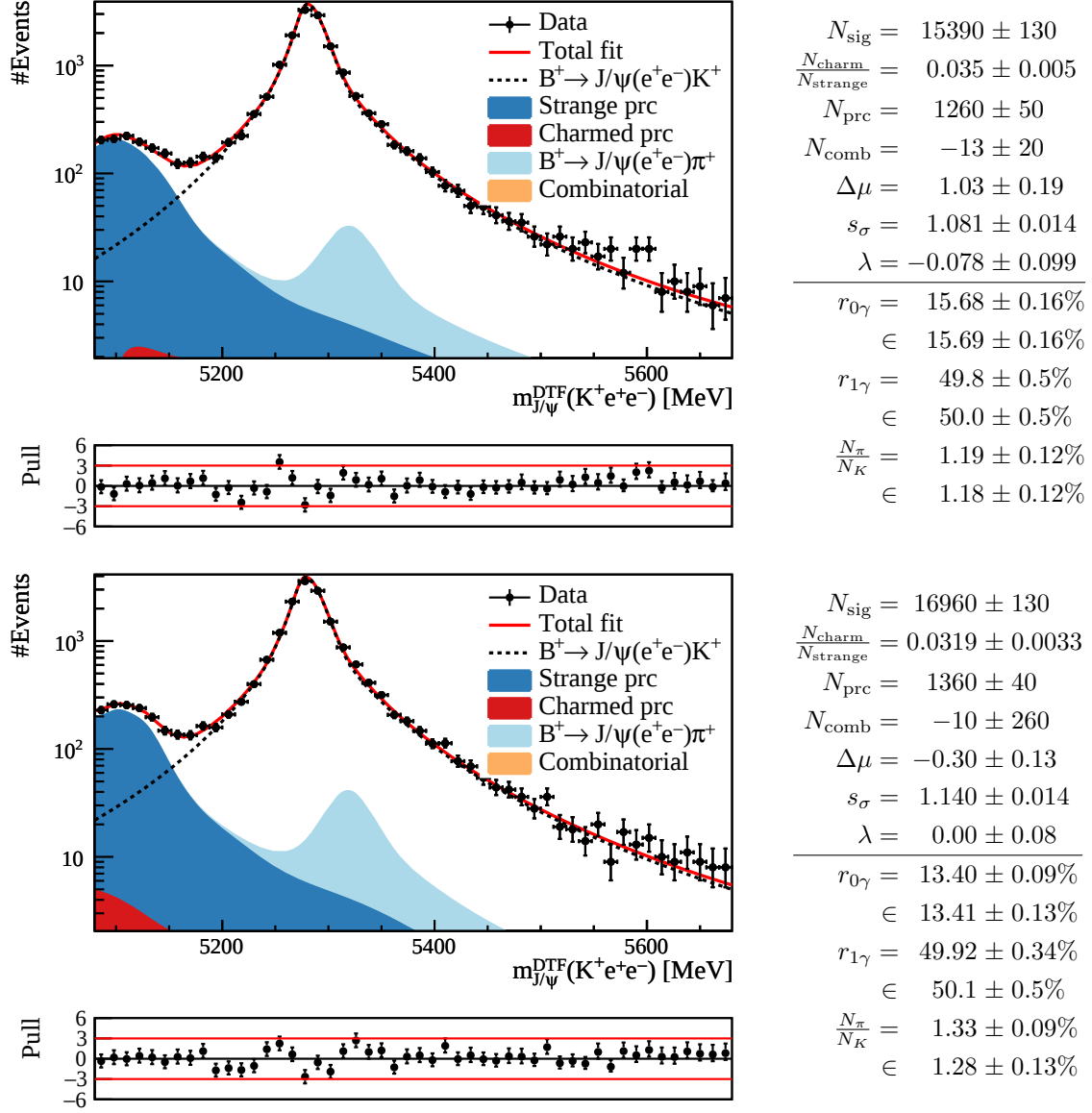


Figure A.2: Fit to $m_{\text{DTF}}^{J/\psi}$ for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the Run 1 (top) and Run 2 (bottom) data-taking conditions selected in the $h\text{TOS!}$ trigger category, and the pull between the fit function and the data points. The different components of the fit are indicated in the legend, where ‘prc’ stands for partially-reconstructed background. The free-floating parameters are displayed in the table to the right of the figure. Constrained parameters are shown at the bottom of the table, where the notation $\in \mu \pm \sigma$ indicates the mean (μ) and standard deviation (σ) of their Gaussian constraint.

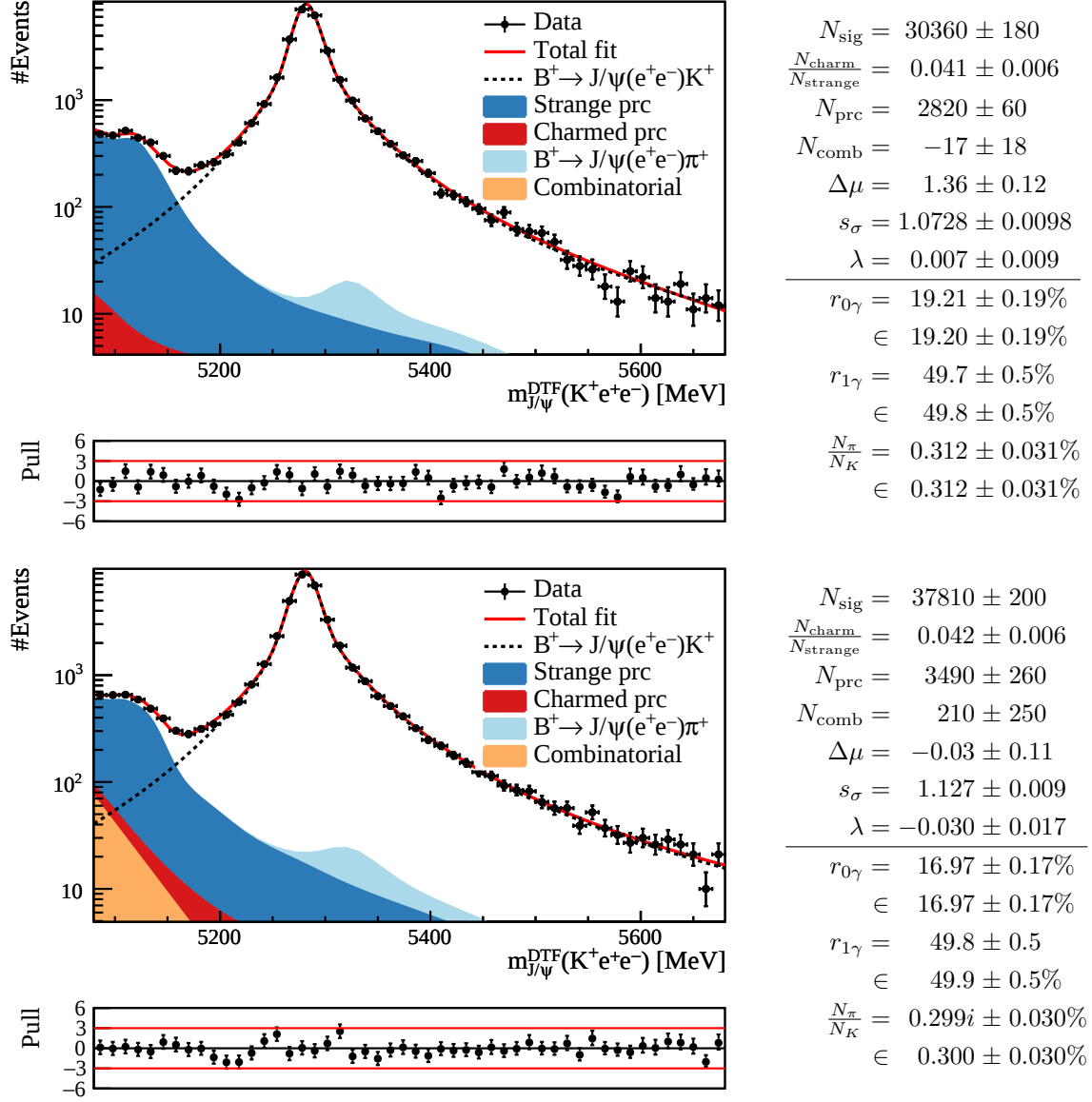


Figure A.3: Fit to $m_{\text{D}^{\text{TF}}}^{J/\psi}$ for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the Run 1 (top) and Run 2 (bottom) data-taking conditions selected in the TIS! trigger category, and the pull between the fit function and the data points. The different components of the fit are indicated in the legend, where ‘prc’ stands for partially-reconstructed background. The free-floating parameters are displayed in the table to the right of the figure. Constrained parameters are shown at the bottom of the table, where the notation $\in \mu \pm \sigma$ indicates the mean (μ) and standard deviation (σ) of their Gaussian constraint.

B. Supplementary kinematic distributions

Distributions of kinematic variables not shown in Sec. 7.7 are shown in Fig. B.1 for the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the Run 2 data-taking conditions, and in Fig. B.2 for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the Run 2 data-taking conditions. Kinematic distributions in the Run 1 data-taking conditions are shown in Fig. B.3 for the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode and in Fig. B.4 for the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode.

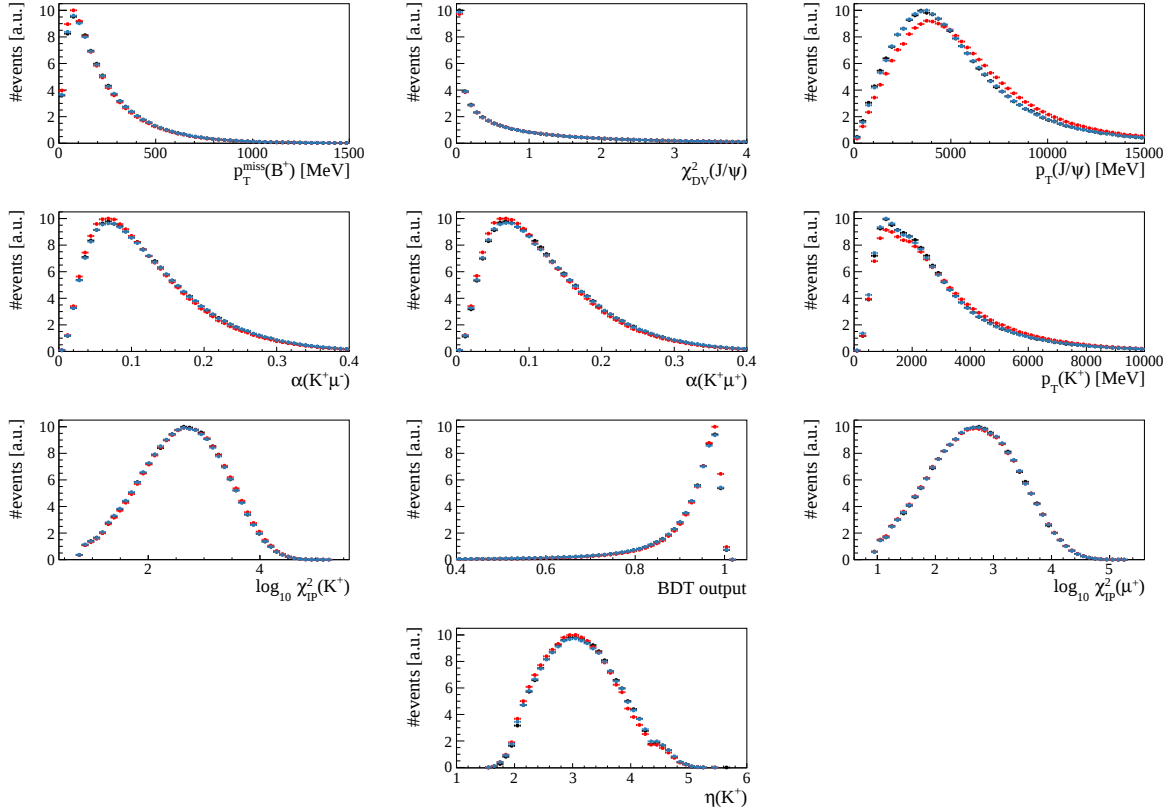


Figure B.1: Kinematic distributions for preselected $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ events in the μ TOS category and in the Run 2 data-taking conditions. Each plot shows the background-subtracted data (black), simulation (red) and simulation with kinematic correction weights (blue).

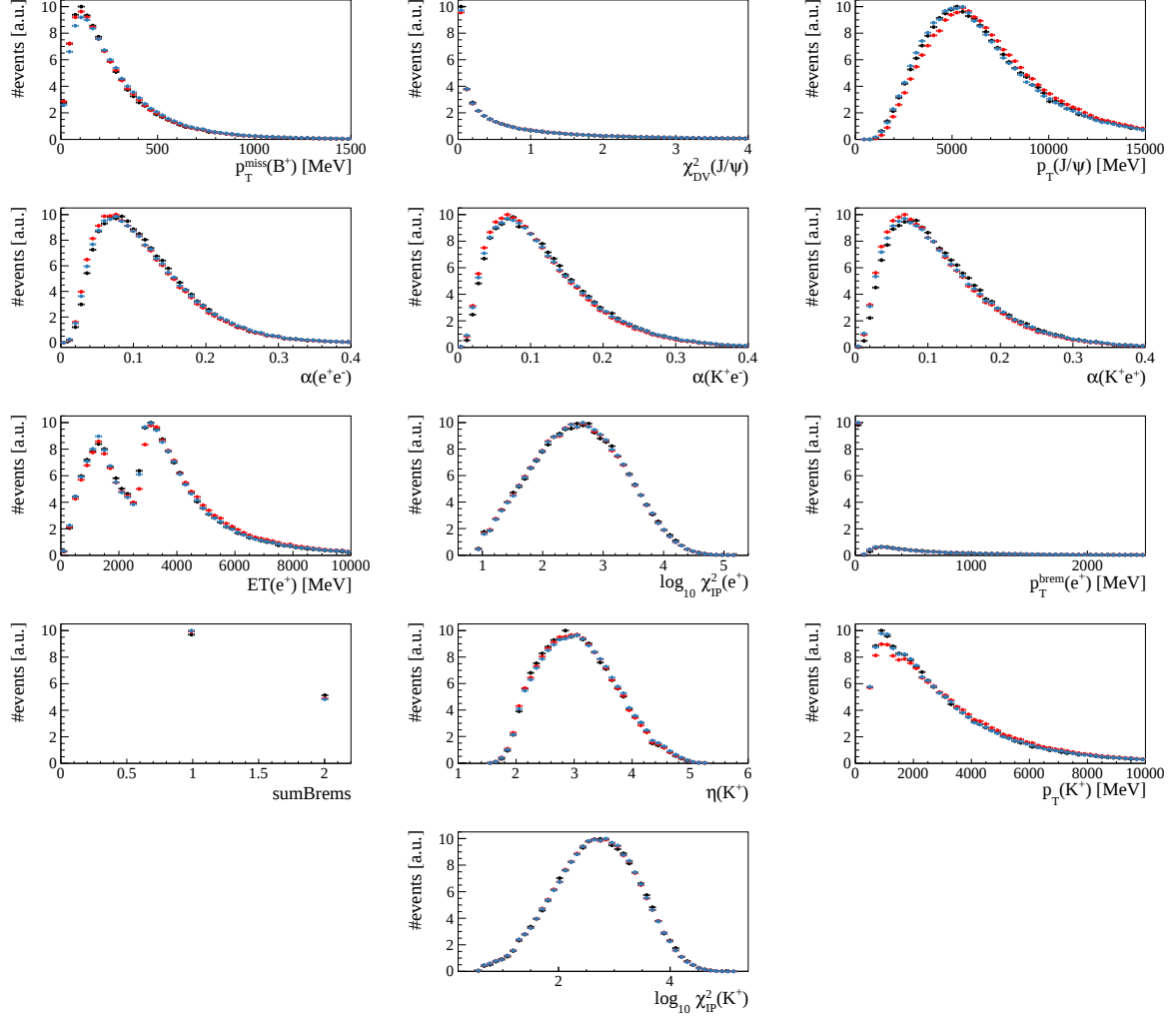


Figure B.2: Kinematic distributions for preselected $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the e TOS category and in the Run 2 data-taking conditions. Each plot shows the background-subtracted data (black), simulation (red) and simulation with kinematic correction weights (blue).

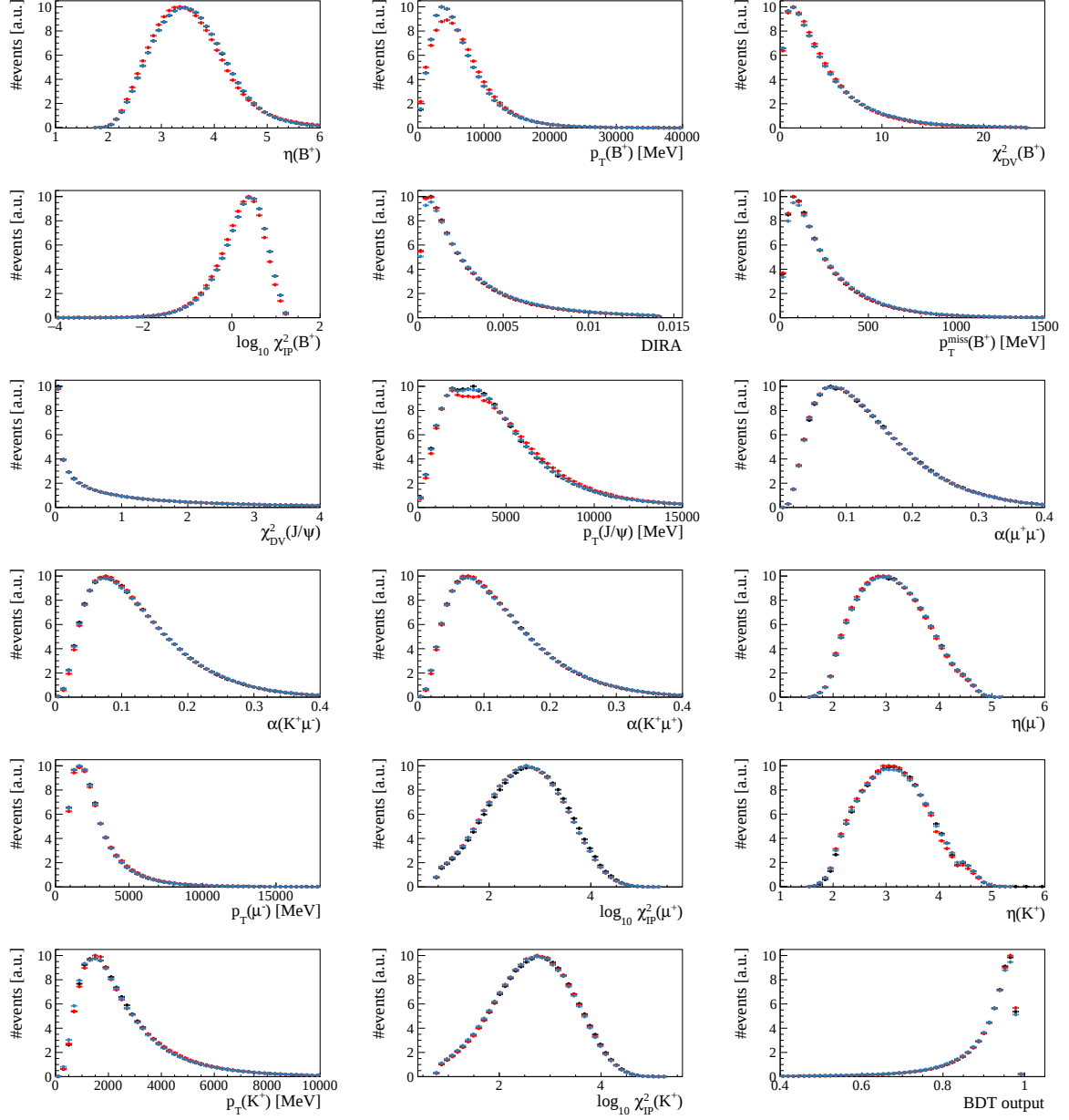


Figure B.3: Kinematic distributions for preselected $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ events in the μ TOS category and in the Run 1 data-taking conditions. Each plot shows the background-subtracted data (black), simulation (red) and simulation with kinematic correction weights (blue).

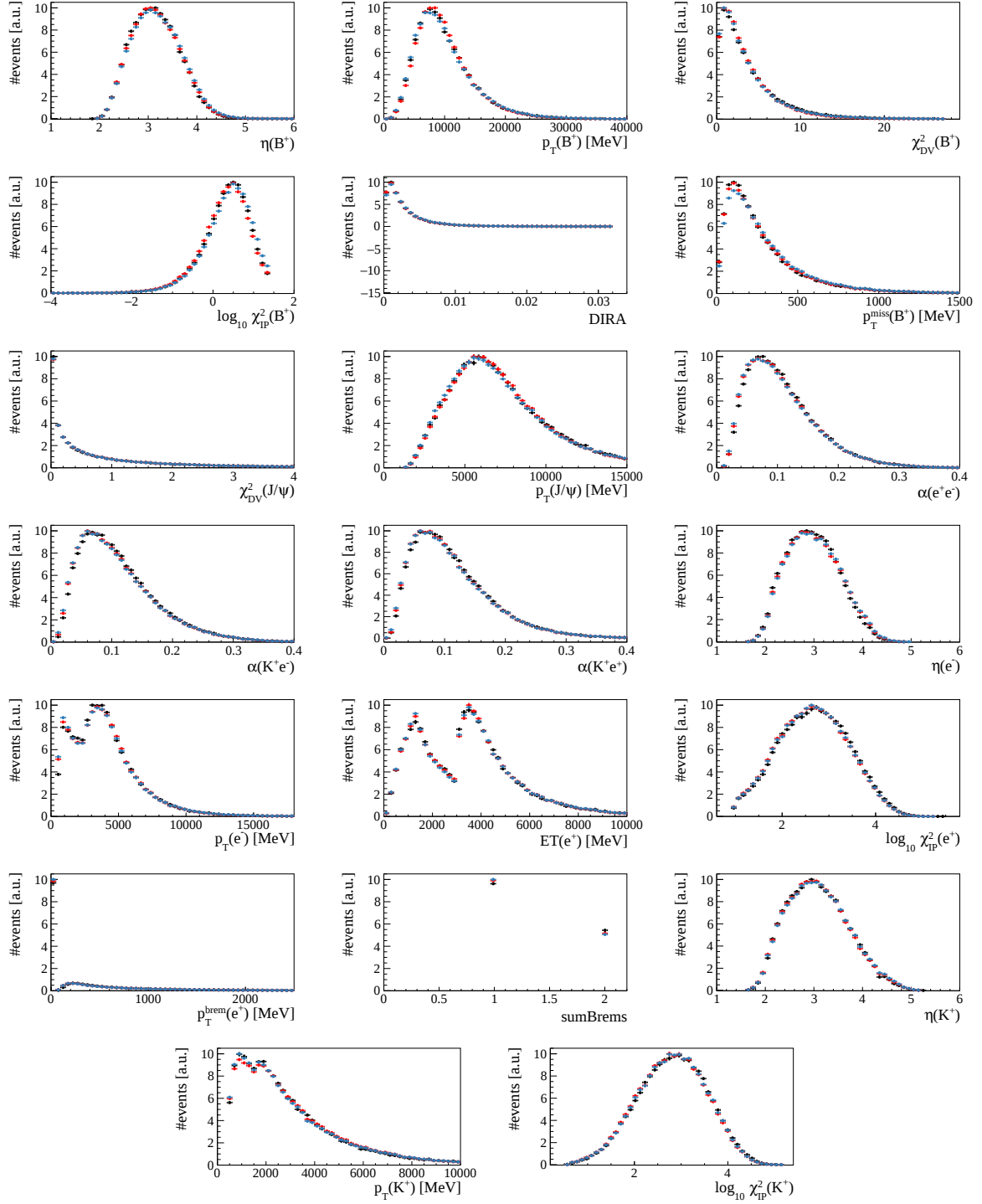


Figure B.4: Kinematic distributions for preselected $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ events in the e TOS category and in the Run 1 data-taking conditions. Each plot shows the background-subtracted data (black), simulation (red) and simulation with kinematic correction weights (blue).

C. Complete list of efficiencies

The complete lists of efficiencies for $B^+ \rightarrow K^+ J/\psi(\ell^+ \ell^-)$ and $B^+ \rightarrow K^+ \ell^+ \ell^-$ decays, factorised according to Eq. 7.1, is presented. The efficiencies are computed without taking the ghosts into account, for the reasons explained in Sec. 7.1. are shown in this section. The efficiencies corresponding to Run 1 (Run 2) computed with no correction are shown in Tab. C.1 (Tab. C.2), the ones computed with trigger and PID corrections and shown in Tab. C.3 (Tab. C.4), and the ones computed with all corrections, that is trigger, PID and kinematics, are shown in Tab. C.5 (Tab. C.6).

Particularly in the case of the electron mode, the trigger and PID efficiencies vary quite significantly depending on whether or not the correction weights are applied. However, the efficiency ratio between the rare and control mode $\varepsilon(B^+ \rightarrow K^+ e^+ e^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(e^+ e^-))$ do not vary by more than $\sim 2\%$ for the *e*TOS and TIS! categories, and by 20% for the *h*TOS! category. However, very few signal events are expected in the latter category. This highlights the robustness of the R_K measurement performed as a double ratio.

Table C.1: Efficiencies for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$, $B^+ \rightarrow K^+ e^+e^-$, $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow K^+ \mu^+\mu^-$ decays, as well as efficiency ratios $\varepsilon(B^+ \rightarrow K^+ e^+e^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(e^+e^-))$ and $\varepsilon(B^+ \rightarrow K^+ \mu^+\mu^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-))$, in the Run 1 data-taking conditions. All efficiencies and ratios are displayed in percent, and the uncertainty quoted is the statistical uncertainty related to the finite size of the simulation samples. The efficiencies are computed with no corrections whatsoever to the simulation.

Run 1 no corrections			
	$J/\psi \ \varepsilon \ [\%]$	rare $\varepsilon \ [\%]$	ratio $\varepsilon \ [\%]$
	electron modes		
$\varepsilon^{\text{geom}}$	16.06 ± 0.03	16.32 ± 0.03	101.57 ± 0.22
$\varepsilon^{\text{rec,strip}}$	13.33 ± 0.02	12.82 ± 0.02	96.18 ± 0.19
$\varepsilon^{\text{presel}}$	76.96 ± 0.06	26.00 ± 0.07	33.79 ± 0.09
ε^{PID}	81.95 ± 0.11	80.00 ± 0.33	97.63 ± 0.42
$\varepsilon^{\text{eTOS}}$	23.47 ± 0.07	19.51 ± 0.14	83.15 ± 0.64
ε^{BDT}	87.89 ± 0.12	86.86 ± 0.27	98.82 ± 0.33
ε^{tot}	0.278 ± 0.001	0.074 ± 0.001	26.49 ± 0.25
$\varepsilon^{h\text{TOS!}}$	4.16 ± 0.03	7.13 ± 0.09	171.41 ± 2.59
ε^{BDT}	96.49 ± 0.16	94.28 ± 0.30	97.71 ± 0.35
ε^{tot}	0.054 ± 0.000	0.029 ± 0.000	53.98 ± 0.87
$\varepsilon^{\text{TIS!}}$	8.44 ± 0.05	8.22 ± 0.10	97.38 ± 1.27
ε^{BDT}	82.60 ± 0.22	78.61 ± 0.50	95.17 ± 0.66
ε^{tot}	0.094 ± 0.001	0.028 ± 0.000	29.87 ± 0.46
	muon modes		
$\varepsilon^{\text{geom}}$	16.14 ± 0.03	16.42 ± 0.03	101.72 ± 0.22
$\varepsilon^{\text{rec,strip}}$	23.98 ± 0.01	23.67 ± 0.03	98.70 ± 0.14
$\varepsilon^{\text{presel}}$	60.22 ± 0.03	14.66 ± 0.05	24.35 ± 0.09
ε^{PID}	94.68 ± 0.08	93.54 ± 0.47	98.79 ± 0.50
$\varepsilon^{\mu\text{TOS}}$	76.49 ± 0.04	72.86 ± 0.17	95.26 ± 0.23
ε^{BDT}	87.78 ± 0.03	89.48 ± 0.14	101.94 ± 0.17
ε^{tot}	1.481 ± 0.003	0.347 ± 0.002	23.46 ± 0.13

Table C.2: Efficiencies for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$, $B^+ \rightarrow K^+ e^+e^-$, $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow K^+ \mu^+\mu^-$ decays, as well as efficiency ratios $\varepsilon(B^+ \rightarrow K^+ e^+e^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(e^+e^-))$ and $\varepsilon(B^+ \rightarrow K^+ \mu^+\mu^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-))$, in the Run 1 data-taking conditions. All efficiencies and ratios are displayed in percent, and the uncertainty quoted is the statistical uncertainty related to the finite size of the simulation samples. The efficiencies are computed with no corrections whatsoever to the simulation.

Run 2 no corrections			
	$J/\psi \ \varepsilon \ [\%]$	rare $\varepsilon \ [\%]$	ratio $\varepsilon \ [\%]$
	electron modes		
$\varepsilon^{\text{geom}}$	17.30 ± 0.03	17.58 ± 0.03	101.60 ± 0.24
$\varepsilon^{\text{rec,strip}}$	14.19 ± 0.01	13.70 ± 0.02	96.58 ± 0.18
$\varepsilon^{\text{presel}}$	75.39 ± 0.05	25.94 ± 0.07	34.40 ± 0.10
ε^{PID}	82.12 ± 0.09	80.87 ± 0.35	98.48 ± 0.44
$\varepsilon^{\text{eTOS}}$	38.37 ± 0.07	31.98 ± 0.17	83.34 ± 0.47
ε^{BDT}	88.34 ± 0.08	87.47 ± 0.22	99.02 ± 0.26
ε^{tot}	0.515 ± 0.002	0.141 ± 0.001	27.43 ± 0.21
$\varepsilon^{h\text{TOS!}}$	4.25 ± 0.03	9.13 ± 0.10	214.86 ± 2.86
ε^{BDT}	90.81 ± 0.20	82.43 ± 0.46	90.76 ± 0.54
ε^{tot}	0.059 ± 0.000	0.038 ± 0.001	64.83 ± 0.99
$\varepsilon^{\text{TIS!}}$	8.62 ± 0.04	8.88 ± 0.10	102.99 ± 1.30
ε^{BDT}	88.65 ± 0.16	84.56 ± 0.44	95.39 ± 0.53
ε^{tot}	0.116 ± 0.001	0.038 ± 0.001	32.66 ± 0.47
	muon modes		
$\varepsilon^{\text{geom}}$	17.38 ± 0.03	17.65 ± 0.03	101.58 ± 0.24
$\varepsilon^{\text{rec,strip}}$	24.10 ± 0.01	23.50 ± 0.03	97.48 ± 0.12
$\varepsilon^{\text{presel}}$	59.64 ± 0.03	14.71 ± 0.05	24.66 ± 0.08
ε^{PID}	95.34 ± 0.07	94.99 ± 0.42	99.63 ± 0.44
$\varepsilon^{\mu\text{TOS}}$	72.24 ± 0.04	68.93 ± 0.16	95.41 ± 0.22
ε^{BDT}	91.38 ± 0.03	93.33 ± 0.10	102.13 ± 0.12
ε^{tot}	1.573 ± 0.003	0.373 ± 0.002	23.71 ± 0.12

Table C.3: Efficiencies for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$, $B^+ \rightarrow K^+ e^+e^-$, $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow K^+ \mu^+\mu^-$ decays, as well as efficiency ratios $\varepsilon(B^+ \rightarrow K^+ e^+e^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(e^+e^-))$ and $\varepsilon(B^+ \rightarrow K^+ \mu^+\mu^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-))$, in the Run 1 data-taking conditions. All efficiencies and ratios are displayed in percent, and the uncertainty quoted is the statistical uncertainty related to the finite size of the simulation samples. The efficiencies are computed including the L0, HLT and PID corrections to the simulation.

Run 1 PID and trigger corrections

	J/ψ ε [%]	rare ε [%]	ratio ε [%]
	electron modes		
$\varepsilon^{\text{geom}}$	16.06 ± 0.03	16.32 ± 0.03	101.57 ± 0.22
$\varepsilon^{\text{rec,strip}}$	13.33 ± 0.02	12.82 ± 0.02	96.18 ± 0.19
$\varepsilon^{\text{presel}}$	76.08 ± 0.06	26.16 ± 0.07	34.38 ± 0.10
ε^{PID}	76.53 ± 0.08	76.61 ± 0.16	100.11 ± 0.24
$\varepsilon^{e\text{TOS}}$	21.76 ± 0.06	17.96 ± 0.12	82.50 ± 0.59
ε^{BDT}	88.10 ± 0.11	86.96 ± 0.24	98.71 ± 0.30
ε^{tot}	0.239 ± 0.001	0.065 ± 0.001	27.39 ± 0.24
$\varepsilon^{h\text{TOS!}}$	3.24 ± 0.03	5.67 ± 0.07	174.60 ± 2.47
ε^{BDT}	96.86 ± 0.13	94.65 ± 0.26	97.72 ± 0.30
ε^{tot}	0.039 ± 0.000	0.022 ± 0.000	56.00 ± 0.84
$\varepsilon^{\text{TIS!}}$	6.55 ± 0.04	6.34 ± 0.07	96.72 ± 1.20
ε^{BDT}	84.34 ± 0.19	80.84 ± 0.43	95.85 ± 0.55
ε^{tot}	0.070 ± 0.000	0.021 ± 0.000	30.69 ± 0.44
	muon modes		
$\varepsilon^{\text{geom}}$	16.14 ± 0.03	16.42 ± 0.03	101.72 ± 0.22
$\varepsilon^{\text{rec,strip}}$	23.98 ± 0.01	23.67 ± 0.03	98.70 ± 0.14
$\varepsilon^{\text{presel}}$	60.22 ± 0.03	14.66 ± 0.05	24.35 ± 0.09
ε^{PID}	91.20 ± 0.03	90.37 ± 0.15	99.09 ± 0.16
$\varepsilon^{\mu\text{TOS}}$	75.81 ± 0.04	72.29 ± 0.17	95.36 ± 0.23
ε^{BDT}	87.81 ± 0.03	89.49 ± 0.14	101.92 ± 0.16
ε^{tot}	1.415 ± 0.003	0.333 ± 0.002	23.54 ± 0.12

Table C.4: Efficiencies for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$, $B^+ \rightarrow K^+ e^+e^-$, $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow K^+ \mu^+\mu^-$ decays, as well as efficiency ratios $\varepsilon(B^+ \rightarrow K^+ e^+e^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(e^+e^-))$ and $\varepsilon(B^+ \rightarrow K^+ \mu^+\mu^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-))$, in the Run 2 data-taking conditions. All efficiencies and ratios are displayed in percent, and the uncertainty quoted is the statistical uncertainty related to the finite size of the simulation samples. The efficiencies are computed including the L0, HLT and PID corrections to the simulation.

Run 2 PID and trigger corrections

	J/ψ ε [%]	rare ε [%]	ratio ε [%]
	electron modes		
$\varepsilon^{\text{geom}}$	17.30 ± 0.03	17.58 ± 0.03	101.60 ± 0.24
$\varepsilon^{\text{rec,strip}}$	14.19 ± 0.01	13.70 ± 0.02	96.58 ± 0.18
$\varepsilon^{\text{presel}}$	74.02 ± 0.05	26.44 ± 0.07	35.71 ± 0.10
ε^{PID}	77.52 ± 0.07	77.50 ± 0.16	99.98 ± 0.23
$\varepsilon^{e\text{TOS}}$	39.05 ± 0.07	32.30 ± 0.16	82.73 ± 0.42
ε^{BDT}	88.53 ± 0.07	87.35 ± 0.20	98.66 ± 0.23
ε^{tot}	0.487 ± 0.001	0.139 ± 0.001	28.60 ± 0.21
$\varepsilon^{h\text{TOS!}}$	4.26 ± 0.03	8.83 ± 0.09	207.30 ± 2.58
ε^{BDT}	91.00 ± 0.19	82.77 ± 0.42	90.95 ± 0.50
ε^{tot}	0.055 ± 0.000	0.036 ± 0.000	65.13 ± 0.93
$\varepsilon^{\text{TIS!}}$	8.67 ± 0.04	8.89 ± 0.09	102.55 ± 1.18
ε^{BDT}	88.99 ± 0.14	84.59 ± 0.40	95.06 ± 0.47
ε^{tot}	0.110 ± 0.001	0.037 ± 0.000	33.76 ± 0.45
	muon modes		
$\varepsilon^{\text{geom}}$	17.38 ± 0.03	17.65 ± 0.03	101.58 ± 0.24
$\varepsilon^{\text{rec,strip}}$	24.10 ± 0.01	23.50 ± 0.03	97.48 ± 0.12
$\varepsilon^{\text{presel}}$	59.64 ± 0.03	14.71 ± 0.05	24.66 ± 0.08
ε^{PID}	93.05 ± 0.03	92.67 ± 0.11	99.59 ± 0.13
$\varepsilon^{\mu\text{TOS}}$	72.46 ± 0.04	69.07 ± 0.15	95.32 ± 0.22
ε^{BDT}	91.31 ± 0.03	93.31 ± 0.10	102.20 ± 0.11
ε^{tot}	1.538 ± 0.003	0.364 ± 0.002	23.69 ± 0.12

Table C.5: Efficiencies for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$, $B^+ \rightarrow K^+ e^+e^-$, $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow K^+ \mu^+\mu^-$ decays, as well as efficiency ratios $\varepsilon(B^+ \rightarrow K^+ e^+e^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(e^+e^-))$ and $\varepsilon(B^+ \rightarrow K^+ \mu^+\mu^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-))$, in the Run 1 data-taking conditions. All efficiencies and ratios are displayed in percent, and the uncertainty quoted is the statistical uncertainty related to the finite size of the simulation samples. The efficiencies are computed with all corrections applied to the simulation, including trigger, PID and kinematic weights, which is the nominal way to compute the efficiencies.

Run 1 all corrections (nominal)			
	$J/\psi \ \varepsilon \ [\%]$	rare $\varepsilon \ [\%]$	ratio $\varepsilon \ [\%]$
	electron modes		
$\varepsilon^{\text{geom}}$	16.09 ± 0.03	16.35 ± 0.03	101.59 ± 0.23
$\varepsilon^{\text{rec,strip}}$	13.12 ± 0.02	12.68 ± 0.02	96.67 ± 0.19
$\varepsilon^{\text{presel}}$	75.51 ± 0.06	26.18 ± 0.07	34.68 ± 0.10
ε^{PID}	76.50 ± 0.08	76.62 ± 0.17	100.15 ± 0.24
$\varepsilon^{e\text{TOS}}$	19.73 ± 0.06	16.49 ± 0.11	83.59 ± 0.62
ε^{BDT}	86.82 ± 0.12	85.61 ± 0.27	98.61 ± 0.34
ε^{tot}	0.209 ± 0.001	0.059 ± 0.000	28.11 ± 0.25
$\varepsilon^{h\text{TOS!}}$	2.88 ± 0.02	5.20 ± 0.06	180.84 ± 2.59
ε^{BDT}	96.44 ± 0.15	93.93 ± 0.29	97.39 ± 0.34
ε^{tot}	0.034 ± 0.000	0.020 ± 0.000	58.64 ± 0.89
$\varepsilon^{\text{TIS!}}$	6.57 ± 0.04	6.37 ± 0.07	97.03 ± 1.20
ε^{BDT}	82.67 ± 0.21	79.09 ± 0.46	95.67 ± 0.61
ε^{tot}	0.067 ± 0.000	0.021 ± 0.000	31.17 ± 0.45
	muon modes		
$\varepsilon^{\text{geom}}$	16.17 ± 0.03	16.46 ± 0.03	101.76 ± 0.23
$\varepsilon^{\text{rec,strip}}$	23.86 ± 0.01	23.46 ± 0.03	98.34 ± 0.14
$\varepsilon^{\text{presel}}$	60.01 ± 0.03	14.52 ± 0.05	24.20 ± 0.09
ε^{PID}	91.32 ± 0.03	90.62 ± 0.14	99.23 ± 0.16
$\varepsilon^{\mu\text{TOS}}$	75.40 ± 0.04	71.56 ± 0.18	94.90 ± 0.24
ε^{BDT}	86.73 ± 0.04	88.37 ± 0.15	101.89 ± 0.18
ε^{tot}	1.383 ± 0.003	0.321 ± 0.002	23.24 ± 0.13

Table C.6: Efficiencies for $B^+ \rightarrow K^+ J/\psi(e^+e^-)$, $B^+ \rightarrow K^+ e^+e^-$, $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow K^+ \mu^+\mu^-$ decays, as well as efficiency ratios $\varepsilon(B^+ \rightarrow K^+ e^+e^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(e^+e^-))$ and $\varepsilon(B^+ \rightarrow K^+ \mu^+\mu^-)/\varepsilon(B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-))$, in the Run 2 data-taking conditions. All efficiencies and ratios are displayed in percent, and the uncertainty quoted is the statistical uncertainty related to the finite size of the simulation samples. The efficiencies are computed with all corrections applied to the simulation, including trigger, PID and kinematic weights, which is the nominal way to compute the efficiencies.

Run 2 all corrections (nominal)			
	$J/\psi \ \varepsilon \ [\%]$	rare $\varepsilon \ [\%]$	ratio $\varepsilon \ [\%]$
	electron modes		
$\varepsilon^{\text{geom}}$	17.58 ± 0.03	17.87 ± 0.03	101.64 ± 0.24
$\varepsilon^{\text{rec,strip}}$	13.66 ± 0.01	13.17 ± 0.02	96.37 ± 0.18
$\varepsilon^{\text{presel}}$	73.92 ± 0.05	26.35 ± 0.07	35.64 ± 0.10
ε^{PID}	77.41 ± 0.07	77.49 ± 0.17	100.10 ± 0.23
$\varepsilon^{e\text{TOS}}$	34.53 ± 0.06	28.23 ± 0.15	81.74 ± 0.45
ε^{BDT}	87.57 ± 0.08	86.41 ± 0.21	98.67 ± 0.26
ε^{tot}	0.416 ± 0.001	0.117 ± 0.001	28.18 ± 0.21
$\varepsilon^{h\text{TOS!}}$	3.61 ± 0.02	7.69 ± 0.08	212.80 ± 2.70
ε^{BDT}	90.28 ± 0.20	81.98 ± 0.44	90.80 ± 0.53
ε^{tot}	0.045 ± 0.000	0.030 ± 0.000	66.50 ± 0.96
$\varepsilon^{\text{TIS!}}$	8.73 ± 0.04	8.97 ± 0.10	102.73 ± 1.18
ε^{BDT}	88.09 ± 0.15	83.70 ± 0.42	95.02 ± 0.50
ε^{tot}	0.107 ± 0.001	0.036 ± 0.000	33.70 ± 0.45
	muon modes		
$\varepsilon^{\text{geom}}$	17.66 ± 0.03	17.94 ± 0.03	101.60 ± 0.24
$\varepsilon^{\text{rec,strip}}$	23.32 ± 0.01	22.70 ± 0.03	97.35 ± 0.12
$\varepsilon^{\text{presel}}$	59.40 ± 0.03	14.38 ± 0.05	24.21 ± 0.08
ε^{PID}	93.19 ± 0.03	92.83 ± 0.11	99.61 ± 0.13
$\varepsilon^{\mu\text{TOS}}$	68.98 ± 0.04	65.53 ± 0.16	94.99 ± 0.24
ε^{BDT}	90.02 ± 0.03	92.20 ± 0.12	102.42 ± 0.14
ε^{tot}	1.415 ± 0.003	0.328 ± 0.002	23.20 ± 0.12

D. $B^+ \rightarrow \pi^+ J/\psi(e^+e^-)$ mis-ID mass variable transformation

For the purpose of the fit, the mass distribution of the background stemming from a $B^+ \rightarrow \pi^+ J/\psi(e^+e^-)$ decay has to be determined. The π^+ in this decay is mis-identified as a K^+ and is assigned a wrong mass. Because of the wrong mass assignment, the B^+ mass distribution becomes very asymmetric, and peaks higher than the real B^+ mass. In this section, it is shown how the B^+ mass spectrum for a $B^+ \rightarrow \pi^+ J/\psi(e^+e^-)$ decay can be estimated using only $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ simulated events.

Consider a $B^+ \rightarrow \pi^+ J/\psi(e^+e^-)$ event reconstructed in two ways: in one case assigning the correct mass hypothesis to the pion track, yielding the measured energy E_π and reconstructed B^+ invariant mass m , and in the second case assigning the kaon mass hypothesis to the pion track, yielding the measured energy E'_π and reconstructed B^+ invariant mass m' . The difference between the squared pion and kaon masses is denoted $\Delta^2 m = m_K^2 - m_\pi^2$. The masses m and m' are given,

$$\begin{aligned} m^2 &= (E_{J/\psi} + E_\pi)^2 - (\mathbf{p}_{J/\psi} + \mathbf{p}_\pi)^2, \\ m'^2 &= (E_{J/\psi} + E'_\pi)^2 - (\mathbf{p}_{J/\psi} + \mathbf{p}_\pi)^2. \end{aligned}$$

Hence:

$$\begin{aligned} m'^2 - m^2 &= 2E_{J/\psi}(E'_\pi - E_\pi) + \Delta^2 m \\ &= 2E_{J/\psi} \left(\sqrt{m_K^2 + p_\pi^2} - \sqrt{m_\pi^2 + p_\pi^2} \right) + \Delta^2 m \\ &= 2E_{J/\psi} \left(\sqrt{m_\pi^2 + \Delta^2 m + p_\pi^2} - \sqrt{m_\pi^2 + p_\pi^2} \right) + \Delta^2 m \\ &\approx 2E_{J/\psi} \left(\frac{\Delta^2 m}{2\sqrt{m_\pi^2 + p_\pi^2}} \right) + \Delta^2 m \\ &= \frac{E_{\text{tot}}}{E_\pi} \Delta^2 m, \end{aligned} \tag{D.1}$$

where, in the fourth line, a Taylor expansion of the square roots at first order in m_K^2 is performed. Hence, this formula holds in the realistic case of $m_K \ll p_\pi$. Therefore, the B^+ mass with the incorrect mass assignment to the pion track can be expressed as a function

of the B^+ mass with the correct mass assignment,

$$m' \approx \left(m^2 + \frac{E_{\text{tot}}}{E_\pi} \Delta^2 m \right)^{\frac{1}{2}}. \quad (\text{D.2})$$

In the case of this analysis, no $B^+ \rightarrow \pi^+ J/\psi(e^+e^-)$ simulation sample is available, and only the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ sample can be used. In the limit $m_K \ll p_\pi$ where Eq. D.2 is valid, the energy of the K^+ , E_K in $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ has a very similar spectrum as the energy of the π^+ , E_π in $B^+ \rightarrow \pi^+ J/\psi(e^+e^-)$. Therefore, E_π can be substituted for E_K in Eq. D.2,

$$m' \approx \left(m^2 + \frac{E_{\text{tot}}}{E_K} \Delta^2 m \right)^{\frac{1}{2}}. \quad (\text{D.3})$$

This shows how the $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ sample can be used to obtain an estimated spectrum for the B mass reconstructed from a $B^+ \rightarrow \pi^+ J/\psi(e^+e^-)$ decay, where the π has been mis-identified as a K .

The impact of the particle ID selection is evaluated using the PIDCalib tool. Efficiency weights for $\pi \rightarrow K$ mis-ID as a function of the pion η and p corresponding to the particle ID requirements in the selection are derived. Subsequently, these weights are applied to the kaon in each $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ candidate. This procedure has been validated using the $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \rightarrow \pi^+ J/\psi(\mu^+\mu^-)$, for which MC simulation is available, see Fig. D.1.

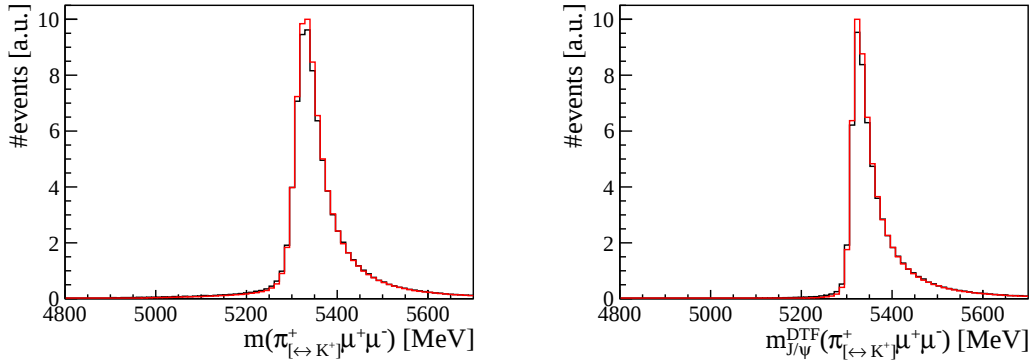


Figure D.1: Comparison of the (left) $m(\pi_{[\leftrightarrow K^+]}^+ \mu^+ \mu^-)$ and (right) $m_{\text{DTF}}^{J/\psi}(\pi_{[\leftrightarrow K^+]}^+ \mu^+ \mu^-)$ distributions obtained from (black) $B^+ \rightarrow \pi^+ J/\psi(\mu^+ \mu^-)$ simulated events with that obtained using (red) $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ through the transformation explained in the text.

E. $r_{J/\psi}$ differential in the h TOS! and TIS! categories

The $r_{J/\psi}$ histograms are shown for the Run 2 conditions, and for the electron mode selected in the h TOS! trigger category in Fig. E.1, and in the TIS! category in Fig. E.2.

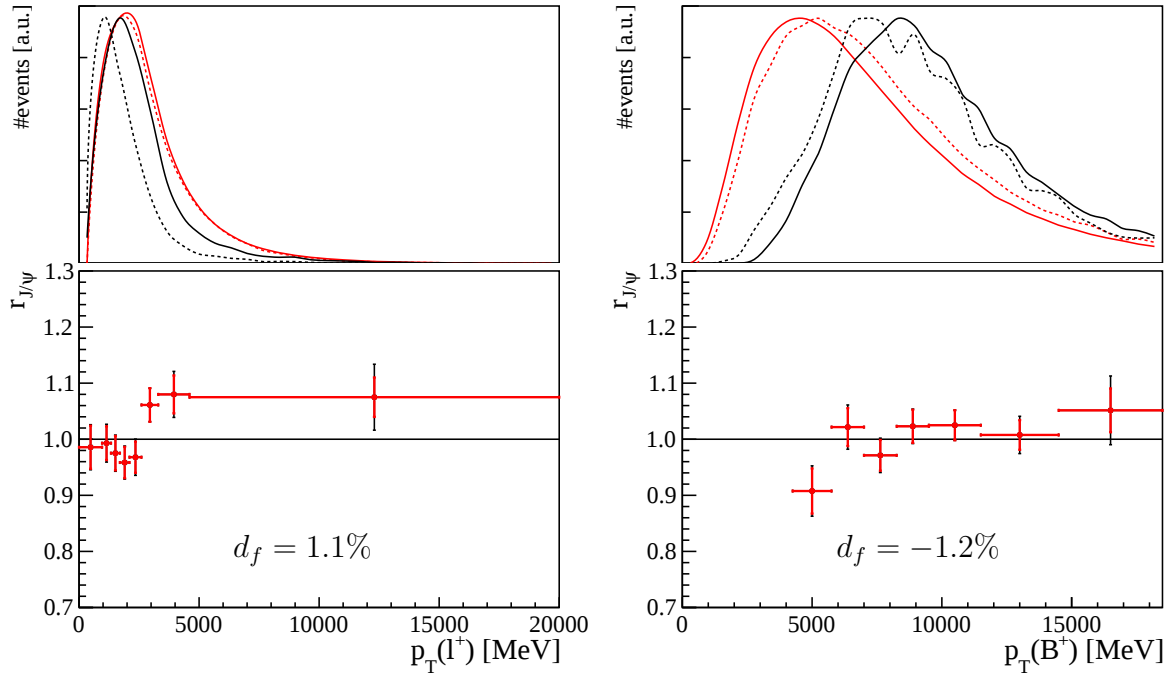


Figure E.1: Binned $r_{J/\psi}$ values (bottom two plots) for the preselected $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ mode in the h TOS! category and the preselected $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ mode in the μ TOS category, in the Run 2 data-taking conditions. The red thick error bars correspond to the statistical uncertainty and the black thin error bars to the systematic uncertainty. A large proportion of the bins have a statistically-dominated uncertainty, in which case the black error bar is hidden under the red. Kinematic distributions (top two plots) for simulated $B^+ \rightarrow K^+ \mu^+ \mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+ e^-$ (dotted black) and $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ (solid black) events are shown. Figure continues onto the next page. The value of the flatness parameter d_f is superimposed on each plot.

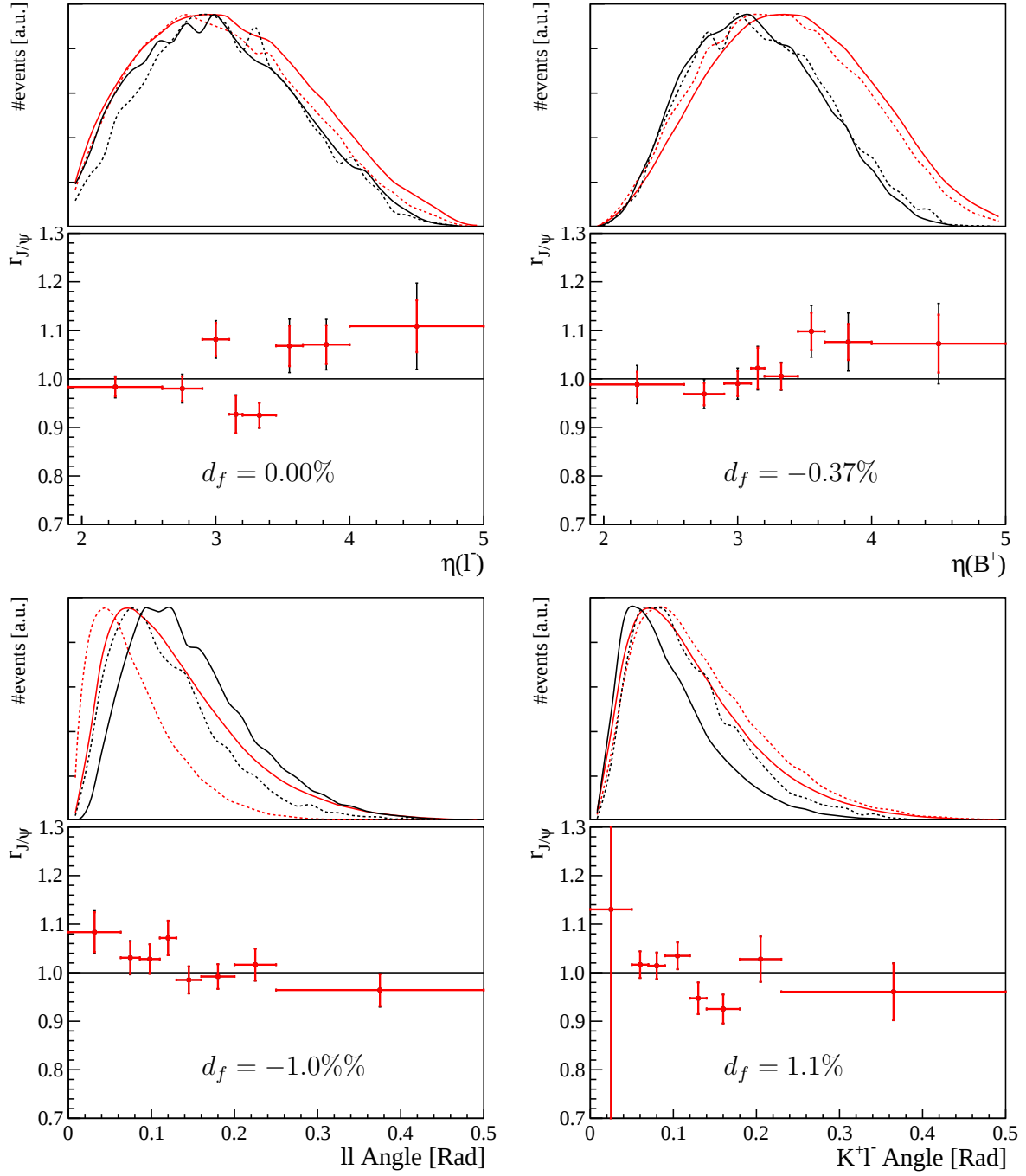


Figure E.1: Continued from the previous page.

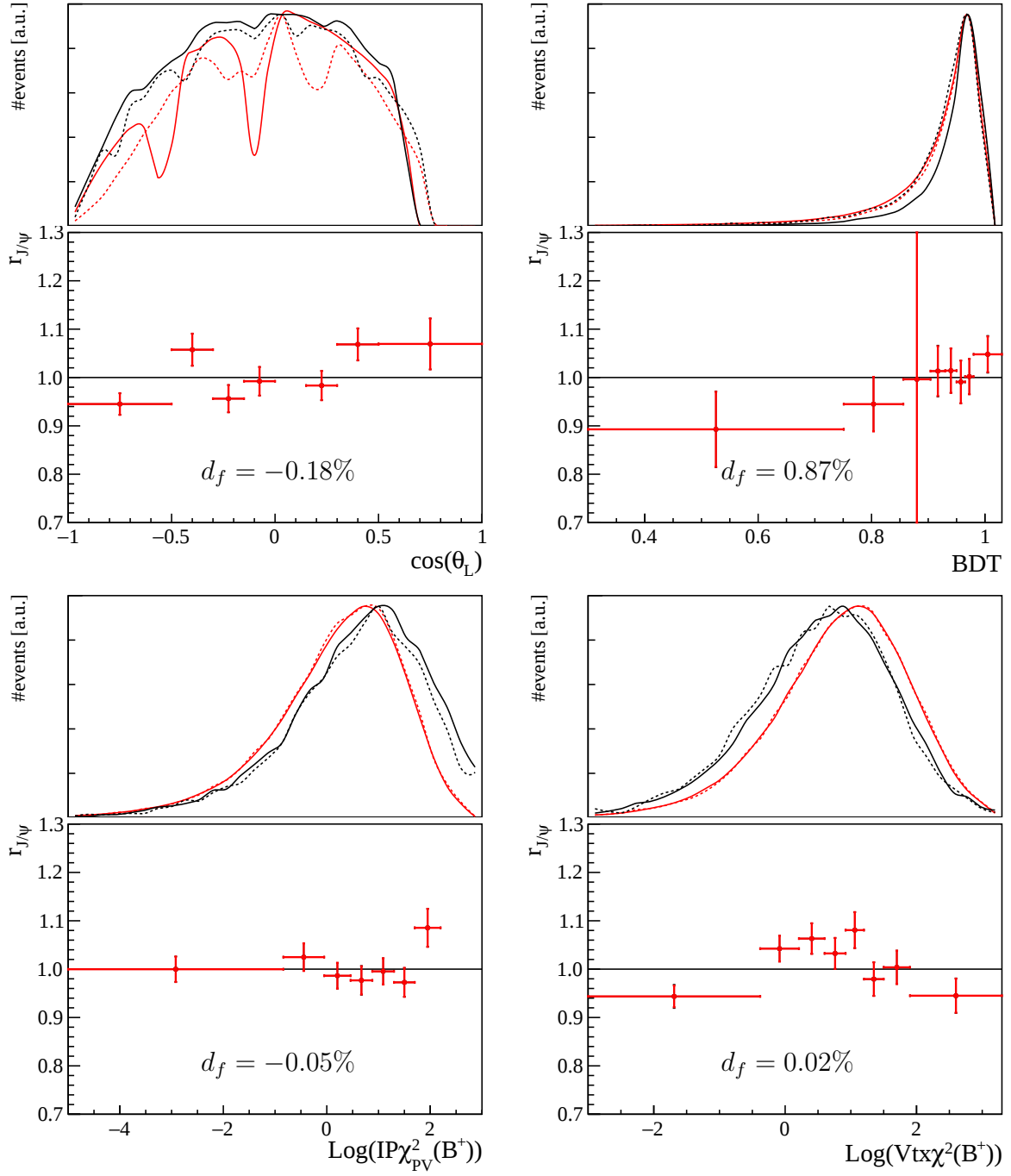


Figure E.1: Continued from the previous page.

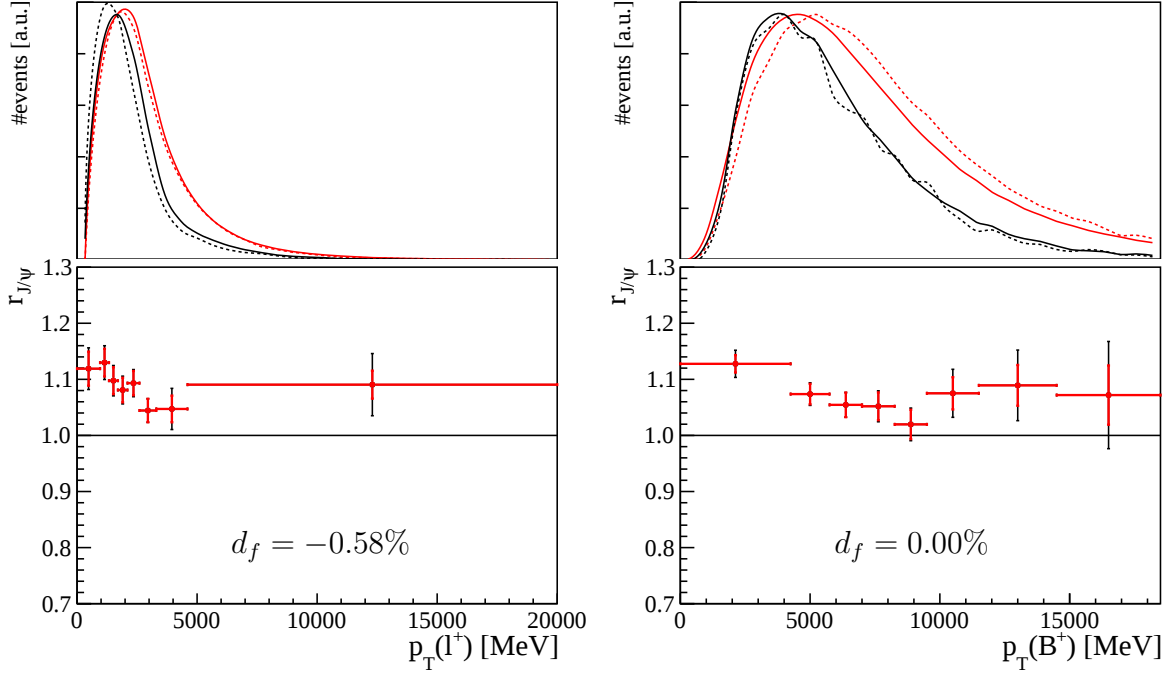


Figure E.2: Binned $r_{J/\psi}$ values (bottom two plots) for the preselected $B^+ \rightarrow K^+ J/\psi(e^+e^-)$ mode in the TIS! category and the preselected $B^+ \rightarrow K^+ J/\psi(\mu^+\mu^-)$ mode in the μ TOS category, in the Run 2 data-taking conditions. The red thick error bars correspond to the statistical uncertainty and the black thin error bars to the systematic uncertainty. A large proportion of the bins have a statistically-dominated uncertainty, in which case the black error bar is hidden under the red. Kinematic distributions (top two plots) for simulated $B^+ \rightarrow K^+ \mu^+ \mu^-$ (dotted red), $B^+ \rightarrow K^+ J/\psi(\mu^+ \mu^-)$ (solid red), $B^+ \rightarrow K^+ e^+ e^-$ (dotted black) and $B^+ \rightarrow K^+ J/\psi(e^+ e^-)$ (solid black) events are shown. Figure continues onto the next page. The value of the flatness parameter d_f is superimposed on each plot.

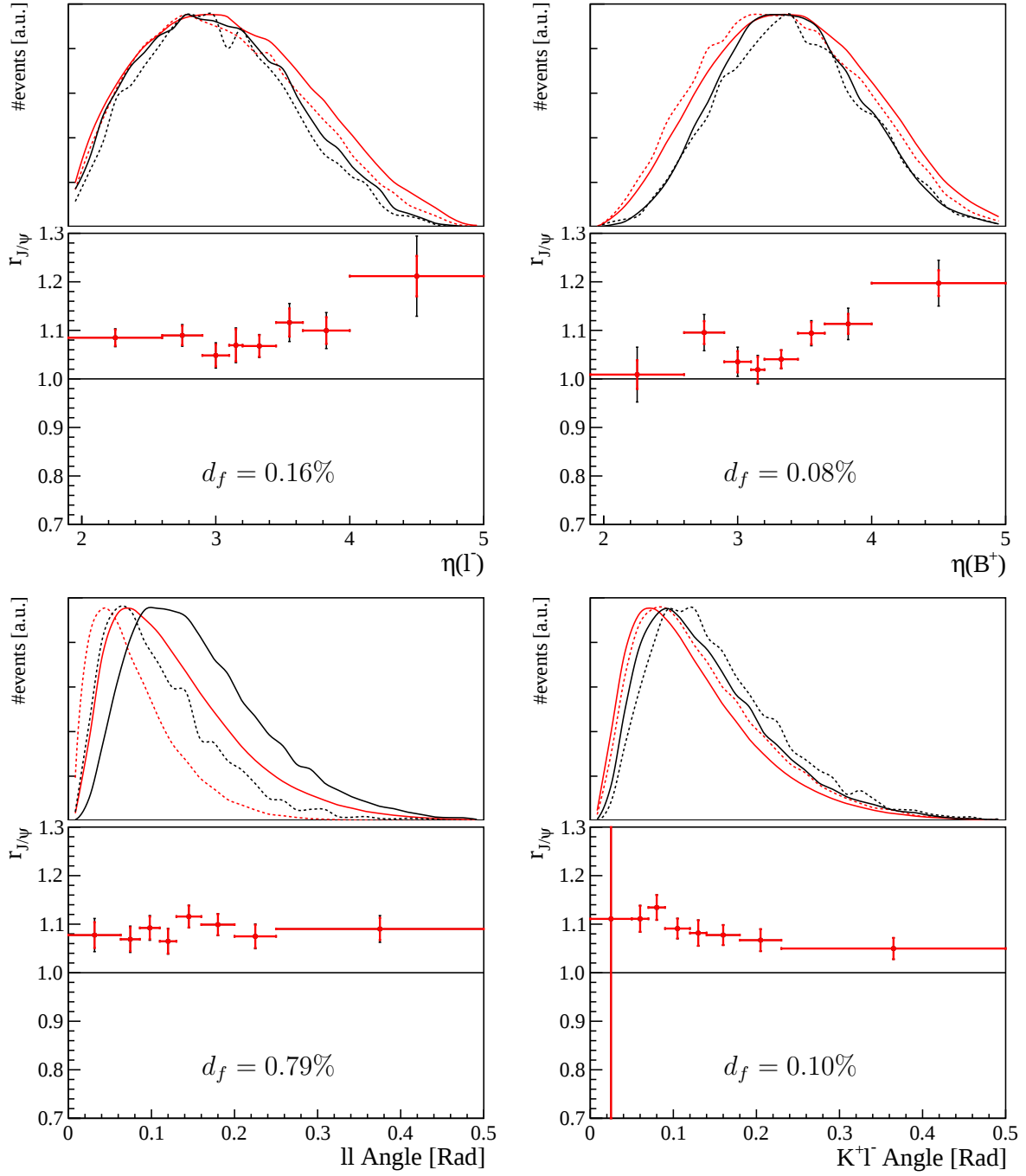


Figure E.2: Continued from the previous page.

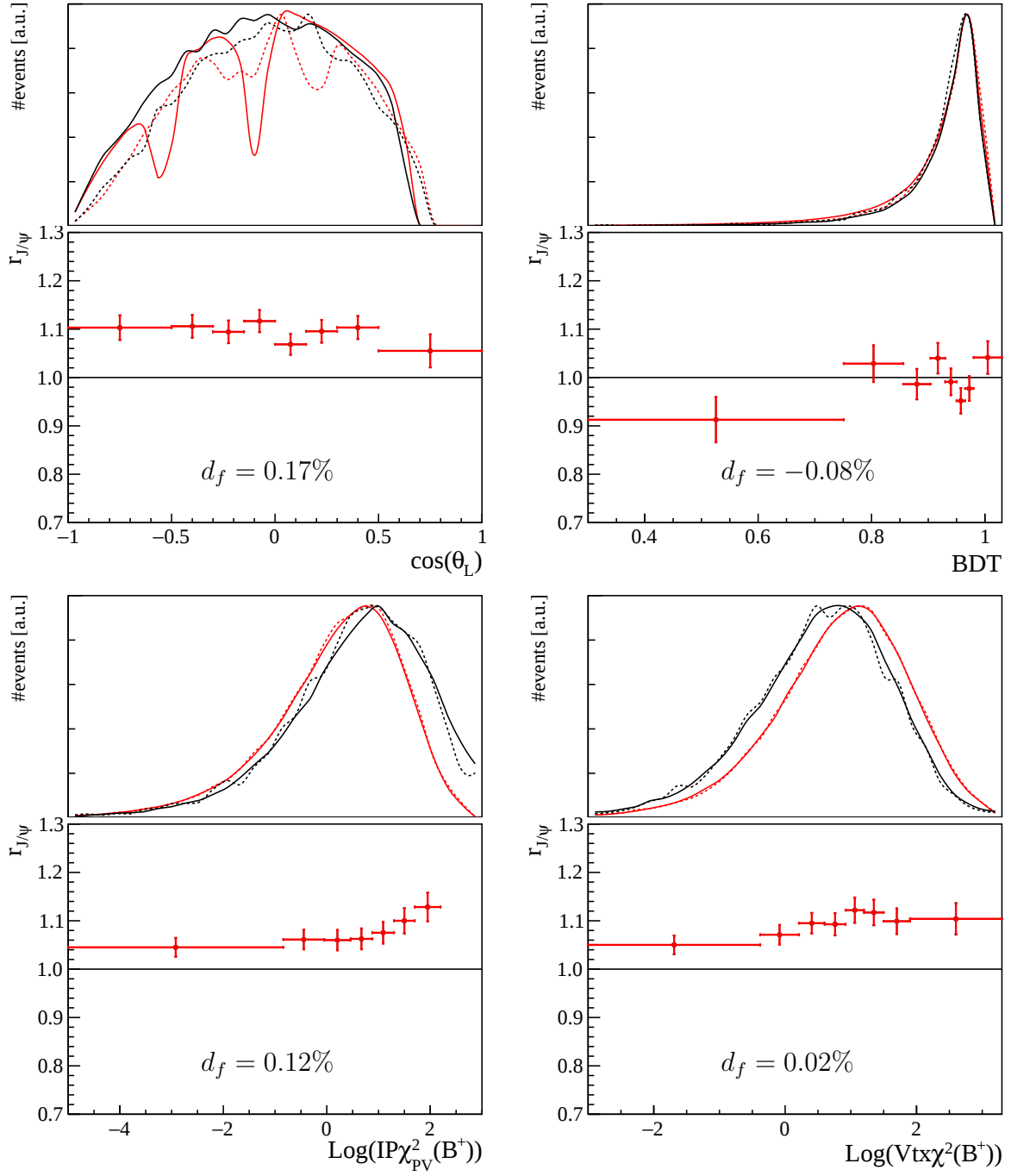


Figure E.2: Continued from the previous page.