

Department of Physics and Astronomy
Heidelberg University

Master thesis in Physics
submitted by

Paul André Günther

born in Kassel (Germany)

October 2018

First observation of the decay $D^*(2007)^0 \rightarrow D^0 e^+ e^-$
in the light of a future Dark Photon search
at the LHCb experiment

This Master thesis has been carried out by
Paul André Günther
at the
Physikalisches Institut of the University of Heidelberg
under the supervision of
Prof. Dr. Stephanie Hansmann-Menzemer
and
Dr. Michel De Cian (EPFL)

Abstract

In this thesis, the first observation of the decay $D^*(2007)^0 \rightarrow D^0 e^+ e^-$ and a measurement of its branching ratio relative to the decay $D^*(2007)^0 \rightarrow D^0 \gamma$ is performed. The dataset was collected by the LHCb experiment in proton-proton collisions at 13 TeV centre-of-mass energy in 2017 and corresponds to an integrated luminosity of approximately 1.87 fb^{-1} . In total, 3253 ± 238 $D^*(2007)^0 \rightarrow D^0 e^+ e^-$ decays are reconstructed and selected, leading to a statistical significance of 12.7 standard deviations. The result of the relative branching fraction measurement is $(9.6 \pm 0.7 \text{ (stat)} \pm 2.3 \text{ (syst)}) \times 10^{-3}$, where the uncertainties are statistical and systematic, respectively. From this, the branching ratio of the decay $D^*(2007)^0 \rightarrow D^0 e^+ e^-$ is inferred to be $(3.38 \pm 0.25 \text{ (stat)} \pm 0.82 \text{ (syst)} \pm 0.09) \times 10^{-3}$, where the third uncertainty is due to the uncertainty on the branching fraction of the normalisation channel $D^*(2007)^0 \rightarrow D^0 \gamma$.

Zusammenfassung

Diese Arbeit präsentiert die erste Beobachtung des Zerfalls $D^*(2007)^0 \rightarrow D^0 e^+ e^-$ und die Messung seines Verzweigungsverhältnisses relativ zum Zerfall $D^*(2007)^0 \rightarrow D^0 \gamma$. Die Daten wurden mit Hilfe des LHCb-Experiments von Proton-Proton Kollisionen im Jahr 2017 bei einer Schwerpunktsenergie von 13 TeV aufgenommen und entsprechen etwa einer integrierten Luminosität von 1.87 fb^{-1} . Insgesamt wurden 3253 ± 238 $D^*(2007)^0 \rightarrow D^0 e^+ e^-$ Zerfälle mit einer statistischen Signifikanz von 12.7 Standardabweichungen rekonstruiert und selektiert. Das gemessene relative Verzweigungsverhältnis ist $(9.6 \pm 0.7 \text{ (stat)} \pm 2.3 \text{ (syst)}) \times 10^{-3}$, wobei die Unsicherheiten statistisch bzw. systematisch sind. Daraus folgt für das Verzweigungsverhältnis des Zerfalls $D^*(2007)^0 \rightarrow D^0 e^+ e^-$ der Wert $(3.38 \pm 0.25 \text{ (stat)} \pm 0.82 \text{ (syst)} \pm 0.09) \times 10^{-3}$, wobei die dritte Unsicherheit durch die Unsicherheit auf das Verzweigungsverhältnis des Zerfalls $D^*(2007)^0 \rightarrow D^0 \gamma$ gegeben ist.

CONTENTS

1	INTRODUCTION	1
2	A SMALL EXCURSION TO THEORY	3
2.1	The Standard Model of Particle Physics	3
2.1.1	Mathematical Formulation of the Standard Model	6
2.1.2	The Dalitz Decay and the Radiative Decay of the $D^*(2007)^0$	7
2.2	Unanswered Questions	9
2.3	A Dark Side of Electromagnetism?	10
3	EXPERIMENTAL SEARCHES FOR DARK PHOTONS	12
4	THE LHCb EXPERIMENT	14
4.1	The Large Hadron Collider	14
4.2	The LHCb Detector and its Components	15
4.2.1	Tracking Down Charged Particles	16
4.2.2	A Particle's Identity	20
4.3	Trigger, Data Acquisition and Data Flow	24
4.3.1	LHCb Software and Simulation	26
5	ANALYSIS STRATEGY AND TOOLS	28
5.1	Strategy	28
5.2	The Method of Maximum Likelihood	29
5.3	Unfolding Distributions	30
5.4	Gradient Boosting and Decision Trees	31
6	SELECTION	34
6.1	Datasets and Trigger	34
6.1.1	Simulation Samples	36
6.2	Background Sources and Pre-selection	37
6.3	Photon Conversion Vetoes	39
6.4	Data vs. Simulation	43
6.4.1	Unfolding a Pure Control Sample	43
6.4.2	Differences between $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+e^-)$ Simulation and Data	48
6.4.3	$D^{*0} \rightarrow D^0e^+e^-$ vs. $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+e^-)$	50
6.4.4	Differences between $D^{*0} \rightarrow D^0e^+e^-$ Simulation and Data	52
6.5	Multivariate Selection using Machine Learning	55
6.6	Optimising the Selection	60
7	DETERMINATION OF THE BRANCHING RATIO	64
7.1	Efficiency Determination	64
7.2	Event Yield Extraction and Signal Significance	66
7.3	Branching Ratio Result	71

8	SYSTEMATIC UNCERTAINTIES	72
9	CONCLUSION	75
9.1	Results	75
9.2	Outlook	76
A	GENERATOR LEVEL CUTS	78
B	PHOTON CONVERSIONS	79
B.1	Photon Conversion Probability	79
B.2	Invariant Mass of the Di-electron System from Photon Conversions	80
B.3	Study of Photon Conversion Background	82
C	COMPARISONS OF SIMULATED AND DATA DISTRIBUTIONS	83
C.1	Control Channel vs. Data	83
C.2	Simulated Signal Channel vs. Control Channel	87
C.3	Re-weighting of Signal Channel Simulation	90
D	MVA CONTROL PLOTS	95
E	FIT EXAMPLE FOR BDT OPTIMISATION	97
F	FINAL SIMULATION FITS	98
G	ERKLÄRUNG	108

The most fascinating notion of fundamental physics is arguably that its ultimate goal is nothing less than a full description of everything that has ever been and will ever be, since the dawn of space-time itself and until the cold and dark destiny of heat death of the Universe. From today's perspective, this goal, if ever reachable, lies in distance future. The two realms of reality, the smallest scales in quantum physics and the largest scales in general relativity, coexist in theory but their link is still missing. Yet observations from all over the observable Universe, where gravity appears to be the driving force, can still influence what we think should be present in the quantum world. Compelling evidence [1] for a kind of matter, in the sense that it interacts at least gravitationally, has led to eager theoretical and experimental efforts of understanding its nature using modern methods of particle physics. Yet in spite of tremendous work that has already been done, no light could be shed on the place this *dark matter* might take in the quantum world. Inspired by the rich interactive structure of luminous matter, however, the idea of similar interactions within the dark sector of matter seems natural and opens a window for experimental searches.

The work presented in this thesis is intended to pave the way for a proposed dark photon search at the LHCb experiment, exploiting decays of the $D^*(2007)^0$ meson [2]. The dark photon (A') is the hypothetical dark sector counterpart to the photon of quantum electrodynamics and might occur via the decay $D^*(2007)^0 \rightarrow D^0 A'$ followed by $A' \rightarrow e^+ e^-$.¹ The so far undiscovered decay $D^{*0} \rightarrow D^0 e^+ e^-$ represents an important background source in a search for dark photon decays in $D^*(2007)^0 \rightarrow D^0 (A' \rightarrow e^+ e^-)$. As a first step towards this dark photon search, in this thesis the branching fraction of the decay $D^{*0} \rightarrow D^0 e^+ e^-$ is determined relative to the well-known radiative decay $D^{*0} \rightarrow D^0 \gamma$, where in both cases the D^0 decays via $D^0 \rightarrow K^- \pi^+$. Measuring the ratio between two branching fractions has the great advantage of cancelling systematic uncertainties arising in the reconstruction and selection of the final state particles. Moreover, quantities related to the production of the D^{*0} , such as cross-sections, charm fragmentation fractions and the luminosity of particle collisions, cancel in the ratio. To fully exploit this advantage, the photon in $D^{*0} \rightarrow D^0 \gamma$ decays is required to create a pair of electrons in a material interaction, such that the final state in the numerator and denominator of the ratio is the same. Yet, using photon conversions gives rise to different challenges that have to be met during data analysis. In particular, the dependence of photon conversions on the distribution of detector material, a low conversion probability and consequential low event yields when simulating photon conversions complicate the analysis. Moreover, the energy released in the $D^{*0} \rightarrow D^0$ transition amounts to only 142 MeV which opens a comparatively small phase space for the electron pair, allowing for only small magnitudes of momenta, and thus makes their distinction from the electron background² difficult. The analysis presented here is the first to use such low-momentum $e^+ e^-$ pairs at the LHCb experiment. Aside from the challenges to conduct this analysis in a hadron collider environment, it benefits from the vast number of D^{*0} mesons produced in proton-proton collisions.

¹ Charge conjugation and $D^{*0} \equiv D^*(2007)^0$ is implied throughout this thesis.

² Vast numbers of electrons are produced in different processes at a high-energy hadron collider, e.g. via $\pi^0 \rightarrow \gamma (\gamma \rightarrow e^+ e^-)$ or $\pi^0 \rightarrow \gamma e^+ e^-$.

In the following three chapters the underlying theory of particle physics is introduced, experimental searches for dark photons are summarised and the LHCb experiment is described. After that, an overview of the analysis strategy is given together with a brief introduction to important analysis tools used in this thesis. Next, the data selection procedure is described in detail, followed by a chapter about the determination of the branching ratio of the $D^{*0} \rightarrow D^0 e^+ e^-$ decay relative to the decay $D^{*0} \rightarrow D^0 \gamma$. A brief evaluation of systematic uncertainties and the conclusion of the presented work make up the last two chapters.

2

A SMALL EXCURSION TO THEORY

This chapter sets the theoretical stage on which this thesis plays. The Standard Model of Particle Physics (SM) is briefly outlined in the first sub-section before the decay $D^{*0} \rightarrow D^0 e^+ e^-$ is discussed and open questions in fundamental physics are summarised. The chapter closes with a small extension of the SM including dark photons. There is a lot of literature about the SM and the underlying toolkit of Quantum Field Theory (QFT), Refs. [3–5] are the foundation of this chapter.

2.1 THE STANDARD MODEL OF PARTICLE PHYSICS

It is an intriguing question to ask what actually can be considered fundamental concerning the smallest constituents of the Universe. Although the SM cannot give a comprehensive answer to this question, observations suggest that *symmetry* is a key concept of Nature. Demanding *local gauge invariance* from a Lagrangian density has turned out to be extremely prolific in describing the dynamics of particles known today. Decades of theoretical and experimental research demonstrated that the symmetry group from which three of the four¹ known fundamental forces arise, is $SU(3)_C \times SU(2)_L \times U(1)_Y$. All elementary particles discovered so far can be classified by their behaviour under gauge transformations from this very group structure and the spin they exhibit. Particles carrying half-integer spin are called *fermions*, while integer spin particles are called *bosons*. Fermions are the building blocks of all familiar matter which is due to the fact that they cannot occupy the same quantum state². This restriction is the seed of the formation of complex structures because it prevents fermions from collapsing all together into the state of lowest energy. Yet, the formation itself and the eventual decay of formed structures is driven by the bosons which convey the fundamental forces.

STRONG INTERACTION The symmetry group $SU(3)_C$ has eight generators and three charges, which are called *gluons* (g) and *colour charges* (red, green and blue)³, respectively. The theory describing the dynamics of coloured particles is called *quantum chromo dynamics* (QCD). Fermions that carry $SU(3)_C$ colour charge are called *quarks* in opposition to *leptons*, which are colourless. The interaction conveyed by the gluons is called *strong force* as its coupling at an energy scale $\Lambda_{\text{QCD}} \sim \mathcal{O}(10^2 \text{ MeV})$ is large compared to the couplings of the other forces. This leads to a hypothesis called *colour confinement*. It states that coloured objects cannot exist individually but hadronise on small time-scales $\sim \hbar/\Lambda_{\text{QCD}}$ into colourless bound states, e.g. *mesons* (bound states of two quarks) or *baryons* (bound states of three

1 General relativity describing gravity is not a quantum theory and there seems to be no obvious way of incorporating it into the context of the SM.

2 This is Pauli's exclusion principle.

3 "The idiot physicists, unable to come up with any wonderful Greek words anymore, call this type of polarization by the unfortunate name of 'color,' which has nothing to do with color in the normal sense." Feynman [6]

quarks).⁴ Qualitatively, this can be understood by figuring what happens when two coloured particles are pulled apart: while the distance between the two colour charges grows, the gluon field in between is squeezed as the gluons themselves attract each other due to their colour charge. This leads to a constant energy density of the field and at some point it is energetically more favourable to form a new quark pair resulting in a colourless hadronic state. However, at energies well above Λ_{QCD} , corresponding to distances smaller than the spacial extent of bound states, the coupling becomes so small that quarks and gluons can be viewed as free particles, which is called *asymptotic freedom*.

EMERGENCE OF WEAK AND ELECTROMAGNETIC INTERACTIONS Supposing a perfectly realised $SU(3)_C \times SU(2)_L \times U(1)_Y$ symmetry also means to accept that all fundamental particles are massless. This is due to the fact that incorporating explicit mass terms into the Lagrangian density describing the particle content of the SM, would break this symmetry. This clearly contradicts the experimental evidence that fundamental particles have mass. Nevertheless, it turns out to be useful to assume that this is true for the sake of an argument. Particles exhibit a property called *chirality*, which, in the case of massless particles, has the simple interpretation of the projection of the spin vector onto the momentum⁵ vector, defining a handedness. The handedness of a particle has consequences for the interactions it can participate in. Left-handed (LH) quarks and leptons form $SU(2)_L$ doublets while right-handed (RH) particles are singlets under this symmetry group. The $SU(2)_L$ group has three generators corresponding to three gauge bosons called W^1 , W^2 and W^3 that couple to the charge called *weak isospin*. The doublet structure of LH fermions reflects that they can interact with the gauge bosons of $SU(2)_L$, *i.e.* are charged under this symmetry, while RH fermions carry no weak isospin and cannot feel this force. This already resembles the historically important observation that particles emitted in nuclear β^- -decay mostly have their spin vector oriented in opposite direction of their momentum vector, *i.e.* are left-handed. However, to accurately describe the observed interactions of fermions much more is needed. The SM gauge group is completed by adding the $U(1)_Y$ symmetry. This introduces another gauge boson called B which couples to *weak hypercharge* (Y). So far the theory presented here seems to have little in common with the particle world that is observed. The SM gauge group requires particles to be massless, there is no electromagnetism and no short-ranged weak interaction. The missing piece is a mechanism that allows to generate particle masses in a gauge-invariant way. This is provided by *spontaneous symmetry breaking* (SSB) which is, in the context of particle physics, also referred to as BEH⁶-mechanism. The idea is to introduce a complex scalar $SU(2)_L$ doublet ϕ (the Higgs doublet) that carries hypercharge. If the potential of this field has a shape as shown in Fig. 2.1, the point at which this field preserves the $SU(2)_L \times U(1)_Y$ symmetry is not stable and the minimum occurs for a non-zero field value (vacuum expectation value). As soon as the field "falls" into the minimum the $SU(2)_L \times U(1)_Y$ symmetry is spontaneously broken to a single $U(1)_Q$ symmetry. This has dramatic consequences for all fields that have couplings to the components of the Higgs doublet and for the Higgs doublet itself. After SSB only one physical Higgs particle remains that acquires its own mass via self-interaction with its vacuum expectation value. The W^1 and W^2 fields are mixed and form the physical states W^+ and W^- , the same happens

4 More exotic bound states such as a combination of four or five quarks are also possible and called *tetraquark* and *pentaquark*, respectively. Because the exact structure of these bound states is still unknown, they are also referred to as *non- $q\bar{q}$* mesons and *exotic baryons* [7].

5 Massless particles have to travel at the speed of light according to special relativity.

6 Brout-Englert-Higgs

with the W^3 and B bosons which result in the physical Z boson and the photon (γ). The photon can be identified as the gauge boson of the remaining $U(1)_Q$ symmetry and thus is massless. It is the force carrier of the *electromagnetic interaction*, coupling to *electric charge* Q . The electric charge of a particle consequently comes from the combination of weak isospin and hypercharge. The QFT describing the dynamics of electrically charged particles is called *quantum electrodynamics* (QED). The W^+ , W^- and Z bosons are the force carriers of the *weak interaction*. They are massive, because they are constantly interacting with the condensate of the scalar Higgs field with its non-zero vacuum expectation value. This interaction can be thought of as constantly flipping a particle's chirality, letting it appear massive. The W^\pm bosons interact only with LH particles which is a property they inherit from the W^1 and W^2 bosons and which is the origin of parity violation in the weak interaction. The Z boson can interact with both LH and RH particles due to its superposition of the W^3 (coupling to LH only) and the B boson (coupling to LH and RH equally), however with different strength. Fermions that are charged under $U(1)_Q$ acquire their masses via the same

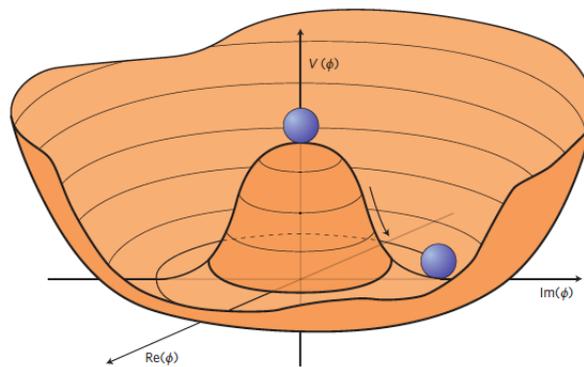


Figure 2.1: "Mexican hat" potential of the complex Higgs field ϕ that leads to spontaneous symmetry breaking. The vacuum state, *i.e.* the minimum of the potential, can be chosen, without loss of generality, from an arbitrary point around the bottom of the brim of the hat [8].

mechanism as the massive bosons. The massiveness of quarks has major consequences for the phenomenology of their weak interactions. After SSB, the weak interaction eigenstates differ from the physical mass eigenstates by a rotation in their abstract vector space, effectively mixing quark interaction eigenstates to form a quark mass eigenstate. The rotation is described by a unitary 3×3 matrix (the CKM⁷ matrix) that can be parametrised by three mixing angles and a complex phase. The complex phase difference between mass and interaction eigenstates leads to a difference between interaction rates of quarks and anti-quarks, *i.e.* breaking or violation of the so-called CP symmetry.

PARTICLE CONTENT Demanding invariance of the SM Lagrangian density under $SU(3)_C \times SU(2)_L \times U(1)_Y$ transformations and SSB via the BEH-mechanism fixes the boson content of the SM to five gauge bosons (γ , g , W^+ , W^- and Z) plus the Higgs boson. The fermion content, however, is somewhat arbitrary. Six quarks, six leptons and the same number of anti-quarks and anti-leptons have been observed and grouped into three generations according to their increasing mass. The first generation comprises the *up* and *down* quark (u , d) on the quark side, and the *electron* and *electron-neutrino* (e , ν_e) on the lepton side. Hence, the first generation contains the particles that make up everyday ordinary

7 Cabbibo-Kobayashi-Maskawa

matter. This is because these particles have the smallest masses and are the products of decays of particles from the other generations. The second generation comprises the *charm* and *strange* quark (c, s) and the *muon* as well as the *muon-neutrino* (μ, ν_μ), the third generation includes the *top* and *beauty*⁸ quark (t, b) and the *tau* and *tau-neutrino* (τ, ν_τ). These different kinds of quarks and leptons are called *flavour*. The mass differences of particles across the generations and flavours are quite remarkable, the τ lepton is roughly 3500 times heavier than the electron, the t quark even has a mass around 75000 times the mass of a u quark, but otherwise exhibiting the exact same properties. This also means that every fermion must have a vastly different coupling⁹ to the Higgs field. The neutrinos are special in this context because no RH neutrino has been observed yet. Without a RH partner to the LH neutrinos, they cannot couple to the Higgs boson and thus remain massless within the SM.

2.1.1 Mathematical Formulation of the Standard Model

The SM is usually mathematically described by defining its Lagrangian density which is constructed to be invariant under the aforementioned gauge symmetries. The full Lagrangian density contains several terms for each fermion, boson and for all the interactions among them and shall not be reproduced here as its complexity exceeds the scope of this thesis. Instead, two important components appearing in the Lagrangian density are briefly explained using the deceptively simple example of QED:

$$\mathcal{L}_{QED} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}(i\not{D} - m)\psi \quad (2.1)$$

The first term¹⁰ represents the dynamics of the free electromagnetic field, *i.e.* the dynamics of the photon in the absence of charge, and is called *kinetic term*. The electromagnetic field tensor is defined as $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$ where A_μ denotes the electromagnetic vector field. Applying the Euler-Lagrange equation to the first term gives the equations of motion for the photon, *i.e.* the Maxwell equations. The second term encodes the dynamics of charged lepton or quark fields ψ with mass m in the absence and presence of photons. The interaction of photons with ψ is hidden in the *gauge covariant derivative* $\not{D} \equiv \gamma^\mu D_\mu \equiv \gamma^\mu(\partial_\mu + ieA_\mu)$, where γ^μ are the gamma matrices and e is the coupling constant of QED. The interaction part can thus be written as:

$$\mathcal{L}_{QED}^{Int} = -eA_\mu J_{EM}^\mu \quad (2.2)$$

where $J_{EM}^\mu \equiv \bar{\psi}\gamma^\mu\psi$ is the electromagnetic current. The interaction of the photon with electromagnetic current is a direct consequence of demanding invariance of the Lagrangian density under local U(1) transformations:

$$\psi(x) \rightarrow e^{-ie\alpha(x)}\psi(x) \quad (2.3)$$

$$A_\mu(x) \rightarrow A_\mu(x) + \partial_\mu\alpha(x) \quad (2.4)$$

where α is an arbitrary real differentiable function of the spacetime coordinate x . Equation 2.2 is explicitly made invariant under these transformations by introducing the gauge covariant

8 Often also called *bottom* quark.

9 These are the *Yukawa* couplings.

10 General notation: greek letters denote 4-vector indices and $\partial_\mu \equiv \partial/\partial x^\mu$ is the partial derivative w.r.t. x^μ . Einstein's summation convention is employed. The adjoint Dirac spinor is defined as $\bar{\psi} \equiv \psi^\dagger \gamma^0$ where γ^0 is the first gamma matrix.

derivative. Following this procedure, the gauge covariant derivative has to be extended by more fields when demanding invariance under other symmetry groups. Eventually, the SM arises from these basic ideas.

2.1.2 The Dalitz Decay and the Radiative Decay of the $D^{*0}(2007)^0$

A good theory has to give accurate predictions to be falsifiable. These predictions can be more qualitative as the postulation of the Higgs boson, or they can provide actual numbers which can then be tested. The SM offers a framework to compute the amplitudes of virtually every possible interaction of SM particles to arbitrary precision, in principle only limited by the computational effort that is made to calculate higher-orders in perturbation theory and by the precision with which SM parameters are known. In reality, calculations are often complicated by non-perturbative¹¹ effects and because elementary particles such as quarks form bound states. Often, however, these complications can be mitigated by clever combination of calculations, *e.g.* by taking ratios of amplitudes. In the case of the $D^{*0} \rightarrow D^0 e^+ e^-$ decay it is useful to treat it relative to the decay $D^{*0} \rightarrow D^0 \gamma$. The first-order Feynman diagrams of both decays are shown in Fig. 2.2. The interaction can be described using QED because photons and electrons are involved in the dynamics. Since the coupling constant of QED is small, it is sufficient to use first-order diagrams to give a good estimate of the rates at which these processes occur relative to each other. The $D^{*0}(2007)^0$ and D^0 are both bound states

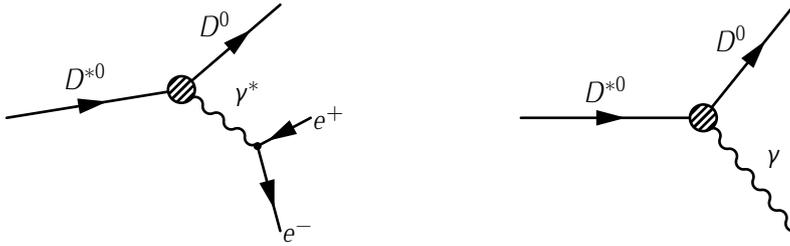


Figure 2.2: A Feynman diagram of the $D^{*0}(2007)^0$ Dalitz decay $D^{*0} \rightarrow D^0 e^+ e^-$ (left) and a Feynman diagram of the radiative decay $D^{*0} \rightarrow D^0 \gamma$ (right). The shaded circle denotes that the QED vertex has to be corrected by a form factor.

of a c and \bar{u} quark, differing only in their total angular momentum J . The D^{*0} exhibits $J = 1$ and thus is a vector particle while the D^0 has no spin and is a pseudo-scalar. Their properties are summarised in Tab. 2.1. The Feynman diagrams in Fig. 2.2 are pictorial representations of the corresponding amplitudes. The amplitude of the decay $D^{*0} \rightarrow D^0 e^+ e^-$ can be written¹² as [9]:

$$M_{Dalitz} = 4\pi\alpha i [f_{D^{*0}D^0}(q^2) \epsilon^{\rho\beta\nu\delta} p_\rho q_\beta \epsilon_\nu] \frac{1}{q^2} [\bar{u}\gamma_\delta u] \quad (2.5)$$

where α is the fine-structure constant, p_ρ and q_β are the four-momenta of the D^0 and the virtual photon in the D^{*0} centre-of-mass system, ϵ_ν the D^{*0} polarisation vector, u and \bar{u} the electron's and positron's spinors giving the leptonic current, $1/q^2$ the virtual photon propagator and $f_{D^{*0}D^0}(q^2)$ the hadronic form factor of the $D^{*0}-D^0$ transition. The form factor takes into account that the bound state of two quarks is not a point-like object but has some

¹¹ These effects often occur in QCD as the strong coupling constant cannot be considered small at low energy scales and thus cannot be treated as perturbation anymore.

¹² Out of convenience and habit all calculations in this sub-section use natural units.

Table 2.1: Properties of the $D^*(2007)^0$ and D^0 [7]. The isospin quantum number is denoted as I , J is the total angular momentum quantum number and P denotes the parity quantum number. The branching ratio of the decays is given in the last column. The sum of the D^{*0} branching ratios is constrained to 100%.

Particle	Mass [MeV/ c^2]	Quant. num. $I(J^P)$	Relevant decays	\mathcal{B}
$D^*(2007)^0$	2006.85 ± 0.05	$\frac{1}{2}(1^-)$	$D^{*0} \rightarrow D^0 \pi^0$	$(64.7 \pm 0.9)\%$
			$D^{*0} \rightarrow D^0 \gamma$	$(35.3 \pm 0.9)\%$
D^0	1864.83 ± 0.05	$\frac{1}{2}(0^-)$	$D^0 \rightarrow K^- \pi^+$	$(3.89 \pm 0.04)\%$

spatial charge distribution. This form factor is a priori unknown and hinders an accurate calculation. However, it similarly appears in the amplitude of the $D^{*0} \rightarrow D^0 \gamma$ decay:

$$M_{radiative} = \sqrt{4\pi\alpha} i [f_{D^{*0}D^0}(0) \epsilon^{\rho\beta\nu\delta} p_\rho q_\beta \epsilon_\nu] \epsilon_\delta^* \quad (2.6)$$

where ϵ_δ^* is the complex conjugate of the photon polarisation. The decay rates can be obtained by squaring the amplitudes and evaluating the two-body and three-body phase space integral of the decays $D^{*0} \rightarrow D^0 \gamma$ and $D^{*0} \rightarrow D^0 e^+ e^-$, respectively. The differential decay rate of $D^{*0} \rightarrow D^0 e^+ e^-$ can then be written relative to the decay rate of $D^{*0} \rightarrow D^0 \gamma$ as [9]:

$$\begin{aligned} \frac{d\Gamma(D^{*0} \rightarrow D^0 e^+ e^-)}{dq^2 \Gamma(D^{*0} \rightarrow D^0 \gamma)} &= \frac{\alpha}{3\pi} \sqrt{1 - \frac{4m_e^2}{q^2}} \times \left[1 + \frac{2m_e^2}{q^2} \right] \times \frac{1}{q^2} \\ &\times \left[\left(1 + \frac{q^2}{m_{D^{*0}}^2 - m_{D^0}^2} \right)^2 - \frac{3m_{D^{*0}}^2 q^2}{(m_{D^{*0}}^2 - m_{D^0}^2)^2} \right]^{\frac{3}{2}} \\ &\times \left| \frac{f_{D^{*0}D^0}(q^2)}{f_{D^{*0}D^0}(0)} \right|^2 \\ &= [\text{QED}] \times |F_{D^{*0}D^0}(q^2)|^2 \end{aligned} \quad (2.7)$$

From Eq. 2.7 follows that it is not necessary to know the absolute value of the form factors as only the ratio contributes. Furthermore, the mass difference between the D^{*0} and D^0 is only $\Delta m_{D^{*0}D^0} = (142.12 \pm 0.07) \text{ MeV}/c^2$ [7], which is small compared to Λ_{QCD} and thus allows to approximate the form factor as constant over the range $q^2 \in [0, \Delta m_{D^{*0}D^0}^2]$. Consequently, it cancels in the ratio of decay rates. The relative branching ratio is now obtained by integrating Eq. 2.7 over the accessible q^2 range, which is bound from below by $4m_e^2$ because two electrons have to be created, while the upper bound $\Delta m_{D^{*0}D^0}^2$ is given by the mass difference of the mesons. A numerical integration yields:

$$\frac{\Gamma(D^{*0} \rightarrow D^0 e^+ e^-)}{\Gamma(D^{*0} \rightarrow D^0 \gamma)} = \frac{\mathcal{B}(D^{*0} \rightarrow D^0 e^+ e^-)}{\mathcal{B}(D^{*0} \rightarrow D^0 \gamma)} = 0.8815\alpha \quad (2.8)$$

In the given range of q^2 , the fine structure constant is well approximated by $\alpha^{-1} \approx 137$ leading to:

$$\frac{\mathcal{B}(D^{*0} \rightarrow D^0 e^+ e^-)}{\mathcal{B}(D^{*0} \rightarrow D^0 \gamma)} = 6.43 \times 10^{-3} \quad (2.9)$$

which is consistent with the number given in Ref. [2].

2.2 UNANSWERED QUESTIONS

The SM is a remarkably useful theory as it allows for precise quantitative predictions of the interactions that can occur between elementary particles. These predictions make the SM highly falsifiable, yet unambiguous and significant deviations have not been found so far. This success stands in conflict with several problems and questions the SM leaves unresolved. To begin with, the SM is dependent on a set of at least 19 parameters that have to be put in from experimental findings. Most of these parameters are the fermion masses together with the vacuum expectation value of the Higgs field, the mass of the Higgs boson itself and the four parameters of the CKM matrix. The masses of fermions within each generation are similar while varying over more than five orders of magnitude between the electron and the t quark. The SM gives no explanation for this pattern despite stating that the higher generations are merely heavier copies of the lightest one. Neutrinos seem to break this pattern as at least two of them are observed to have a small mass. The absolute scale of neutrino masses is currently estimated to be below $2 \text{ eV}/c^2$ [7]. Whether or not this mass is generated via a BEH-mechanism, as it is the case for all other fermions, is still unclear. The discovery of the Higgs boson in 2012 put the last piece of the SM puzzle only to reveal new questions. The mass of the Higgs boson is of the same order as the masses of the W and Z bosons which is surprising from a theoretical point of view. If there is a higher energy scale at which new physics phenomena occur, the natural scale of the Higgs mass will be of the order of this higher scale. The Planck scale $M_P \sim 10^{19} \text{ GeV}$ at which gravitational effects start to play a role in the quantum world is an example of such a scale. The fact that the Higgs mass is much lower than the Planck mass M_P could point towards a larger puzzle in which new effects tame the Higgs mass. Even if there were no new physics phenomena on the way up the energy ladder to the Planck scale, the SM would eventually break down when reaching the last rung because there is no quantum theory of gravity. In addition to the lack of quantised gravity, there is a problem with the cosmological constant. This constant can be thought of as the energy of the vacuum. After SSB of $SU(2) \times U(1)$, the Higgs potential acquires a non-zero expectation value at its minimum. As long as gravity is not included in the SM, this constant term in the Higgs potential has no physical consequences. However, when gravity is considered, this term contributes to the cosmological constant driving it to a value that is $\sim 10^{50}$ to $\sim 10^{120}$ times larger than the observed value¹³, depending on what contributions are counted [11]. In general, cosmological observations involving the cosmic microwave background radiation, acceleration of the Universe's expansion and large scale distribution of galaxies and clusters revealed that matter described by the SM only represents $\sim 5\%$ of the energy density of the Universe. The remaining 95% of the energy density are accounted for by roughly 69% *dark energy* and 26% *dark matter* [10]. The word 'dark' can be understood figuratively because it reflects that little is known so far about this kind of energy and matter, as well as literally because dark matter either does not or does only very weakly interact with photons and thus appears to be dark. Furthermore, the main part of the observed SM matter in the Universe consists of baryons rather than anti-baryons. With violation of the CP symmetry via the complex phase of the CKM matrix, the SM provides a mechanism to produce an asymmetry between particles and anti-particles, however, this mechanism cannot sufficiently explain the observed asymmetry. Aside from these big unanswered questions, there are experimental results from precision tests of the SM that

¹³ The most recent value for the cosmological constant comes from the Planck Collaboration [10] and is incredibly small: $\Lambda = (4.24 \pm 0.11) \times 10^{-66} \text{ eV}^2$ in natural units.

show some tension between prediction and measurement. A prominent example of such a deviation comes from the measurement of the anomalous magnetic moment of the muon which currently disagrees with SM prediction by $\sim 3.5\sigma$ [7]. It can be hoped that such discrepancies will eventually lead the way to answers of the fundamental open questions of the SM.

2.3 A DARK SIDE OF ELECTROMAGNETISM?

Under the assumption that dark matter consists of particles which are in principle similar to those described by the SM but have no or very suppressed couplings to the known forces, it is imaginable that this dark sector has its own rich interaction phenomenology. These interactions could serve as a portal to an exploration of the dark sector. One idea for such a portal, that has become increasingly popular in recent years, is the *dark photon*¹⁴ [12–15]. In a minimal scenario of this *vector* portal, the SM can be extended by a new "dark" gauge group $U(1)'$ the gauge boson of which is the dark photon. Imposing this new symmetry on the Lagrangian density of the SM and writing down all allowed terms yields:

$$\mathcal{L}_{SM+U(1)'} = \mathcal{L}_{SM} - \frac{1}{4}F'_{\mu\nu}F'^{\mu\nu} + \frac{\epsilon}{2}F'_{\mu\nu}B^{\mu\nu} + \dots \quad (2.10)$$

where \mathcal{L}_{SM} is the Lagrangian density of the SM, $F'_{\mu\nu} = \partial_\mu A'_\nu - \partial_\nu A'_\mu$ is the field strength tensor of $U(1)'$ with the dark photon field A'_μ , the field strength tensor $B_{\mu\nu}$ belongs to the SM $U(1)_Y$ hypercharge field and the ellipses stands for other possible terms such as dark matter particles coupling to the new gauge boson. The combination of the $U(1)'$ and $U(1)_Y$ field tensors into one kinetic term is known as *kinetic mixing* and was proposed in Ref. [16]. The parameter ϵ is the strength of this mixing. As in the SM, the $SU(2)_L \times U(1)_Y \times U(1)'$ symmetry can be spontaneously broken¹⁵ leading to an effective Lagrangian density:

$$\mathcal{L}_{eff} = -\frac{1}{4}F'_{\mu\nu}F'^{\mu\nu} + \frac{1}{2}m_{A'}^2 A'_\mu A'^\mu + \epsilon e A'_\mu J_{EM}^\mu \quad (2.11)$$

where $m_{A'}$ is the mass of the dark photon, e is the elementary electric charge and J_{EM}^μ is the electromagnetic current of charged leptons (*cf.* Eq. 2.2). Hence, kinetic mixing leads to mixing between the ordinary and the dark photon. This opens a non-gravitational window into the dark sector as, depending on $m_{A'}$, the dark photon can decay into and be produced from any electrically charged particle, which makes this theory testable at a multitude of experiments. The dark photon is well motivated as a possible solution for the discrepancy between the measured and calculated value of the anomalous magnetic moment of the muon [17, 18] as it contributes to higher-order corrections in which the photon is involved. It has further been proposed as an explanation for astrophysical anomalies related to dark matter [19], the dynamics of which can be influenced by a dark photon coupling to it. In the minimal scenario presented in Equations 2.10 and 2.11, the rates at which a dark photon is produced and decays, and thus also its lifetime, depend only on a two-dimensional parameter space $(m_{A'}, \epsilon)$. Broadly, smaller values of ϵ and $m_{A'}$ lead to longer lifetimes of the dark photon, while larger masses open more on-shell decay channels and thus combined with a large mixing strength ϵ yield shorter lifetimes. The differing lifetimes in different parts of the parameters space lead to distinct search strategies, *i.e.* searches for prompt (short lifetime) and displaced

¹⁴ Also called hidden, heavy, para- or secluded photon.

¹⁵ Possibly by a dark Higgs sector.

(long lifetime) decays of a dark photon. The experimental accessibility of a wide area of this parameter space has led to a multitude of searches for dark photons, which are summarised in the next chapter.

3

EXPERIMENTAL SEARCHES FOR DARK PHOTONS

This chapter gives a brief summary of searches for massive dark photons that have already been conducted at various experiments. The searches are put in order according to the exploited production mechanism of the dark photon. Only experiments that searched for visible decays of the dark photon are listed. The areas of the parameter space excluded by these experiments are shown in Figure 3.1.

DARK PHOTONS FROM BREMSSTRAHLUNG The E137, E141, E774, KEK and Orsay electron beam-dump experiments [20–24] looked for signs of long-lived dark photons in the bremsstrahlungs process $eZ \rightarrow eZA'$ where Z denotes a highly charged nucleus and A' is the irradiated dark photon with e^+e^- as final state. The A1, APEX and NA64 [25–27] experiments used the same production channel and final state to search for prompt dark photon decays. The ν -CAL experiment [28] used proton beam dumps in iron to produce dark photons via $p\text{Fe} \rightarrow p\text{Fe}A'$, looking for displaced $A' \rightarrow e^+e^-$. The Heavy Photon Search (HPS) experiment [29] is a dedicated dark photon experiment searching for both, long-lived and promptly decaying dark photon using electron beam-dumps on tungsten.

e^+e^- ANNIHILATION The BaBar [30] and BESIII [31] collaborations searched for prompt $A' \rightarrow e^+e^-$ and $A' \rightarrow \mu^+\mu^-$ via the production channel $e^+e^- \rightarrow A'\gamma$. The KLOE collaboration [32] used the same production mechanism, however, exploiting the prompt hadronic final state $A' \rightarrow \pi^+\pi^-$. The Belle collaboration [33] searched for prompt dark photon decays in all of the aforementioned final states, using the production mechanism $e^+e^- \rightarrow A'h'$ where h' denotes a hypothetical Higgs boson associated to the dark sector.

DRELL-YAN PROCESS The LHCb experiment [34] searched for signs of both long-lived and prompt dark photons in $q\bar{q} \rightarrow A'$ with the $A' \rightarrow \mu^+\mu^-$ final state, putting bounds on the until then unexplored mass region $m_{A'} > 10 \text{ GeV}/c^2$.

MESON DECAYS Decays of light unflavoured mesons (π^0 , η , ω , ϕ etc.) are well suited to search for dark photons as they often involve virtual or on-shell photons. The ν -CAL, NOMAD, and PS191 experiments [28, 35–37] used proton beam dumps to search for $\pi^0 \rightarrow A'\gamma$ with long-lived $A' \rightarrow e^+e^-$, excluding small values of ϵ and $m_{A'}$. The CHARM collaboration used the same meson production mechanism and dark photon decay, however, exploiting the $\eta^{(\prime)} \rightarrow A'\gamma$ decay. The KLOE experiment searched for $\phi \rightarrow A'\eta$ followed by a prompt $A' \rightarrow e^+e^-$ or $A' \rightarrow \mu^+\mu^-$ decay [38]. The LHCb experiment also exploited meson decays ($\eta \rightarrow \mu^+\mu^-\gamma$, $\eta \rightarrow \mu^+\mu^-$, $\omega \rightarrow \mu^+\mu^-\pi^0$, $\omega \rightarrow \mu^+\mu^-$, $\rho^0 \rightarrow \mu^+\mu^-$, $\phi \rightarrow \mu^+\mu^-$) probing the mass region below $3 \text{ GeV}/c^2$ [34].

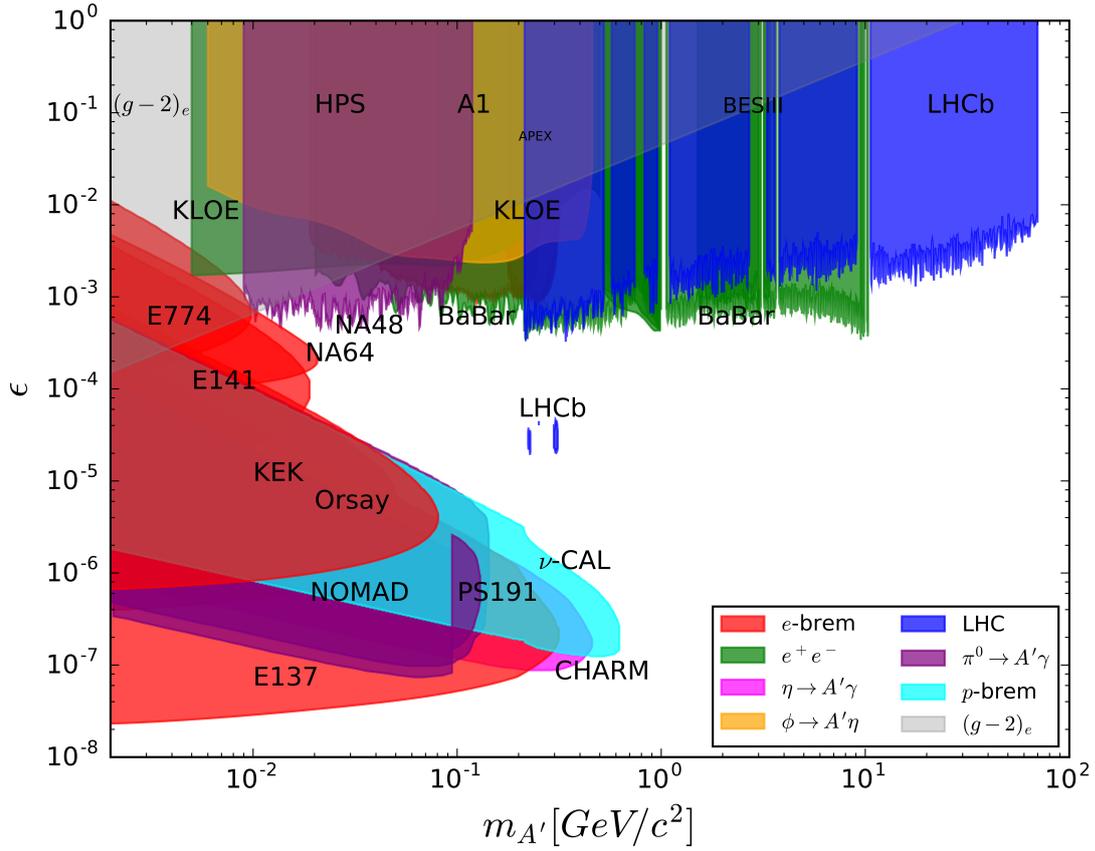


Figure 3.1: Constraints at 90% CL on visible A' decays from different production mechanisms. The mass range for visible searches is bound from below by the di-electron invariant mass and reaches up to invariant masses currently accessible in collider experiments. The bound coming from the precise measurement of the electron's anomalous magnetic moment $(g-2)_e$ is shown in grey. This plot was created using Darkcast [39].

4

THE LHCb EXPERIMENT

This chapter gives a brief overview of the the Large Hadron Collider (LHC) [40] at the Conseil Européen pour la Recherche Nucléaire (CERN) and a more detailed one of the LHCb experiment (the 'b' stands for beauty) and its sub-detector components. The focus lies on the components that are crucial for the analysis of the decay $D^{*0} \rightarrow D^0 e^+ e^-$ relative to the decay $D^{*0} \rightarrow D^0 (\gamma \rightarrow e^+ e^-)$. The LHCb detector and its performance have been described in more detail in [41] and [42], respectively.

4.1 THE LARGE HADRON COLLIDER

The arguably largest playground for high energy particle physics to this day is located near Geneva at CERN close to and below the French-Swiss border. The LHC is a hadron-hadron collider and accelerator composed of two superconducting magnet-rings with 26.7 km circumference, which store counter-rotating particle beams. These particle beams collide in four interaction points, around which the four large experiments - ATLAS, CMS, ALICE and LHCb - are built. ATLAS and CMS use independently designed general-purpose detectors to search for direct hints of yet undiscovered particles, ALICE uses a detector setup specialised on heavy-ion physics while the LHCb detector is built to measure particles containing b or c quarks. The LHC was designed for a maximum centre-of-mass energy of $\sqrt{s} = 14$ TeV at a maximum luminosity of $\mathcal{L} = 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ for proton-proton collisions at ATLAS and CMS [40]. The operation of the LHC is divided into runs, in which particles are brought to collision and physics data is taken, and shutdowns during which maintenance can be conducted at the LHC and at the experiments. Run I took place from 2009 to 2013. During Run II of operation of the LHC, which started in 2015, the peak luminosity reached more than $\mathcal{L} = 2 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ at ATLAS and CMS at $\sqrt{s} = 13$ TeV in 2017 and 2018. The peak luminosity received by LHCb is about two orders of magnitude lower [43], which is achieved by defocussing of the beams at LHCb's collision point. The accelerator complex is shown in Figure 4.1. Protons are taken from hydrogen, which is ionised by an electric field, and are injected in bunches into the initial linear accelerator (LINAC2), which accelerates them to a kinetic energy of 50 MeV before bringing them into the Proton Synchrotron Booster (BOOSTER). The BOOSTER brings the protons up to an energy of 1.4 GeV and transfers them into the Proton Synchrotron (PS) that accelerates them to 25 GeV. From the PS the protons are injected into the Super Proton Synchrotron (SPS) that speeds them up to 450 GeV before inserting them in the LHC, where the protons are finally accelerated to multi-TeV energies [44].

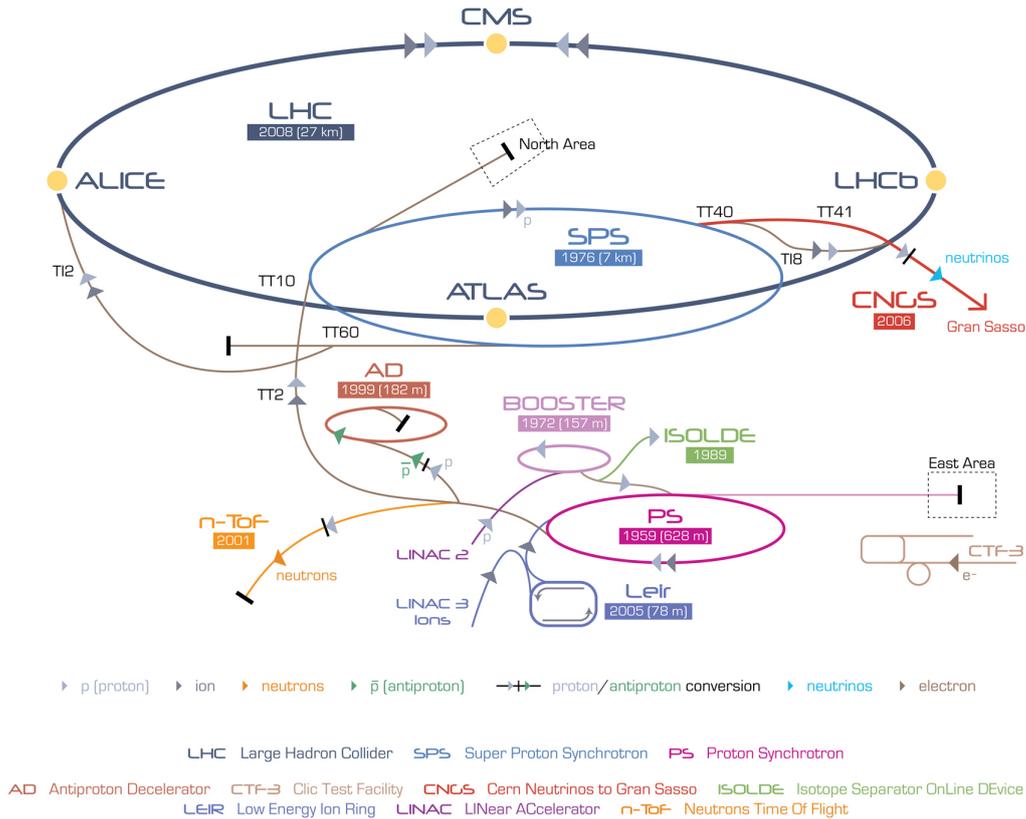


Figure 4.1: Accelerator complex at CERN [45].

4.2 THE LHCb DETECTOR AND ITS COMPONENTS

The LHCb detector is a single-arm forward spectrometer covering approximately 10 mrad to 300 (250) mrad of the angle in forward direction with respect to the beam line in the x - z (y - z) plane, corresponding to a pseudorapidity¹ range of $2 < \eta < 5$ (see Figure 4.2). Its coordinate system is defined by the z -axis pointing along the beam pipe into the detector (downstream), the y -axis vertically oriented towards the surface and the x -axis horizontally pointing towards the centre of the accelerator ring. The single-arm design of the detector is unique among the experiments at the LHC and allows for large yields of B and D mesons to be collected as light quarks are predominantly boosted in the forward² direction. At the centre of LHCb's physics programme stands the rigorous measurement of rare and CP symmetry violating decays of hadrons containing b and c quarks. These hadrons typically have long lifetimes of the order of 10^{-12} seconds because their decays are driven by the weak interaction. Thus displaced (secondary) vertices can be observed aside the original (primary) vertex where the proton-proton collision took place. It is one of LHCb's largest strengths to reconstruct these vertices with very high accuracy. To do so in an environment rife with all kinds of light particles sprouting from the high-energetic proton collisions, LHCb employs a chain of different detector components.

¹ Pseudorapidity is defined as $\eta \equiv -\ln(\tan(\frac{\theta}{2}))$ where θ is the angle between a particle's momentum vector and the beam axis.

² "Forward" here means towards the detector, the same is of course true for the backward direction.

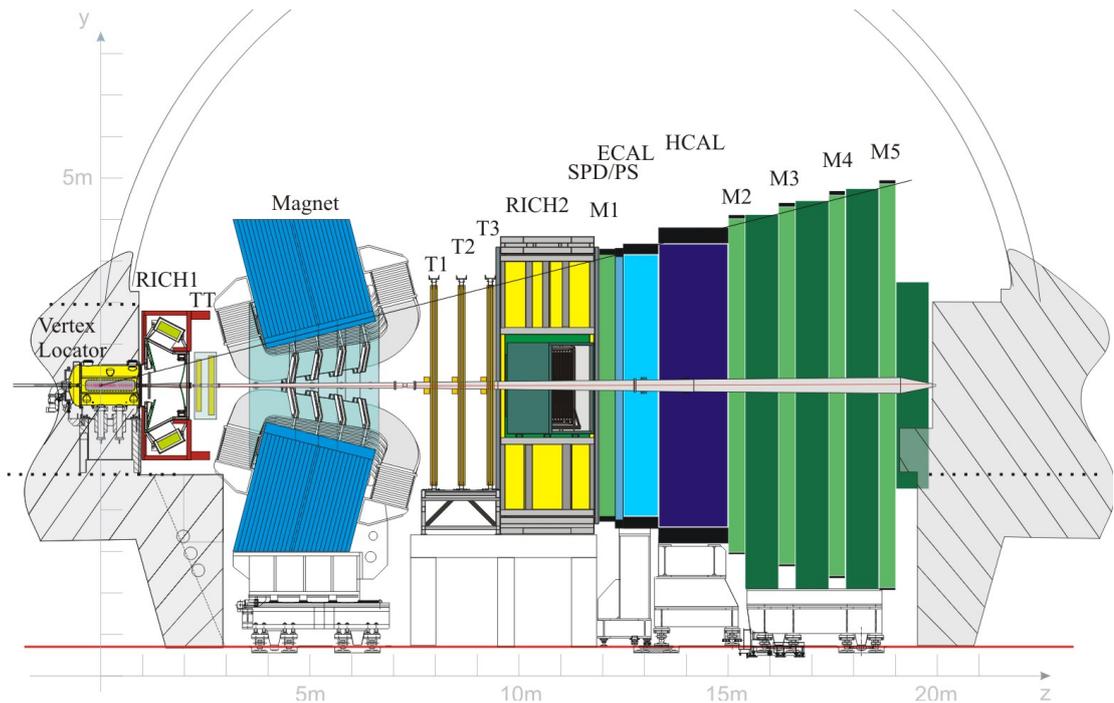


Figure 4.2: Layout of the LHCb spectrometer shown from the side [42]. The Vertex Locator (VELO) is located around the collision point. RICH1 and RICH2 are ring-imaging Cherenkov detectors. TT and T1-3 are the high-precision tracking stations. SPD, PS, ECAL and HCAL are parts of the calorimeter system. M1-5 are the muon stations.

4.2.1 Tracking Down Charged Particles

The tracking system of LHCb uses two different detector technologies – silicon micro-strips and straw tubes – in the vertex locator (VELO) and in the four planar tracking stations, of which three are located downstream of the magnet and one is placed upstream. In Figure 4.2 they are labelled TT, T1, T2, T3.

THE VERTEX LOCATOR The VELO is arranged around the proton-proton interaction point and is therefore the first sub-detector that is traversed by particles, while only charged particles leave tracks in its material. Its purpose is to identify the primary and displaced secondary vertices that are distinctive for b and c hadron decays. Furthermore, it allows to precisely measure the impact parameter (IP) of charged particles' trajectories. The IP is defined as the distance between the track and the primary vertex at the track's point of closest approach to the primary vertex (see Fig. 4.3). The VELO consists of 21 stations each of which has a silicon micro-strip module on the left and right of the beam axis. The modules are composed of two almost semi-circular parts measuring the distance R to the beam line and the azimuthal angle ϕ in the x - y plane, respectively (see Figure 4.4). The modules are arranged such that they have a small overlap along the y -axis to avoid gaps in the detector. The full diameter of a module is 90.5 mm, 300 μm in thickness and the minimum pitch between the strips in the inner region is around 40 μm . The sensors are placed at a radial distance of about 7 mm from the beam axis at a known position on the z -axis. The best hit resolution is around 4 μm and the resolution of the position of the primary vertex is around 15 μm in the presence of 20 tracks. The resolution of the IP in x - and y -direction is roughly

$50\ \mu\text{m}$ at a momentum of $10\ \text{GeV}/c$ (see Fig. 4.3). All modules are mounted in a vessel that maintains a vacuum separated from that of the LHC machine. The boundary between the two vacua consists of a thin-walled corrugated aluminium sheet (called *RF-foil* see Figure 4.5) to minimize the material that charged particles have to traverse before hitting the sensors. Aside from protecting the LHC vacuum from out-gassing of the VELO modules, the RF-foil also shields the detector against RF pick-up from the LHC beams. The RF-foil makes up the major part of the material budget present in the VELO and leads to measurable effects of multiple scattering and material interaction as shown in Figure 4.5. Knowledge about the material distribution in the VELO can be crucial for analyses that search for long-lived particles to reduce background from material interactions such as photon conversions. For this purpose the material of the VELO was mapped using secondary interactions of hadrons produced in beam-gas collisions collected special runs where helium gas was injected into the VELO. This is documented in Ref. [46].

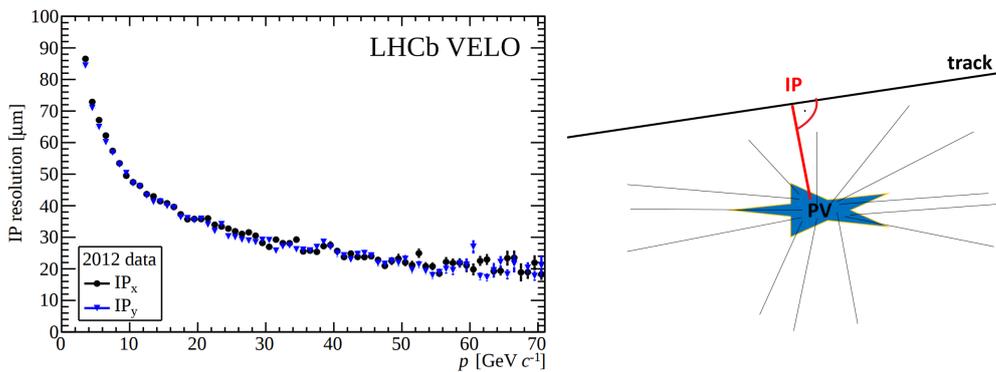


Figure 4.3: Right: Schematic of the impact parameter of a charged particle. Left: Impact parameter resolution as a function of momentum for the x - and y -component, respectively [47].

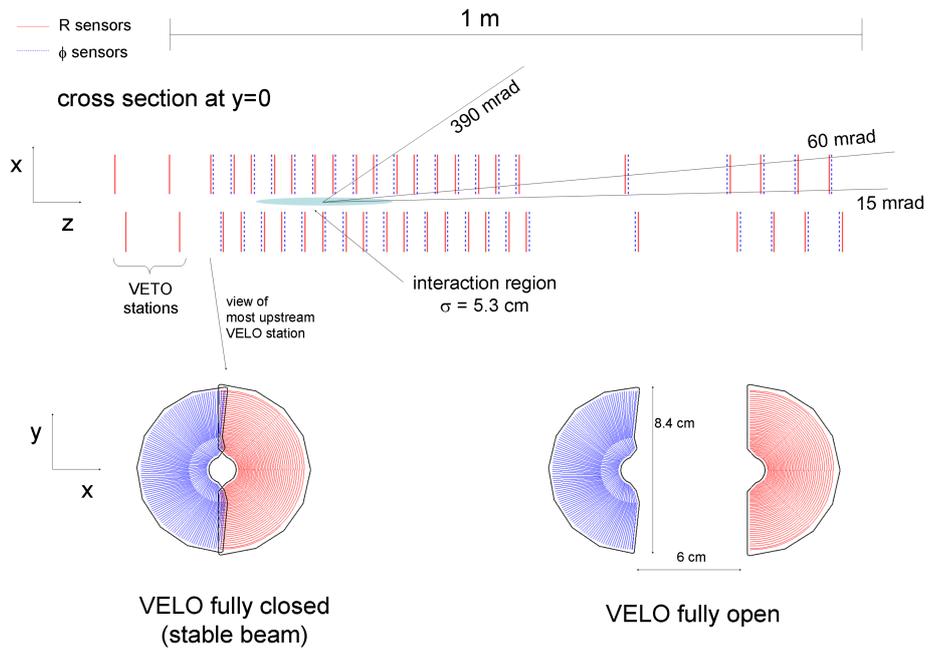


Figure 4.4: Setup of the VELO silicon modules along the beam axis, at $y = 0$, with the detector in the fully closed position [41]. The front face of the first modules is also illustrated in both the closed and open positions. "Closed" and "open" refers to the necessity to move the detector away from the interaction region (open) until a stable beam is established. As soon as stable beam conditions are obtained, the detector is placed into the optimal position (closed). Also shown are the angles for which at least three VELO modules are traversed by particles.

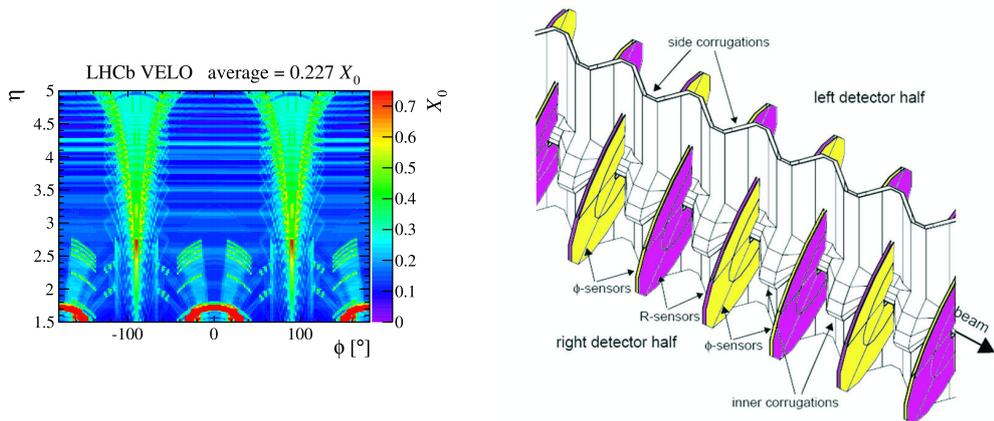


Figure 4.5: Left: Average radiation length X_0 seen by particles passing through the VELO as a function of the azimuthal angle ϕ and pseudorapidity η representing the material budget integrated from the interaction point to $z=835$ mm. Right: Inside of the RF-foil as modelled in GEANT4. The R - and ϕ -sensors are indicated by different shading. [41, 47]

THE SILICON TRACKER In addition to the VELO, there are two more silicon detectors that – as a whole – are called the Silicon Tracker (ST): the Trigger Tracker³ (TT) and the Inner Tracker (IT). Both, TT and IT, use silicon micro-strip sensors with a strip pitch of about $200\ \mu\text{m}$, resulting in a spatial resolution of about $50\ \mu\text{m}$. They are both composed of four layers arranged in an $(x-u-v-x)$ configuration: vertical strips in the first and the last layer, and strips rotated by a stereo angle of -5° and $+5^\circ$ in the $x-y$ -plane in the second and the third layer, respectively. The TT is located upstream in front of the magnet. It is a 150 cm wide and 130 cm high planar tracking station, covering the full acceptance of the experiment. The IT is installed downstream of the magnet at the centre of the tracking stations T1–T3, close to the beam pipe. It is a 120 cm wide and 40 cm high silicon micro-strip detector covering only approximately 2% of LHCb’s acceptance (see Figure 4.6). This technology was opted for in this region because of the high track density close to the beam pipe. It is estimated from simulation that $\sim 20\%$ of tracks within LHCb’s acceptance correspond to particles passing through the IT.

THE OUTER TRACKER The Outer Tracker (OT) covers the outer region of the three tracking stations T1–T3 (see Figure 4.6) where the particle flux is smaller compared to the inner region (IT). The gas-tight straw-tube modules, that are employed here, contain two layers of drift-tubes with inner diameters of 4.9 mm allowing to reconstruct tracks with a spatial resolution of about $200\ \mu\text{m}$. Charged particles passing through the straw tubes ionise the gas along their trajectory. By measuring the drift time of the created charges to the anode wire in the centre of the tube with respect to the bunch crossing time it is possible to determine the distance between the charged particle’s trajectory and the wire. The layout of the OT is similar to the one of the IT and TT– four layers in an $(x-u-v-x)$ arrangement. With an active area of $(5971 \times 4850)\ \text{mm}^2$, the outer boundary corresponds to the full acceptance of LHCb.

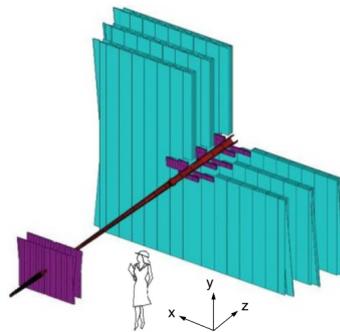


Figure 4.6: Schematic view of the LHCb tracking system [41]. The TT is shown on the left front-side, the stations T1–T3 in the back right-side. Purple marks the ST and turquoise marks the OT. The beam pipe passes through in the middle.

MAGNET LHCb’s dipole magnet consists of two saddle-shaped aluminium coils in a window-frame steel yoke with a total weight of roughly 1600 tons. The magnet poles are tilted towards the interaction point, following the acceptance of LHCb (cf. Figure 4.2). It is a water-cooled warm magnet with a bending power of about 4 Tm. The deflection of charged particles takes place mostly in the $x-z$ -plane. The two coils are identical and placed

³ Also known as Tracker Turicensis.

mirror-symmetrically to each other. Their polarity is reversed periodically during data taking which on average mitigates the folding of detection asymmetries of charged particles into the data. The polarities are denoted *MagUp* and *MagDown* for positive and negative polarity, respectively.

TRACK RECONSTRUCTION The quality of the reconstructed track strongly depends on the number of clusters a charged particle leaves in the tracking system. The most precise estimate of momentum and trajectory of a particle can be given for so-called *long tracks*. These are tracks that have hits in the VELO and in the stations T1–3, and optionally in the TT (see Figure 4.7 right). The relative momentum resolution for long tracks below 20 GeV/*c* is about 0.5% (see Figure 4.7). They are reconstructed by building tracks from hits in the VELO and TT. Because of a small fringe field between the VELO and TT originating from the magnet, a rough estimate of the momentum and charge of these track is possible. They are then subsequently combined with hits from the T stations and fitted using a Kalman filter algorithm [48], taking into account multiple scattering and energy loss by ionisation. The analysis presented here makes use only of tracks that fulfil the requirements for long tracks. Other types of tracks (*e.g.* only hits in TT and T stations, called *downstream track* in Figure 4.7) suffer from larger background rates and worse momentum resolution, particularly in the case of low-momentum electrons as present in the $D^{*0} \rightarrow D^0 e^+ e^-$ decay.

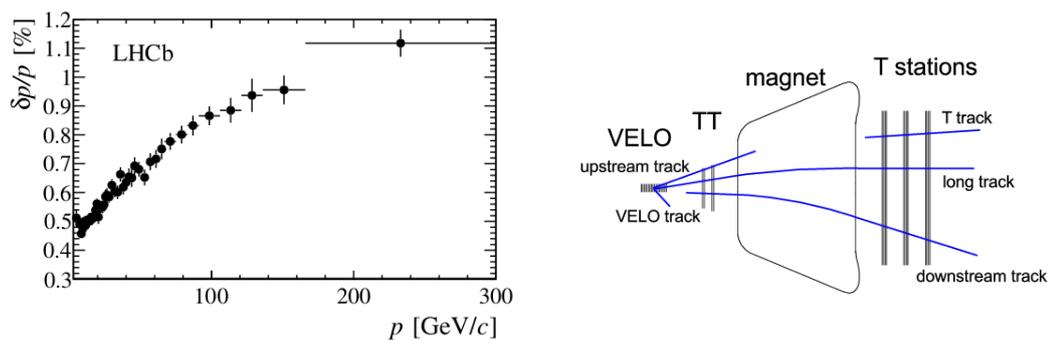


Figure 4.7: Left: Relative momentum resolution versus momentum for long tracks as obtained from data of J/ψ decays [42]. Right: Track types reconstructed by the track finding algorithms [49].

4.2.2 A Particle's Identity

The only measurable quantity that unambiguously distinguishes particles of different kinds is their mass. Hence, the goal of particle identification (PID) is to test a hypothesis about the measured particle's mass based on information from the sub-detectors described below. For particle candidates used in data analysis the mass is then set to its nominal value. Combining the momentum information from the track reconstruction with this mass gives the full four-momentum vector and thus full knowledge about the particle's kinematic state. The PID information is provided by four different detector systems: the two RICH detectors, the calorimeter system and the muon stations.

THE RICH SYSTEM The RICH system is built up of two ring-imaging Cherenkov detectors, RICH1 and RICH2, which identify charged hadrons (essentially π^\pm , K^\pm and p) over a momentum range from $2\text{ GeV}/c$ to $100\text{ GeV}/c$. Especially the K - π discrimination is crucial, since these particles are often produced in decays of B and D mesons and also occur in the final state studied in this analysis via $D^0 \rightarrow K^-\pi^+$. RICH1 is located downstream of the VELO and provides PID information for low and intermediate momentum particles, while RICH2, located downstream of the magnet, aims at identification of particles with higher momentum. Two radiators, C_4F_{10} and CF_4 which have different refractive indices, are used in RICH1 and RICH2, respectively. The distinction of particles is achieved by evaluating the relation between the Cherenkov angle and the momentum of the particles as shown in Figure 4.8. The separation between kaons and pions is clearly visible and leads to very efficient identification of these kind of particles together with a low mis-identification rate in the typical momentum region as shown in Figure 4.9. In addition to the radiators mentioned above, Figure 4.8 also shows the momentum dependence of the Cherenkov angle for aerogel which covers an even lower momentum region than the gas currently used in RICH1 and in principle allows for discrimination between very low-momentum electrons, muons and light mesons. This technology was used in Run I of the LHC, however, the aerogel was removed for Run II as its ability to provide PID for particles with an energy below the Cherenkov threshold of C_4F_{10} for kaons⁴ is compromised by the total number of photons in RICH1 in the higher track multiplicity environment of Run II [50]. Nevertheless, the RICH system still provides useful information for low-momentum particles. Especially electrons with momenta below $5\text{ GeV}/c$, which frequently occur in $D^{*0} \rightarrow D^0 e^+ e^-$ decays, can be distinguished well from pions and muons as is visible in Figure 4.8 (left) in the uppermost left corner. Electrons with higher momenta already irradiate their Cherenkov light very close to the maximum Cherenkov angle for C_4F_{10} and can thus barely be differentiated from other high energy particles. For these electrons another detector technology has to be employed as explained in the next paragraph.

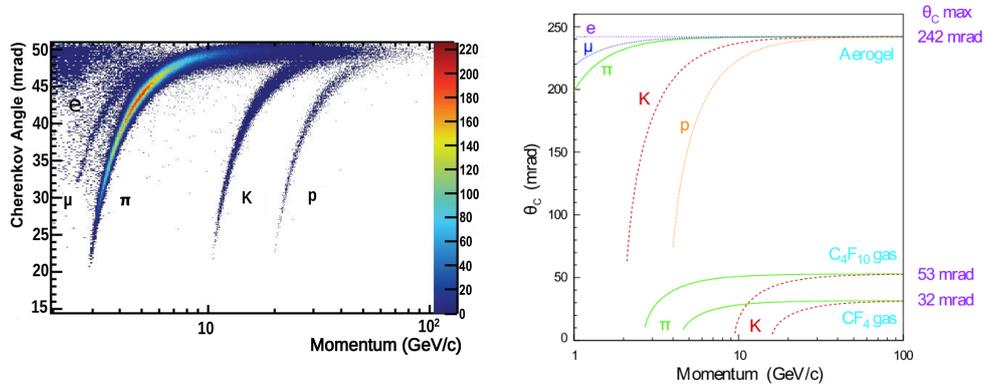


Figure 4.8: Left: Cherenkov angle versus momentum in C_4F_{10} for isolated tracks in data [42]. The curved bands clearly show the different particle types. Right: Calculated Cherenkov angles versus particle momentum for different radiators [41].

⁴ The refractive index of C_4F_{10} is $n=1.0014$ [41], which gives $E_{th}(K^\pm) = \frac{m_{K^\pm} c^2}{\sqrt{1-1/n^2}} \approx 9.3\text{ GeV}$.

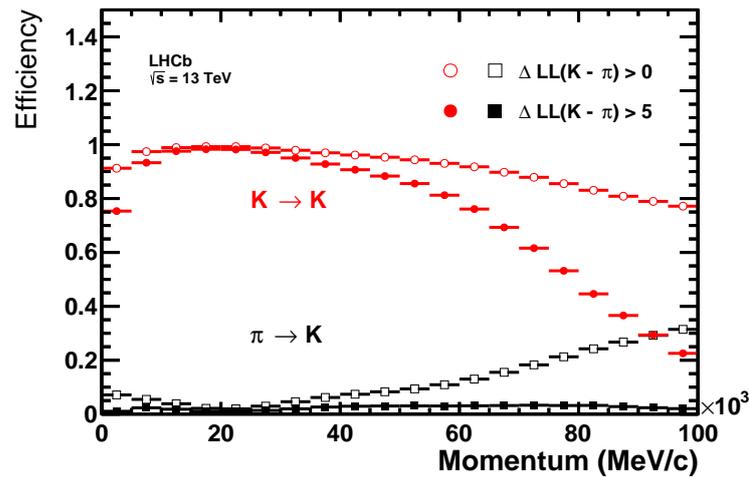


Figure 4.9: Efficiency of correctly identifying a kaon and efficiency of mis-identifying pions as kaons versus momentum for two different cuts in the likelihood difference (variable on the basis of which the mass hypothesis is accepted/rejected) for 2015 data [50].

THE CALORIMETER SYSTEM Located downstream of RICH2, the calorimeter system forms a huge bulk of material in which hadrons, electrons and photons show their characteristic particle showers. The calorimeters measure positions and energies of the particles that interact with their material, providing information that is essential for PID and efficient trigger strategies. In particular, the identification of electrons and photons is a crucial task of the calorimeter system. This is done by the Electromagnetic CALorimeter (ECAL) together with two additional detectors, the PreShower detector (PS) and the Scintillator Pad Detector (SPD), which are placed in front of it. The SPD detects charged particles that pass through and thus allows to distinguish between electrons and neutral particles like photons and neutral pions that subsequently produce similar shower shapes in the calorimeters. It is also used to measure the number of charged particles or tracks per event that enter the calorimeter system, the information of which can be used to discard events that are too crowded to be measured accurately. The SPD is followed by the PS that mainly causes electrons and photons to shower with its 15 mm of lead. Since hadrons have a longer interaction length than electrons, the PS provides a longitudinal segmentation which is useful to distinguish electrons from charged hadrons. The combination of information from the PS with the measurement of the ECAL improves the localisation of electrons and photons. The ECAL itself is composed of several absorption layers, each with 2 mm of lead and 4 mm thick scintillator tiles, which form a 42 cm deep stack of material corresponding to 25 radiation lengths. This structure ensures containment of the whole or at least the largest part of electromagnetic showers caused by electrons and photons and it is designed to have an energy resolution of $\sigma_E/E = 10\%/\sqrt{E} \oplus 1\%$ (E in GeV). The last station in the calorimeter system is the Hadronic Calorimeter (HCAL). It is built from iron and scintillating tiles as absorber and active material, respectively. The thickness of the HCAL corresponds to 5.6 interaction lengths which is limited by the space available in the detector cavern. The upstream ECAL, however, adds another 1.2 interactions lengths. The energy resolution of the HCAL determined during test beams is $\sigma_E/E = (69 \pm 5)\%/\sqrt{E} \oplus (9 \pm 2)\%$ (E in GeV). Because all parts of the calorimeter system give crucial information about the identity of the particles that traverse it, it is widely used for triggering on these very particles. This is only possible

because the combination of scintillation tiles and photo-multiplier tubes, that are placed directly on the modules of the ECAL and HCAL, allows for fast readout.

THE MUON STATIONS Muons play a key role in many physics analyses at LHCb because they often occur in final states of interesting decays and leave clean signatures in the detector. The sub-detectors M1-M5 in Figure 4.2 are designed to allow for efficient and fast triggering on muons and subsequent identification. M1 is located upstream of the PS, to improve transverse momentum measurements in the trigger. M2-5 sit downstream of the calorimeters. Placed between these stations are iron absorbers, each 80 cm thick to reduce the hadronic background. M1-5 mainly use multi-wire-proportional chambers as detector technology allowing for fast readout. Only the inner region of M1 is made of gas electron multiplier chambers because of the higher particle flux. Since there are no muons present in the decay studied in this analysis, the muon stations do not play an important role here.

GLOBAL PARTICLE IDENTIFICATION All of the detector systems mentioned above can contribute to give a final estimate about the identity of the particle under consideration. The sub-detectors that provide the most useful information will vary depending on the kind of particle that is measured and in general also on its kinematics. The RICH system mostly helps to identify charged hadrons, especially kaons and pions, if these are above threshold and in the right momentum region. For electrons and photons the PS and ECAL give the most vital information, while the RICH can only be useful for electrons with comparatively low momentum. Since the response of every sub-detector on its own is not meaningful enough to give sufficient information about the likelihood of the particle that has been observed, the available information is integrated into more powerful combined likelihoods [42, 51]. Two different approaches are used for this combination. The simpler method adds the likelihood information of each sub-system linearly to form *Combined Differential Log-Likelihoods* (CombDLL), which is the likelihood ratio $\Delta\mathcal{L}_{comb}(X - \pi) = \log(\mathcal{L}(X)/\mathcal{L}(\pi))$ [52], where X represents the mass hypothesis of the electron, muon, kaon or proton. The CombDLL is then a measure of how likely the mass hypothesis of the considered particle is relative to the null hypothesis of the most abundant particle type, the pion. This variable is widely used in LHCb data analyses and also often referred to as PIDX with $X \in (e, \mu, K, p)$. The simple approach, however, mostly disregards the deeper connection between a particles motion through the detector, the global event structure such as number of tracks and hits in the system, and the response they trigger in the PID detectors. Finding and using these correlations is a task well suited for machine learning algorithms. The output of the simple approach is further exploited by feeding it as input into an artificial neural network (ANN) together with momentum variables obtained from the tracking system, track quality variables and additional information from the PID detectors not entering the likelihood computation. Being trained on simulated events, the ANN provides a single probability⁵ (typically called *ProbNNX*) for each particle hypothesis. The training samples usually contain decays of heavy hadrons that feature the kinematic distributions of processes studied in data analyses. The exact composition, quality and number of simulated events used changes the response of the ANN classifiers and allows for different and more specialised tunes of ProbNN variables. In addition to the standard ProbNN variables, more sophisticated machine learning algorithms such as deep neural networks [53] and special gradient boosting algorithms like CatBoost [54] can be used to improve the PID performance. These novel variables are often referred to

⁵ Not in a sense of a frequentist probability but rather a "degree of belief" probability.

as *Yandex PID* (yPID) and make use of even more input variables. First performance studies of yPIDs show good results especially for the identification of electrons (see Figure 4.10), which makes these variables of interest for this analysis.

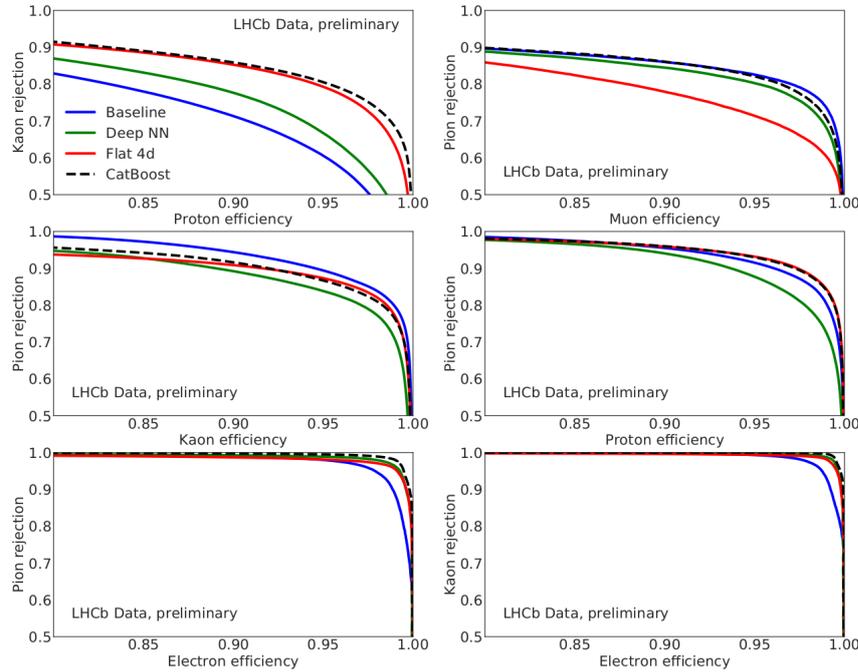


Figure 4.10: Receiver operating characteristic curves for comparison between yPID variables obtained from a deep neural network (green), gradient boosting (dashed), with flat response in kinematic variables (red) and the baseline standard ProbNN variable (blue) for different particle types [55].

4.3 TRIGGER, DATA ACQUISITION AND DATA FLOW

Most interesting physics processes for LHCb occur at rates much lower than the rate at which protons are brought to collision. The efficient selection of these interesting decays from the ~ 30 M proton–proton collisions per second⁶ is indispensable due to limited bandwidth of the detector readout and limited computing resources. This challenge is addressed at LHCb using a three-level trigger system consisting of the level-0 hardware trigger (L0) that reduces the event output rate to the maximum allowed by the front-end electronics (1.1 MHz), a first stage of the software trigger (HLT1)⁷, which further reduces the rate to ~ 150 kHz, and the last stage of the software trigger (HLT2) the final output rate of which is then around 10 kHz [56]. The L0 decision is based on separate FPGA⁸ algorithms that use information from the ECAL, HCAL and muon stations, respectively. Events that triggered this hardware stage are passed to the HLT1 software which reconstructs trajectories of charged particles using information from the VELO, TT and tracking stations. Events that fulfil all requirements from at least one HLT1 trigger line are in turn passed to the HLT2. The HLT2 accesses information from all sub-detector systems and already performs a full event reconstruction. This is called *online*

⁶ This number is limited by the bunch crossing rate of 40 MHz and takes into account gaps in the bunch trains.

⁷ HLT stands for High Level Trigger.

⁸ Field Programmable Gate Array

reconstruction in contrast to the *offline reconstruction* which is performed on raw data that has already been written to permanent storage. Between Run I and Run II of the LHC the online reconstruction has been improved such that its quality is identical to the one of the offline reconstruction. This is possible due to a buffer that was included between HLT1 and HLT2 allowing for direct event reconstruction with high precision because of real-time alignment and calibration of the detector. The Run-II data and calibration flow is shown in Figure 4.11. With the high quality online reconstruction it is now possible to skip the offline reconstruction and directly use the output of the HLT2 for physics analyses. This is called Turbo stream [57, 58] since it delivers data usable for physics analysis faster than the traditional Full stream. A comparison between the streams is depicted in Fig. 4.12. An event written to permanent storage by the Turbo stream contains only the information that is requested by the trigger line, *i.e.* only information that is deemed useful to analyse a certain set of decays. On the one hand, this reduces the event size by an order of magnitude, then again if all of the information about an event is needed the analyst still has to fall back to the Full stream⁹. The analysis presented here exploits the opportunity given by the Turbo stream to start analysing the data shortly after it has been recorded.

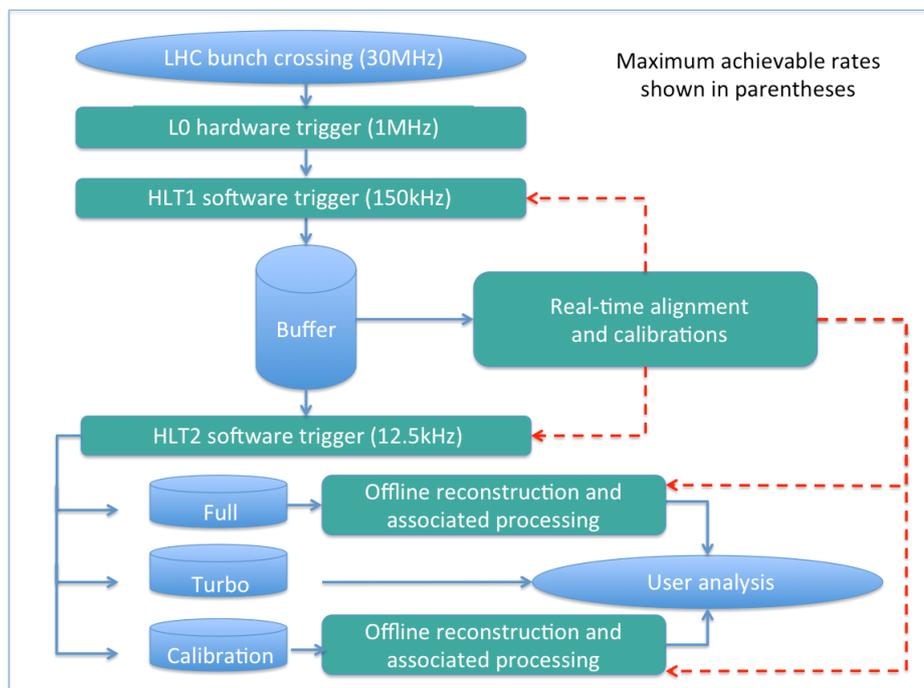


Figure 4.11: LHCb data processing in Run II [58]. The blue solid lines represent flow of data and the red dashed line the propagation of calibrations.

⁹ Additionally, there is also Turbo++ which saves the whole event reconstruction performed by the HLT2 when the trigger line fires.

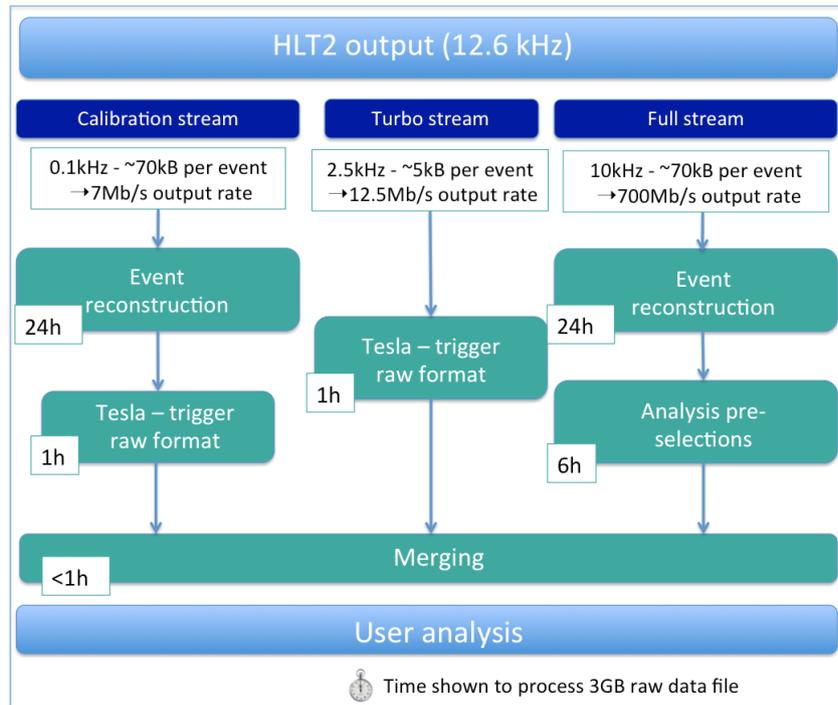


Figure 4.12: Comparison between Turbo data processing and the traditional Full stream [58]. The time in hours is given for each step to process 3 Gbytes of raw data. Also shown is the calibration stream which separates events to calculate data-driven efficiencies for both the Full and Turbo Stream. The software application that is used to format the raw output of the trigger for user analysis is called TESLA.

4.3.1 LHCb Software and Simulation

Each data-processing step after the L0 trigger is controlled by a dedicated software application that processes the data event-by-event. The data flow and the associated applications are depicted in Fig. 4.13. Events that pass the L0 trigger are handled by the software application of the HLT1 and HLT2 called MOORE. The event reconstruction done by MOORE is practically identical to the one subsequently performed by BRUNEL in the offline reconstruction step. This allows for splitting the data flow into the Turbo stream and Full stream as mentioned above. In the Full stream, after the offline reconstruction by BRUNEL, the data is further filtered through a set of selections called *stripping*, controlled by the DAVINCI application. These ‘stripped’ data sets are stored and made available for user analysis. In the Turbo stream, the output of MOORE is formatted by TESLA for direct availability to users, *i.e.* without stripping. The actual data set (“Ntuple”) containing the desired variables used for a specific analysis can be created individually using DAVINCI. Usually, analyses do not only rely on the actual data that has been recorded but also make use of simulated events, called Monte Carlo (MC) data. Simulated events should be as similar to data as possible. Their processing is therefore very similar to the one of real data. Two main software applications replace the proton-proton collisions and the detector response. The event simulation is controlled by GAUSS which calls PYTHIA for the generation of inelastic proton-proton collisions. If a particle of interest was produced in PYTHIA, it is forced to decay to the desired final state, simulating its kinematics with EVTGEN. Subsequently, the whole event is propagated through a virtual model of the detector, simulating the particle’s interactions with the detector material

using GEANT4. The BOOLE application takes the simulated energy depositions in the detector material and converts them to signals mimicking the real detector response. After BOOLE, the simulated output is similar enough to the real detector output, such that the simulated events can run through the usual data processing chain.

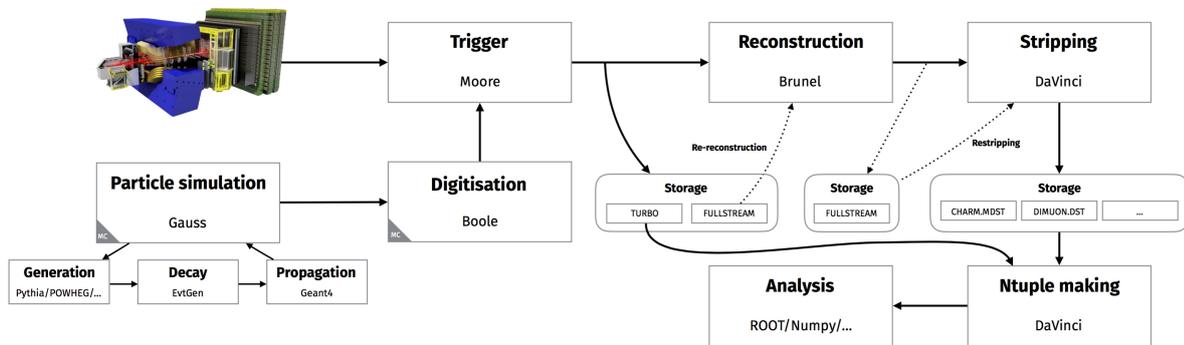


Figure 4.13: LHCb data and simulation flow [59]. Each step of the data flow is shown with its associated software application. Steps that are only taken for simulated events are marked with an "MC" in the lower left corner of the boxes.

5

ANALYSIS STRATEGY AND TOOLS

This chapter outlines the overall strategy employed to analyse the data recorded by the LHCb experiment and summarises the most important statistical tools that are used in the analysis process.

5.1 STRATEGY

The goal of the analysis presented in this thesis is to perform the first observation and branching ratio measurement of the Dalitz decay $D^*(2007)^0 \rightarrow D^0 e^+ e^-$ relative to the well-established radiative decay $D^*(2007)^0 \rightarrow D^0 \gamma$, where in both channels the D^0 decays via $D^0 \rightarrow K^- \pi^+$ (cf. Tab. 2.1). The Dalitz decay and the radiative decay are in the following also referred to as *signal* channel and *control* or *normalisation* channel, respectively. The benefits of considering ratios of decay widths has been brought up already in Sec. 2.1.2 where the form factor of the $D^{*0}-D^0$ transition can be dropped in the calculation of the ratio of decay widths. Similarly, experimental quantities with potentially large uncertainties such as the cross section of producing a D^{*0} meson in proton-proton collisions, the exact value of the data's integrated luminosity and uncertainties related to the measurement of $D^0 \rightarrow K^- \pi^+$ cancel in the ratio, allowing for more precise determination of the ratio than of the individual branching fractions. Yet knowing this ratio, the branching fraction of the decay $D^{*0} \rightarrow D^0 e^+ e^-$ can still be determined because the branching fraction of the normalisation channel has already been measured [60]. As this thesis aims at a first observation of the signal channel, the analysis is performed with a blinded¹ signal region during the selection of signal candidates. The selection is explicitly explained in Chap. 6. The analysis steps can be outlined as follows:

0. **Trigger selection:** Events recorded by the LHCb detector are filtered through a trigger line, selecting events with signatures that are expected for the signal and control channel.
1. **Pre-selection:** A loose selection is applied to the dataset, reducing the amount of combinatorial background such that the control channel becomes visible in data.
2. **Signal selection:** The combinatorial background is further reduced by applying a multivariate selection using gradient boosting of decision trees with a special loss function allowing for uniform classification of signal or background (see Sec. 5.4). The multivariate classifier is trained on simulated data of the signal channel that has been corrected for discrepancies between simulation and data, using information obtained from the control channel via the *sPlot* technique (see Sec. 5.3). The background training sample is taken from a data sideband.

¹ Blinding is a common procedure in data analysis to avoid observer bias by hiding the final result until all analysis steps are fixed.

3. **Yield extraction:** After the selection, the yielded number of candidates for the decays $D^{*0} \rightarrow D^0 e^+ e^-$ and $D^{*0} \rightarrow D^0 \gamma$ are obtained by fitting the corresponding mass distributions using extended maximum likelihood fits (see Sec. 5.2).
4. **Efficiency determination:** The selection efficiencies are estimated using simulated samples of both the signal and the control channel. The samples are corrected for data-simulation discrepancies to obtain a better estimate of the efficiencies.
5. **Branching fraction determination:** The ratio of signal and control channel yields eventually gives the branching fraction of both channels relative to each other, taking into account the corresponding efficiencies $\varepsilon_{D^0 e^+ e^-}$ and $\varepsilon_{D^0 \gamma}$:

$$\frac{N_{D^0 e^+ e^-}}{N_{D^0 \gamma}} = \frac{\mathcal{B}(D^{*0} \rightarrow D^0 e^+ e^-) \times \varepsilon_{D^0 e^+ e^-}}{\mathcal{B}(D^{*0} \rightarrow D^0 \gamma) \times \varepsilon_{D^0 \gamma}} \quad (5.1)$$

The branching fraction of the $D^{*0} \rightarrow D^0 e^+ e^-$ decay can then be inferred by multiplication with the known branching ratio of the decay $D^{*0} \rightarrow D^0 \gamma$.

5.2 THE METHOD OF MAXIMUM LIKELIHOOD

Any measurement of a quantity can be viewed as sampling from an underlying probability density function (PDF) which contains the information about how likely the outcome of that measurement in comparison to all possible outcomes is. Although the exact PDF from which an experiment samples is hardly ever known, it can usually be approximated by a function depending on some parameters. In particle physics, these parameters often correspond to fundamental physical quantities such as masses and natural widths of particles, but also to experimental parameters such as the resolution of a detector or the number of particles that have been observed. Hence, it is a common problem to estimate these parameters given a finite sample of data and it can be tackled using the method of maximum likelihood. Supposing the data contains n independent measurements of the random variable x , *i.e.* x_1, \dots, x_n , and a functional form of the PDF $f(x; \theta)$ is known or assumed, yet depending on an unknown parameter θ . Then, a so-called *likelihood function* can be defined as

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta) \quad (5.2)$$

which is just the joint PDF for the x_i treated as a function of θ as the x_i are fixed measurements [61]. The best estimator for θ is the one that maximises the likelihood function, *i.e.* the parameter θ giving the highest probability to measure the data that has actually been measured. In cases where $L(\theta)$ is differentiable, this estimator can be easily found by evaluating the null of the likelihood function's derivative. In particle physics, the number n of independent measurements itself often is a random variable of interest with a mean value ν , *e.g.* corresponding to the number of certain particle decays that have been observed. In this case, the maximum likelihood method can be extended by a Poisson distribution according to which n is distributed, leading to the *extended* likelihood function:

$$L(\nu, \theta) = \frac{\nu^n}{n!} e^{-\nu} \prod_{i=1}^n f(x_i; \theta) = \frac{e^{-\nu}}{n!} \prod_{i=1}^n \nu f(x_i; \theta) \quad (5.3)$$

In practice, the negative log-likelihood function

$$NLL(\theta) \equiv -\ln L(\theta) = -\sum_{i=1}^n \ln f(x_i; \theta) \quad (5.4)$$

is often used as it is numerically more stable to transform potentially small numbers such as probabilities into larger negative values via the logarithm and to subsequently sum them instead of taking the product. Furthermore, since it is convention in optimisation methods to minimise functions rather than maximising them, the log-likelihood is made negative. All fits performed in this analysis make use of the (extended) maximum likelihood method to estimate the fit parameters.

5.3 UNFOLDING DISTRIBUTIONS

The situation described in Sec. 5.2 can easily be generalised to a more realistic case where x is not a single random variable but represents a multidimensional random vector \vec{x} , *i.e.* the outcome of an observation is characterised by several quantities (*e.g.* mass and momenta). Similarly, the extended likelihood function in Eq. 5.3 can be generalised to describe multiple sources of observed events (*e.g.* signal and background) with several parameters $\vec{\theta}$. The extended log-likelihood function can then be given as [61]:

$$\ln L(\vec{\mu}) = -\sum_{j=1}^m \mu_j + \sum_{i=1}^n \ln \left(\sum_{j=1}^m \mu_j f_j(\vec{x}_i) \right) \quad (5.5)$$

where n is the total number of events considered, m is the number of species of events populating the data sample, $\mu_j = \theta_j v$ is the number of events expected on average for the j^{th} species. In a common particle physics case, the species yield the expected number of signal events μ_S and background events μ_B , which are folded inside the dataset. As it is often of interest to study the \vec{x} -distributions of one species, signal and background have to be unfolded. This can be achieved by applying the *sPlot* technique [62]. In a nutshell, the idea is to perform an extended maximum likelihood fit to a *discriminating* variable y (*e.g.* the invariant mass distribution of a particle) to find the average number of signal and background events present in the dataset. This information is subsequently used to calculate event weights for each species such that the weighted distributions of *control* variables \vec{x} reflect the genuine distribution of one species only. In a simplified manner, the *sPlot* technique can be sketched mathematically as

$$(\mu_S + \mu_B)f(\vec{x}, y) = \mu_S f_S(\vec{x}, y) + \mu_B f_B(\vec{x}, y) \quad (5.6)$$

where f_S and f_B are the PDF of the signal and background component, respectively, and f is the combined PDF of both. The goal now is to transform $f(\vec{x}, y)$ such that the distribution $f_S(\vec{x})$ is obtained:

$$\mu_S f_S(\vec{x}) = (\mu_S + \mu_B) \int f(\vec{x}, y) w(y) dy \quad (5.7)$$

The weight function $w(y)$ projects out the signal PDF for one fixed \vec{x} and is in principle arbitrary. However, to generalise the transformation for all \vec{x} it has to be assumed that f_S and f_B factorise:

$$f_S(\vec{x}, y) = f_S(\vec{x})f_S(y) \quad (5.8)$$

$$f_B(\vec{x}, y) = f_B(\vec{x})f_B(y) \quad (5.9)$$

i.e. that \vec{x} and y are uncorrelated. The weight function is then further constrained by

$$\int w(y)f_S(y)dy = 1 \quad (5.10)$$

$$\int w(y)f_B(y)dy = 0 \quad (5.11)$$

and optimised for minimal variance by requesting:

$$\int w^2(y)f(\vec{x}, y)d\vec{x}dy \stackrel{!}{=} \min \quad (5.12)$$

The constrained minimisation gives the optimal weight function

$$w(y) = \frac{\alpha f_S(y) + \beta f_B(y)}{\mu_S f_S(y) + \mu_B f_B(y)} \quad (5.13)$$

where α and β are Lagrange multipliers which can be determined using the constraints given by Eq. 5.10 and Eq. 5.11. The weight function $w(y)$ assigns a weight to each event according to its y value. The weighted events are consequently effectively distributed according the PDF of the desired species in the control variables \vec{x} .

5.4 GRADIENT BOOSTING AND DECISION TREES

A major part of data analysis consists of the attempt to classify samples of data into categories of interest. In the case of a search for a specific particle decay, the categories are signal and background. It is thus desired to find a strong test statistic² that allows to distinguish between signal and background and subsequently to select the signal category. According to the Neyman–Pearson lemma [52], the test statistic with the highest power is given by the likelihood ratio, *i.e.* by the ratio of the PDFs describing each category. A category's full PDF, though, is typically not known. Yet the desired strong test statistic can still be approximated by using general parametrisations such as deep/artificial neural networks, which are described in detail *e.g.* in Ref. [63], or decision trees. Decision trees are especially suited for binary classification problems as they split the data at each node according to a binary decision on the measured variables (see Fig. 5.1). The decision is based on *training* data that has to be passed to the algorithm implementing the decision tree structure. Training data typically comes from simulation or reference processes for which the category of each sample³ is known. If the size of the decision tree is not restricted, it is in principle possible to grow it until all samples in the training data are classified correctly. This classifier, however, does in general not approximate the sought-after optimal test statistic as it also parametrises statistical

² A test statistic is a function of measured variables.

³ The word "sample" can also be replaced with "event" in the context of particle collisions or decays.

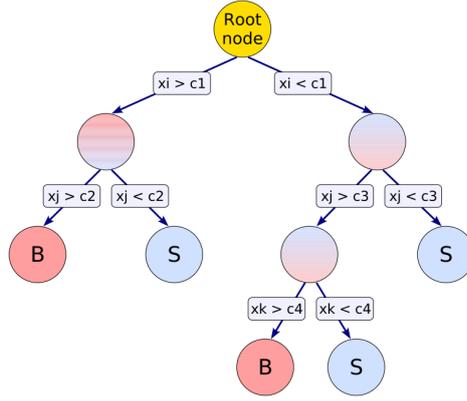


Figure 5.1: Schematic view of a decision tree [64]. At each node a decision is made using one of the input variables in \vec{x} . Leaf nodes are denoted "S" and "B", respectively, in the style of a signal-background classification.

fluctuations within the training data. In the field of particle physics this is often referred to as *overtraining*⁴ and leads to loss of the classifier's performance when it is applied to new data samples. On the other hand, restricting the size of a decision tree heavily weakens its ability to approximate the optimal test statistic. A solution to this problem is the so-called *boosting*. Instead of growing one large decision tree, boosting algorithms grow a forest of several shallow decision trees or in general a series of so-called *weak learners*. During the process of boosting the importance of a single sample varies according to the classification error of the previous trees, such that successive trees pay more attention to previously falsely classified samples. The combination of weak learners to a single strong learner is typically more stable against overtraining than a single large decision tree. A multi-purpose boosting algorithm that is exploited in this thesis is called *gradient boosting* [66]. The sought-after optimal test statistic can be parametrised by a weighted sum

$$s(\vec{x}; P) = \sum_{j=0}^m \beta_j f(\vec{x}; \vec{a}_j); \quad P \in \{\beta_j; \vec{a}_j\}_0^m \quad (5.14)$$

where $f(\vec{x}; a_j)$ corresponds to a weak learner, *e.g.* a shallow decision tree, the structure of which depends on the input variables \vec{x} and the parameters \vec{a}_j , which are the splitting variables, split locations and the leaves of the individual trees. The boosting adjusts the parameters P such that the deviation between $s(\vec{x}; P)$ and the true value γ , defining the true class of the sample obtained from training data, is minimised [64]. This deviation is measured by the *loss function* $L(s, \gamma)$. A popular choice of loss function for a classification problem is the so-called AdaLoss function [67]

$$L_{ada} = \sum_{i=1}^n \exp(-\gamma_i s_i) \quad (5.15)$$

⁴ Machine learning people rather call this *overfitting* and it is related to a more general problem called the *bias-variance tradeoff* [65].

where the sum runs over all samples. The i^{th} sample has $\gamma_i = \pm 1$, e.g. +1 for a signal and -1 for a background sample, and s_i is given by Eq. 5.14. The minimisation is accomplished using the gradient descent method with the gradients of the negative loss function⁵

$$-\frac{\partial L_{ada}}{\partial s_i} = \gamma_i \exp(-\gamma_i s_i) \quad (5.16)$$

which are positive and negative for signal and background, respectively, and have larger moduli for poorly classified samples. The gradient boosting technique is multi-purpose as it can be used with different loss functions serving special purposes. Aside from classification as in the case of AdaLoss, such a purpose can be the uniformity of $s(\vec{x})$ with respect to certain variables \vec{u} , e.g. the invariant mass of a particle. A loss function suited for this purpose is the *flatness* loss function [68]:

$$L_{flat} = \sum_b w_b \int |F_b(s) - F(s)|^2 ds \quad (5.17)$$

where b denotes a bin of data in \vec{u} , w_b is the fraction of e.g. signal samples in bin b , $F_b(s)$ is the cumulative distribution of classifier responses in bin b and $F(s)$ their global cumulative distribution. The gradients of this loss function are

$$-\frac{\partial L_{flat}}{\partial s_k} \approx -2w_b(F_b(s_k) - F(s_k)) \quad (5.18)$$

where b contains the k^{th} sample. To approximate the test statistic that does not only efficiently classify samples but also tends to a uniform distribution in \vec{u} , both loss functions L_{ada} and L_{flat} can be combined:

$$L_{ada+flat} = L_{flat} + \alpha L_{ada} \quad (5.19)$$

where α is a real-valued parameter that controls the trade-off between classification quality and uniformity.

One of the key strengths of boosted decision trees and machine learning techniques for function parametrisation in general is that they are able to exploit correlations among the given variates. This is not only powerful in the case of a classification problem but it is also extremely useful in matching multi-variate distributions to each other. A typical application of such a matching is the correction of simulated distributions by distributions of data. The simple approach of re-weighting one-dimensional variable distributions, such that they agree, is technically not sufficient since the joint PDF of all variables can still be wrong. To ensure that correlations are respected in the re-weighting procedure, gradient boosted decision trees can be applied also for this purpose. A description of this gradient-boosting re-weighter can be found in Ref. [69].

⁵ The gradient of the (negative) loss function is also called *pseudo-residual*.

6 | SELECTION

This chapter documents the various steps taken to separate the signals of the $D^{*0} \rightarrow D^0 e^+ e^-$ decay and the control channel decay $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+ e^-)$ from background events. Furthermore, it describes the study of differences between signal and control channel.

6.1 DATASETS AND TRIGGER

The data analysed in this work has been recorded by the LHCb experiment from proton-proton collisions at a centre-of-mass energy of $\sqrt{s} = 13 \text{ TeV}$ during the year 2017, and corresponds to an integrated luminosity¹ of 1.87 fb^{-1} . As a first selection step, events are required to pass the three consecutive trigger stages described in Sec. 4.3. No specific requirement is put on the selection performed by the L0 hardware- and HLT1 software-stage.² Examples of hardware-trigger decisions that might let $D^{*0} \rightarrow D^0 e^+ e^-$ and $D^{*0} \rightarrow D^0 \gamma$ candidates pass are LOELECTRON or LOHADRON, which require electron and hadron candidates, respectively, to be above a certain transverse energy threshold and remove events with high multiplicity in the SPD detector ($n\text{SPDHits} < 450$). The first software stage typically triggers on track and trajectory properties of charged particles, e.g. the line HLT1TRACKALLLO requires a certain minimum momentum p and transverse momentum p_T of tracks that additionally have to be detached from the primary interaction vertex, which is usually fulfilled by $D^0 \rightarrow K^- \pi^+$ decays. Events passing any of the first two trigger selections are further processed by the HLT2. In 2017, a dedicated HLT2 trigger line was implemented as preparation for the analysis of D^{*0} decays involving electrons. This new trigger selection uses existing lines that select $D^0 \rightarrow K^- \pi^+$ candidates and adds selected di-electron candidates. Promising D^{*0} candidates are further selected and categorised as *detached* or *prompt*. This distinction is made because in $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+ e^-)$ decays the di-electron system has a vertex that is detached from the primary vertex, while in $D^{*0} \rightarrow D^0 e^+ e^-$ the di-electron systems appears promptly. The selection performed by HLT2 is summarised in Table 6.1 and the names of used trigger lines are given in Table 6.2. Aside from common requirements on good vertex and track quality (χ_{vtx}^2 , χ_{trk}^2 , TrkGhostProb), identification of particles (PID) and consistency of a track with originating from the primary vertex (χ_{IP}^2 , χ_{VD}^2)³, the trigger line has very low thresholds of the electron p_T . These low thresholds have been chosen as the mass difference, and hence the phasespace accessible to decay products, between D^{*0} and D^0 is small ($\Delta m_{D^{*0} D^0} = 142.12 \text{ MeV}/c^2$).

The trigger selection is conducted online during data taking and events that pass it are promptly made available via the Turbo Stream (see Sec. 4.3) for the analysis steps presented in following sections. Since no significant differences are observed between data taken with

- 1 No calibrated value of the integrated luminosity was available until the end of this project. The given value is taken from Ref. [43] for the time between May and December 2017.
- 2 The first two trigger stages are predominantly fired by the $D^0 \rightarrow K^- \pi^+$ decay as the electrons often have not enough momentum to pass L0 and HLT1 electron requirements. Therefore, specific trigger lines are not necessary but could be used in the future.
- 3 χ_{VD}^2 : distance between a particle's end-vertex and the primary vertex in units of χ^2 .

Table 6.1: HLT2 trigger selection criteria to collect $D^{*0} \rightarrow D^0 e^+ e^-$ and $D^{*0} \rightarrow D^0 \gamma$ candidates. Cuts that are only applied to the prompt or detached category are marked accordingly. No entry in the Prompt/Detached column means that the cut is applied to both categories.

Candidate	Selection	Prompt/Detached
D^{*0}	$1850 \text{ MeV}/c^2 < M < 2100 \text{ MeV}/c^2$ $\chi_{\text{vtx}}^2/\text{ndf} < 10$	Prompt
D^0	$1715 \text{ MeV}/c^2 < M < 2015 \text{ MeV}/c^2$ $\chi_{\text{vtx}}^2/\text{ndf} < 10$ $p_T > 1000 \text{ MeV}/c$ DOCA ^a < 0.1 mm DIRA ^b > cos(17.3 mrad) $\chi_{VD}^2 > 25$	
K^-/π^+	$p > 5000 \text{ MeV}/c$ $p_T > 800 \text{ MeV}/c$ $\chi_{\text{IP}}^2 > 4$ TrkGhostProb < 0.4 $\chi_{\text{trk}}^2/\text{ndf} < 3$	
K^-	PIDK > 5	
π^+	PIDK < 5	
γ/e^+e^-	$M < 250 \text{ MeV}/c^2$ $\chi_{\text{vtx}}^2/\text{ndf} < 15$ $p_T > 200 \text{ MeV}/c$ DOCA < 1 mm	Prompt
e^\pm	$p_T > 50 \text{ MeV}/c$ $p_T > 80 \text{ MeV}/c$ PIDe > 0 TrkGhostProb ^c < 0.15	Prompt Detached

^a Distance Of Closest Approach between two daughter tracks.

^b DIRection Angle: Angle between this particle's reconstructed momentum vector and the vector pointing from the primary vertex to the decay vertex of the particle.

^c Ghost probability of a particle's track. The probability is the output of an ANN trained to identify fake tracks that are built from hits not corresponding to a true particle's trajectory.

MagUp and *MagDown*, respectively, the corresponding samples are merged and further analysed as one. As already mentioned in Sec. 4.2.1, only long tracks are used in this analysis. To be more precise, the used HLT2 trigger lines only build di-electron and D^0 candidates from long tracks of their daughter particles.

Table 6.2: HLT2 trigger lines used in the analysis. The lines that define the categories *Prompt* and *Detached* are marked accordingly. No entry in the Prompt/Detached column means that the line contributes to both categories.

Name of trigger line	Prompt/Detached
HLT2CHARMHADDo2HHL00SE	
HLT2CHARMHADSHAREDDETACHEDDoToHHCHILDPI/K	
HLT2SHAREDTRACKFITTEDDETACHEDDIELECTRONNoIP	Detached
HLT2SHAREDTRACKFITTEDVERYSOFTDIELECTRON	Prompt
HLT2DSTDoGAMMAEXTRASEL	Detached
HLT2DSTDoEEEXTRASEL	Prompt

6.1.1 Simulation Samples

In addition to the genuine data recorded by the LHCb detector, simulated samples of the signal process $D^{*0} \rightarrow D^0 e^+ e^-$ and control process $D^{*0} \rightarrow D^0 (\gamma \rightarrow e^+ e^-)$ are used throughout the analysis, most importantly for the determination of the selection efficiencies. Furthermore, two simulated samples of the decay $D^{*0} \rightarrow D^0 \pi^0$ are used, where the π^0 decays via $\pi^0 \rightarrow \gamma\gamma$ and $\pi^0 \rightarrow \gamma e^+ e^-$, respectively. These decays exhibit the same final state as the signal and control processes if the additional photon is not reconstructed (denoted by χ in $D^{*0} \rightarrow D^0 [\chi(\gamma \rightarrow e^+ e^-)]_{\pi^0}$ and $D^{*0} \rightarrow D^0 [\chi e^+ e^-]_{\pi^0}$). The simulation of proton-proton collisions and especially the subsequent propagation of all simulated particles through the detector are computationally very involved. Therefore, simulated samples are usually centrally produced by the LHCb collaboration using the computing resources available at CERN and via the Worldwide LHC Computing Grid. However, as the demand for simulations is generally quite high, no official simulation for the data-taking conditions in 2017 of the aforementioned decays was available until the end of this project. Instead, small samples of these decays have been produced privately. The samples' statistics are summarised in Table 6.3. The

Table 6.3: Number of generated events and number of entries in the privately produced simulation samples. The numbers correspond to the sum of both magnet polarities, *MagUp* and *MagDown*. For $D^{*0} \rightarrow D^0 e^+ e^-$ and $D^{*0} \rightarrow D^0 [\gamma e^+ e^-]_{\pi^0}$ only the entries of the prompt trigger category are shown, for the other two processes only entries of the detached category are shown. The total number of generated events with and without photon conversion is not known. "True candidates" refers to decays for which the information about true identities of particles and their origins could be matched to reconstructed particles.

Decay	Generated (EVTGEN)	Ntuple (DAVINCI)	True candidates
$D^{*0} \rightarrow D^0 e^+ e^-$	17.983×10^6	129173	33458
$D^{*0} \rightarrow D^0 \gamma$		28061	8377
$D^{*0} \rightarrow D^0 [\gamma e^+ e^-]_{\pi^0}$	5.9165×10^6	35803	8926
$D^{*0} \rightarrow D^0 [\gamma\gamma]_{\pi^0}$		4658	837

private production of $D^{*0} \rightarrow D^0 (\gamma \rightarrow e^+ e^-)$ and $D^{*0} \rightarrow D^0 [\chi(\gamma \rightarrow e^+ e^-)]_{\pi^0}$ turned out to be particularly challenging as the final state can only be reached by converting a photon and thus only by simulating the full detector. This is problematic since the probability that

the photon converts is only of the order of a few percent (see Appendix B.1), *i.e.* in most cases the complex detector simulation is run without yielding the decay of interest. As a first improvement on this, a photon conversion filter has been implemented in GAUSS, such that only events in which a photon conversion happened are written to storage in order to save disk space [70]. The problem of the resource-consuming detector simulation itself is tackled by using cuts at generator level, *i.e.* on the output of EVTGEN (see Appendix A), by a fast simulation option⁴ called *ReDecay* [71] implemented in GAUSS and by a novel, still experimental method called *Split-Sim* [72]. The development of the later one has been triggered by the difficulties arising when simulating $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+e^-)$ in this analysis. ReDecay splits the underlying event produced by PYTHIA off the signal and reuses it with multiple newly generated and simulated signal decays. Split-Sim takes this idea one step further and splits off the photon in $D^{*0} \rightarrow D^0\gamma$ and runs the detector simulation with the photon only. If the photon converts, the rest of the event is simulated and recombined with the photon. Using these methods significantly speeds up the simulation process, albeit the number of candidates in the resulting samples is still low due to limitations of available computing resources when simulating privately. After the detector simulation and subsequent digitisation of the detector response (see Sec. 4.3.1), the simulated events have to pass the same trigger selection as the data. The effort made to configure, test and produce the private simulation was an important step in the development of this thesis.

6.2 BACKGROUND SOURCES AND PRE-SELECTION

After the trigger selection, the 2017 dataset is still dominated by background processes, such that the $K\pi ee$ invariant mass spectrum has no visible contribution from the signal and control processes. The largest portion of background comes from random combinations of the final state particles. This combinatorial background can be roughly grouped into (a) combination of two random electrons with a D^0 candidate, (b) combination of a random photon conversion with a D^0 candidate and (c) random combination of K^- and π^+ to a D^0 candidate. The latter background source can be reduced efficiently by requesting D^0 candidates to have an invariant mass in the window $m_{\text{DTF}} \in [1840, 1900] \text{ MeV}/c^2$. The decay tree fitter [73] (DTF) invariant mass m_{DTF} is obtained by refitting the reconstructed decay chain with the D^{*0} vertex constrained to the primary vertex.⁵ The additional constraint improves the resolution of invariant mass spectra. Further selection on the D^0 candidate includes requesting a significantly detached decay vertex and a tightened transverse momentum cut. The remaining combinatorial background consists predominantly of (a) and (b) and their reduction represents the main challenge of the selection process. All cuts applied in the pre-selection are summarised in Table 6.4. As a first step, D^{*0} candidates with small impact parameters are selected, the transverse momentum cut and the PID requirement of the electrons are tightened, and the invariant mass of the di-electron candidate is restricted to the physically allowed upper bound given by the mass difference between D^{*0} and D^0 . Cutting on the di-electron invariant mass is powerful in rejecting combinatorial background of kind (a) but almost useless

⁴ It is in general good practice to employ methods that allow for more efficient simulation as computing resources are limited. This is not restricted to the analyses presented here.

⁵ The DTF can be used to define additional variables such as the four-momenta of final state particles, which are re-calculated using the given constraint and can be used together with variables that are not constrained, *e.g.* the impact parameter of the D^{*0} can be exploited even if the DTF constrains to the primary vertex.

for reducing (b). Therefore, only a loose cut is applied but will be tightened later in final selection steps to reduce remaining background of kind (a). Other cuts require good quality of the decay tree fit (χ_{DTF}^2) and remove events with too many hits in the SPD. While traversing the detector, electrons irradiate bremsstrahlung and thus lose energy. If the resulting photons are detected and matched to the corresponding electron, it is in principle possible to correct for the energy loss. This correction is not fully reliable as sometimes photons are falsely matched to electrons, leading to a change of shape in the invariant mass spectrum of the D^{*0} . When using the bremsstrahlung recovery, the data thus is often analysed in bremsstrahlung categories depending on how many photons are added to an electron (usually none, one or two). For this analysis only electrons are considered for which no bremsstrahlung photon is added (HasBremAdded is false) because of low statistics in the other categories.

An important variable throughout this analysis is the reconstructed mass difference between the D^{*0} and D^0 , denoted Δm_{DTF} . This variable is advantageous as resolution effects that propagate from the D^0 mass into the reconstructed D^{*0} mass largely cancel in the difference and the remaining resolution is dominated by the resolution of the di-electron system. In the pre-selection, this mass difference is roughly constrained to the region of interest around the nominal mass difference.

In addition to the combinatorial background, the decays $D^{*0} \rightarrow D^0[\chi(\gamma \rightarrow e^+e^-)]_{\pi^0}$ and $D^{*0} \rightarrow D^0[\chi e^+e^-]_{\pi^0}$ occur as partially reconstructed background in Δm_{DTF} . Since one photon is not reconstructed in these decays, they peak at an invariant mass difference below the nominal one. The mass difference between D^{*0} and D^0 and the invariant mass of the D^0 candidate after pre-selection are shown in Fig. 6.1.

Table 6.4: Pre-selection cuts applied to candidates that pass the trigger selection. "Global" denotes cuts which are not associated with a specific candidate but with the whole event.

Candidate	Selection
D^{*0}	$1900 \text{ MeV}/c^2 < m_{\text{DTF}}(K^-\pi^+e^+e^-) < 2100 \text{ MeV}/c^2$ $IP < 0.1 \text{ mm}$ $\Delta m_{\text{DTF}}(K^-\pi^+e^+e^-, K^-\pi^+) < 250 \text{ MeV}/c^2$
D^0	$1840 \text{ MeV}/c^2 < m_{\text{DTF}}(K^-\pi^+) < 1900 \text{ MeV}/c^2$ $p_T > 3500 \text{ MeV}/c$ $BPVDLS^a > 4$
γ/e^+e^-	$m(e^+e^-) < 143 \text{ MeV}/c^2$
e^\pm	$p_T > 100 \text{ MeV}/c$ $y\text{PIDe} > 0.5$ HasBremAdded == False
Global	$\chi_{\text{DTF}}^2 < 50$ $n\text{SPDHits} < 450$

^a Best Primary Vertex Decay Length Significance

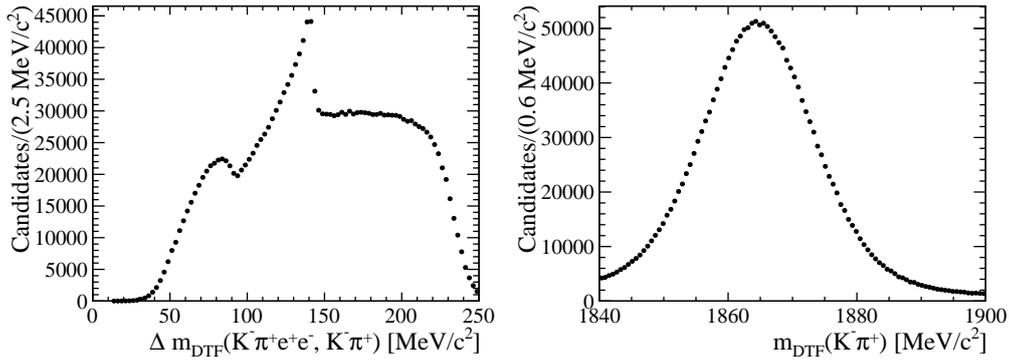


Figure 6.1: Left: Difference between the invariant mass of the D^{*0} and D^0 candidate after pre-selection. The peak at 140 MeV/c^2 consists of $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+e^-)$ decays with a clearly visible bremsstrahlung tail to the left. The partially reconstructed background is visible as a broad structure around 80 MeV/c^2 . Right: Invariant mass distribution of the D^0 candidate after applying the pre-selection.

6.3 PHOTON CONVERSION VETOES

A crucial task of this analysis is to distinguish the signal decay $D^{*0} \rightarrow D^0 e^+ e^-$ from the normalisation decay $D^{*0} \rightarrow D^0 \gamma$. The typical approach⁶ to remove intermediate resonances present in control channels by cutting out the corresponding mass regions is not efficiently applicable in the case of real versus internal photon conversions. As can be seen in Fig. 6.2, the di-electron invariant masses of $D^{*0} \rightarrow D^0 e^+ e^-$ and $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+ e^-)$ decays have large overlapping regions below 30 MeV/c^2 (~ 1.5 on the log-scale). Moreover, it is a known problem of GEANT4 that the di-electron invariant mass obtained from simulated photon conversions is approximately correct for invariant masses below $\sim 20 \text{MeV}/c^2$, while the abundance of higher invariant mass values is underestimated (see Appendix B.2). Several approaches have been tested in the course of this thesis to separate real from internal photon conversions, *e.g.* using the angle between two electrons, using cases where both electrons have the same VELO track or exploiting the uncertainty of the conversion vertex along the z -axis. However, the best separation performance is reached when using the fact that photons have to hit material before they convert. Two implementations of a veto against material interaction of a photon are tested. The first approach compares the position of the electrons' origin vertex with locations of material within the VELO. For this purpose, a detailed map of the VELO's material budget is available together with predefined distance metrics, giving distances to the RF-foil, the silicon modules and an uncertainty-weighted average distance to VELO material [46]. The second approach tested in this thesis evaluates the distance between primary and origin vertex of the electrons. This distance is projected onto the x - y -plane and referred to as *transverse flight distance* (FD_T or transverse FD) of the photon (see Fig. 6.3). The uncertainty of this quantity is denoted σ_{FD_T} or transverse σ_{FD} . The idea behind this approach is to exploit the material-free region between the proton-proton interaction point and the RF-foil shielding the VELO sensors. As the resolution of primary vertices is usually very high, the transverse flight distance can be used to efficiently distinguish between prompt electrons from a region close to the primary vertex, as in the case of $D^{*0} \rightarrow D^0 e^+ e^-$

⁶ A prominent example at LHCb is the analysis of the decay $B^0 \rightarrow K^{*0} \ell^+ \ell^-$ where an intermediate decay $J/\psi \rightarrow \ell^+ \ell^-$ is suppressed by cutting out the J/ψ mass region.

decays, and detached electrons as in the case of $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+e^-)$ decays. It is crucial in both approaches to take the uncertainty of the conversion vertex into account. The electrons created in photon conversions as well as a large fraction of electrons from $D^{*0} \rightarrow D^0e^+e^-$ decays have highly collimated trajectories. This makes it hard to precisely reconstruct the origin vertex of the electrons and leads to large vertex uncertainties.

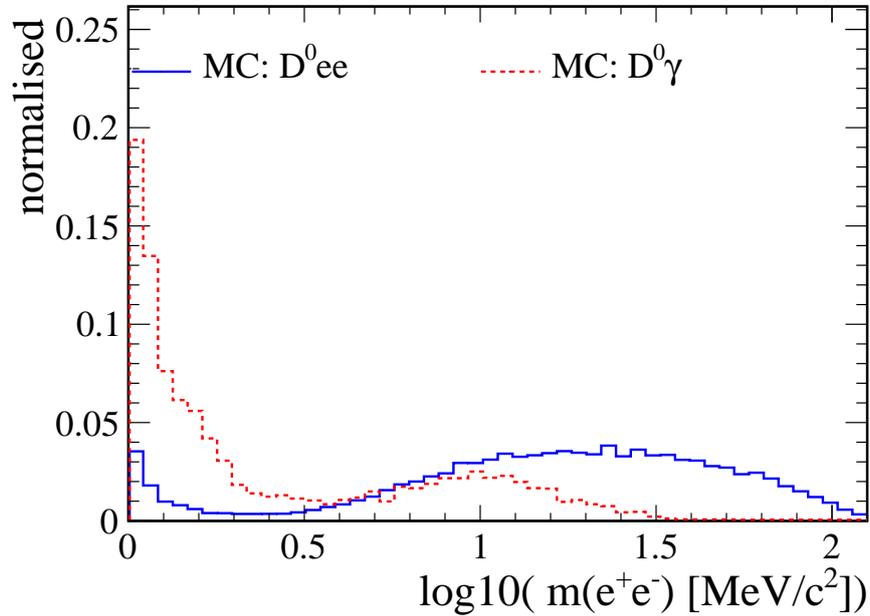


Figure 6.2: Simulated di-electron invariant mass spectrum for $D^{*0} \rightarrow D^0e^+e^-$ (blue solid line) and $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+e^-)$ (red dashed line) after pre-selection. The distributions' minimum around 0.5 is due to the transition between both electrons sharing a track within the VELO and being resolved as two tracks.

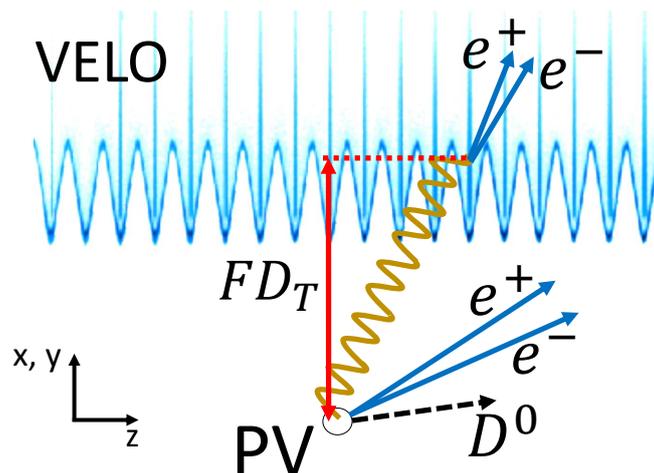


Figure 6.3: Sketch of the decays $D^{*0} \rightarrow D^0e^+e^-$ and $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+e^-)$. The image of the VELO consists of reconstructed secondary vertices from beam-gas collision during Run I [46]. The sketch should be seen as symmetric with respect to rotations around the z -axis.

The performance of both approaches is compared in Fig. 6.4 and Fig. 6.5 by plotting the purity of signal and control channel, respectively, against the efficiency of the veto. The purity of the $D^{*0} \rightarrow D^0 e^+ e^-$ decay is defined as

$$\text{Purity} = 1 - \frac{\mathcal{B}(D^{*0} \rightarrow D^0 \gamma) \times \varepsilon_{\text{veto}}^{\text{con,MC}} \times \xi_{\text{con}}}{\mathcal{B}(D^{*0} \rightarrow D^0 e^+ e^-) \times \varepsilon_{\text{veto}}^{\text{sig,MC}} \times \xi_{\text{sig}}} \quad (6.1)$$

where $\varepsilon_{\text{veto}}^{\text{sig,MC}}$ and $\varepsilon_{\text{veto}}^{\text{con,MC}}$ are the efficiencies of the applied veto cut obtained from simulation for signal and control channel, respectively, and ξ_{sig} and ξ_{con} are the corresponding generator-level-cut efficiencies of the simulated samples (see Appendix A). The ratio of branching fractions is taken from Eq. 2.9. The veto is then chosen as the cut for which the purity is above 99%. To select the control channel, *i.e.* vetoing the signal channel, it is sufficient to cut only on the transverse FD because the signal channel is much more rare and hence the relative pollution of the control channel is small per default. The chosen cut to select the control channel is $FD_T(\gamma) > 7$ mm which has an efficiency of $(96 \pm 2)\%$ for the decay $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+ e^-)$ and a purity of $(99.07 \pm 0.03)\%$, with the numbers derived on simulation. The selection of

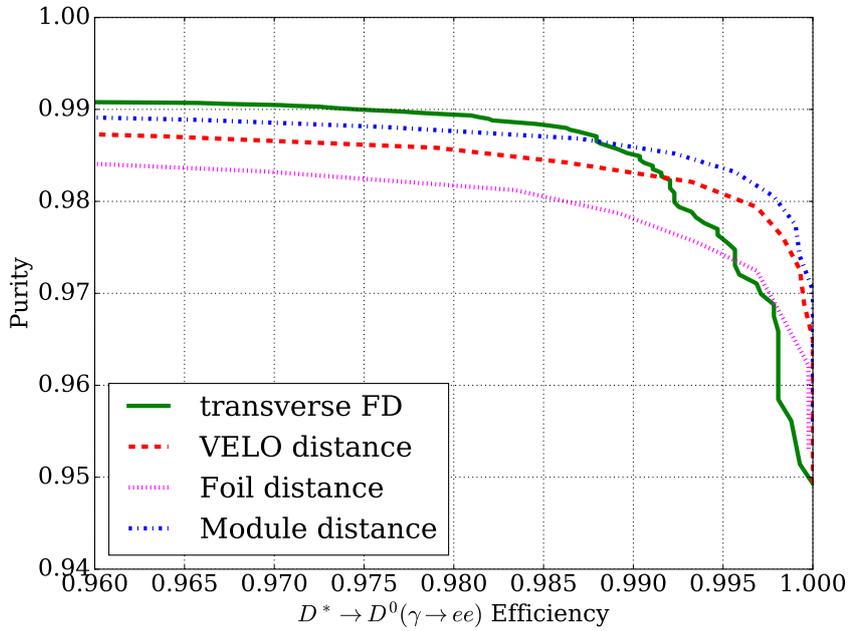


Figure 6.4: Purity as a function of $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+ e^-)$ efficiency for a scan through cuts on different veto variables using simulation. The best performance for high purity is given by cuts on the transverse flight distance (solid green line), followed by the distance to VELO modules (dashed-dotted blue line), the uncertainty-weighted average VELO distance (dashed red line) and the distance to the RF-foil (tiny-dashed magenta line).

the signal channel, *i.e.* vetoing the decay $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+ e^-)$, on the contrary suffers from the large branching fraction of the control channel. To avoid the pollution of the signal channel by control channel events with large conversion vertex uncertainty, a combined cut on the transverse flight distance and its error is applied. The chosen cut is $FD_T < 4.7$ mm and $\sigma_{FD_T} < 10$ mm and has an efficiency of $(61.5 \pm 0.6)\%$ for the decay $D^{*0} \rightarrow D^0 e^+ e^-$ and a purity of $(99.2 \pm 0.8)\%$, with the numbers obtained from simulation. This cut is indicated as a box with black edges in Fig. 6.6.

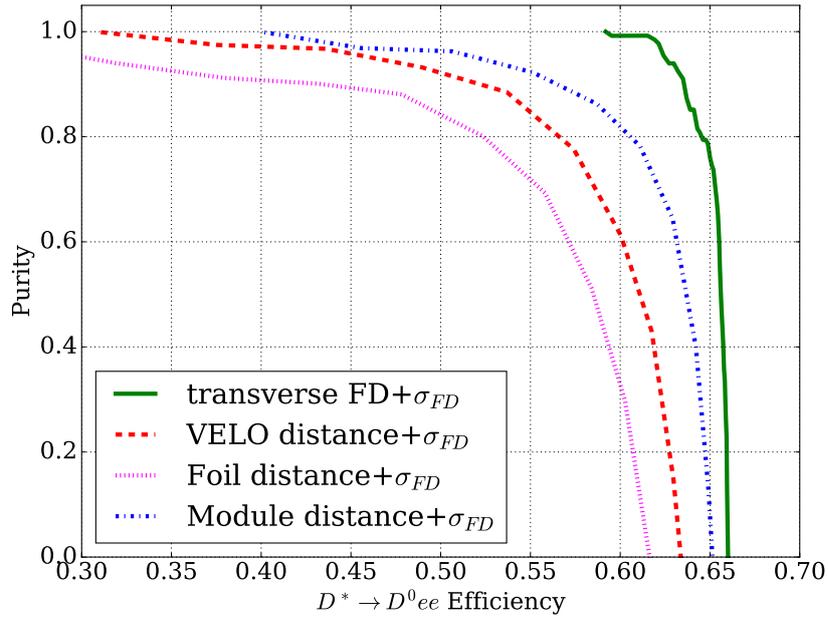


Figure 6.5: Purity as a function of $D^{*0} \rightarrow D^0 e^+ e^-$ efficiency for a scan through cuts on different veto variables and the uncertainty of the transverse flight distance using simulation. The best performance for high purity is given by cuts on the transverse flight distance (solid green line), followed by the distance to VELO modules (dashed-dotted blue line), the uncertainty-weighted average VELO distance (dashed red line) and the distance to the RF-foil (tiny-dashed magenta line). Note that zero purity means that there are as many control channel events in the sample as signal events, according to equation Eq. 6.1.

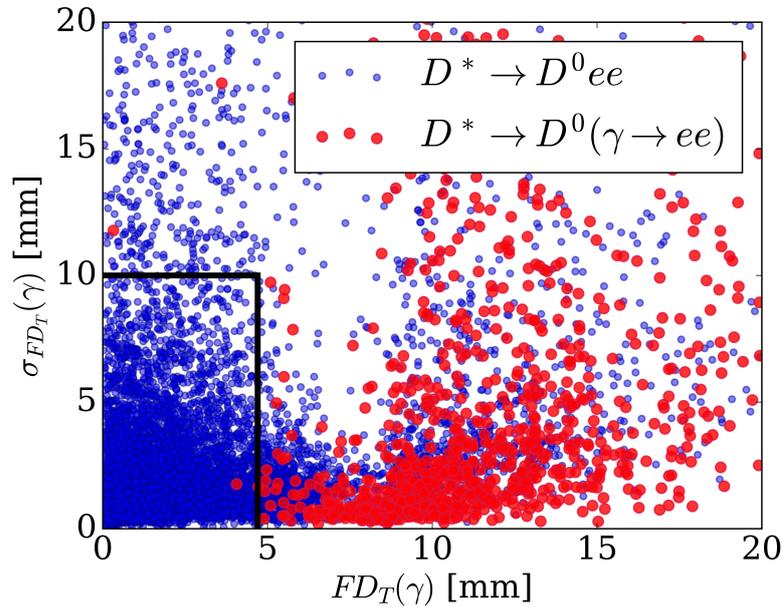


Figure 6.6: Distribution of $\sigma_{FD_T}(\gamma)$ versus $FD_T(\gamma)$ of the photon/di-electron for $D^{*0} \rightarrow D^0 e^+ e^-$ (small blue dots) and $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+ e^-)$ (large red dots) simulation, zoomed-in on the region of small flight distances and errors. The control channel veto boundaries are indicated as black lines, enclosing the selected signal events. The number of dots corresponds to the actual number of entries in the simulation samples in this region and does *not* represent the expected ratio of signal and control channel events.

6.4 DATA VS. SIMULATION

The simulated decays are used throughout the analysis to find differences between signal and background processes, and to estimate the selection efficiencies. It is therefore crucial to ensure that simulated distributions agree with the data distributions or to estimate by how much they differ. In order to do so, the real signal, control and background distributions have to be unfolded in data. The statistical technique that is used to perform the unfolding is the *sPlot* technique (see Sec. 5.3). Since the mass difference shown in Fig. 6.1 has a visible contribution of the control channel, it can be used as discriminating variable to unfold a pure control channel sample (see Sec. 6.4.1). However, obtaining a pure signal sample is not feasible in the same way, first of all because the decay $D^{*0} \rightarrow D^0 e^+ e^-$ cannot be expected to be cleanly visible in the mass difference at this stage of the selection, and secondly because the signal is blinded. This is a common problem, often handled by simply using the unfolded control channel distributions to verify the signal distributions. However, this is only valid if signal and control channel have very similar control variable distributions. In the case of the $D^{*0} \rightarrow D^0 e^+ e^-$ and $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+ e^-)$ decays, it turns out that this is not given for all variables of interest. To still verify the simulated signal distributions, the difference between the decays $D^{*0} \rightarrow D^0 e^+ e^-$ and $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+ e^-)$ is studied in Sec. 6.4.3, with the goal to correct for it and subsequently be able to still use the control channel to check the signal channel distributions. Discrepancies between simulation and data are corrected by applying event weights as described in Sects. 6.4.2 and 6.4.4

6.4.1 Unfolding a Pure Control Sample

To unfold a pure control channel sample, the distribution of Δm_{DTF} is fitted after applying the signal veto derived in Sec. 6.3. The fit is shown in Fig. 6.8 with its parameters given in Table 6.5. The shape of the control channel component is modelled by the sum of a Crystal Ball⁷ (*CB*) function [74] and a Gaussian function (*G*), the partially reconstructed background component from $D^{*0} \rightarrow D^0[\chi(\gamma \rightarrow e^+ e^-)]_{\pi^0}$ is modelled by the sum of two bifurcated⁸ Gaussian functions (*BifurG*) and the combinatorial background is described by a model built from four sigmoid functions (*S*). The CB function consists of a Gaussian core with a power-law tail. The tail accounts for energy losses such as bremsstrahlung. The CB function is continuous, it has four parameters and it is defined as

$$CB(x; \alpha, n, \mu, \sigma) = N \times \begin{cases} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), & \text{for } \frac{x-\mu}{\sigma} > -\alpha \\ A\left(B - \frac{x-\mu}{\sigma}\right)^{-n}, & \text{for } \frac{x-\mu}{\sigma} \leq -\alpha \end{cases} \quad (6.2)$$

⁷ Named after the Crystal Ball Collaboration – more precisely after the Crystal Ball detector – at the Stanford Linear Accelerator Center.

⁸ A Gaussian with two different widths, σ_L to the left and σ_R to the right, giving an asymmetric shape.

where

$$\begin{aligned} A &= \left(\frac{n}{|\alpha|} \right)^n \exp \left(-\frac{|\alpha|^2}{2} \right), \\ B &= \frac{n}{|\alpha|} - |\alpha|, \\ N &= \frac{1}{\sigma(C+D)}, \\ C &= \frac{n}{\alpha(n-1)} \exp \left(-\frac{|\alpha|^2}{2} \right), \\ D &= \sqrt{\frac{\pi}{2}} \left(1 + \operatorname{erf} \left(\frac{|\alpha|}{\sqrt{2}} \right) \right), \end{aligned}$$

N is a normalisation factor and $\operatorname{erf}(x)$ is the error function. The parameter α marks the transition between the Gaussian function and the power-law tail, n determines the shape of the power-law, μ is the peak position and σ is the width of the Gaussian. The model for $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+e^-)$ is then given by:

$$\begin{aligned} \mathcal{P}_{con}(\Delta m_{\text{DTF}}; f_{con}, \alpha, n, \mu_{CB}, \sigma_{CB}, \mu_G, \sigma_G) &= f_{con} \times CB(\Delta m_{\text{DTF}}; \alpha, n, \mu_{CB}, \sigma_{CB}) \\ &+ (1 - f_{con}) \times G(\Delta m_{\text{DTF}}; \mu_G, \sigma_G) \end{aligned} \quad (6.3)$$

where f_{con} is the fraction of the normalisation belonging to the CB function. To include detector resolution effects, the CB function was at first folded with a Gaussian function. However, a better fit performance was obtained by simply adding a Gaussian component to take small resolution effects into account, which mostly play a role at the right side of the CB function where bremsstrahlung does not distort the signal peak. The tail parameters of the CB function as well as width and fraction of the Gaussian function are fixed by fits to the corresponding simulated distribution (see Fig. 6.7).

The sigmoid function \mathcal{S} can be defined as

$$\mathcal{S}(x; \mu, k) = \frac{1}{1 + \exp\left(-\frac{x-\mu}{k}\right)} \quad (6.4)$$

where μ determines the inflection point of the curve and k controls the slope. The combinatorial background component is modelled by:

$$\begin{aligned} \mathcal{P}_{combBG}(\Delta m_{\text{DTF}}; f_{comb}, \vec{\mu}, \vec{k}) &= f_{comb} \times \mathcal{S}(\Delta m_{\text{DTF}}; \mu_1, k_1) \times \mathcal{S}(\Delta m_{\text{DTF}}; \mu_2, k_2) \\ &+ (1 - f_{comb}) \times \mathcal{S}(\Delta m_{\text{DTF}}; \mu_2, k_2) \\ &\times \mathcal{S}(\Delta m_{\text{DTF}}; \mu_3, k_4) \end{aligned} \quad (6.5)$$

with $\vec{\mu} \equiv \{\mu_1, \mu_2, \mu_3, \mu_4\}$, $\vec{k} \equiv \{k_1, k_2, k_3, k_4\}$ and f_{comb} the normalisation fraction belonging to the product of the first two sigmoid functions. This *ad hoc* background model is used because a large variety of other, more common background models failed to describe the assumed combinatorial background shape below the control channel peak and the partially reconstructed background.

The advantage of this model is its flexibility in describing different shapes. Its downside in the case of the distribution presented in Fig. 6.8, however, lies in the necessity to constrain or fix a large number of its parameters for the fit to converge. The determination of these parameters turned out to be difficult as neither the shape of the partially reconstructed

background nor the one of the combinatorial background is *a priori* known. It has been attempted to study the combinatorial background shape by looking at a so-called same-sign sample of 2016 data in which D^0 candidates are combined with electrons that have the same electric charge. This approach did not succeed, as the present combinatorial background does predominantly not consist of random combinations of electrons with a D^0 candidate but rather comes from random combinations of D^0 candidates with genuine photon conversions. To better reproduce the shape of these combinations, two methods have been tested. The first method has been applied in a $\chi_c \rightarrow J/\psi (\gamma \rightarrow e^+ e^-)$ analysis [75] to study the shape of combinatorial photon conversion background by producing "fake photons" using the data sample. Fake photons are generated by setting the photon energy equal to twice the energy of one electron. This leads to a spread of fake photon energies, while the angular distribution between the D^0 and the photon stays the same. The Δm_{DTF} distribution resulting from combining the fake photons with the D^0 candidates (see Fig. B.3), however, does not resemble the distribution observed for large mass differences and is thus not trusted to give the correct shape of the combinatorial background for the fit. It is though taken as a hint on how the distribution could roughly look like for small Δm_{DTF} . In the second method, a D^0 candidate in one event is combined with a di-electron candidate from a different event. This so-called *event mixing* has the advantage that it leaves the kinematic variables of the photon unchanged instead of altering them as in the aforementioned method. However, the distribution of the angle between D^0 and γ , if not flat in the first place, changes and has a great impact on the Δm_{DTF} distribution because of the strong correlation between the angle and the invariant mass of the D^{*0} candidate. Indeed, the angular distribution obtained from the sideband $\Delta m_{\text{DTF}} \in [150, 250] \text{ MeV}/c^2$ is not flat but increases towards small angles. Weighting the angular distribution in the event-mixed sample such that it agrees with the angular distribution of the sideband gives a Δm_{DTF} distribution similar to the one obtained using the first method (see Fig. B.4). The discrepancy between shapes obtained from the two methods and the observed shape for large Δm_{DTF} could be due to the selection made by the trigger. The generation of fake photons as well as the event-mixing are performed on events that have already passed the trigger and it is conceivable that the trigger selection has an impact on the exact shape of the combinatorial background. This is not straightforward to study as the Turbo stream only stores events that passed the trigger line. Yet in principle, the event mixing method should be able to reproduce the correct shape of the combinatorial background and it has to be studied in more detail why it does not in this case. For the scope of this thesis, the combinatorial background model is chosen such that it is flexible enough to describe not only the shape found via event mixing but also describes the observed sideband shape. Several parameters of the background model are fixed to values that occurred in non-converging fits but were found to still describe the observed shape well (see Table 6.5). This is clearly not a satisfactory solution and represents a major caveat of this thesis.

The model of the partially reconstructed background is given by:

$$\begin{aligned} \mathcal{P}_{\text{partBG}}(\Delta m_{\text{DTF}}; f_{\text{part}}, v_1, \sigma_{L1}, \sigma_{R1}, v_2, \sigma_{L2}, \sigma_{R2}) \\ = f_{\text{part}} \times \text{BifurG}(\Delta m_{\text{DTF}}; v_1, \sigma_{L1}, \sigma_{R1}) \\ + (1 - f_{\text{part}}) \times \text{BifurG}(\Delta m_{\text{DTF}}; v_2, \sigma_{L2}, \sigma_{R2}) \end{aligned} \quad (6.6)$$

where v_i is the mean, σ_{Li} and σ_{Ri} are the widths to the left and to the right, respectively, of the i^{th} bifurcated Gaussian function. The shape of the partially reconstructed background can in principle be determined from the simulation sample of $D^{*0} \rightarrow D^0 [\chi (\gamma \rightarrow e^+ e^-)]_{\pi^0}$, the statistics of which is yet so low that no useful information about model parameters can be

extracted. Instead, the parameters of Eq. 6.6 are fixed to reasonable values such that the fit converges.

The full model fitted in Fig. 6.8 is given by:

$$\mathcal{P}_{conFull} = N_{con} \times \mathcal{P}_{con} + N_{combBG} \times \mathcal{P}_{combBG} + N_{partBG} \times \mathcal{P}_{partBG} \quad (6.7)$$

where N_{con} , N_{combBG} and N_{partBG} are the yields of the extended maximum likelihood fit for the corresponding fit component. The yields are used to calculate so-called *sWeights* for each component via the *sPlot* technique. Control variable distributions, in which events are multiplied with their *sWeights* of one component, resemble the pure distribution of this component.

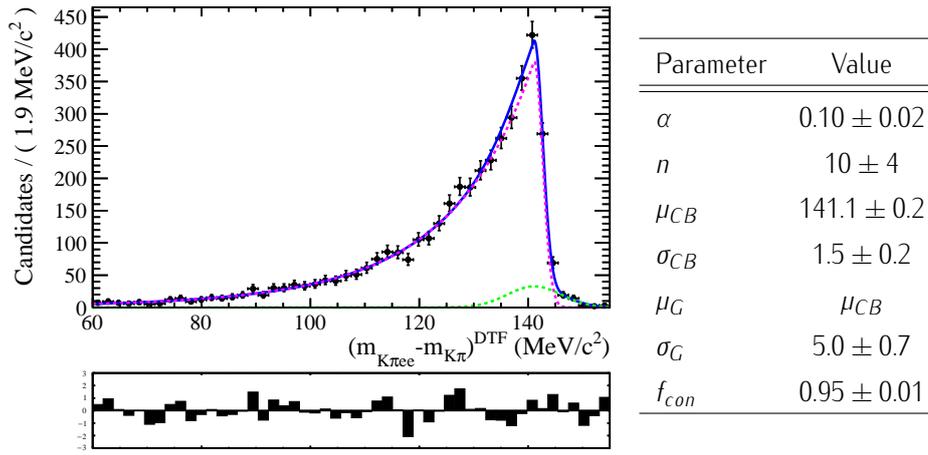


Figure 6.7: Fit of the Δm_{DTF} distribution obtained from simulation of $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+e^-)$. The Gaussian component is shown as a dashed green line, the CB component as dashed magenta line. The corresponding fit parameters can be found in the table on the right.

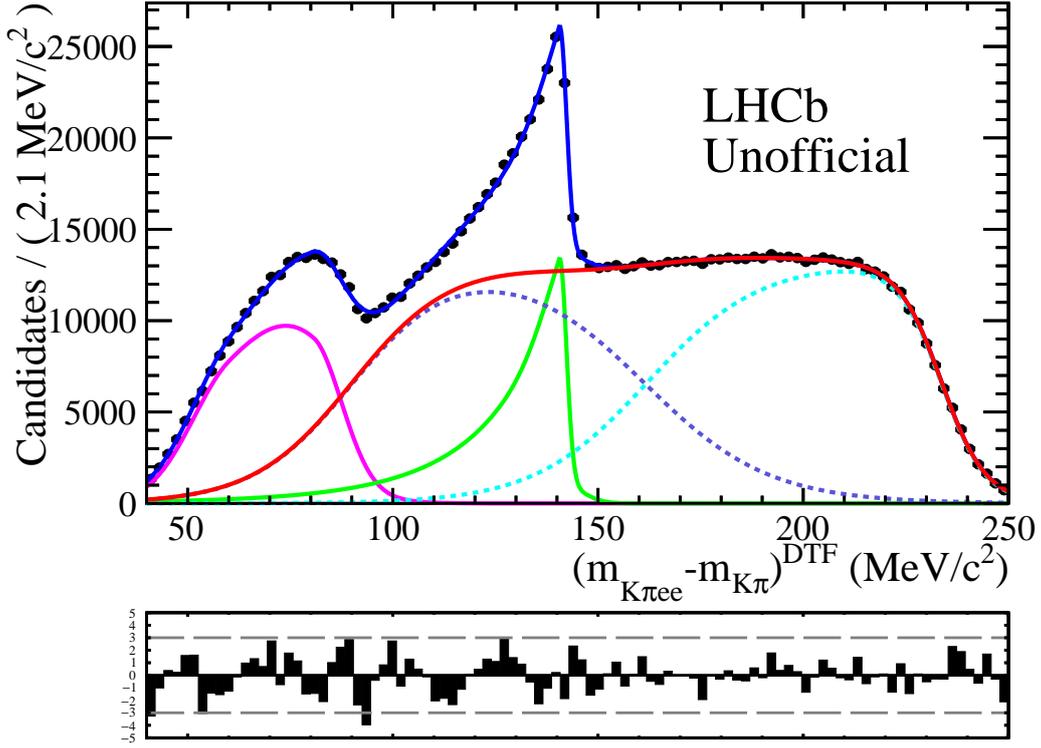


Figure 6.8: Fit of the Δm_{DTF} spectrum after pre-selection and signal channel veto cut. The combinatorial background model is shown as solid red curve, its two additive components are the dashed blue and cyan curves, respectively. The model of the partially reconstructed background is shown in magenta and the model for $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+e^-)$ is depicted in green. The model parameters can be found in Table 6.5.

Table 6.5: Parameters of the fit shown in Fig. 6.8. Parameters given without error are fixed in the fit.

\mathcal{P}_{con}		\mathcal{P}_{combBG}		\mathcal{P}_{partBG}	
Parameter	Value	Parameter	Value	Parameter	Value
α	0.1	μ_1	90.4 ± 0.3	ν_1	60
n	10	k_1	12.1 ± 0.3	σ_{L1}	10.2
μ_{CB}	140.66 ± 0.03	μ_2	160.9 ± 0.3	σ_{R1}	16
σ_{CB}	1.65 ± 0.02	k_2	-15.2 ± 0.2	ν_2	80.7
μ_G	μ_{CB}	μ_3	162	σ_{L2}	13.8
σ_G	5.0	k_3	13.7	σ_{R2}	7.2
f_{con}	0.95	μ_4	233.63	f_{part}	0.49
N_{con}	125112 ± 1451	k_4	-5.39	N_{partBG}	169025 ± 1667
		f_{combBG}	0.5		
		N_{combBG}	900716 ± 2076		

6.4.2 Differences between $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+e^-)$ Simulation and Data

The sWeights obtained in Sec. 6.4.1 can now be used to compare the simulated control channel distributions to the corresponding unfolded data distributions. The focus of this comparison lies on variables that are exploited in the subsequent analysis steps. Especially variables used in the multivariate analysis (Sec. 6.5) are required to be well described by the simulation. As can be seen in Fig. 6.9 (left column), the transverse momentum of the D^{*0} candidate, the electron transverse momentum, the momentum in y -direction and the pseudo-rapidity of the electron show some tension between data and simulation. The kinematic distributions of mother hadrons produced promptly in the proton-proton collision, *e.g.* the D^{*0} , are sometimes not well described in simulation as it is difficult to precisely model the fragmentation happening after the collision. The difference visible in Fig. 6.9 (a) is yet comparatively small. The electron distributions in Fig. 6.9 (c), (e) and (g) are more concerning. It has been checked that this difference is not due to the aforementioned improper modelling of the di-electron invariant mass in GEANT4 by re-weighting the simulated di-electron invariant mass distribution with the one obtained from the Bethe-Heitler formula. The Bethe-Heitler formula gives the exact description of the di-electron invariant mass resulting from photon conversions (for more details see Appendix B.2). However, the impact of this correction was found to be negligible. Another possible explanation is the discrepancy between the VELO model in GEANT4 and the actual geometry of the detector components. In particular, the shape of the RF-foil, which is the first material in the detector seen by the photons, is slightly different to the real one. Shape differences have been studied in Ref. [47] using hadronic interactions in the VELO material, however, the question of whether or not these differences can lead to such large deviations between data and simulation requires a more dedicated study of photon conversions. For the present analysis, the simulated distributions are re-weighted using gradient-boosting (see Sec. 5.4) to compensate for the deviation from the real data distributions. The gradient-boosting re-weighter is fed with the transverse momenta of the D^{*0} and the electrons, and with the y -component of the electrons' momentum. Additionally, the three spatial distributions of the photon's conversion vertex are re-weighted to tackle potential deviations due to inaccuracies in the detector shape model. The result of the re-weighting is shown in the right column of Fig. 6.9. More distributions can be found in the Appendix C.1. The overall agreement after re-weighting is good. Persisting differences are mainly due to difficulties arising when re-weighting is performed in more than two variables with little statistics available in the training of the gradient-boosting re-weighter. The lack of statistics makes it difficult for the re-weighter to correctly adhere the variables' mutual correlations. The set of re-weighting variables is thus chosen such that it is as small as possible while leading to an acceptable agreement between data and simulation. For all further analysis steps, the re-weighted $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+e^-)$ simulation is used if not stated differently.

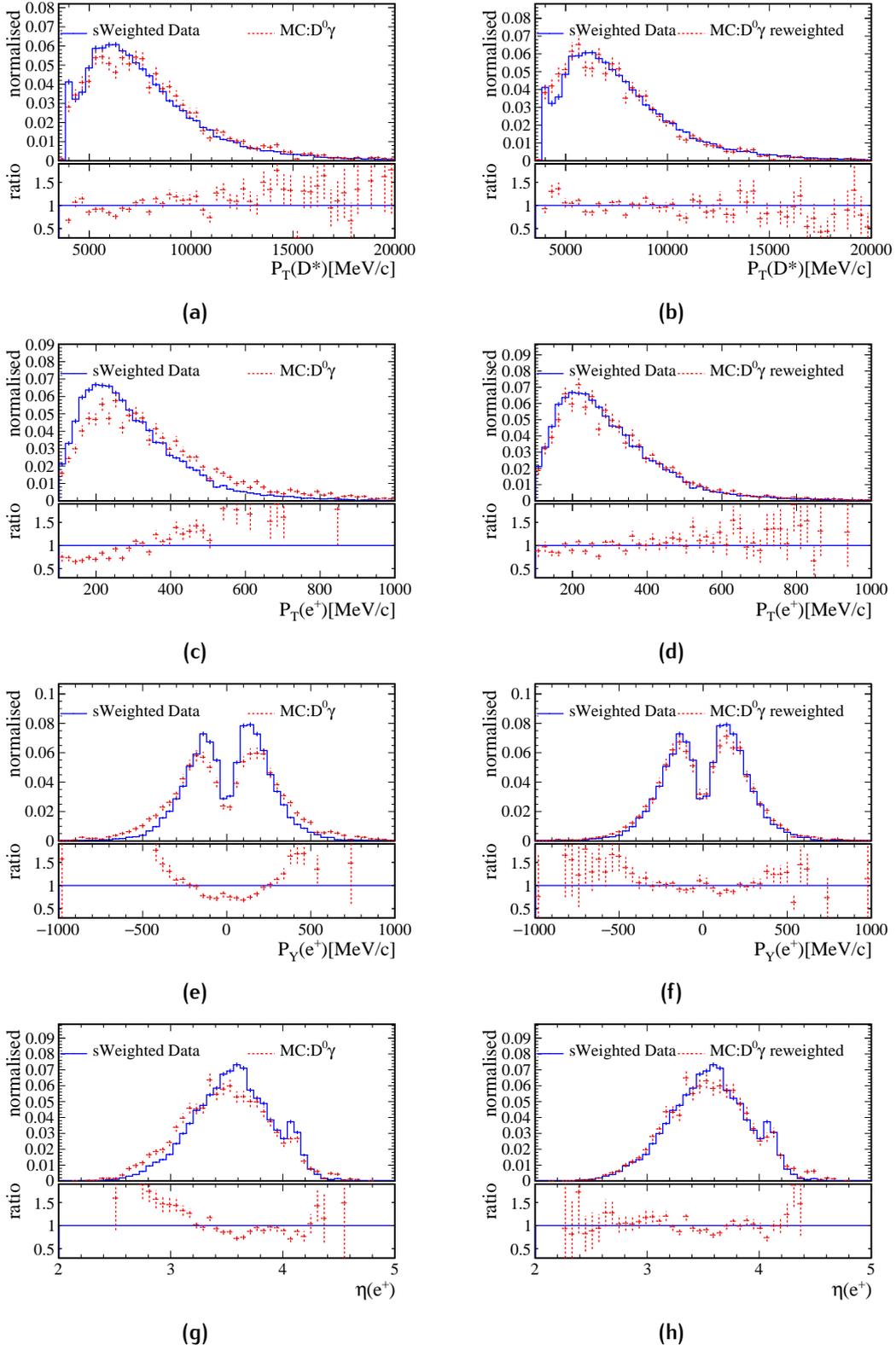


Figure 6.9: Comparison between simulated $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+e^-)$ distributions (dashed red) and sWeighted data (solid blue) shown in the left column. The right column shows the corresponding re-weighted distributions.

6.4.3 $D^{*0} \rightarrow D^0 e^+ e^-$ vs. $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+ e^-)$

To be able to investigate differences between the signal channel simulation and data from the control channel, the intrinsic difference between the decays $D^{*0} \rightarrow D^0 e^+ e^-$ and $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+ e^-)$ has to be understood. The underlying physics of both decays, *i.e.* the matrix elements of the plain D^{*0} decays, is very similar, differing only by an additional QED vertex. It is therefore reasonable to assume that differences are mainly due to the photon conversion. Qualitatively, a photon is more likely to produce a pair of electrons when it is travelling through parts of the VELO where a large amount of material is present. On the contrary, large amounts of material block the free path of electrons coming directly from the primary interaction region, leading to multiple scattering, bremsstrahlung and thus to a lower reconstruction efficiency. Indeed, this effect can be observed by comparing the topology of both decays in the η - ϕ -plane, where η is the pseudo-rapidity and ϕ the azimuthal angle with its zero defined on the x -axis. Fig. 6.10a shows the η - ϕ -plane of the signal channel, which resembles the average radiation length seen by particles passing through the VELO (*cf.* Fig. 4.5 left). The main bulk of material that causes the drop in efficiency around $\phi = \pm 90^\circ$ ($\pm y$ -direction) belongs to the RF-foil, but also the small overlap of VELO modules in this area (*cf.* Fig. 4.4 bottom left) increases the amount of integrated material seen by the electrons. The narrow band visible at $\phi \approx 0^\circ$ (x -direction) corresponds to a drop in efficiency due to electron trajectories crossing the beam line. This narrow band is also

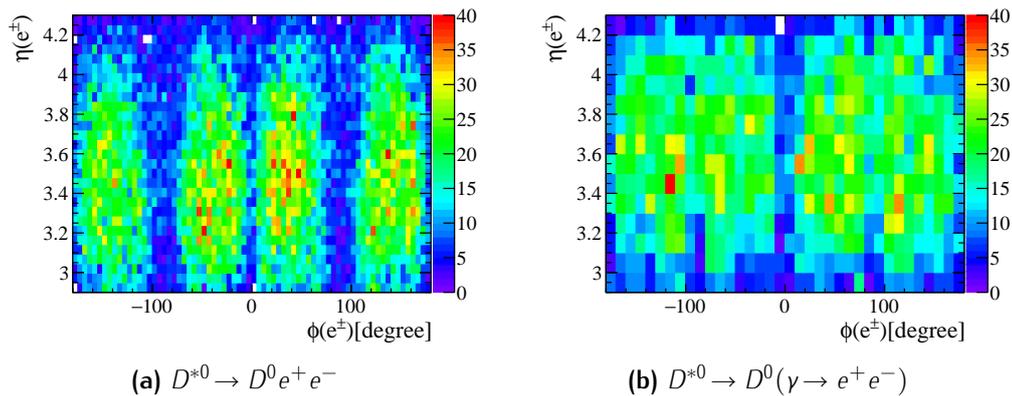


Figure 6.10: Simulated η - ϕ -plane of electrons from signal channel (a) and from control channel (b).

visible in the η - ϕ -plane of electrons produced in photon conversions, shown in Fig. 6.10b. However, regions within the VELO where there is a large amount of material, *i.e.* around $\phi = \pm 90^\circ$, are beneficial for photon conversions and thus more control channel electrons occur in these regions. The same effect can be seen at $\eta \approx 4.2$ in Fig. 6.10b and Fig. 6.9h as for this value in pseudo-rapidity (equivalently the polar angle $\theta \approx 30$ mrad) the photon traverses the VELO almost in its full length seen from the interaction point (*cf.* Fig. 4.4). These acceptance differences result in differences between the kinematic distributions of the decays $D^{*0} \rightarrow D^0 e^+ e^-$ and $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+ e^-)$. Particularly the x - and y -components of the momenta are affected as shown in Fig. 6.11 in the left column (more figures can be found in the Appendix C.2). Re-weighting the simulated signal η and ϕ distributions of the electrons such that they agree with the corresponding uncorrected control channel distributions resolves the differences in the kinematic variables as shown in the right column of Fig. 6.11. It can therefore be assumed that distributions of the signal channel are comparable to distributions of the control channel as long as they are compensated for material effects.

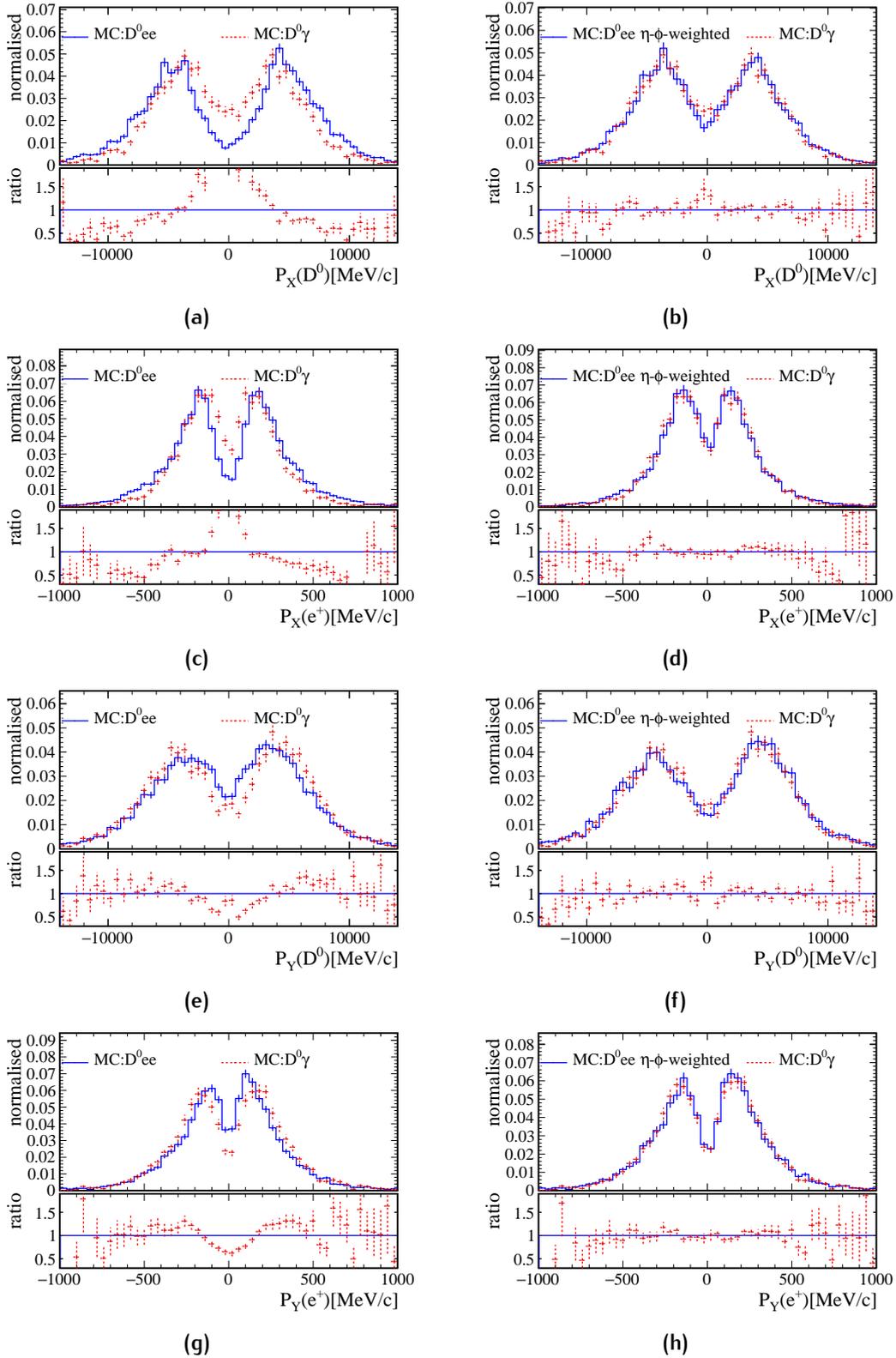


Figure 6.11: Comparison between simulated control channel distributions (dashed red) and simulated signal (solid blue) shown in the left column. The right column shows the distributions where signal is re-weighted in η and ϕ of the electrons.

6.4.4 Differences between $D^{*0} \rightarrow D^0 e^+ e^-$ Simulation and Data

The strategy to verify the simulated distributions of the decay $D^{*0} \rightarrow D^0 e^+ e^-$ is now as follows:

1. Re-weight the signal channel simulation in η and ϕ of the electrons to match the simulated control channel distributions. This time, however, the previously re-weighted $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+ e^-)$ distributions are used as re-weighting target.
2. Investigate differences between η - ϕ -weighted signal simulation and re-weighted control channel simulation.
3. Correct for newly formed differences by re-weighting the signal simulation in the respective variables.
4. Remove η - ϕ -weights from the simulated signal distributions and take weights found in step 3 as correction factor for differences between $D^{*0} \rightarrow D^0 e^+ e^-$ simulation and data.

The underlying assumption for this strategy to work is that the differences between data and simulation are similar in both signal and control channel apart from the obvious differences due to the photon conversion. Another approach, which would get around this assumption, is to find a different known particle decay exhibiting enough similarities to the $D^{*0} \rightarrow D^0 e^+ e^-$ decay such that it can be used to verify the simulated sample. However, to find such a decay is not straightforward and thus for the scope of this thesis the assumption is taken as acceptable.

The comparison between the η - ϕ -weighted signal simulation and re-weighted control channel simulation is shown in right column of Fig. 6.12 (more figures in Appendix C.3). The variables $p_T(D^{*0})$ and $p_T(e^\pm)$ are re-weighted to match the corrected control channel distributions as shown in the left column of Fig. 6.12. The result of the re-weighting procedure for the simulated sample of $D^{*0} \rightarrow D^0 e^+ e^-$ is shown in Fig. 6.13 (more figures in C.3). In the following selection steps, the re-weighted $D^{*0} \rightarrow D^0 e^+ e^-$ simulation is used.

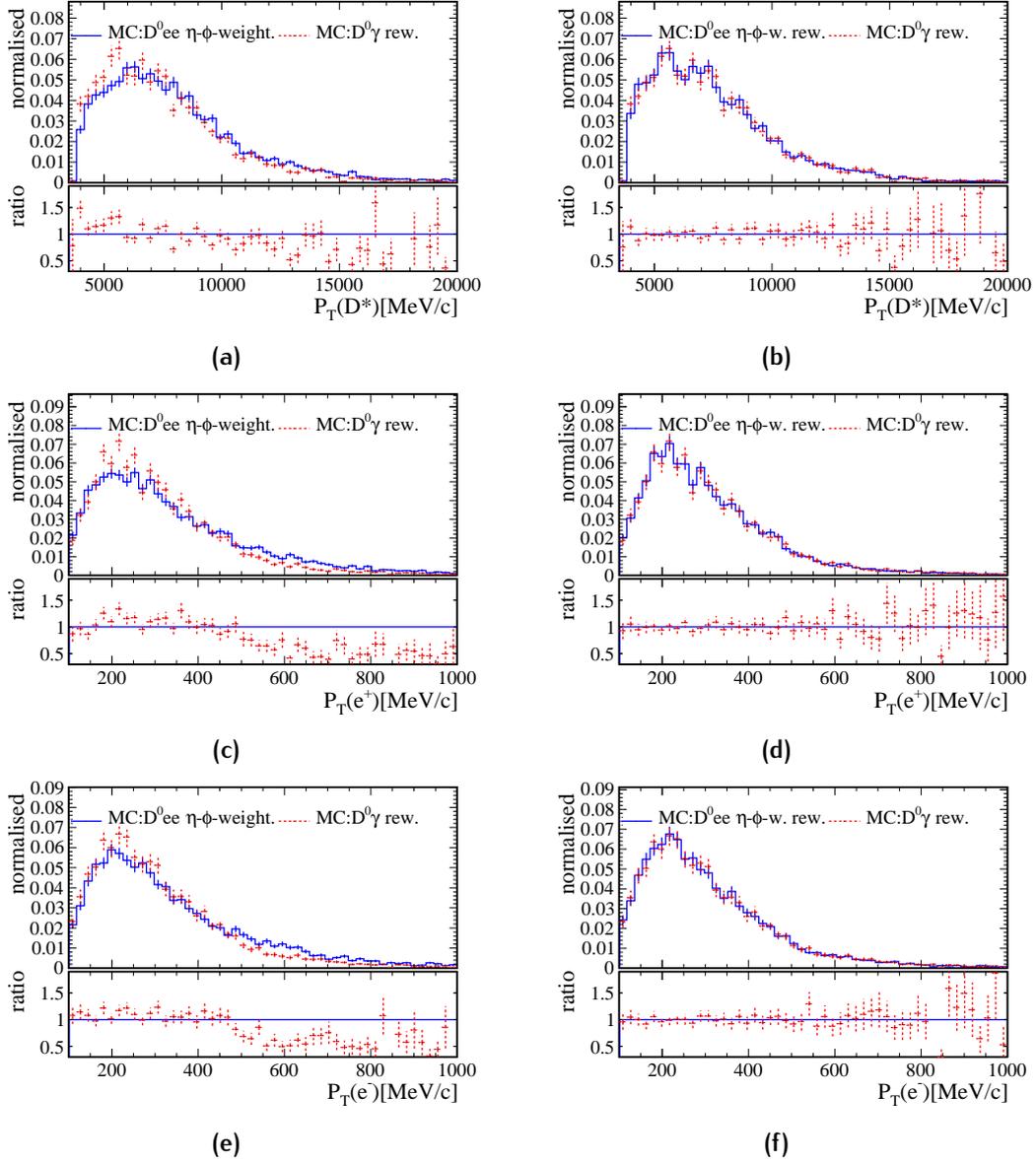


Figure 6.12: Comparison between simulation of signal re-weighted in η and ϕ of the electrons (blue solid line) and corrected control channel (dashed red) (left column). Result of the additional signal simulation re-weighting in $p_T(D^{*0})$ and $p_T(e^{\pm})$ to match corrected control channel distributions (right column).

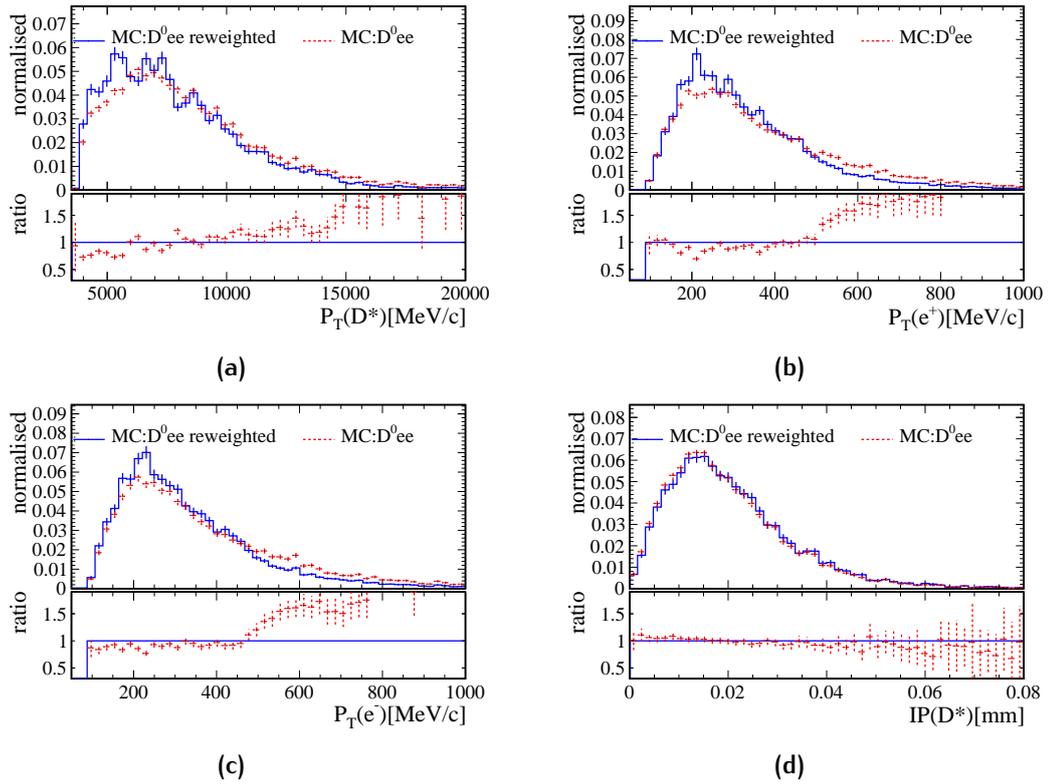


Figure 6.13: Result of the re-weighting in $p_T(D^{*0})$ and $p_T(e^\pm)$ without η - ϕ -weights. The re-weighted signal simulation is shown as solid blue line, the original uncorrected simulation is shown as dashed red points.

6.5 MULTIVARIATE SELECTION USING MACHINE LEARNING

After correction of the simulated signal sample, it can be used as training sample in a machine learning algorithm that further separates $D^{*0} \rightarrow D^0 e^+ e^-$ from background events. For the algorithm to learn how signal looks like in the presence of background events, it has to also be given information about how the background distributions look like. Therefore, the upper sideband of the mass difference between D^{*0} and D^0 , defined as $\Delta m_{\text{DTF}} \in [150, 250] \text{ MeV}/c^2$, is taken as pure background sample. Furthermore, variables are needed that allow the algorithm to find differences between signal and background and additionally are not or only very weakly correlated with Δm_{DTF} . The latter requirement is particularly important for the background sample as otherwise an artificial signal could be modelled out of the background by the multivariate selection exploiting correlations. A correlation of input variables with Δm_{DTF} in the signal sample on the other hand leads to a non-flat signal selection efficiency in this variable which in turn might distort the signal shape. The choice of machine learning algorithm itself is in principle arbitrary and thus several artificial neural networks as well as boosted decision trees have been tested. The set of input variables shown in Table 6.6 is found by subsequently including variables with good separation power and keeping those which significantly improve the separation power of the resulting classifier. Not many variables have been found that fulfil both criteria, independence of Δm_{DTF} and good separation power. Especially kinematic variables of the electrons exhibit correlations with Δm_{DTF} in the signal sample. The cause of this correlation is presumably the dependence of energy loss due to bremsstrahlung on the momentum of the electron, *i.e.* the migration of an event into the bremsstrahlung tail of the signal peak depends on the momentum of the electrons present in this event. For an efficient separation of signal and background, however, using kinematic variables of the electrons is inevitable. The machine learning algorithm that is opted for in this analysis is therefore gradient boosting on decision trees using the special loss function given in Eq. 5.18. This allows to use variables correlated to Δm_{DTF} while also retaining flatness of the selection efficiency in this variable. A comparison between signal and background

Table 6.6: Variables exploited by the uniform gradient boosted decision tree.

Candidate	Input Variable
D^{*0}	$\log(p_T)$
	$\log(IP)$
	$\log(\chi_{\text{IP}}^2)$
e^\pm	$\log(\min[p_T(e^+), p_T(e^-)])$
misc.	$\min[\arctan(p_X(e^\pm) \times p_X(D^0))]$
	$\min[\arctan(p_Y(e^\pm) \times p_Y(D^0))]$

distributions of the chosen input variables is shown in Fig. 6.14. Input variables are smoothed and scaled to small values using the logarithm and arc tangent function. This makes it easier for the gradient boosting classifier to approximate the optimal test statistic in regions that are otherwise sparsely populated by events. Linear correlations among the input variables and with Δm_{DTF} in the signal sample are shown in Fig. 6.15. Linear correlations with Δm_{DTF} in the background are found to be negligible (see Appendix Fig. D.1). Samples of the signal simulation and the upper sideband from data are randomly split into five training samples.

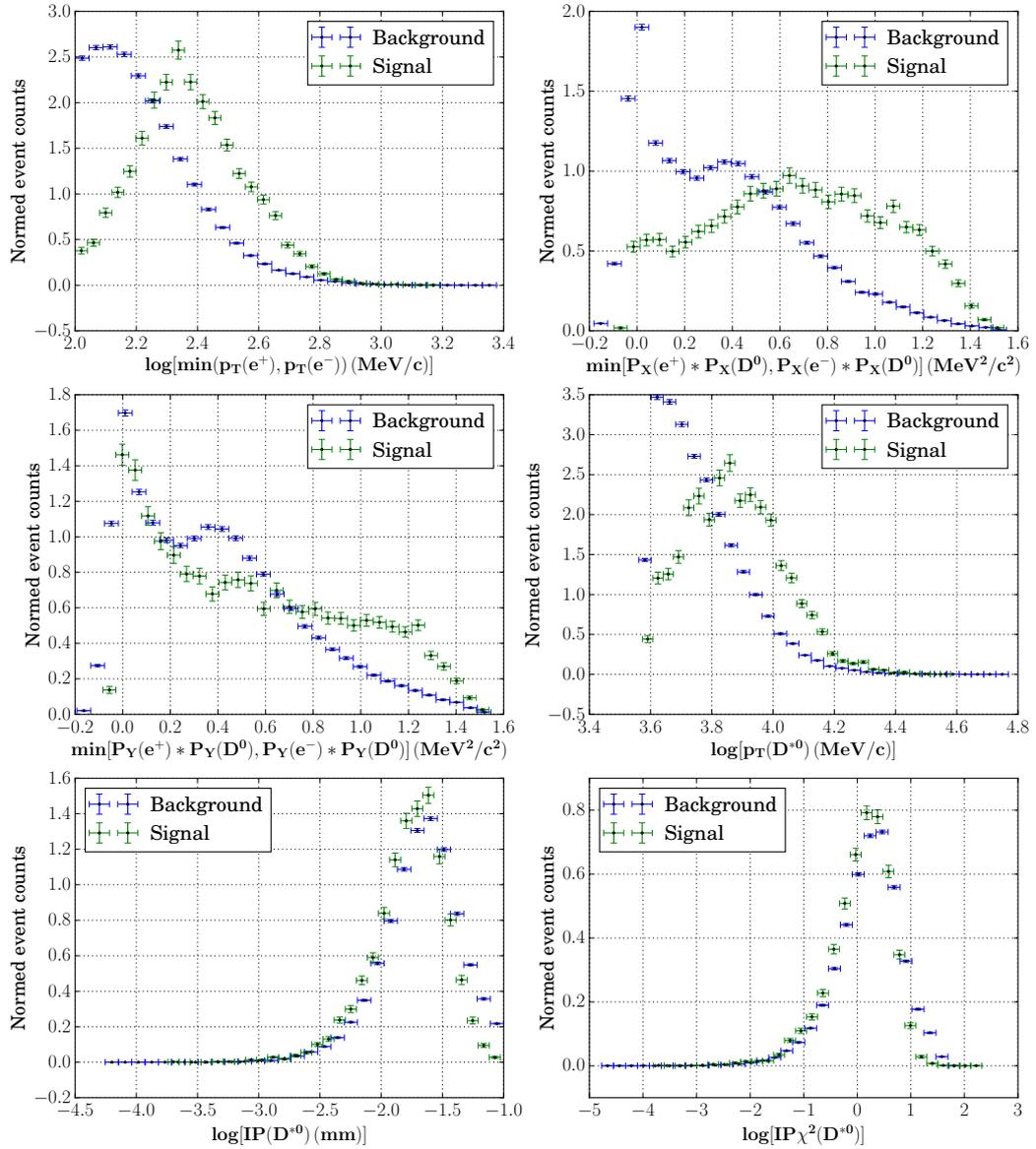


Figure 6.14: Input variable distributions from simulated signal (green) and Δm_{DTF} upper sideband (blue).

During the training process, a so-called *5-fold cross-validation* [76] is applied. Of the five sub-samples, one sub-sample is retained as test data to validate the classifier trained on the remaining four sub-samples. This process is repeated five times such that all available data has been used for training and testing likewise. This way, the limited statistics available in the simulated signal sample is most efficiently used. If data, which is not used in training process, is presented to the classifier, the median of the five single predictions is taken as global response of the classifier. The response of the classifier is a number between 0 and 1,

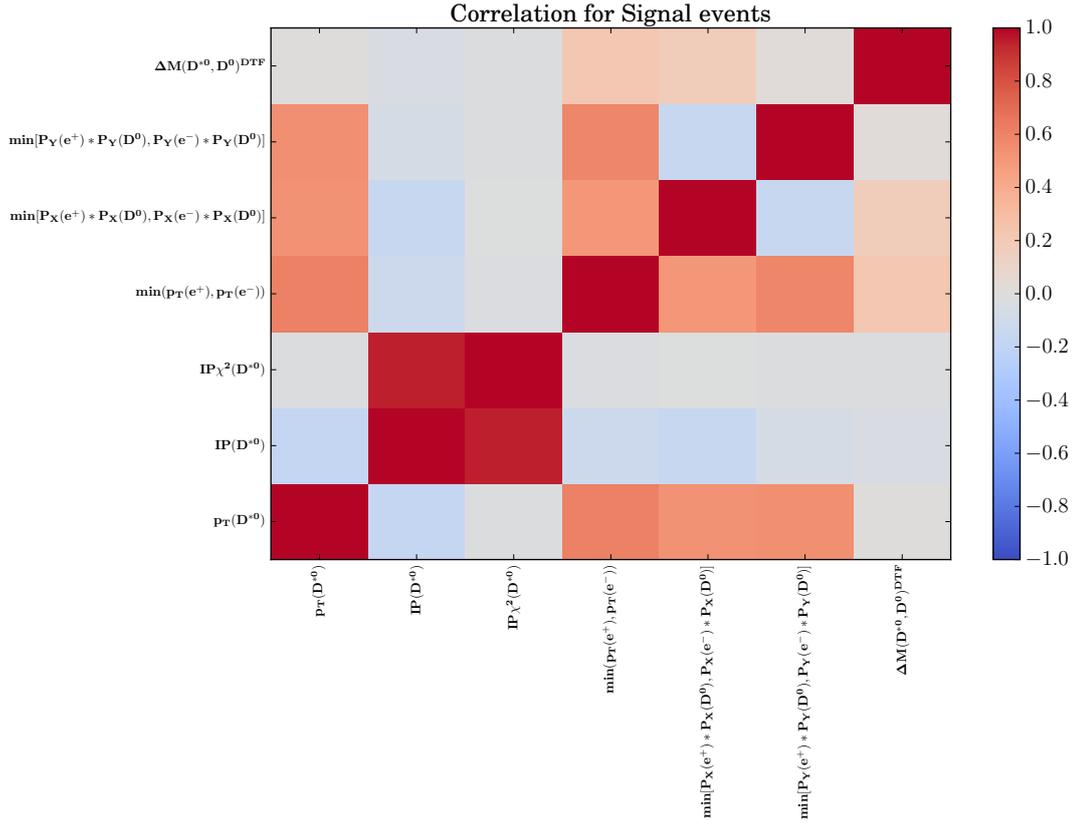


Figure 6.15: Linear correlation of input variables and Δm_{DTF} obtained from signal simulation.

where events with a prediction close to 0 are classified as background-like and a prediction closer to 1 identifies an event as more signal-like. The predictions for the training and test samples are shown in Fig. 6.16. No significant overtraining is observed as both distributions roughly agree within their statistical uncertainty. Performances of classifiers can be compared using the *receiver operating characteristic* (ROC) curve. In Fig. 6.17 the performance of both gradient boosting classifiers with loss functions introduced in Sec. 5.4 is shown. It is expected that the classifier using only the AdaLoss function performs slightly better than the one using both, AdaLoss and flatness loss function. The observed performance difference is small with an area under the ROC curve of 0.85 for AdaLoss and 0.84 for uniform gradient boosting. The resulting selection efficiencies as a function of Δm_{DTF} are shown in Fig. 6.18 for the signal and background test samples. The background sample shows no prominent dependence of the selection efficiency on the mass difference between D^{*0} and D^0 for various cuts on the classifier response for both loss functions. In the signal test sample, the expected correlation between selection efficiency and Δm_{DTF} is visible as an almost linear slope towards higher mass differences when using the AdaLoss function. This correlation is clearly mitigated by additionally using the flatness loss function. In a direct comparison between the selection efficiencies of both classifiers for equal cuts on their response, the better performance of AdaBoost is again visible.

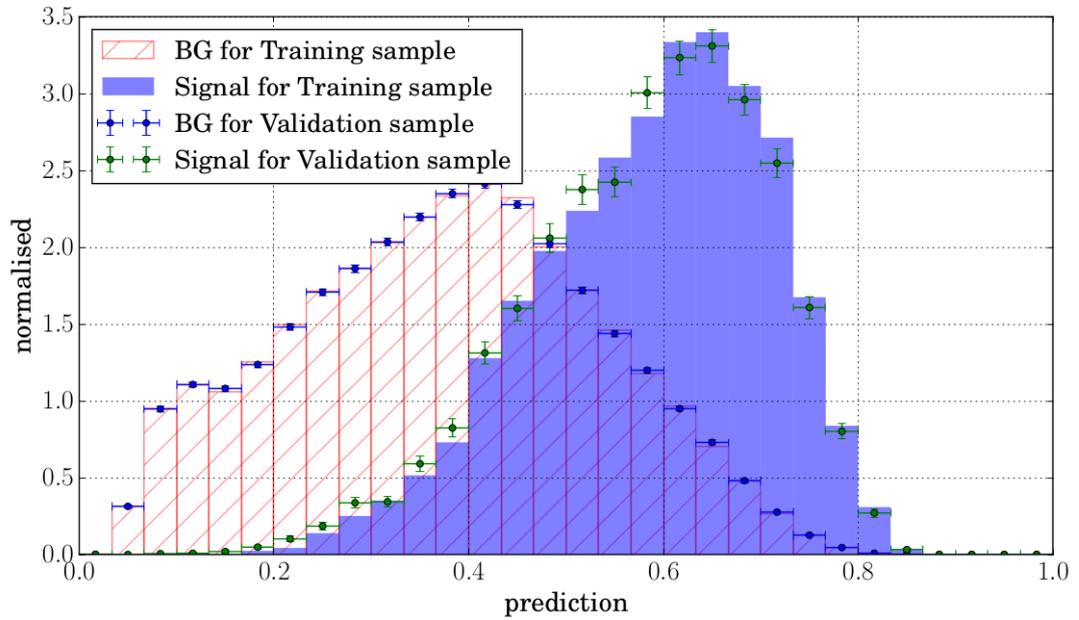


Figure 6.16: Predictions made by the uniform gradient-boosting (uGB) classifier for test (data points, blue for background and green for signal) and training (columns, shaded red for background and solid blue for signal) sample. The comparison between a classifier's prediction for the training sample and an independent validation sample serves as overtraining check.

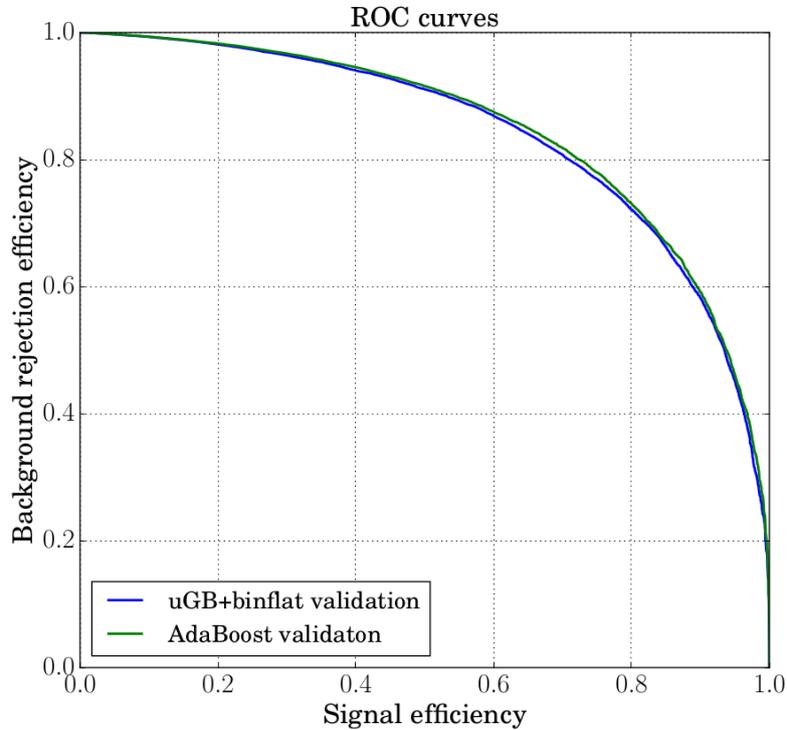


Figure 6.17: ROC curves of gradient-boosting classifier with AdaLoss function (green) and flatness loss function evaluated on the test sample.

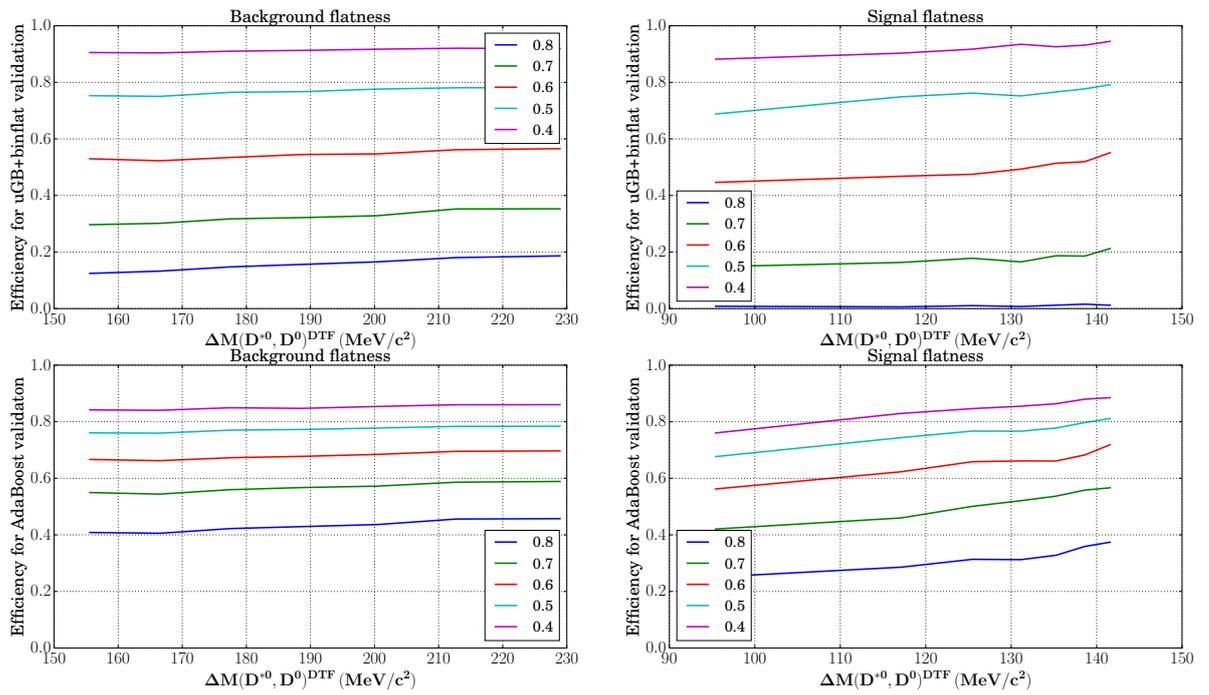


Figure 6.18: Selection efficiency as a function of Δm_{DTF} for different cuts on the flatness classifier response evaluated on the background test sample (top left) and signal test sample (top right). For comparison, the corresponding selection efficiencies for AdaBoost are shown in the bottom row.

6.6 OPTIMISING THE SELECTION

After the training process, the classifier response (BDT response) is a powerful selection variable, bundling the ability of its input variables to suppress combinatorial background. Another powerful variable to reject background originating from random combinations of two electrons to a di-electron candidate is the di-electron mass as already mentioned in the pre-selection Sec. 6.2. Applying a single cut on each of these two variables constitutes the final step in the selection process. As the goal of the work presented here is the first observation of the decay $D^{*0} \rightarrow D^0 e^+ e^-$, the cuts are chosen such that a metric representing the statistical significance of the expected signal becomes maximal. A popular choice for the metric is the figure of merit (FoM)

$$\mathcal{F} = \frac{S}{\sqrt{S+B}} \quad (6.8)$$

where S is the expected signal yield of the $D^{*0} \rightarrow D^0 e^+ e^-$ decay and B the expected number of background events in the signal region. The cuts that maximise this quantity are found by evaluating the yield $N_{D^0\gamma}$ obtained in the control channel after applying the cuts and scaling it according to

$$S = N_{D^0\gamma} \times \frac{\mathcal{B}(D^{*0} \rightarrow D^0 e^+ e^-) \times \epsilon_{sel}^{sig,MC} \times \zeta_{sig}}{\mathcal{B}(D^{*0} \rightarrow D^0 \gamma) \times \epsilon_{sel}^{con,MC} \times \zeta_{con}} \quad (6.9)$$

where $\epsilon_{sel}^{sig,MC}$ and $\epsilon_{sel}^{con,MC}$ are the selection efficiencies obtained from signal and control channel simulation, respectively, ζ_{sig} and ζ_{con} are the simulation specific generator-level-cut efficiencies (see Appendix A) and the ratio of branching ratios is taken from Eq. 2.9. The yield $N_{D^0\gamma}$ is extracted as parameter of a control channel fit performed for each set of cuts under consideration. The quantity S then corresponds to the expected number of $D^{*0} \rightarrow D^0 e^+ e^-$ events for a given set of selection cuts. The number of background events B in the signal region, defined as $\Delta m_{DTF} \in [100, 150] \text{ MeV}/c^2$, is determined by fitting the upper sideband of the Δm_{DTF} distribution after applying the set of cuts under consideration and extrapolating the found background yield into the signal region. An example of these fits is shown in Appendix E. The determination of the selection efficiencies from simulation is discussed in Chap. 7.1. The optimal working point is derived by first performing a coarse scan through the space spanned by possible cuts on the classifier response and on the di-electron mass. The resulting values of the FoM are shown in Fig. 6.19. The region exhibiting the highest values of the FoM is then scanned again using finer steps. This procedure is repeated until shrinking the step size does not give more information about the optimal working point considering the uncertainty of the FoM. The uncertainty of the FoM is calculated from the statistical uncertainty of the yields obtained from the fits and the uncertainty of the efficiencies. The final scan is shown in Fig. 6.20, according to which the optimal working point is given by the selection: $\text{BDTresponse} > 0.45$ and $m(e^+ e^-) < 40 \text{ MeV}/c^2$. These two cuts finalise the selection and thus mark the last step taken before the signal region is un-blinded. The final Δm_{DTF} distributions with applied control and signal channel veto, respectively, are shown in Fig. 6.21 and Fig. 6.22. In both distributions, a clear peak is visible around $\Delta m_{DTF} = 142 \text{ MeV}/c^2$.

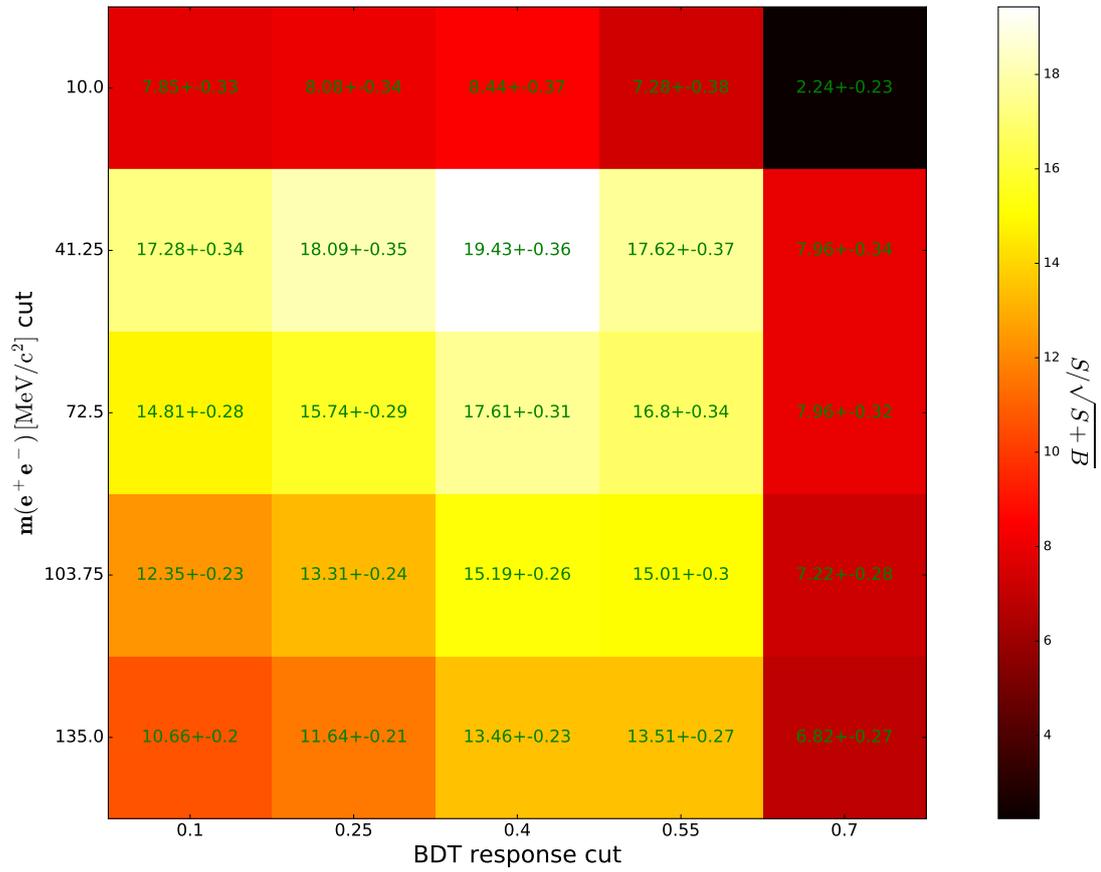


Figure 6.19: Coarse scan through cuts on the invariant mass of the di-electron system and the BDT classifier response. The number in each field shows the value of the FoM for the corresponding cuts.

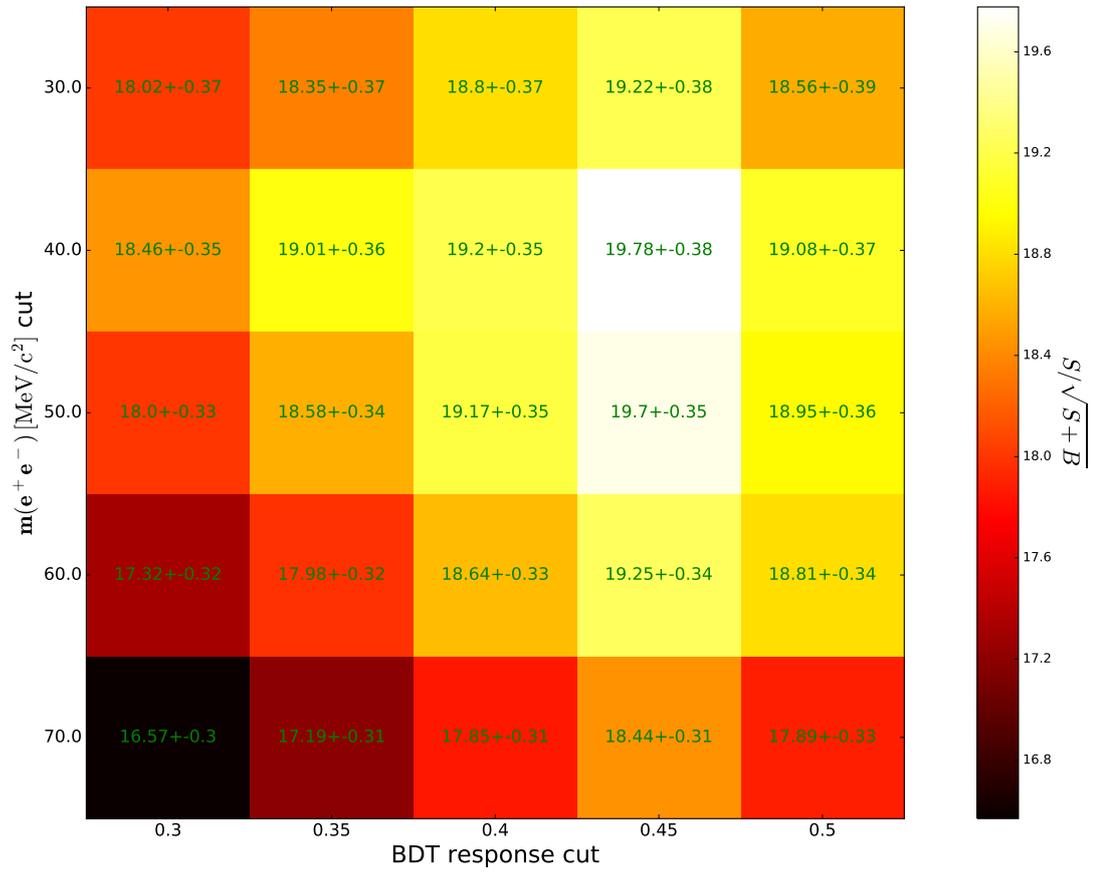


Figure 6.20: Fine scan through cuts on the invariant mass of the di-electron system and the BDT classifier response. The number in each field shows the value of the FoM for the corresponding cuts.

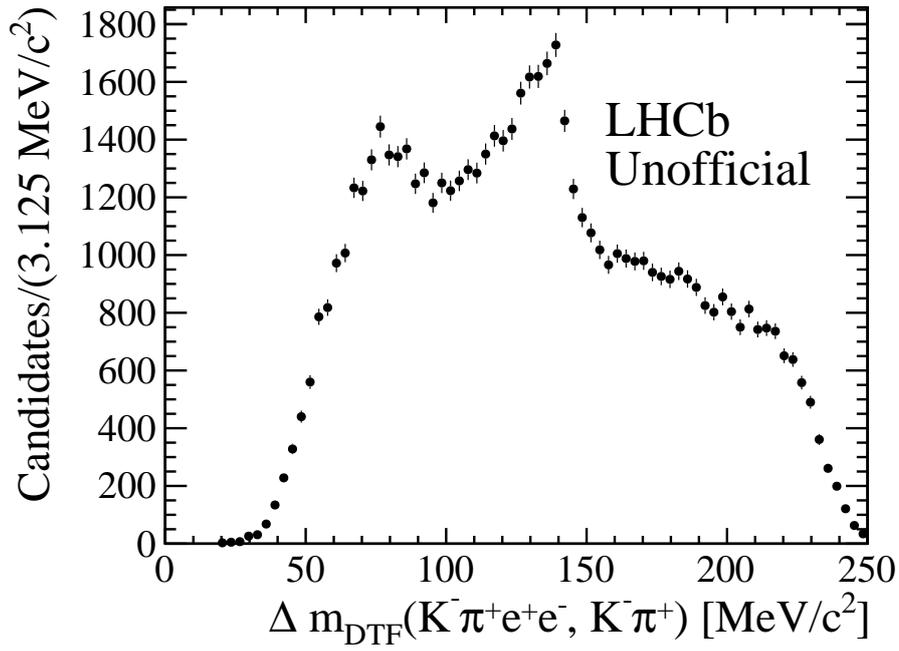


Figure 6.21: Distribution of Δm_{DTF} after full selection and applied control channel veto.

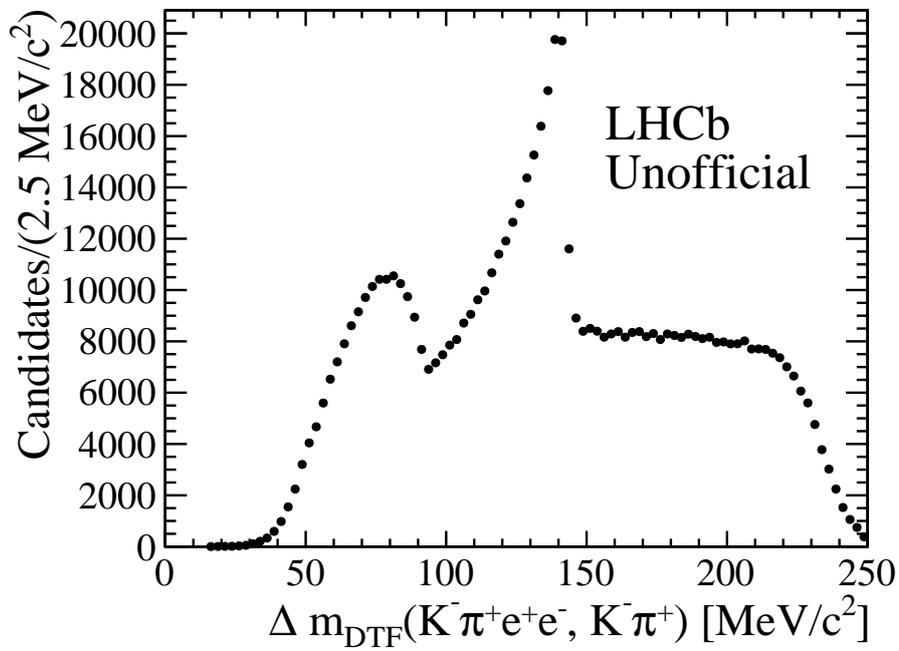


Figure 6.22: Distribution of Δm_{DTF} after full selection and applied signal channel veto.

7

DETERMINATION OF THE BRANCHING RATIO

In this chapter, the branching ratio of the decay $D^{*0} \rightarrow D^0 e^+ e^-$ relative to the decay $D^{*0} \rightarrow D^0 \gamma$ is determined using data and simulation selected as discussed in the previous chapter. First, efficiencies of the full selection are obtained from simulation, before the final yields of $D^{*0} \rightarrow D^0 e^+ e^-$ and $D^{*0} \rightarrow D^0 (\gamma \rightarrow e^+ e^-)$ decays are extracted by fitting the final Δm_{DTF} distributions.

7.1 EFFICIENCY DETERMINATION

The total efficiency estimates the number of selected candidates relative to all candidate decays that occurred during data taking. It is composed of partial efficiencies specifying fractions of candidates that end up in LHCb's detector acceptance, that are reconstructed, that pass the trigger selection and that pass the signal selection. Since simulated events run through the same selection chain as data, the total efficiency is in principle directly given by the ratio of selected events and the number of events that have been generated in the simulation. The generated number of events, however, already includes the set of generator level cuts which select only candidates that decay in LHCb's acceptance and have kinematic properties that allow for a proper reconstruction. Furthermore, the simulated samples are re-weighted after the pre-selection such that they better describe the recorded data. The re-weighting has an impact on the efficiency determination. The total efficiency obtained from simulation for signal and control channel is therefore split into four partial efficiencies:

$$\begin{aligned}\epsilon_{tot}^{sig,MC} &= \epsilon_{rec+trg+presel}^{sig,MC} \times \epsilon_{veto}^{sig,MC} \times \epsilon_{BDT+M}^{sig,MC} \times \bar{\zeta}_{sig} \\ \epsilon_{tot}^{con,MC} &= \epsilon_{rec+trg+presel}^{con,MC} \times \epsilon_{veto}^{con,MC} \times \epsilon_{BDT+M}^{con,MC} \times \bar{\zeta}_{con}\end{aligned}\quad (7.1)$$

where $\bar{\zeta}_{sig}$ and $\bar{\zeta}_{con}$ are the generator level cut efficiencies for signal and control channel, respectively, $\epsilon_{rec+trg+presel}$ is the efficiency of all selection steps up to and including pre-selection and is given by

$$\begin{aligned}\epsilon_{rec+trg+presel}^{sig,MC} &= \frac{N_{rec+trg+presel}^{MC}(D^{*0} \rightarrow D^0 e^+ e^-)}{N_{gen}^{MC}(D^{*0} \rightarrow D^0 e^+ e^-)} \\ \epsilon_{rec+trg+presel}^{con,MC} &= \frac{N_{rec+trg+presel}^{MC}(D^{*0} \rightarrow D^0 (\gamma \rightarrow e^+ e^-))}{N_{gen}^{MC}(D^{*0} \rightarrow D^0 \gamma)}\end{aligned}\quad (7.2)$$

where $N_{rec+trg+presel}^{MC}$ and N_{gen}^{MC} are the number of selected and generated candidates, respectively. The generator level cut efficiencies (*cf.* Appendix A) are known from the production of the simulated samples. Also known from the production of simulated samples is the number of generated $D^{*0} \rightarrow D^0 e^+ e^-$ events (see Table 6.3). The number of generated $D^{*0} \rightarrow D^0 \gamma$ events, however, cannot be simply taken from the full simulation production. As described in Sec. 6.1.1 only events are kept in which the photon actually produced a pair of electrons.

This leads to the technicality that sometimes not a single photon from a set¹ of $D^{*0} \rightarrow D^0 \gamma$ decays, passing the generator level cuts, converts. No record of these events is kept in the private simulation production used in this analysis, albeit that they contribute to the total number of generated events. To work around this problem, a dedicated simulation production of $D^{*0} \rightarrow D^0 \gamma$ decays is performed with sets large enough that they definitely contain photon conversions. The efficiency $\varepsilon_{rec+trg+presel}^{con,MC}$ is then given by:

$$\varepsilon_{rec+trg+presel}^{con,MC} = \varepsilon_{rec+trg} \times \frac{N_{rec+trg+presel}^{MC}(D^{*0} \rightarrow D^0(\gamma \rightarrow e^+ e^-))}{N_{rec+trg}^{MC}(D^{*0} \rightarrow D^0(\gamma \rightarrow e^+ e^-))} \quad (7.3)$$

with

$$\varepsilon_{rec+trg} = (1.99 \pm 0.04) \times 10^{-4} \quad (7.4)$$

where $\varepsilon_{rec+trg}$ now contains the efficiency of reconstruction and trigger selection obtained from the dedicated simulation production and $N_{rec+trg}^{MC}(D^{*0} \rightarrow D^0(\gamma \rightarrow e^+ e^-))$ is the number of events present in the tuple after reconstruction and trigger selection (see Table 6.3). The efficiency $\varepsilon_{rec+trg}$ also implicitly contains the conversion probability of the photon. The efficiency ε_{veto}^{MC} of the veto cuts is calculated as the ratio of events passing the pre-selection and additionally passing the veto requirement. After pre-selection, applying veto cuts and re-weighting of the simulated samples, the efficiency of the multivariate selection and the final cut on the di-electron invariant mass has to be determined using the sum of weights² of the simulated samples:

$$\varepsilon_{BDT+M}^{sig,MC} = \frac{\sum_{i=1}^{N_{sel}} w_i(D^{*0} \rightarrow D^0 e^+ e^-)}{\sum_{i=1}^{N_{rec+trg+presel}} w_i(D^{*0} \rightarrow D^0 e^+ e^-)} \quad (7.5)$$

$$\varepsilon_{BDT+M}^{con,MC} = \frac{\sum_{i=1}^{N_{sel}} w_i(D^{*0} \rightarrow D^0(\gamma \rightarrow e^+ e^-))}{\sum_{i=1}^{N_{rec+trg+presel}} w_i(D^{*0} \rightarrow D^0(\gamma \rightarrow e^+ e^-))}$$

The result of the efficiency determination for the full selection is given in Table 7.1. The difference between the selection efficiency of $D^{*0} \rightarrow D^0 e^+ e^-$ and $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+ e^-)$ decays is dominated by the photon conversion probability contributing to the efficiency $\varepsilon_{rec+trg+presel}^{MC}$ in the control channel. Yet, the efficiency ε_{BDT+M}^{MC} of the signal decay is significantly smaller than the one of the decay $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+ e^-)$. This is due to the cut on the invariant di-electron mass, which has almost no effect in the case of photon conversions but removes a part of the $D^{*0} \rightarrow D^0 e^+ e^-$ signal.

-
- 1 The simulation production is split in computing jobs requesting a certain number of decays passing the generator level cuts. "Set" refers to all generated decays of one job.
 - 2 Using the sum of weights cannot increase or decrease the statistical power of the simulated sample. Therefore the uncertainty of the sum of weights is scaled with effective weights given by $w_{eff} = \sum w_i / \sum w_i^2$.

Table 7.1: Selection efficiencies of signal and control channel as obtained from simulated samples.

(a) Signal Channel		(b) Control channel	
Efficiency	Estimated Value [%]	Efficiency	Estimated Value [%]
ξ	1.81 ± 0.03	ξ	8.00 ± 0.02
$\varepsilon_{rec+trg+presel}^{MC}$	$(1.03 \pm 0.01) \times 10^{-1}$	$\varepsilon_{rec+trg+presel}^{MC}$	$(2.94 \pm 0.07) \times 10^{-3}$
ε_{veto}^{MC}	61.5 ± 0.6	ε_{veto}^{MC}	96 ± 2
ε_{BDT+M}^{MC}	55.2 ± 0.5	ε_{BDT+M}^{MC}	80.9 ± 0.7
ε_{tot}^{MC}	$(6.33 \pm 0.13) \times 10^{-4}$	ε_{tot}^{MC}	$(1.83 \pm 0.05) \times 10^{-4}$

7.2 EVENT YIELD EXTRACTION AND SIGNAL SIGNIFICANCE

The final ingredients for the branching ratio determination of the decay $D^{*0} \rightarrow D^0 e^+ e^-$ relative to the decay $D^{*0} \rightarrow D^0 \gamma$ are the corresponding measured event yields. They are extracted by fitting the Δm_{DTF} data distributions after applying all veto and selection cuts. The fit model used to describe the $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+ e^-)$ candidate distribution is the one defined in Eq. 6.7. The fit is shown in Fig. 7.1 with its fit parameters given in Table 7.2. Tail parameters of the $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+ e^-)$ model are fixed using a fit to the simulated distribution (see Fig. F.1). The number of measured $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+ e^-)$ events is estimated as

$$N_{con} = 97931 \pm 1479 \quad (7.6)$$

where the uncertainty is the statistical one obtained from the fit. For the extraction of the $D^{*0} \rightarrow D^0 e^+ e^-$ yield, a different background model has been developed as the combinatorial background predominantly consists of random combinations of two prompt electrons to a di-electron candidate rather than di-electron candidates from random photon conversions. The combinatorial background model is inspired by a shape used to describe the soft³ pion background in the $D^{*0}-D^0$ mass difference of reconstructed $D^{*0} \rightarrow D^0 \pi^0$ decays in Ref. [77]. This model is combined with a Gaussian function and given by

$$\begin{aligned} & \mathcal{P}_{combBGsig}(\Delta m_{DTF}; f_{combBGsig}, \Delta m_0, C, B, \mu, \sigma) \\ &= f_{combBGsig} \times \left[1 - \exp\left(-\frac{\Delta m_0 - \Delta m_{DTF}}{C}\right) + B \times \left(\frac{\Delta m_{DTF}}{\Delta m_0} - 1\right) \right] \\ &+ (1 - f_{combBGsig}) \times G(\Delta m_{DTF}; \mu, \sigma) \end{aligned} \quad (7.7)$$

where Δm_0 is the threshold at which the first background component starts to act, C determines the curvature of the shape, B steers the models behaviour for large Δm_{DTF} and G is the Gaussian function. This model is tested using a 2016 data same-sign sample of $D^{*0} \rightarrow D^0 e^+ e^-$ candidates. As can be seen in Fig. 7.2, the model describes the soft di-electron background quite well. The model describing the partially reconstructed background $D^{*0} \rightarrow D^0[\chi e^+ e^-]_{\pi^0}$ is build from two sigmoid functions (see Eq. 6.4) according to:

$$\mathcal{P}_{partBGsig}(\Delta m_{DTF}; \mu_1, k_1, \mu_2, k_2) = \mathcal{S}(\Delta m_{DTF}; \mu_1, k_1) \times \mathcal{S}(\Delta m_{DTF}; \mu_2, k_2) \quad (7.8)$$

³ Soft refers to the fact that pions from $D^{*0} \rightarrow D^0 \pi^0$ decays populate a very small phasespace and thus have very little energy.

This model is tested using the simulated $D^{*0} \rightarrow D^0[\chi e^+ e^-]_{\pi^0}$ sample (see Fig. F.3) and the two inflection point parameters μ_1 and μ_2 are fixed to the values obtained from the fit to simulation. The model \mathcal{P}_{sig} describing the $D^{*0} \rightarrow D^0 e^+ e^-$ component is the same as for $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+ e^-)$, given by Eq. 6.3. Its parameter set is completely fixed to values obtained from a fit to simulation (see Fig. F.2). The full model for the final signal fit is given by:

$$\mathcal{P}_{sigFull} = N_{sig} \times \mathcal{P}_{sig} + N_{combBGsig} \times \mathcal{P}_{combBGsig} + N_{partBGsig} \times \mathcal{P}_{partBGsig} \quad (7.9)$$

The fit is shown in Fig. 7.3 with its parameters given in Table 7.3. The number of measured $D^{*0} \rightarrow D^0 e^+ e^-$ events is estimated as

$$N_{sig} = 3253 \pm 238 \quad (7.10)$$

where the uncertainty is the statistical one obtained from the fit.

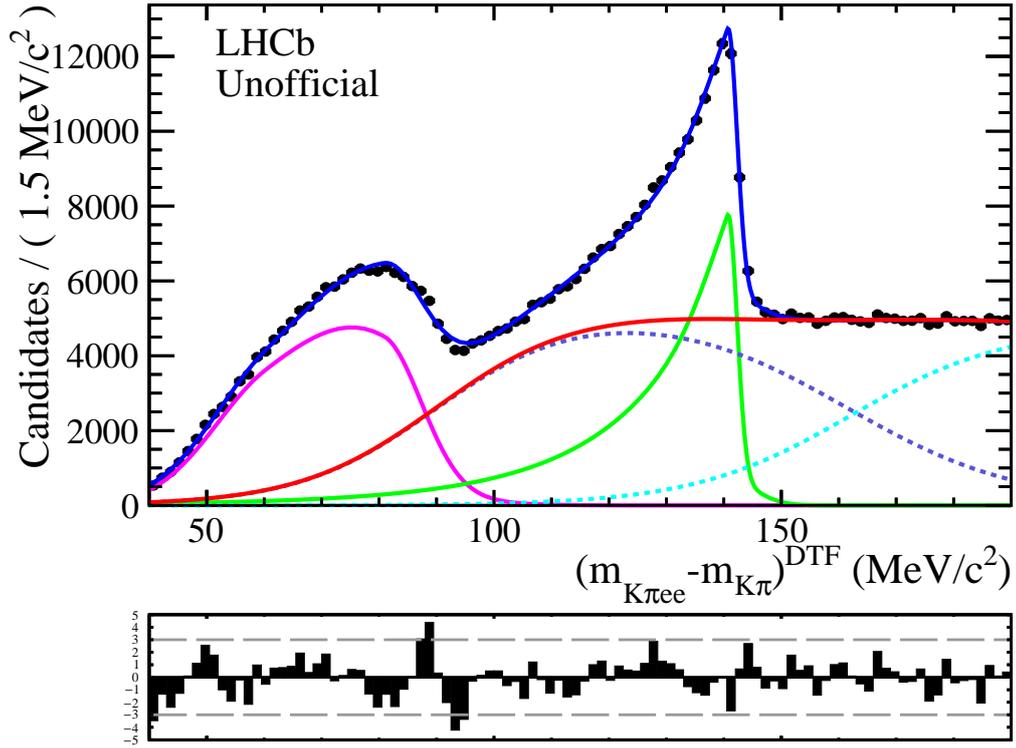


Figure 7.1: Fit of the $\Delta m_{D_{TF}}$ distribution after the full selection and applied signal channel veto. The combinatorial background model is shown as solid red curve, its two additive components are the dashed blue and cyan curves, respectively. The model of the partially reconstructed background is shown in magenta and the model for $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+e^-)$ is depicted in green. The parameters of the fit are given in Table 7.2.

Table 7.2: Parameters of the fit shown in Fig. 7.1. Parameters given without error are fixed in the fit.

\mathcal{P}_{con}		\mathcal{P}_{combBG}		\mathcal{P}_{partBG}	
Parameter	Value	Parameter	Value	Parameter	Value
α	0.1	μ_1	90.4	ν_1	60
n	11	k_1	12.1	σ_{L1}	10.2
μ_{CB}	140.74 ± 0.03	μ_2	160.9	σ_{R1}	16
σ_{CB}	1.57 ± 0.03	k_2	-15.2	ν_2	80.9
μ_G	μ_{CB}	μ_3	162	σ_{L2}	13.8
σ_G	4.6	k_3	13.7	σ_{R2}	6.7
f_{con}	0.95	f_{combBG}	0.720 ± 0.003	f_{part}	0.524 ± 0.004
N_{con}	97931 ± 1479	N_{combBG}	338343 ± 1508	N_{partBG}	112579 ± 432

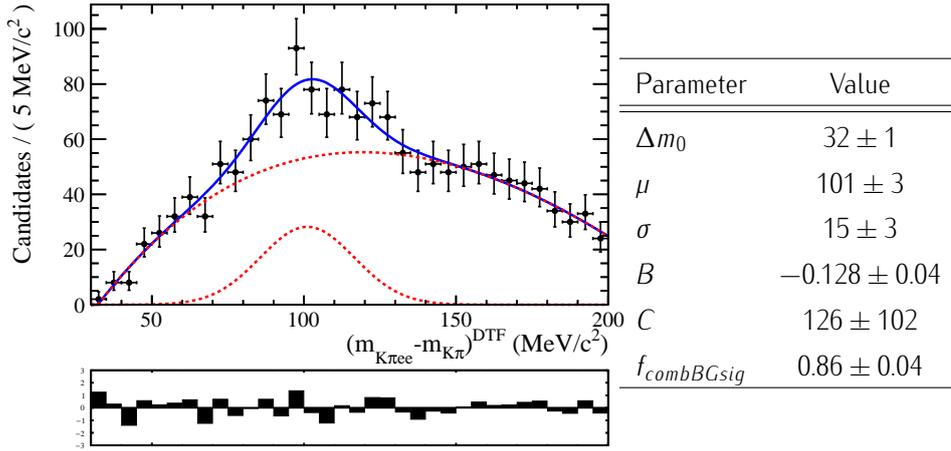


Figure 7.2: Fit of 2016 data $D^{*0} \rightarrow D^0 e^+ e^+$ candidates after full selection and control channel veto cut (left). The fitted model is defined by Eq. 7.7. The dashed red curves depict the Gaussian and "soft pion" component, respectively. The table (right) shows the corresponding fit parameters.

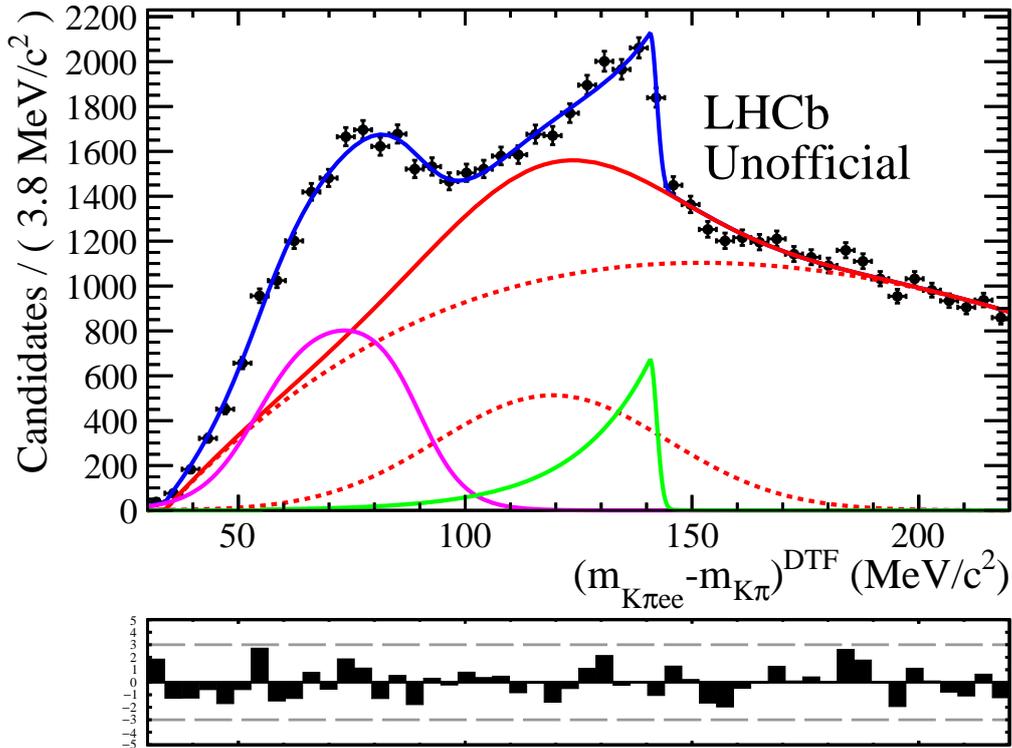


Figure 7.3: Fit of the Δm_{DTF} distribution after the full selection and applied control channel veto. The combinatorial background model is shown as solid red curve, its two additive components are the dashed red curves. The model of the partially reconstructed background is shown in magenta and the model for $D^{*0} \rightarrow D^0 e^+ e^-$ is depicted in green. The parameters of the fit are given in Table 7.3.

Table 7.3: Parameters of the fit shown in Fig. 7.3. Parameters given without error are fixed in the fit.

\mathcal{P}_{sig}		$\mathcal{P}_{combBGsig}$		$\mathcal{P}_{partBGsig}$	
Parameter	Value	Parameter	Value	Parameter	Value
α	0.087	Δm_0	33.75 ± 0.05	μ_1	54.3
n	100	B	-0.1048 ± 0.003	k_1	6.3 ± 0.2
μ_{CB}	140.8	C	126	μ_2	89.9
σ_{CB}	1.43	μ	119 ± 1	k_2	-5.0 ± 0.6
μ_G	μ_{CB}	σ	25 ± 2	$N_{partBGsig}$	8147 ± 340
σ_G	10	$f_{combBGsig}$	0.83 ± 0.02		
f_{sig}	0.997	$N_{combBGsig}$	50799 ± 404		
N_{sig}	3253 ± 238				

The model presented in Eq. 7.9 can be viewed as an alternative hypothesis H to the null hypothesis H_0 stating that only background is present in the distribution shown in Fig. 7.3. This hypothesis is tested by comparing the minimal value of the negative log-likelihood obtained from the fit with the signal component, to the one obtained from a fit with the background model only. The optimal test statistic according to the Neyman–Pearson lemma [52] is as usual given by the likelihood ratio Λ . Moreover, according to Wilk’s theorem [78], the test statistic $-2\ln(\Lambda)$ approximately follows a χ^2 -distribution when the sample size approaches infinity. The degrees of freedom are given by the difference between the number of free parameters in the full model and in the background model. The χ^2 value is given by

$$\begin{aligned}\chi^2 &= -2\ln(\Lambda) \\ &= -2[NLL(H) - NLL(H_0)]\end{aligned}\quad (7.11)$$

where NLL is the negative log-likelihood as defined in Eq. 5.4. Plugging in the negative log-likelihoods $NLL(H) = -303742.7$ and $NLL(H_0) = -303662.2$ obtained from the fits gives $\chi^2 = 161$. The only additional free parameter in the full model compared to the background model is the signal yield. Therefore, only one degree of freedom lies between H and H_0 . Integrating the corresponding χ^2 -distribution from $\chi^2 = 161$ to infinity gives the p -value of the null hypothesis:

$$p = 6.5 \times 10^{-37}\quad (7.12)$$

corresponding to a significance⁴ of the alternative hypothesis of

$$z = 12.7\sigma.\quad (7.13)$$

⁴ Conveniently, the significance in units of σ is also approximately given by $\sqrt{\chi^2}$.

7.3 BRANCHING RATIO RESULT

The branching ratio of $D^{*0} \rightarrow D^0 e^+ e^-$ relative to $D^{*0} \rightarrow D^0 \gamma$ is now simply given by:

$$\frac{\mathcal{B}(D^{*0} \rightarrow D^0 e^+ e^-)}{\mathcal{B}(D^{*0} \rightarrow D^0 \gamma)} = \frac{N_{sig} \times \epsilon_{tot}^{con,MC}}{N_{con} \times \epsilon_{tot}^{sig,MC}} \quad (7.14)$$

Using the numbers for the efficiencies and yields given above, this results in:

$$\frac{\mathcal{B}(D^{*0} \rightarrow D^0 e^+ e^-)}{\mathcal{B}(D^{*0} \rightarrow D^0 \gamma)} = (9.6 \pm 0.7 \text{ (stat)}) \times 10^{-3} \quad (7.15)$$

The statistical uncertainty is propagated from the statistical error of the event yields. The uncertainties of the efficiencies are treated as systematic uncertainty in the next chapter. Using the known branching ratio of $D^{*0} \rightarrow D^0 \gamma$ from Table 2.1, the branching ratio of $D^{*0} \rightarrow D^0 e^+ e^-$ can be inferred from Eq. 7.15. This leads to

$$\mathcal{B}(D^{*0} \rightarrow D^0 e^+ e^-) = (3.38 \pm 0.25 \text{ (stat)}) \times 10^{-3} \quad (7.16)$$

8

SYSTEMATIC UNCERTAINTIES

The branching ratio measurement of the decay $D^{*0} \rightarrow D^0 e^+ e^-$ relative to the decay $D^{*0} \rightarrow D^0 \gamma$ is subject to various sources of systematic uncertainties arising from imperfections of the used methods and incomplete knowledge about the exact behaviour of the detector. A full study of systematic uncertainties is out of the scope of this thesis, therefore only uncertainty sources that are not expected to largely cancel in the ratio of branching fractions and contribute the most are quantified. A summary of the quantified systematic uncertainties is given in Table 8.1.

FINITE SIZE OF SIMULATED SAMPLES The selection efficiencies determined in Sec. 7.1 have uncertainties due to the finite size of the simulated samples they are calculated from. The uncertainties are obtained by treating the application of selection criteria as a binomial process with the probability given by the true selection efficiency. Since the true selection efficiency is unknown, the estimated efficiency $\hat{\epsilon} = k/N$ is taken as probability instead, where k and N denote the number of events after and before the selection, respectively. The uncertainty of the selection efficiency is then given by

$$\Delta\hat{\epsilon} = \sqrt{\frac{\hat{\epsilon}(1-\hat{\epsilon})}{N}} \quad (8.1)$$

The efficiencies and their uncertainties are given in Table 7.1. The relative error of the efficiencies directly propagates to a relative error of the branching fraction.

SIMULATED VELO MATERIAL DESCRIPTION Photon conversions in simulated control channel decays rely on an accurate description of the material within the VELO. In particular, the probability that a photon conversion occurs depends on the amount of material that is traversed by the photon, which can be expressed in terms of the fraction of a radiation length X_0 . The relative uncertainty of X_0 in simulation amounts to 6% [47] and is mostly due to limits on the accuracy of shape modelling in GEANT4. The uncertainty of X_0 is propagated to the photon conversion probability (see Appendix B.1) which is implicitly contained in the total selection efficiency of $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+ e^-)$. In a first order approximation, the relative uncertainty of X_0 directly contributes as relative uncertainty of the branching ratio.

SIMULATION CORRECTION BY RE-WEIGHTING The simulated control channel sample is corrected for deviations from data by re-weighting it using the *sPlot* technique. The set of variables used in the re-weighting has been chosen to be as small as possible while showing a good correction performance. Other variable sets, however, can have similar correction power and their impact on the efficiency determination is studied by exchanging variables in the set with variables that are expected to have a comparable impact on the re-weighting, e.g. exchanging $p_T(D^{*0})$ with $p_T(D^0)$. Furthermore, the size of the set is varied by removing or adding single variables. The same procedure is applied to the re-weighting of simulated $D^{*0} \rightarrow D^0 e^+ e^-$ decays. Since the correction of the simulated $D^{*0} \rightarrow D^0 e^+ e^-$ sample is not fully independent of the correction performed on $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+ e^-)$, each variation of the

re-weighting variables in the signal channel is repeated for each tested re-weighting variable set of the control channel. This gives a distribution of efficiency ratios $\varepsilon_{tot}^{con,MC} / \varepsilon_{tot}^{sig,MC}$, the standard deviation of which is assigned as systematic uncertainty. The found 0.5% uncertainty is small compared to others, which is due to cancellation of re-weighting uncertainties in the ratio of efficiencies.

SIGNAL MODEL The yields of signal and control channel are extracted by fitting a Crystal Ball function together with a Gaussian function to the shapes observed in the Δm_{DTF} distributions (see Eq. 6.3). To evaluate the influence of the fit model on the branching ratio result, an alternative model is fitted to the final signal and control channel distributions. The Crystal Ball function is replaced by a Gaussian function with an exponential tail to the left.¹ As the nominal model has its parameters fixed to values obtained from simulation, also the parameters of the alternative model are determined by fitting the simulated distributions. The yield extraction is then performed exactly as described in Sec. 7.2. The obtained deviation from the measured branching ratio is assigned as systematic uncertainty, which amounts to 17.7% in the signal channel and 0.3% in the control channel. The systematic uncertainty derived for the control channel is neglected because it is small compared to the one from the signal channel.

COMBINATORIAL BACKGROUND MODEL The systematic uncertainty due to the choice of a combinatorial background model in the signal channel is estimated by fitting an alternative model to the final Δm_{DTF} distribution. The alternative model is the sum of a Gaussian function and the product of two sigmoid functions (*cf.* Eq. 7.7 and Eq. 7.8). As for the nominal model, the alternative model has been tested on the same-sign sample. The difference between the yield obtained by the nominal fit and using the alternative model is assigned as systematic uncertainty, amounting to 1.6%. The systematic uncertainty due to the combinatorial background model in the control channel fit is estimated by varying its fixed parameters as follows: the Δm_{DTF} distribution of the control channel is fitted 50 times with each fixed parameter of the combinatorial background randomly sampled from a Gaussian distribution. The mean of the Gaussian distribution is given by the originally fixed parameter value and the width is chosen such that sampled parameters on average lead to fits of acceptable quality. This procedure gives a sample of 50 control channel yields, the standard deviation of which is assigned as systematic uncertainty. The standard deviation of the yields is large, giving a 15.2% uncertainty on the ratio of branching fractions. The uncertainty of the signal and control channel combined amounts to 15.3%.

A more detailed study of systematic uncertainties has to investigate more potential uncertainty sources and would utilise more thorough methods to quantify them. A typical, though time-consuming approach are pseudo-experiments. To estimate the systematic uncertainty due to the choice of a particular fit model, a large number of pseudo-datasets could be created by sampling from the fit model. Fitting each of these pseudo-datasets with the nominal and alternative model and estimating the uncertainty from averaged quantities gives a more robust uncertainty assessment. This way, it could be checked whether or not the large uncertainty obtained by the alternative signal model is an extreme case. Similarly, the effect of fixing bremsstrahlung tail parameters to values obtained from simulation should be studied. As mentioned above, the material description of the detector simulation is an uncertainty

¹ This model is sometimes called "RooExpAndGauss".

Table 8.1: Sources of systematic uncertainty and their relative contribution to the measured ratio of branching fractions. The total systematic uncertainty is the quadratic sum of all given systematic uncertainties.

Source	Relative uncertainty [%]
Simulation statistics	3.4
Material description in simulation	6
Data/Simulation correction	0.5
Signal Model	17.7
Background Model	15.3
Total	24.4

source on its own and it might accordingly lead to an uncertainty on the simulated electron bremsstrahlung, which does not fully cancel in the ratio of branching fraction as conversion electrons traverse slightly less material than prompt ones. Another source of systematic uncertainty that could be studied is due to the veto that separates signal from control channel. This uncertainty is neglected here because it is small compared to others, yet with more statistics in the control channel simulation a more precise estimate than $(0.8 \pm 0.8)\%$ (see Sec. 6.3) for the expected $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+e^-)$ pollution in the final $D^{*0} \rightarrow D^0e^+e^-$ yield could be given.

In general, efficiencies of tracking and reconstructing electrons are a subject of on-going research at LHCb. Especially for such low momentum electrons as present in the analysed decays, tracking and reconstruction efficiencies are not known yet. It is expected, however, that these efficiencies and their uncertainties largely cancel in the ratio of branching fractions. A potential exception of this cancellation might be a difference of tracking efficiency between di-electron candidates that have one track within the VELO and those which exhibit two separate tracks. The former predominantly occurs for electrons from the decay $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+e^-)$ while the latter is the dominant case in $D^{*0} \rightarrow D^0e^+e^-$ decays. This effect has to be studied in a future analysis.

The first observation of the decay $D^{*0} \rightarrow D^0 e^+ e^-$ and a measurement of its branching ratio relative to the decay $D^{*0} \rightarrow D^0 \gamma$ is presented in this thesis. The analysis results are summarised in the first section of this chapter, before future prospects are pointed out as closing remarks.

9.1 RESULTS

The branching fraction of the decay $D^{*0} \rightarrow D^0 e^+ e^-$ relative to the decay $D^{*0} \rightarrow D^0 \gamma$, followed by $D^0 \rightarrow K^- \pi^+$ in both channels, has been measured using a dataset corresponding to an integrated luminosity of 1.87 fb^{-1} collected at 13 TeV centre-of-mass energy in proton-proton collisions with the LHCb detector. The result of the relative branching fraction measurement is

$$\frac{\mathcal{B}(D^{*0} \rightarrow D^0 e^+ e^-)}{\mathcal{B}(D^{*0} \rightarrow D^0 \gamma)} = (9.6 \pm 0.7 \text{ (stat)} \pm 2.3 \text{ (syst)}) \times 10^{-3}$$

where the uncertainties are statistical and systematic, respectively. With the known branching fraction of the normalisation channel, the total branching ratio of the decay $D^{*0} \rightarrow D^0 e^+ e^-$ is given by

$$\mathcal{B}(D^{*0} \rightarrow D^0 e^+ e^-) = (3.38 \pm 0.25 \text{ (stat)} \pm 0.82 \text{ (syst)} \pm 0.09) \times 10^{-3}$$

where the third uncertainty is due to normalisation to the decay $D^{*0} \rightarrow D^0 \gamma$. The number of reconstructed $D^{*0} \rightarrow D^0 e^+ e^-$ decays is

$$N_{sig} = 3253 \pm 238$$

which leads to 12.7σ statistical significance of the signal hypothesis, justifying the claim of an observation.

The result shows no significant deviation from the SM prediction given in Eq. 2.9 as

$$\left[\frac{\mathcal{B}(D^{*0} \rightarrow D^0 e^+ e^-)}{\mathcal{B}(D^{*0} \rightarrow D^0 \gamma)} \right]_{SM} = 6.43 \times 10^{-3}.$$

Several challenges have been met in the course of this analysis by using modern techniques such as the Turbo Stream for data acquisition, gradient boosting with a flatness loss function to treat correlations and gradient boosting as re-weighting algorithm. Moreover, the presented analysis is the first to exploit prompt low-momentum electrons at the LHCb experiment.

9.2 OUTLOOK

For a future publication of the analysis presented in this thesis, several improvements can be made, the priority of which should be the reduction and further study of systematic uncertainties. The statistical uncertainty is expected to already decrease roughly by a factor of $1/\sqrt{2}$ when the available 2018 dataset is included. The selection process should be further optimised by requiring specific L0 and HLT1 trigger decisions, which is expected to render cleaner variable distributions. The combinatorial background due to random combinations of a D^0 candidate with a di-electron candidate from a photon conversion has to be studied in more detail to obtain a better fit model, which should largely reduce the systematic uncertainty assigned to it. The small private simulation productions will be replaced by larger official ones, which enable a more precise determination of selection efficiencies and possibly a better handle on the shapes of the partially reconstructed backgrounds. With a better background model in the control channel, also the study of discrepancies between data and simulation via the *sPlot* technique becomes more robust. There are also efforts at LHCb to improve the treatment of photon conversions in GEANT4 with the goal to obtain a better approximation of the di-electron invariant mass distribution. Eventually, the analysis can be promoted to a dark photon search, for which the experience gained during the analysis of $D^{*0} \rightarrow D^0 e^+ e^-$ is of great value. The big leap, however, will be the LHCb detector upgrade during the long shutdown of the LHC between 2019 and 2021. For Run III of the LHC, a triggerless-readout system will be employed at LHCb, which in particular allows efficient software triggering on low-momentum signatures [79] such as the electrons occurring in this analysis. Moreover, the redesign of the VELO to use pixels is expected to decrease its material budget [80] and therefore the probability of photon conversions, which will be beneficial for a further extension of the analysis to a search for long-lived dark photons in decays of the $D^*(2007)^0$ meson.

A

GENERATOR LEVEL CUTS

Propagating simulated events through the detector simulation is only reasonable if the signal particles exhibit kinematic properties which can be reconstructed afterwards. Furthermore, signal decay products that are not within the acceptance of the LHCb detector can also be discarded before handing them to GEANT4. Therefore, only signal candidates passing a set of cuts run through the detector simulation. The cuts applied to simulated samples used in this analysis are listed in Table A.1. They are chosen such that particles have enough momentum to make it into long tracks. The momentum cut of the γ is tuned such that electrons coming from a conversion of this photon are likely to fulfil the requirements put on prompt electrons. The efficiencies of these cuts are given in Table A.2.

Table A.1: Cuts applied at generator level in the simulation. The cuts on the photon are only applied for $D^{*0} \rightarrow D^0 \gamma$. The angle θ is defined as the angle between the particles momentum vector and the z -axis.

Particle	Gen. Lvl. Cut
D^{*0}	$p > 6000 \text{ MeV}/c$
D^0	$p > 3000 \text{ MeV}/c$
K^-	$p > 1500 \text{ MeV}/c$ $p_T > 50 \text{ MeV}/c$ $10 \text{ mrad} < \theta < 400 \text{ mrad}$
π^+	$p > 1500 \text{ MeV}/c$ $p_T > 50 \text{ MeV}/c$ $10 \text{ mrad} < \theta < 400 \text{ mrad}$
e^\pm	$p > 1500 \text{ MeV}/c$ $p_T > 50 \text{ MeV}/c$ $10 \text{ mrad} < \theta < 400 \text{ mrad}$
γ	$p > 2500 \text{ MeV}/c$ $10 \text{ mrad} < \theta < 400 \text{ mrad}$

Table A.2: Generator-level-cut efficiencies for signal and control channel simulation.

Simulated decay	Generator-level-cut efficiency ξ
$D^{*0} \rightarrow D^0 e^+ e^-$	$(1.81 \pm 0.03)\%$
$D^{*0} \rightarrow D^0 \gamma$	$(8.00 \pm 0.02)\%$

B | PHOTON CONVERSIONS

B.1 PHOTON CONVERSION PROBABILITY

The probability P that a photon interaction with matter results in a creation of an electron pair, without first undergoing Compton scattering, is given by [7]:

$$P = 1 - \exp\left(-\frac{x}{\lambda}\right) \quad (\text{B.1})$$

where x is the distance the photon travelled through the material and λ is the photon attenuation length. For photons, the attenuation length is related to the radiation length by $\frac{7}{9}\lambda = X_0$. Hence, the conversion probability can be expressed in terms of the radiation length as

$$P = 1 - \exp\left(-\frac{7x}{9X_0}\right) \quad (\text{B.2})$$

Since the photon energies present in $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+e^-)$ decays produced in proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$ are $\mathcal{O}(\text{GeV})$, the dominant photon interaction process with matter is pair creation. The conversion probability of photons analysed in this thesis can therefore be roughly estimated by Eq. B.2. The distance a photon has to travel before it hits the first active material within the VELO can be expressed in fractions of the radiation length and amounts to $0.042X_0$. The average amount of material traversed by the photon before leaving the VELO at $z = 835 \text{ mm}$ is $0.227X_0$ [47]. With these two numbers approximate upper and lower bounds for the conversion probability can be calculated using Eq. B.2:

$$3\% \leq P \leq 16\% \quad (\text{B.3})$$

For small probabilities, Eq. B.2 can be well approximated by the first term of its Taylor expansion and thus a relative uncertainty on X_0 directly gives the relative uncertainty on P .

B.2 INVARIANT MASS OF THE DI-ELECTRON SYSTEM FROM PHOTON CONVERSIONS

The di-electron invariant mass from converting photons is not properly modelled in GEANT4. The simulation distributes the energy of the photon to the electron and positron correctly but then approximates the distribution of the polar angle with respect to the momentum vector of the photon as the sum of two exponential functions [81, 82]. The approximation is opted for out of efficiency reasons, however, it leads to a di-electron invariant mass distribution that is approximately correct for low masses but decreases too fast for higher masses. The exact di-electron invariant mass distribution can be obtained from the Bethe-Heitler formula as [83]:

$$\frac{d\sigma}{dx} \propto \frac{1}{x^3} \left[F_1(x) \ln \left(\frac{216xZ^{-1/3}}{\sqrt{1 + \left(\frac{216x^2Z^{-1/3}}{2y} \right)^2}} \right) - F_2(x) \right] \quad (\text{B.4})$$

with

$$F_1(x) = \left(2 + \frac{2}{x^2} - \frac{1}{x^4} \right) \cosh^{-1}(x) - \left(1 + \frac{1}{x^2} \right) \sqrt{1 - \frac{1}{x^2}}$$

$$F_2(x) = \frac{1}{6} \left(16 + \frac{21}{x^2} - \frac{17}{x^4} \right) \cosh^{-1}(x) - \frac{1}{12} \left(28 + \frac{17}{x^2} \right) \sqrt{1 - \frac{1}{x^2}}$$

where $x \equiv \frac{q}{2m_e}$, $y \equiv \frac{E_\gamma}{2m_e c^2}$, Z is atomic number of the material in which the conversion happens, q is equivalent to the di-electron invariant mass, m_e is the electron mass, E_γ is the photon energy and c is the speed of light. The distribution is shown in Fig. B.1 for different values of Z and E_γ . The values correspond to typical photon energies present in $D^{*0} \rightarrow D^0 \gamma$

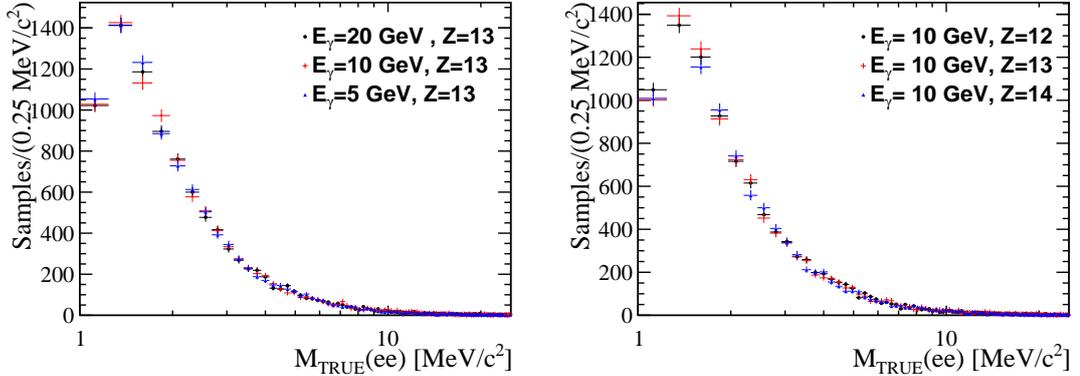


Figure B.1: Di-electron invariant mass distribution as obtained from Eq. B.4 for typical photon energies present in this analysis (left) and for atomic numbers of magnesium, aluminium and silicon which are typical VELO materials (right).

decays from proton-proton collisions at $\sqrt{s} = 13$ TeV and to atomic numbers of material present in the VELO. The impact of different possible E_γ and Z is negligible. The di-electron invariant mass distribution from the Bethe-Heitler formula compared to the one from GEANT4 is shown in Fig. B.2. It is clearly visible that higher invariant masses are more abundant in the distribution obtained from the Bethe-Heitler formula.

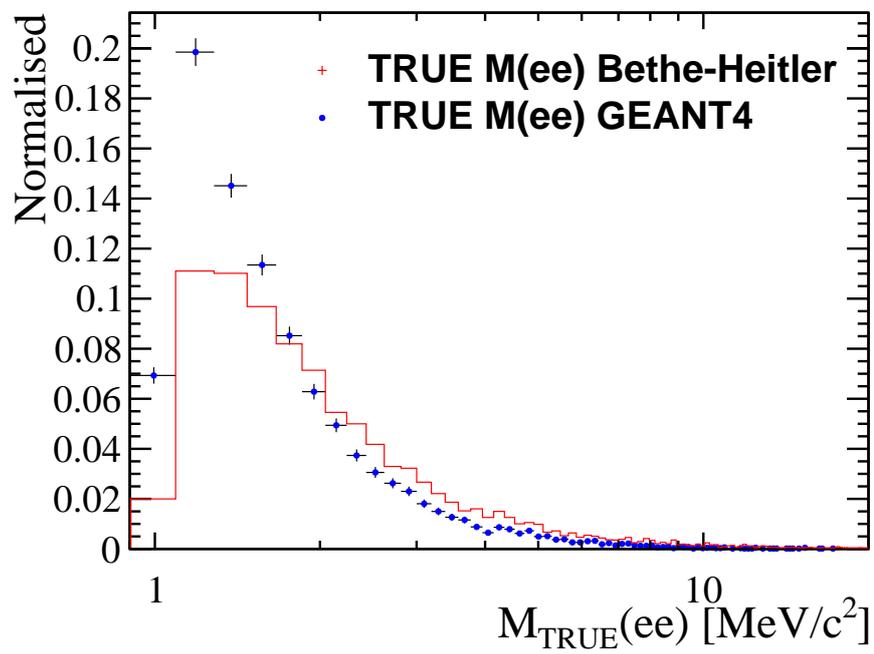


Figure B.2: Comparison between di-electron invariant mass distribution obtained from Eq. B.4 (red solid line) and produced by GEANT4 (blue points).

B.3 STUDY OF PHOTON CONVERSION BACKGROUND

The combinatorial background arising from random combinations of a D^0 candidate with a di-electron candidate from a genuine photon conversion is studied by producing "fake photons" and by event mixing as described in Sec. 6.4.1.

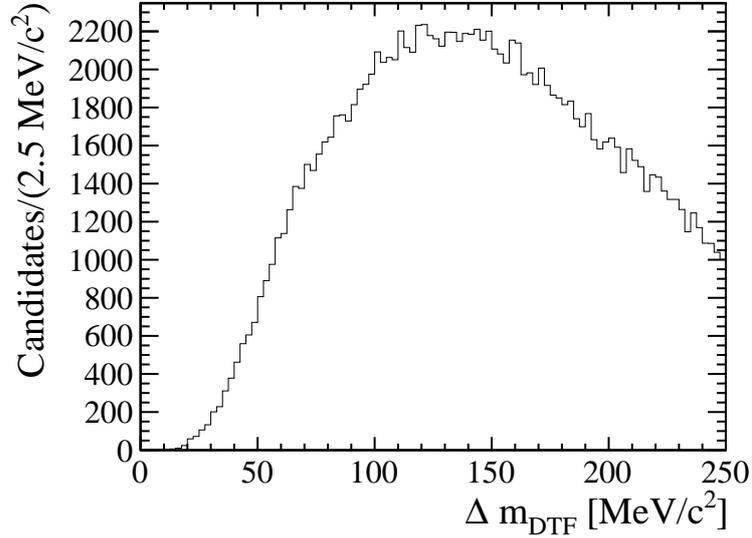


Figure B.3: "Fake photon" Δm_{DTF} distribution.

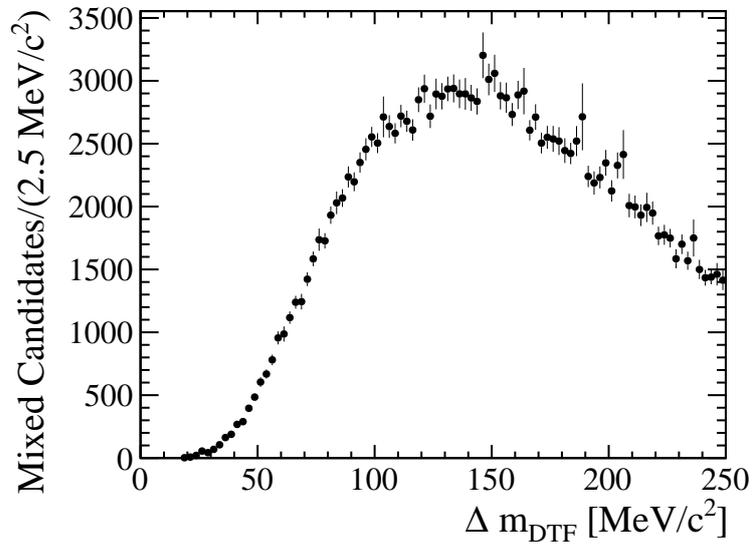


Figure B.4: Δm_{DTF} distribution obtained by event mixing and subsequent re-weighting of the angle between D^0 and γ to match the angular distribution of data without event mixing.

C

COMPARISONS OF SIMULATED AND DATA DISTRIBUTIONS

C.1 CONTROL CHANNEL VS. DATA

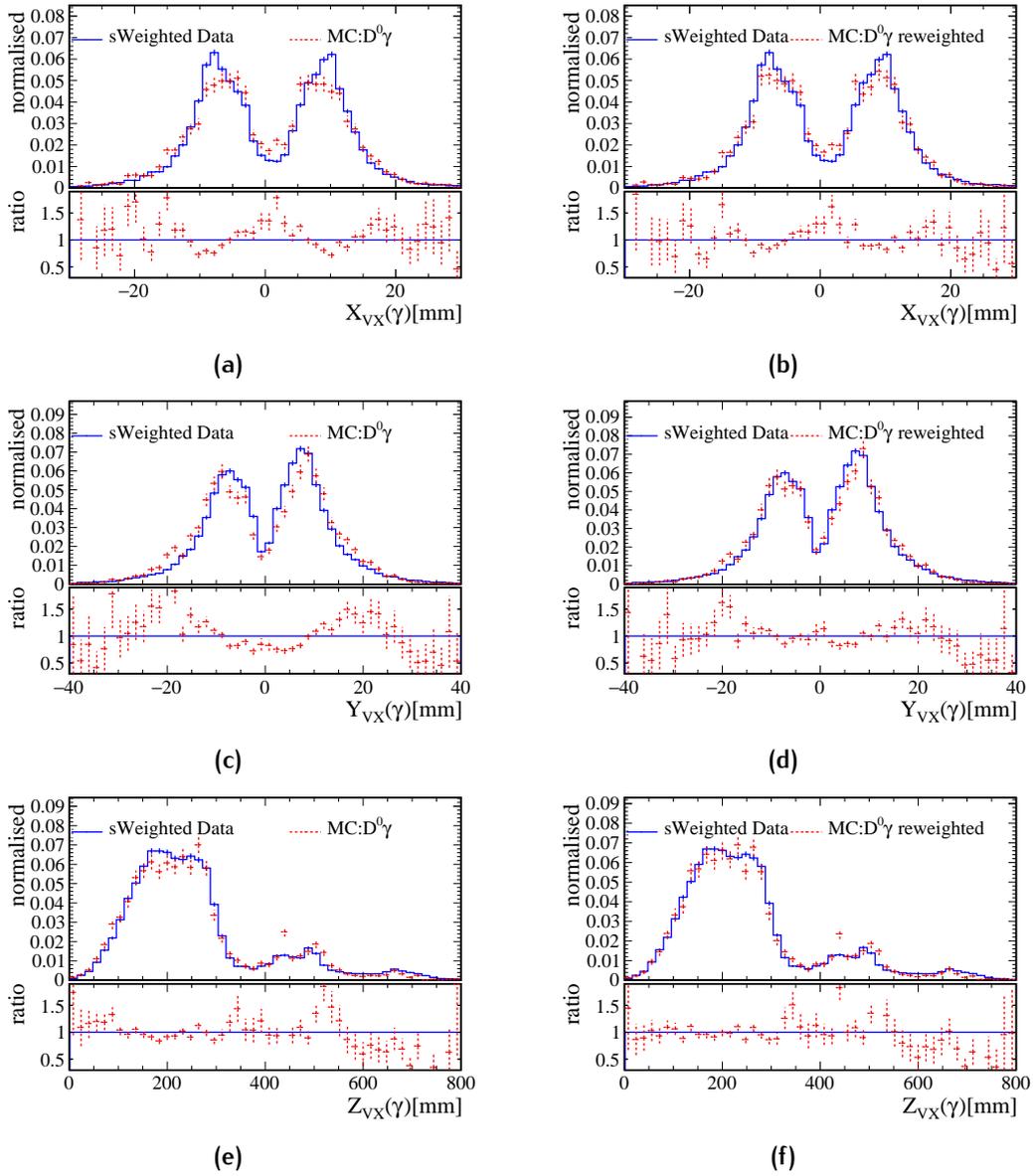


Figure C.1: Comparison between simulated $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+e^-)$ distributions (dashed red) and sWeighted data (solid blue) shown in the left column. The right column shows the corresponding re-weighted distributions.

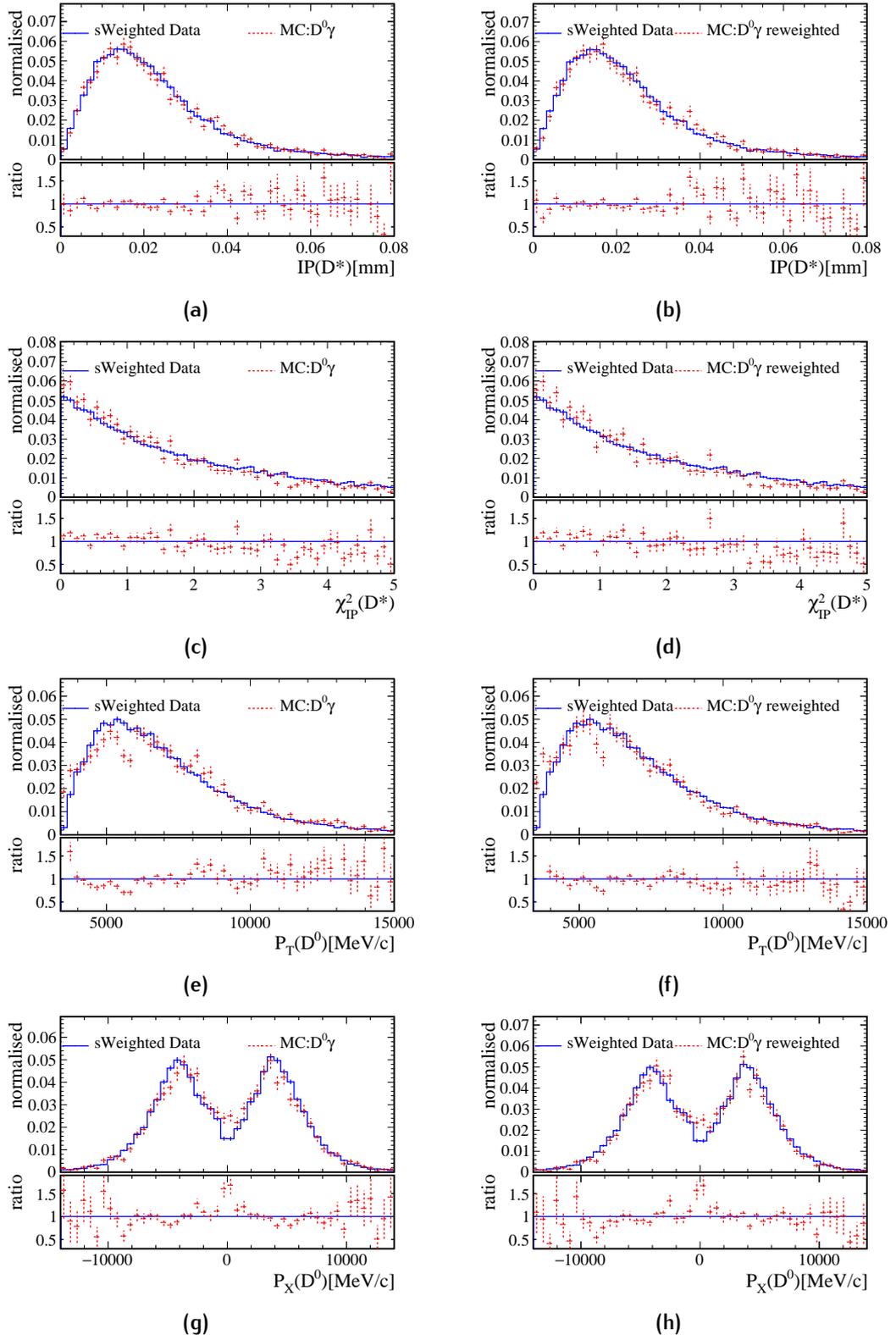


Figure C.2: Comparison between simulated $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+e^-)$ distributions (dashed red) and sWeighted data (solid blue) shown in the left column. The right column shows the corresponding re-weighted distributions.

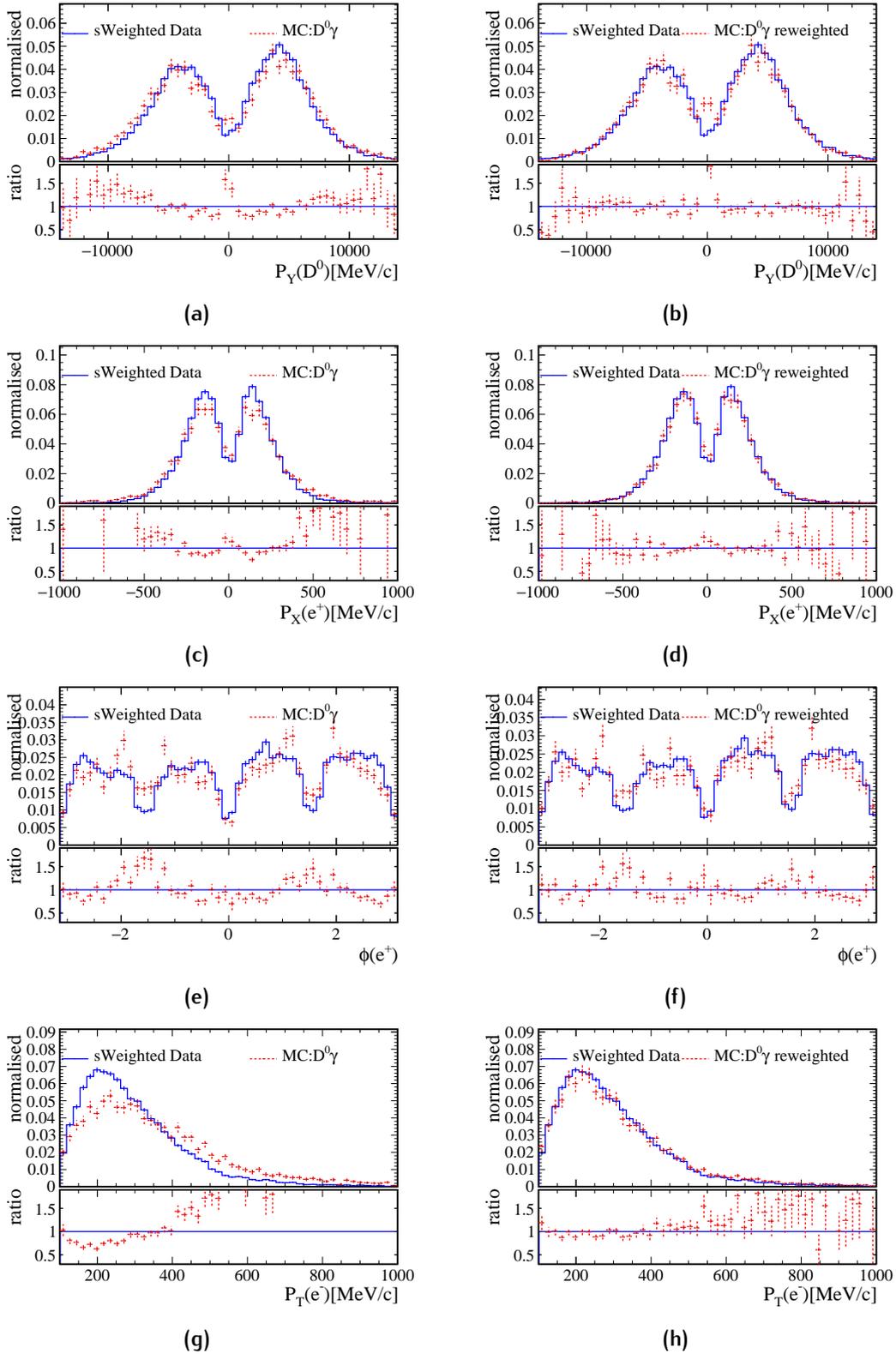


Figure C.3: Comparison between simulated $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+e^-)$ distributions (dashed red) and sWeighted data (solid blue) shown in the left column. The right column shows the corresponding re-weighted distributions.

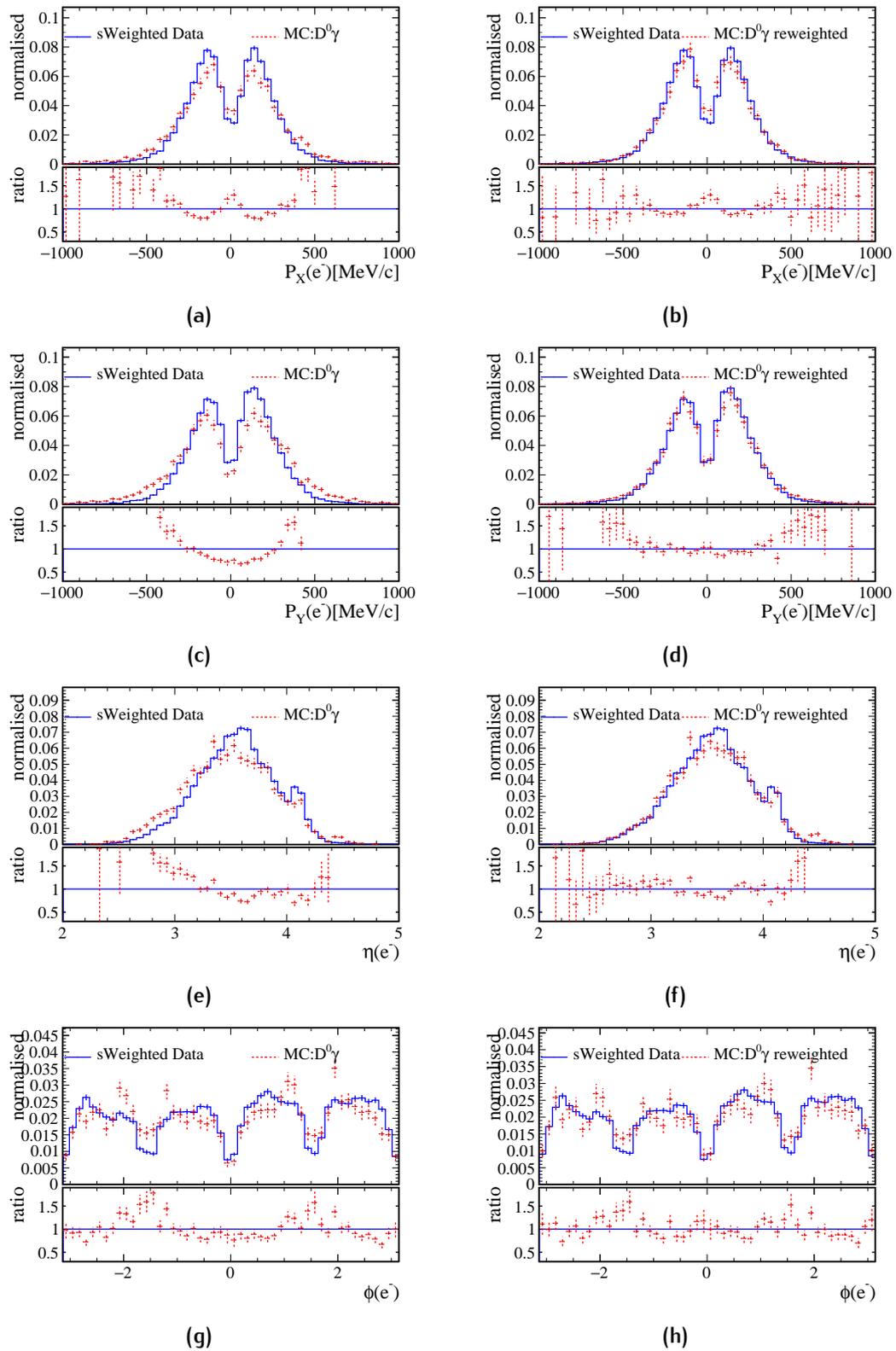


Figure C.4: Comparison between simulated $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+e^-)$ distributions (dashed red) and sWeighted data (solid blue) shown in the left column. The right column shows the corresponding re-weighted distributions.

C.2 SIMULATED SIGNAL CHANNEL VS. CONTROL CHANNEL

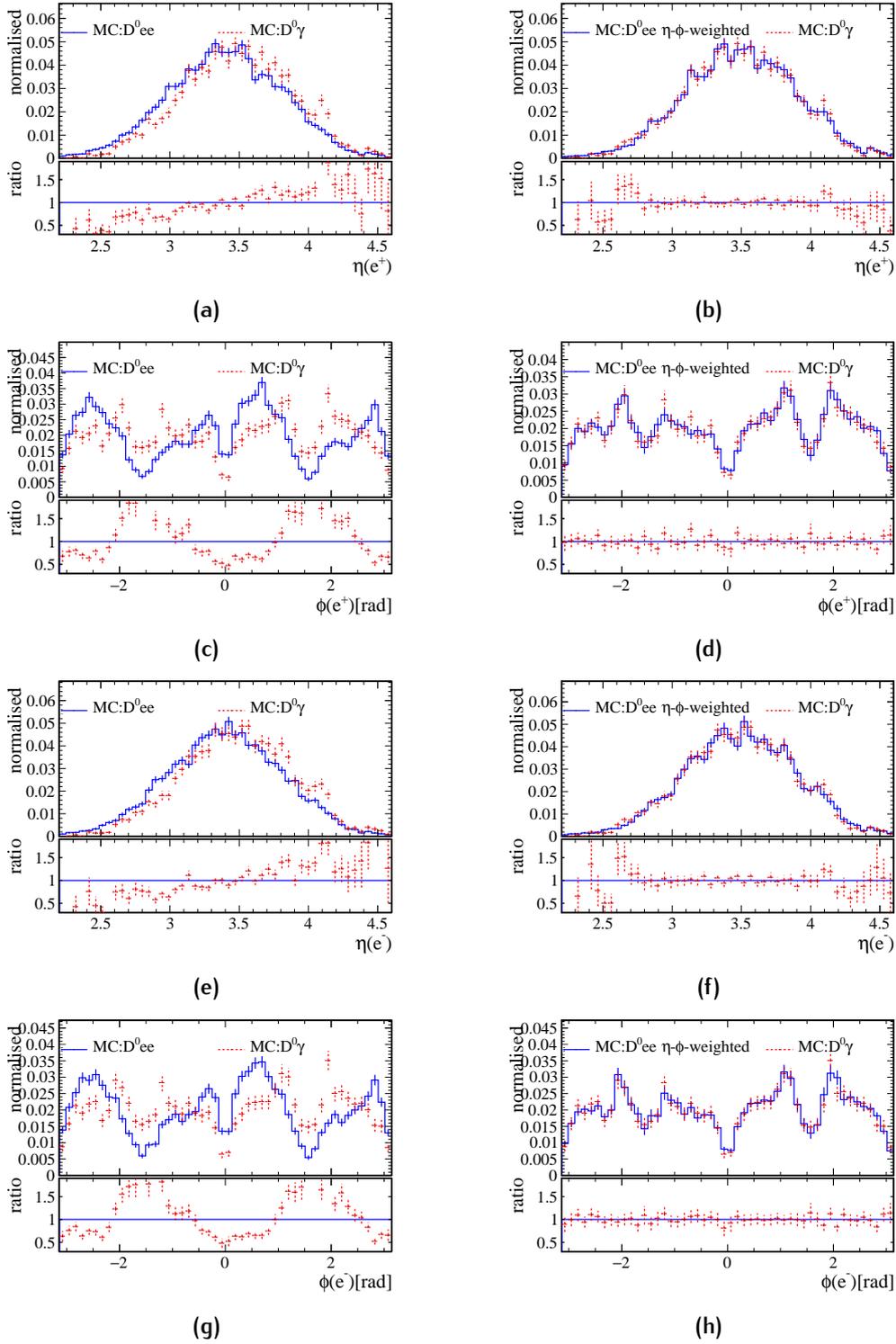


Figure C.5: Comparison between simulated control channel distributions (dashed red) and simulated signal (solid blue) shown in the left column. The right column shows the distributions where signal is re-weighted in η and ϕ of the electrons.

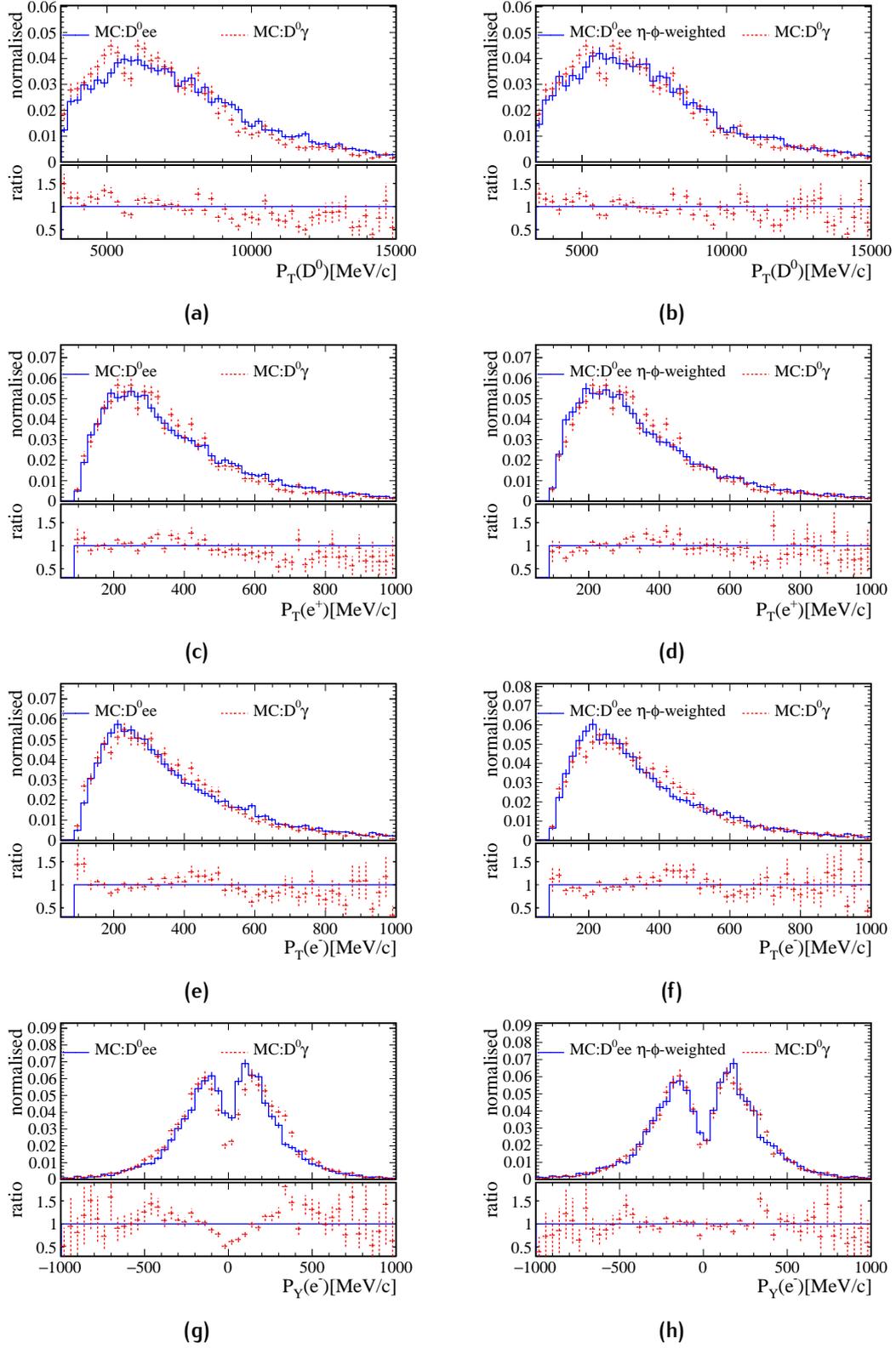


Figure C.6: Comparison between simulated control channel distributions (dashed red) and simulated signal (solid blue) shown in the left column. The right column shows the distributions where signal is re-weighted in η and ϕ of the electrons.

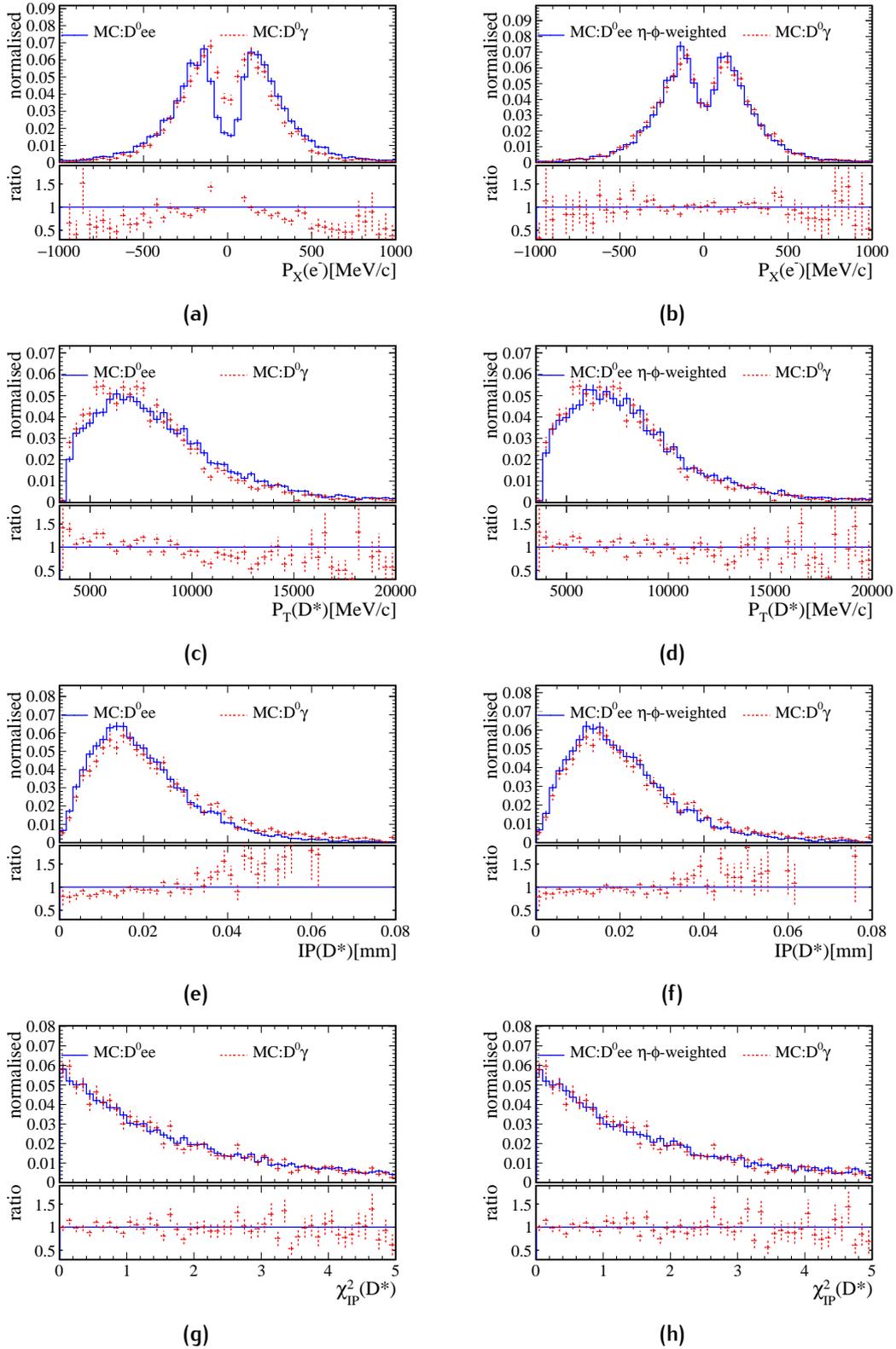


Figure C.7: Comparison between simulated control channel distributions (dashed red) and simulated signal (solid blue) shown in the left column. The right column shows the distributions where signal is re-weighted in η and ϕ of the electrons.

C.3 RE-WEIGHTING OF SIGNAL CHANNEL SIMULATION

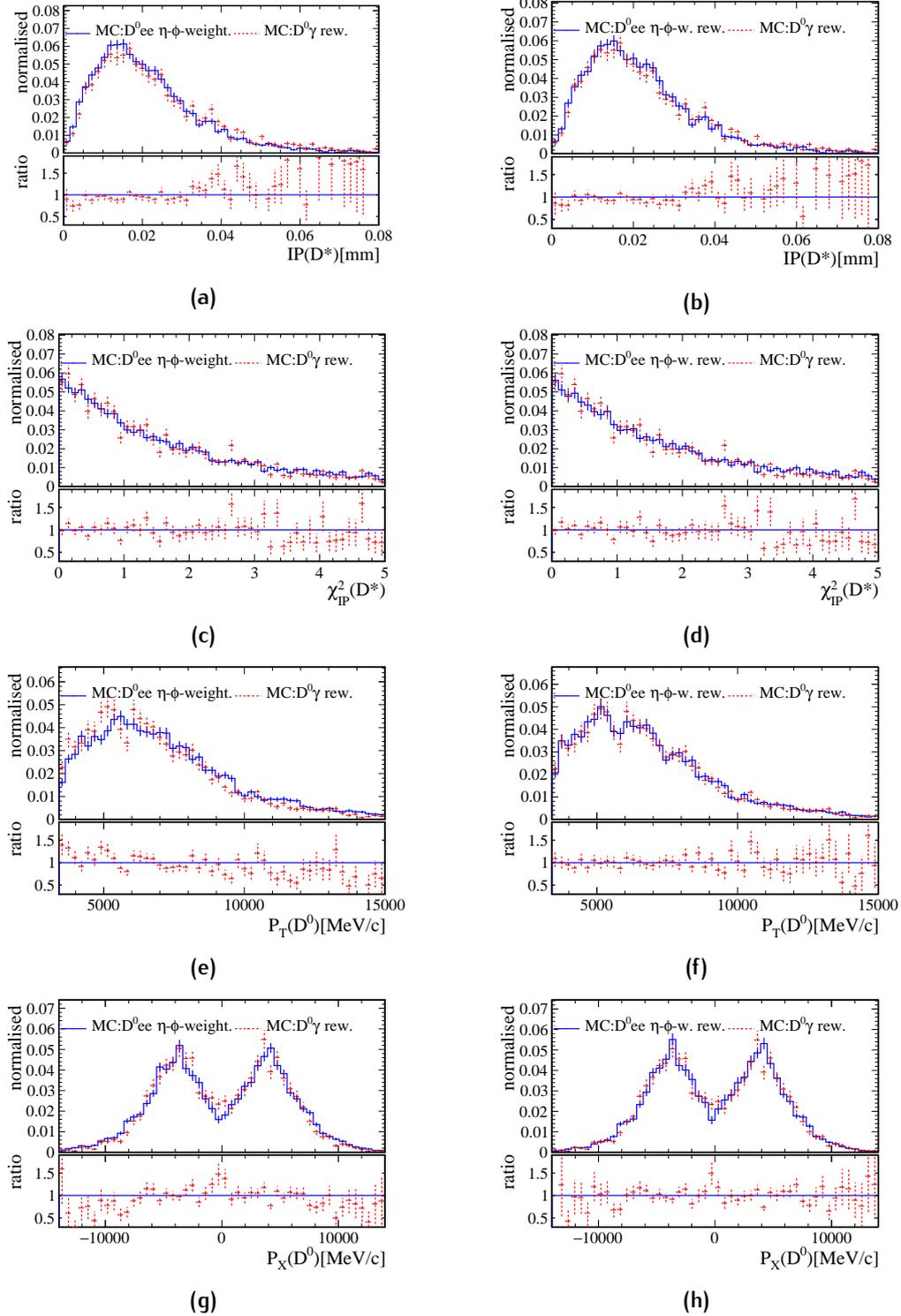


Figure C.8: Comparison between simulation of signal re-weighted in η and ϕ of the electrons (blue solid line) and corrected control channel (dashed red) (left column). Result of the additional signal simulation re-weighting in $p_T(D^{*0})$ and $p_T(e^\pm)$ to match corrected control channel distributions (right column).

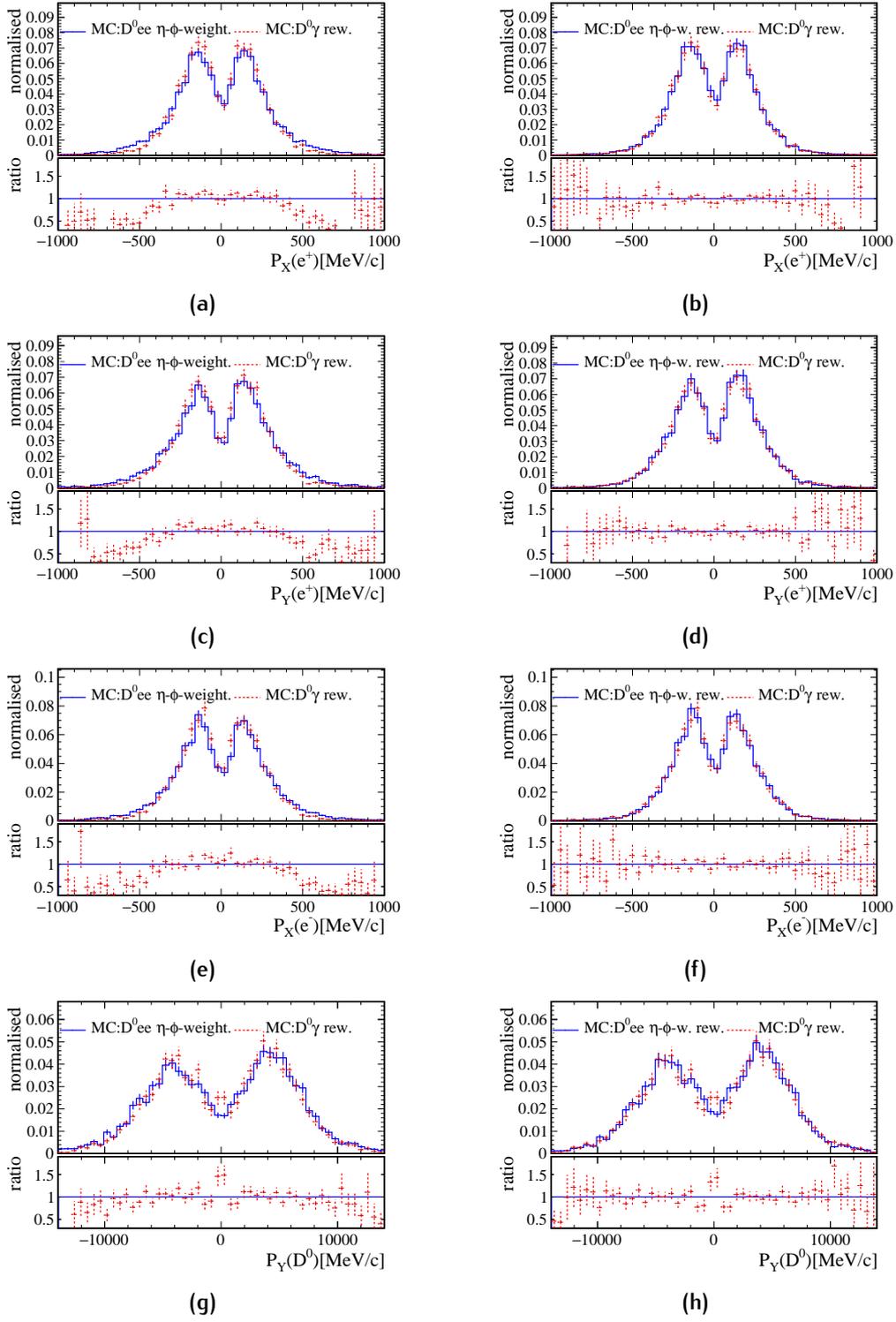


Figure C.9: Comparison between simulation of signal re-weighted in η and ϕ of the electrons (blue solid line) and corrected control channel (dashed red) (left column). Result of the additional signal simulation re-weighting in $p_T(D^{*0})$ and $p_T(e^\pm)$ to match corrected control channel distributions (right column).

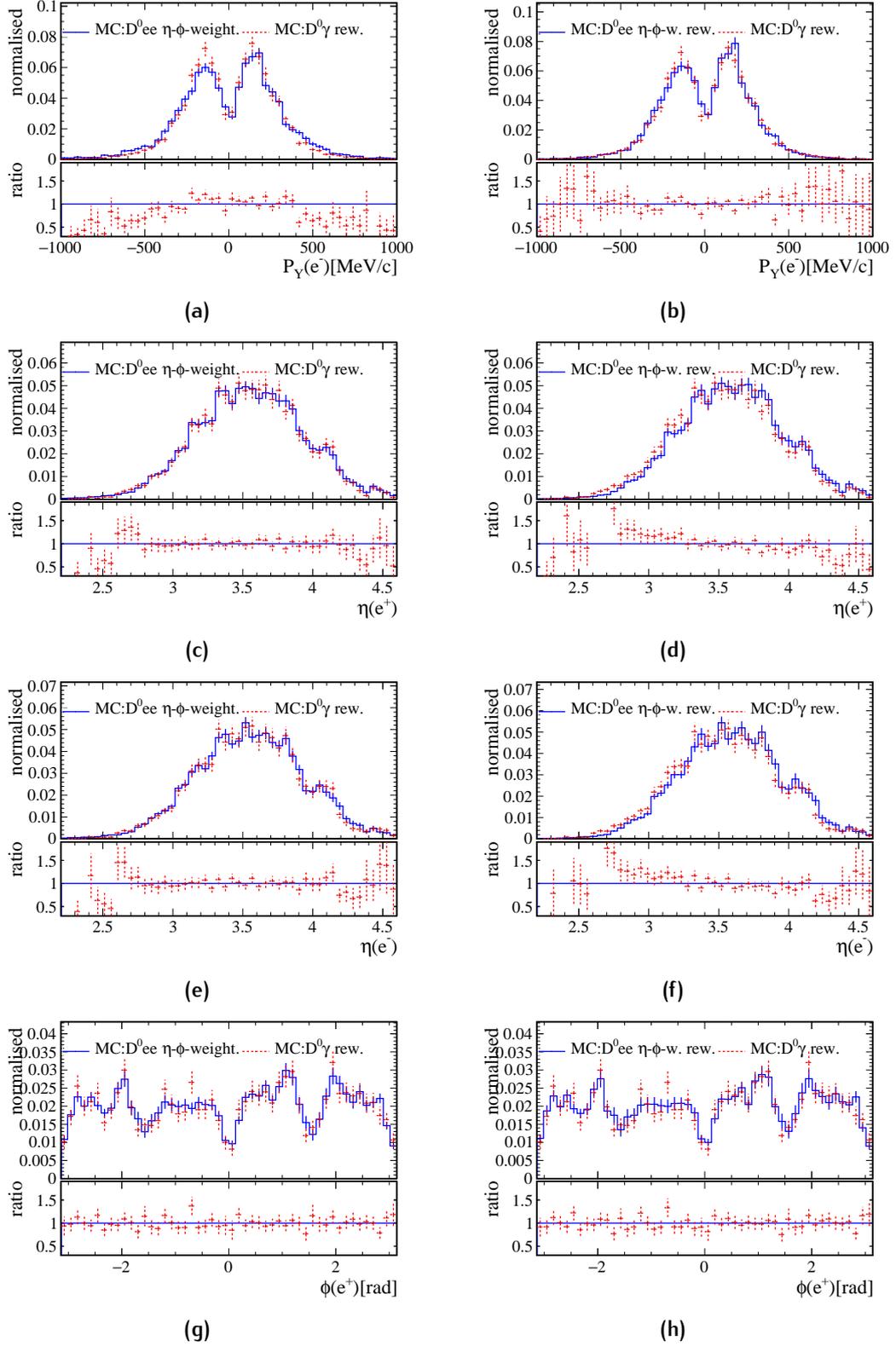


Figure C.10: Comparison between simulation of signal re-weighted in η and ϕ of the electrons (blue solid line) and corrected control channel (dashed red) (left column). Result of the additional signal simulation re-weighting in $p_T(D^{*0})$ and $p_T(e^\pm)$ to match corrected control channel distributions (right column).

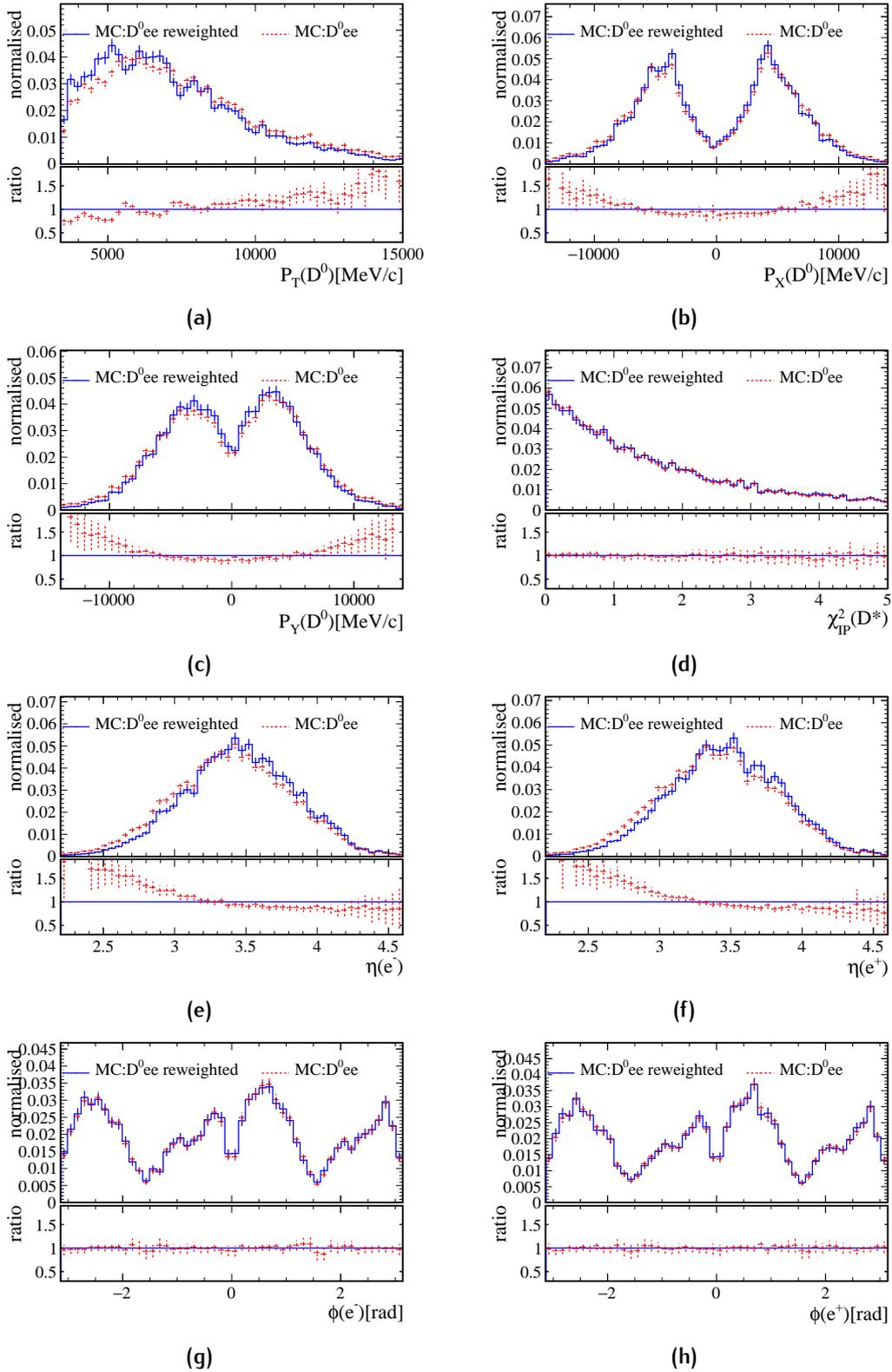


Figure C.11: Result of the re-weighting in $p_T(D^{*0})$ and $p_T(e^\pm)$ without η - ϕ -weights. The re-weighted signal simulation is shown as solid blue line, the original uncorrected simulation is shown as dashed red points.

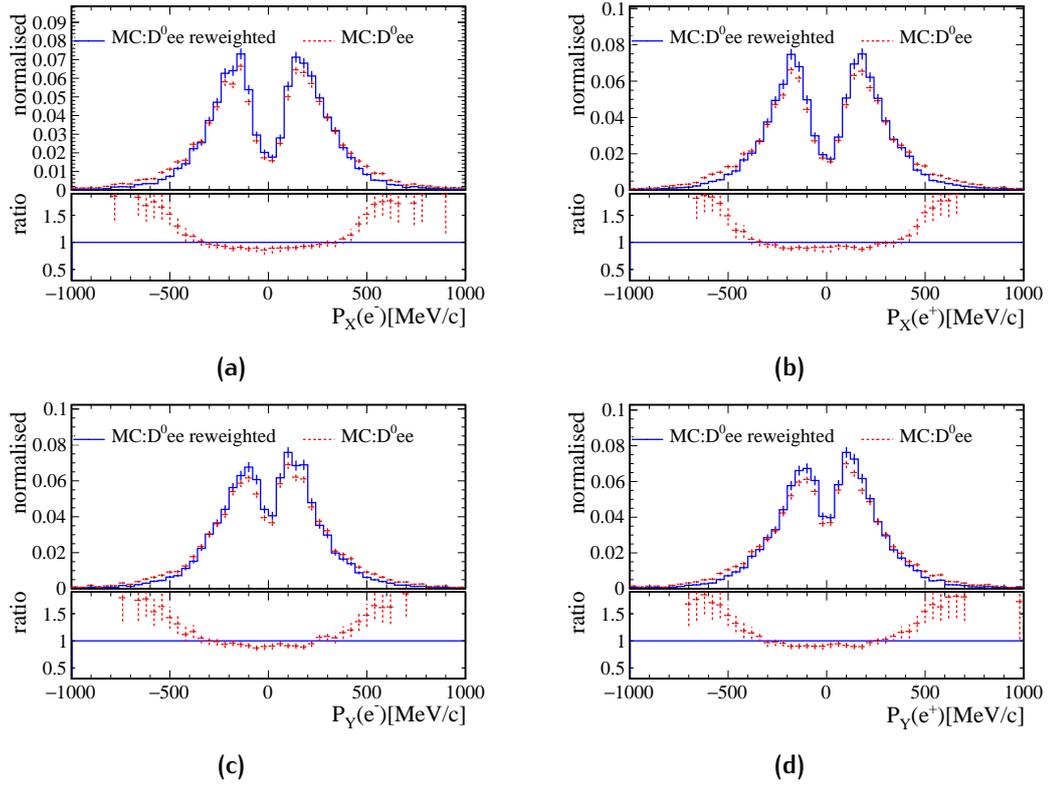


Figure C.12: Result of the re-weighting in $p_T(D^{*0})$ and $p_T(e^\pm)$ without η - ϕ -weights. The re-weighted signal simulation is shown as solid blue line, the original uncorrected simulation is shown as dashed red points.

D

MVA CONTROL PLOTS

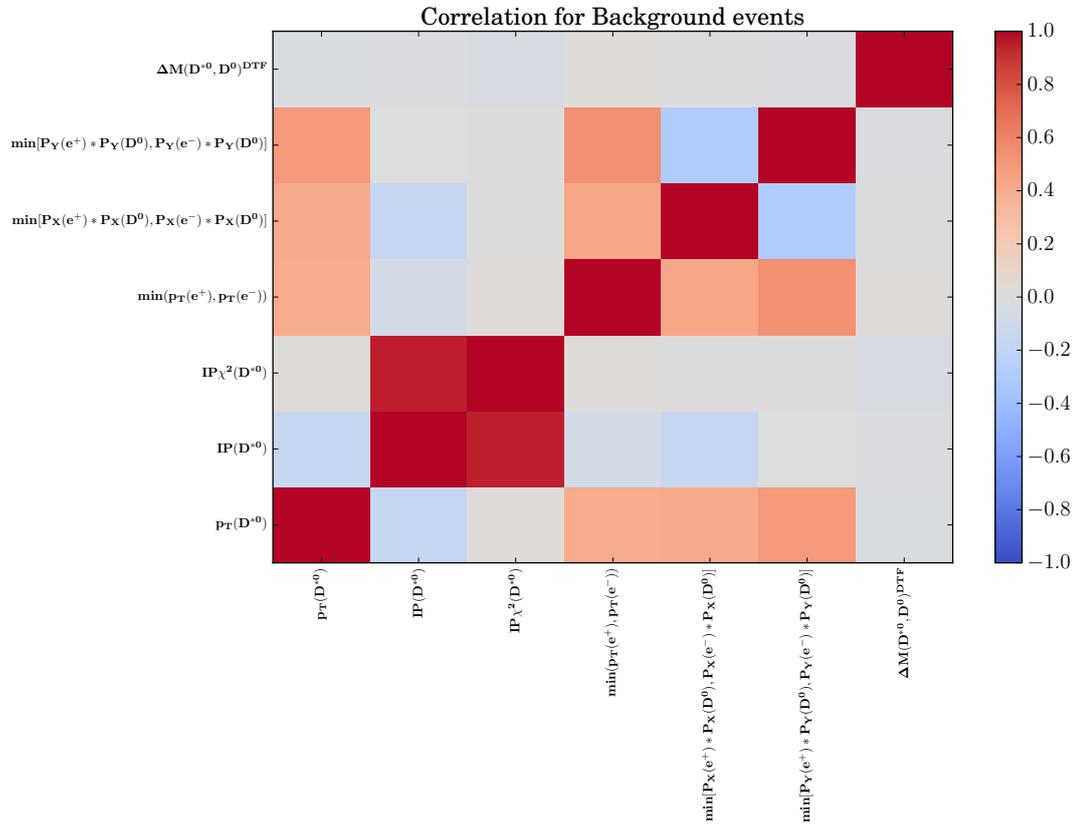


Figure D.1: Linear correlation of input variables and Δm_{DTF} obtained from upper sideband of Δm_{DTF} .

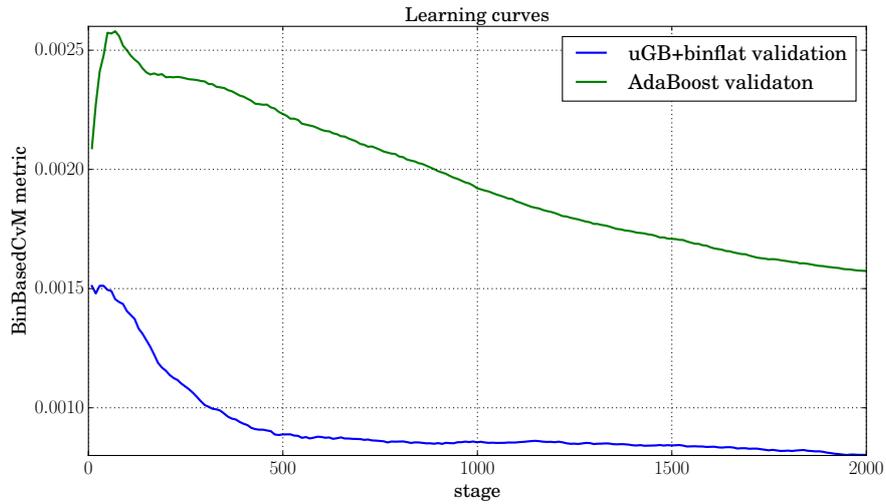


Figure D.2: Cramer-von Mises (CvM) similarity between distributions as a function of number of trained decision trees. This metric gives an estimate on how flat the signal efficiency is in the Δm_{DTF} variable for both, gradient boosting with AdaLoss (green) and flatness loss (blue). Smaller values of this metric correspond to better flatness.

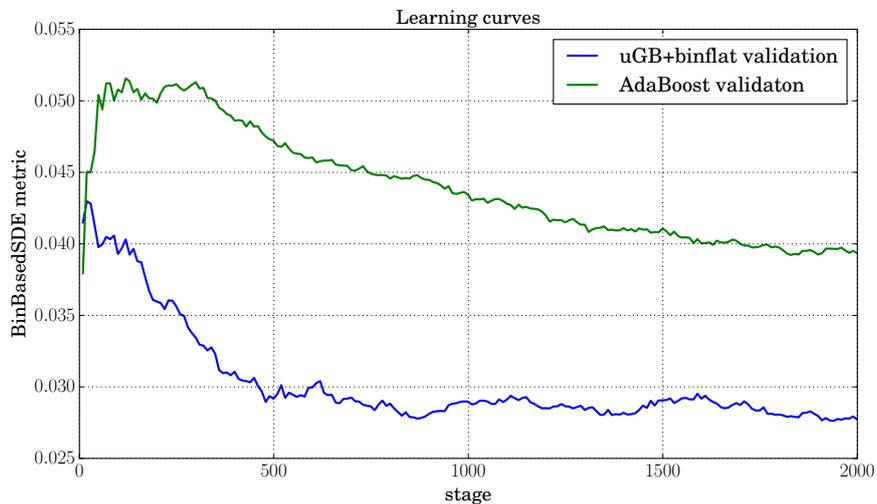


Figure D.3: Standard deviation of efficiency (SDE) between distributions as a function of number of trained decision trees. This metric gives an estimate on how flat the signal efficiency is in the Δm_{DTF} variable for both, gradient boosting with AdaLoss (green) and flatness loss (blue). Smaller values of this metric correspond to better flatness.

E

FIT EXAMPLE FOR BDT OPTIMISATION

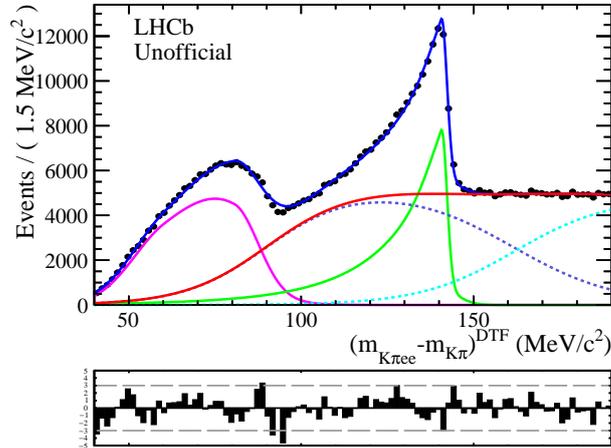


Figure E.1: Fit of the $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+e^-) \Delta m_{\text{DTF}}$ distribution obtained from data after requiring $\text{BDT}_{\text{response}} > 0.45$ and $m(e^+e^-) < 40 \text{ MeV}/c^2$. The fit model is the one described in Eq. 6.7 except for the upper part of the combinatorial background which is not fitted here.

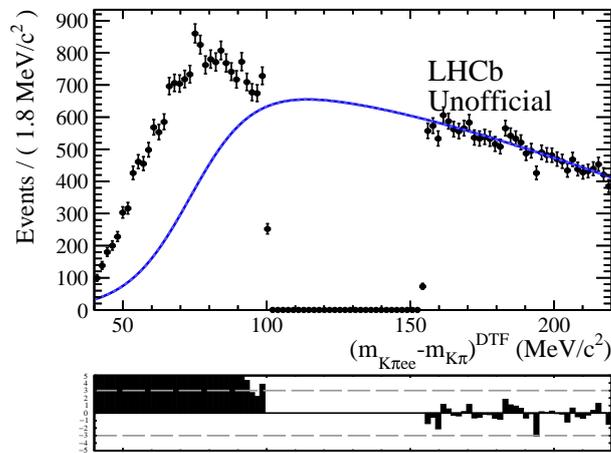
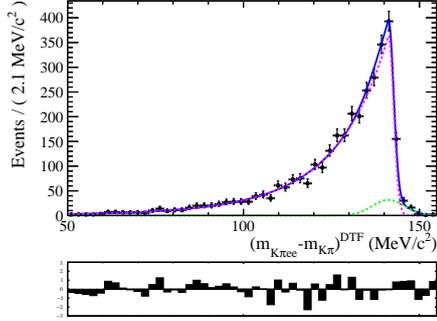


Figure E.2: Fit of the $D^{*0} \rightarrow D^0 e^+e^- \Delta m_{\text{DTF}}$ combinatorial background distribution obtained from data after requiring $\text{BDT}_{\text{response}} > 0.45$ and $m(e^+e^-) < 40 \text{ MeV}/c^2$. The fit model is the one described in Eq. 6.5 except for the upper part of the combinatorial background which is not fitted here. The left-flank part of the model is fixed to the shape obtained from a 2016 same-sign sample of $D^{*0} \rightarrow D^0 e^+e^-$ candidates.

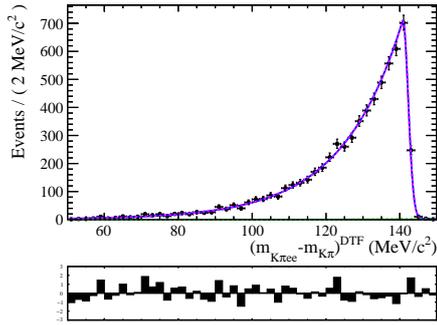
F

FINAL SIMULATION FITS



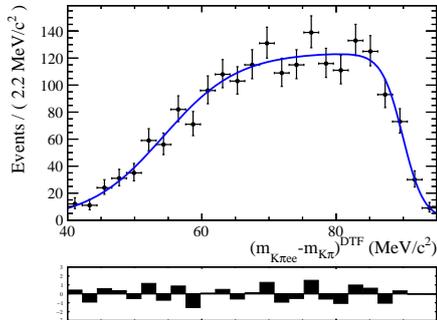
Parameter	Value
α	0.10 ± 0.02
n	11 ± 4
μ_{CB}	141.3 ± 0.3
σ_{CB}	1.4 ± 0.2
μ_G	μ_{CB}
σ_G	4.6 ± 0.6
f_{con}	0.95 ± 0.01

Figure F.1: Fit of simulated $D^{*0} \rightarrow D^0(\gamma \rightarrow e^+e^-)$ after all selection cuts and applied signal channel veto (left). Parameter table (right) of the fitted model defined in 6.3.



Parameter	Value
α	0.087 ± 0.006
n	100
μ_{CB}	140.8 ± 0.1
σ_{CB}	1.43 ± 0.09
μ_G	μ_{CB}
σ_G	10 ± 5
f_{con}	0.997 ± 0.002

Figure F.2: Fit of simulated $D^{*0} \rightarrow D^0 e^+ e^-$ after all selection cuts and applied control channel veto (left). Parameter table (right) of the fitted model defined in 6.3.



Parameter	Value
μ_1	54.3 ± 0.8
k_1	5.6 ± 0.5
μ_2	89.9 ± 0.3
k_2	-1.6 ± 0.2

Figure F.3: Fit of simulated $D^{*0} \rightarrow D^0[\chi e^+e^-]_{\tau^0}$ after all selection cuts and applied control channel veto (left). Parameter table (right) of the fitted model defined in Eq. 7.8.

BIBLIOGRAPHY

- [1] P. J. E. Peebles. "Growth of the nonbaryonic dark matter theory." In: *Nature Astronomy* 1, 0057 (Mar. 2017), p. 0057. arXiv:[1701.05837](#).
- [2] Philip Ilten et al. "Dark photons from charm mesons at LHCb." In: *Phys. Rev. D* 92.11 (2015), p. 115017. arXiv:[1509.06765 \[hep-ph\]](#).
- [3] Paul Langacker. *The standard model and beyond*. eng. Second edition. Series in high energy physics, cosmology and gravitation. Boca Raton ; London ; New York: CRC Press, Taylor & Francis Group, 2017, xiii, 636 Seiten.
- [4] Michael Edward Peskin and Daniel V. Schroeder. *An introduction to quantum field theory*. eng. 1. Indian ed. Kolkata: Levant Books, 2005, XXII, 842 S.
- [5] Mark Thomson. *Modern particle physics*. eng. Cambridge [u.a.]: Cambridge University Press, 2013, XVI, 554 S.
- [6] Richard Phillips Feynman. *QED. the strange theory of light and matter*. Expanded ed. Princeton science library. Princeton, NJ [u.a.]: Princeton Univ. Press, 2006, XXIV, 158 S.
- [7] M. Tanabashi et al. "Review of Particle Physics." In: *Phys. Rev. D* 98 (3 2018). (Particle Data Group), p. 030001.
- [8] John Ellis, Mary K. Gaillard, and Dimitri V. Nanopoulos. "A Historical Profile of the Higgs Boson." In: *The standard theory of particle physics: Essays to celebrate CERN's 60th anniversary*. Ed. by Luciano Maiani and Luigi Rolandi. 2016, pp. 255–274. arXiv:[1504.07217 \[hep-ph\]](#).
- [9] L.G. Landsberg. "Electromagnetic decays of light mesons." In: *Physics Reports* 128.6 (1985), pp. 301–376.
- [10] N. Aghanim et al. "Planck 2018 results. VI. Cosmological parameters." In: (2018). arXiv:[1807.06209 \[astro-ph.CO\]](#).
- [11] Steven Weinberg. "The cosmological constant problem." In: *Rev. Mod. Phys.* 61 (1 1989), pp. 1–23.
- [12] Jim Alexander et al. "Dark Sectors 2016 Workshop: Community Report." In: 2016. arXiv:[1608.08632 \[hep-ph\]](#).
- [13] Marco Battaglieri et al. "US Cosmic Visions: New Ideas in Dark Matter 2017: Community Report." In: (2017). arXiv:[1707.04591 \[hep-ph\]](#).
- [14] Lotty Ackerman et al. "Dark Matter and Dark Radiation." In: *Phys. Rev. D* 79 (2009). [277(2008)], p. 023519. arXiv:[0810.5126 \[hep-ph\]](#).
- [15] Rouven Essig et al. "Working Group Report: New Light Weakly Coupled Particles." In: *Proceedings, 2013 Community Summer Study on the Future of U.S. Particle Physics: Snowmass on the Mississippi (CSS2013): Minneapolis, MN, USA, July 29–August 6, 2013*. 2013. arXiv:[1311.0029 \[hep-ph\]](#).
- [16] Bob Holdom. "Two U(1)'s and Epsilon Charge Shifts." In: *Phys. Lett.* 166B (1986), pp. 196–198.

- [17] Motoi Endo, Koichi Hamaguchi, and Go Mishima. “Constraints on Hidden Photon Models from Electron $g-2$ and Hydrogen Spectroscopy.” In: *Phys. Rev. D* 86 (2012), p. 095029. arXiv:[1209.2558 \[hep-ph\]](#).
- [18] Hooman Davoudiasl, Hye-Sung Lee, and William J. Marciano. “Dark Side of Higgs Diphoton Decays and Muon $g-2$.” In: *Phys. Rev. D* 86 (2012), p. 095009. arXiv:[1208.2973 \[hep-ph\]](#).
- [19] Nima Arkani-Hamed et al. “A Theory of Dark Matter.” In: *Phys. Rev. D* 79 (2009), p. 015014. arXiv:[0810.0713 \[hep-ph\]](#).
- [20] J. D. Bjorken et al. “Search for Neutral Metastable Penetrating Particles Produced in the SLAC Beam Dump.” In: *Phys. Rev. D* 38 (1988), p. 3375.
- [21] E. M. Riordan et al. “A Search for Short Lived Axions in an Electron Beam Dump Experiment.” In: *Phys. Rev. Lett.* 59 (1987), p. 755.
- [22] A. Bross et al. “A Search for Short-lived Particles Produced in an Electron Beam Dump.” In: *Phys. Rev. Lett.* 67 (1991), pp. 2942–2945.
- [23] A. Konaka et al. “Search for Neutral Particles in Electron Beam Dump Experiment.” In: *Phys. Rev. Lett.* 57 (1986), p. 659.
- [24] M. Davier and H. Nguyen Ngoc. “An Unambiguous Search for a Light Higgs Boson.” In: *Phys. Lett.* B229 (1989), pp. 150–155.
- [25] H. Merkel et al. “Search at the Mainz Microtron for Light Massive Gauge Bosons Relevant for the Muon $g-2$ Anomaly.” In: *Phys. Rev. Lett.* 112.22 (2014), p. 221802. arXiv:[1404.5502 \[hep-ex\]](#).
- [26] S. Abrahamyan et al. “Search for a New Gauge Boson in Electron-Nucleus Fixed-Target Scattering by the APEX Experiment.” In: *Phys. Rev. Lett.* 107 (2011), p. 191804. arXiv:[1108.2750 \[hep-ex\]](#).
- [27] D. Banerjee et al. “Search for a new X(16.7) boson and dark photons in the NA64 experiment at CERN.” In: (2018). arXiv:[1803.07748 \[hep-ex\]](#).
- [28] J. Blumlein et al. “Limits on the mass of light (pseudo)scalar particles from Bethe-Heitler e^+e^- and $\mu^+\mu^-$ pair production in a proton - iron beam dump experiment.” In: *Int. J. Mod. Phys. A* 7 (1992), pp. 3835–3850.
- [29] P. H. Adrian et al. “Search for a Dark Photon in Electro-Produced e^+e^- Pairs with the Heavy Photon Search Experiment at JLab.” In: (2018). arXiv:[1807.11530 \[hep-ex\]](#).
- [30] J. P. Lees et al. “Search for a Dark Photon in e^+e^- Collisions at BaBar.” In: *Phys. Rev. Lett.* 113.20 (2014), p. 201801. arXiv:[1406.2980 \[hep-ex\]](#).
- [31] M. Ablikim et al. “Dark Photon Search in the Mass Range Between 1.5 and 3.4 GeV/c^2 .” In: *Phys. Lett.* B774 (2017), pp. 252–257. arXiv:[1705.04265 \[hep-ex\]](#).
- [32] A. Anastasi et al. “Limit on the production of a new vector boson in $e^+e^- \rightarrow U\gamma$, $U \rightarrow \pi^+\pi^-$ with the KLOE experiment.” In: *Phys. Lett.* B757 (2016), pp. 356–361. arXiv:[1603.06086 \[hep-ex\]](#).
- [33] I. Jaegle. “Search for the dark photon and the dark Higgs boson at Belle.” In: *Phys. Rev. Lett.* 114.21 (2015), p. 211801. arXiv:[1502.00084 \[hep-ex\]](#).
- [34] Roel Aaij et al. “Search for Dark Photons Produced in 13 TeV pp Collisions.” In: *Phys. Rev. Lett.* 120.6 (2018), p. 061801. arXiv:[1710.02867 \[hep-ex\]](#).

- [35] J. Blumlein et al. "Limits on neutral light scalar and pseudoscalar particles in a proton beam dump experiment." In: *Z. Phys. C* 51 (1991), pp. 341–350.
- [36] P. Astier et al. "Search for heavy neutrinos mixing with tau neutrinos." In: *Phys. Lett. B* 506 (2001), pp. 27–38. arXiv:[hep-ex/0101041](https://arxiv.org/abs/hep-ex/0101041) [[hep-ex](#)].
- [37] G. Bernardi et al. "Search for Neutrino Decay." In: *Phys. Lett.* 166B (1986), pp. 479–483.
- [38] D. Babusci et al. "Limit on the production of a light vector gauge boson in phi meson decays with the KLOE detector." In: *Phys. Lett.* B720 (2013), pp. 111–115. arXiv:[1210.3927](https://arxiv.org/abs/1210.3927) [[hep-ex](#)].
- [39] Philip Ilten et al. "Serendipity in dark photon searches." In: *JHEP* 06 (2018), p. 004. arXiv:[1801.04847](https://arxiv.org/abs/1801.04847) [[hep-ph](#)].
- [40] Lyndon Evans and Philip Bryant. "LHC Machine." In: *JINST* 3 (2008), S08001.
- [41] A. A. Alves Jr. et al. "The LHCb detector at the LHC." In: *JINST* 3 (2008), S08005.
- [42] R. Aaij et al. "LHCb detector performance." In: *Int. J. Mod. Phys. A* 30 (2015), p. 1530022. arXiv:[1412.6352](https://arxiv.org/abs/1412.6352) [[hep-ex](#)].
- [43] CERN. *Accelerator Performance and Statistics*. 2018. URL: <http://acc-stats.web.cern.ch/acc-stats/#lhc/overview-panel> (visited on 08/12/2018).
- [44] *The accelerator complex*. 2012. URL: <http://cds.cern.ch/record/1997193>.
- [45] Christiane Lefèvre. "The CERN accelerator complex. Complexe des accélérateurs du CERN." 2008.
- [46] M. Alexander et al. "Mapping the material in the LHCb vertex locator using secondary hadronic interactions." In: (2018). arXiv:[1803.07466](https://arxiv.org/abs/1803.07466) [[physics.ins-det](#)].
- [47] R. Aaij et al. "Performance of the LHCb Vertex Locator." In: *JINST* 9 (2014), P09007. arXiv:[1405.7808](https://arxiv.org/abs/1405.7808) [[physics.ins-det](#)].
- [48] Wouter D. Hulsbergen. "Decay chain fitting with a Kalman filter." In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 552.3 (2005), pp. 566–575.
- [49] R. Aaij et al. "Measurement of the track reconstruction efficiency at LHCb." In: *JINST* 10 (2015), P02007. arXiv:[1408.1251](https://arxiv.org/abs/1408.1251) [[hep-ex](#)].
- [50] A. Papanestis and C. D'Ambrosio. "Performance of the LHCb RICH detectors during the LHC Run II." In: *Nucl. Instrum. Meth.* A876 (2017), pp. 221–224. arXiv:[1703.08152](https://arxiv.org/abs/1703.08152) [[physics.ins-det](#)].
- [51] R. Aaij et al. "Selection and processing of calibration samples to measure the particle identification performance of the LHCb experiment in Run 2." In: (2018). arXiv:[1803.00824](https://arxiv.org/abs/1803.00824) [[hep-ex](#)].
- [52] J. Neyman and E.S. Pearson. "IX. On the problem of the most efficient tests of statistical hypotheses." In: *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 231.694–706 (1933), pp. 289–337. eprint: <http://rsta.royalsocietypublishing.org/content/231/694-706/289.full.pdf>.
- [53] François Chollet et al. *Keras*. <https://keras.io>. 2015.
- [54] Anna Veronika Dorogush et al. "Fighting biases with dynamic boosting." In: *CoRR* abs/1706.09516 (2017). arXiv:[1706.09516](https://arxiv.org/abs/1706.09516).

- [55] Mikhail Hushchyn. "Machine Learning based Global Particle Identification Algorithms at the LHCb Experiment." CERN-Poster-2018-647. 2018.
- [56] R. Aaij et al. "The LHCb trigger and its performance in 2011." In: *JINST* 8 (2013), P04022. arXiv:[1211.3055 \[hep-ex\]](#).
- [57] Sean Benson et al. "The LHCb Turbo Stream." In: *Journal of Physics: Conference Series* 664.8 (2015), p. 082004.
- [58] R. Aaij et al. "Tesla: an application for real-time data analysis in High Energy Physics." In: *Comput. Phys. Commun.* 208 (2016), pp. 35–42. arXiv:[1604.05596 \[physics.ins-det\]](#).
- [59] *LHCb Starterkit Lessons*. <https://lhcb.github.io/starterkit-lessons/first-analysis-steps/run-2-data-flow.html>. 2018.
- [60] M. Ablikim et al. "Precision measurement of the D^{*0} decay branching fractions." In: *Phys. Rev. D* 91.3 (2015), p. 031101. arXiv:[1412.4566 \[hep-ex\]](#).
- [61] Glen Cowan. *Statistical data analysis. [with applications from particle physics]*. Oxford science publications. Oxford: Clarendon Press, 1998, XIV, 197 S.
- [62] Muriel Pivk and Francois R. Le Diberder. "sPlot: A statistical tool to unfold data distributions." In: *Nucl. Instrum. Meth.* A555 (2005), pp. 356–369. arXiv:[physics/0402083 \[physics.data-an\]](#).
- [63] C Peterson and T S Rognvaldsson. "An introduction to artificial neural networks." In: LU-TP-91-23. LUTP-91-23 (1991), 74 p.
- [64] A. Hoecker et al. "TMVA – Toolkit for Multivariate Data Analysis." In: *ArXiv Physics e-prints* (Mar. 2007). eprint: [physics/0703039](#).
- [65] P. Mehta et al. "A high-bias, low-variance introduction to Machine Learning for physicists." In: *ArXiv e-prints* (Mar. 2018). arXiv:[1803.08823](#).
- [66] Jerome H. Friedman. "Greedy Function Approximation: A Gradient Boosting Machine." In: *The Annals of Statistics* 29.5 (2001), pp. 1189–1232.
- [67] Yoav Freund and Robert E. Schapire. "A decision-theoretic generalization of on-line learning and an application to boosting." In: *J. Comput. Syst. Sci.* 55 (1997), p. 119.
- [68] Alex Rogozhnikov et al. "New approaches for boosting to uniformity." In: *JINST* 10.03 (2015), T03002. arXiv:[1410.4140 \[hep-ex\]](#).
- [69] A. Rogozhnikov. "Reweighting with Boosted Decision Trees." In: *J. Phys. Conf. Ser.* 762.1 (2016). https://github.com/arogozhnikov/hep_ml. arXiv:[1608.05806 \[physics.data-an\]](#).
- [70] Michel De Cian. *Implementation of a Conversion Filter in the Gauss application*. private communication. 2018.
- [71] Dominik Müller and Benedetto Gianluca Siddi. "Fast simulation options in LHCb from ReDecay to fully parametrised." CERN-Poster-2017-580. 2017.
- [72] Dominik Müller. *Splitting the simulation into multiple steps*. private communication. 2018.
- [73] Wouter D. Hulsbergen. "Decay chain fitting with a Kalman filter." In: *Nucl. Instrum. Meth.* A552 (2005), pp. 566–575. arXiv:[physics/0503191 \[physics\]](#).

- [74] Tomasz Skwarnicki. "A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances." [DESY-F31-86-02](#). PhD thesis. Institute of Nuclear Physics, Krakow, 1986.
- [75] R. Aaij et al. "Measurement of the relative rate of prompt χ_{c0} , χ_{c1} and χ_{c2} production at $\sqrt{s} = 7\text{TeV}$." In: *JHEP* 10 (2013), p. 115. arXiv:[1307.4285 \[hep-ex\]](#).
- [76] Avrim Blum, Adam Kalai, and John 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 '99. Santa Cruz, California, USA: ACM, 1999, pp. 203–208.
- [77] Bernard Aubert et al. "Measurement of the branching ratios $\Gamma(D_s^{*+} \rightarrow D_s^+ \pi^0)/\Gamma(D_s^{*+} \rightarrow D_s^+ \gamma)$ and $\Gamma(D^{*0} \rightarrow D^0 \pi^0)/\Gamma(D^{*0} \rightarrow D^0 \gamma)$." In: *Phys. Rev. D* 72 (2005), p. 091101. arXiv:[hep-ex/0508039 \[hep-ex\]](#).
- [78] S. S. Wilks. "The large-sample distribution of the likelihood ratio for testing composite hypotheses." In: *Ann. Math.Stat.* 9 (1938), pp. 60–62.
- [79] *LHCb Trigger and Online Technical Design Report*. CERN-LHCC-2014-016. Geneva: CERN, 2014.
- [80] *LHCb VELO Upgrade Technical Design Report*. CERN-LHCC-2013-021. Geneva: CERN, 2013.
- [81] S. Agostinelli et al. "Geant4: A simulation toolkit." In: *Nucl. Instrum. Meth.* A506 (2003), p. 250.
- [82] John Allison et al. "Geant4 developments and applications." In: *IEEE Trans.Nucl.Sci.* 53 (2006), p. 270.
- [83] A. Borsellino. "Momentum Transfer and Angle of Divergence of Pairs Produced by Photons." In: *Phys. Rev.* 89 (5 1953), pp. 1023–1025.

ACKNOWLEDGEMENTS

After roughly one year of work to create this thesis, I want to thank the many people who supported me and always gave me good advice when it comes to hard or sometimes stupid physics and life questions. In particular, I want to thank my supervisor Stephanie Hansmann-Menzemer who already supported me in my time as a Bachelor student and again gave me the opportunity to join the LHCb collaboration to work on a cool project. I am also grateful to Klaus Reyggers who, after my Bachelor thesis, again agreed to be my second referee.

Special thanks (also again) go to Michel De Cian, from who I learned most of the stuff I believe to know about data analysis at LHCb and particle physics in general and who always had the patience to explain and think through issues, even over the long distance to Lausanne. I also want to thank the whole LHCb Heidelberg group for taking me in and always having an open ear for my questions.

The most special thanks go to my family, my friends and my Schnatz who endured my moods during harder times and always put my feet back on the ground.

G | ERKLÄRUNG

Ich versichere, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den