TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

DIPLOMARBEIT

Search for an invisibly decaying Z' boson and study of particle identification at the Belle II experiment

ausgeführt an der
Fakultät für Physik der Technischen Universität Wien
in Zusammenarbeit mit dem
Institut für Hochenergiephysik der Österreichischen Akademie der Wissenschaften

zur Erlangung des akademischen Grades
Diplom-Ingenieur

im Rahmen des Studiums
Technische Physik

eingereicht von

Paul Feichtinger, BSc
Matrikelnummer: 01425593

Betreuung: Privatdoz. Dipl.-Ing. Dr.techn. Christoph Schwanda
Mitbetreuung: Dr. Gianluca Inguglia

Wien, am 5. Februar 2021

Unterschrift des Betreuers

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub

Kurzfassung

Diese Arbeit handelt von der Suche nach einem hypothetischen Z' Vektorboson in Elektron-Positron Kollisionen ($e^+e^- \to \mu^+\mu^-$ ($Z' \to$ invisible)) am Belle II Detektor im KEK Forschungszentrum für Hochenergiephysik in Tsukuba, Japan. Das Z' ergibt sich aus einer Erweiterung des Standardmodells mit einer abelschen Eichsymmetrie, basierend auf der $L_\mu - L_\tau$ Zahl. Daraus resultiert eine Wechselwirkung an der nur Leptonen der zweiten und dritten Generation teilnehmen, weshalb es von vielen Experimenten, die nur die erste Generation betrachten, nicht wahrgenommen werden kann. Dieses neue Teilchen könnte die Unstimmigkeiten bezüglich des anomalen magnetischen Dipolmoments des Muons und die von LHCb gemessenen Diskrepanzen in $b \to s\mu^+\mu^-$ Prozessen erklären. Außerdem könnte es als Vermittler einer neuen Kraft Mechanismen zur Produktion von Dunkler Materie ermöglichen.

Nach einer anfänglichen Trennung von Signal- und Hintergrundprozessen wird ein künstliches neuronales Netz für weitere Klassifizierung verwendet. Das Netz wird mit einer speziellen Kostenfunktion trainiert, die in direktem Zusammenhang zur Sensitivität der Suche steht. Basierend auf diesen Ergebnissen werden die zu erwartenden 90 % Konfidenzintervalle der Kopplungsstärke eines Z' mit einer Masse bis zu 9 GeV/c^2 berechnet.

Weiters wird die Teilchenidentifikation von Belle II mit Pionen von dem Zerfall eines τ-Leptons im Prozess $e^+e^- \to \tau^+\tau^-$ untersucht. Damit kann die Wahrscheinlichkeit, dass ein Pion falsch als Elektron oder Muon identifiziert wird, zwischen simulierten und real aufgenommenen Daten verglichen werden. Daraus wird ein Korrekturfaktor berechnet, um etwaige Diskrepanzen in der Simulation auszugleichen.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Abstract

The subject of this thesis is the search for a hypothetical Z' vector boson in events with two muons plus missing energy in the final state ($e^+e^- \to \mu^+\mu^-$ ($Z' \to$ invisible)) at the Belle II detector, located at the SuperKEKB accelerator in Tsukuba, Japan. The Z' arises through a natural extension of the Standard Model gauge group with a gauged $L_\mu - L_\tau$ number. The resulting abelian interaction only couples to leptons of the second and third generation, thus hiding from most experiments only sensitive to first-generation leptons. This new particle could explain the discrepancies concerning the anomalous magnetic dipole moment of the muon, the tensions in $b \to s\mu^+\mu^-$ transitions observed by LHCb, and act as a mediator between the interactions of the Standard Model and the Dark Sector.

After an initial separation of signal from background, an artificial neural network (ANN) is used for additional classification. It is optimised directly on a figure of merit designed to maximise the sensitivity of the search. Based on these results, the expected 90% confidence intervals on the coupling constant for a Z' with a mass up to 9 GeV/c^2 are calculated.

Furthermore, the particle identification of Belle II is studied with pions originating from one of the τ leptons in $e^+e^- \to \tau^+\tau^-$ events. This pion sample is then used to estimate the probability that pions are misidentified as electrons or muons. This probability is compared between simulation and collision data, and a correction factor is calculated, which can be used to account for discrepancies in the simulation.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table of Contents

Abbreviations **vii**

1 Introduction **1**

2 Theoretical overview **3**

2.1 The Standard Model of particle physics . . . 3

2.2 Dark matter . . . 5

2.3 The $L_\mu - L_\tau$ gauge boson Z' . . . 7

2.3.1 The invisibly decaying Z' . . . 8

2.3.2 Detection at Belle II . . . 9

2.4 Statistical methods . . . 10

2.4.1 Inferential statistics . . . 10

2.4.2 Parameter estimation . . . 11

2.4.3 Goodness-of-fit . . . 13

2.4.4 Nuisance parameters . . . 13

2.4.5 Hypothesis tests and confidence intervals . . . 13

2.4.6 Look elsewhere effect . . . 16

2.4.7 Optimising the search region . . . 17

2.4.8 Bayesian statistics . . . 18

2.5 Artificial neural networks . . . 20

3 Experimental setup **24**

3.1 Luminosity . . . 24

3.2 SuperKEKB . . . 25

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

3.3 The Belle II experiment . . . 26
3.3.1 Subdetectors . . . 29
3.3.2 Beam induced backgrounds . . . 33
3.3.3 Trigger system . . . 34
3.3.4 Data taking periods . . . 35
3.4 Software . . . 36
3.5 Simulation . . . 37

4 Particle identification studies **38**
4.1 Methodology . . . 38
4.2 Particle identification variables . . . 38
4.3 Monte Carlo and data samples . . . 40
4.4 Event selection . . . 41
4.5 Calculation of efficiencies . . . 47
4.6 Results . . . 51

5 Analysis procedure **57**
5.1 Monte Carlo samples . . . 57
5.1.1 Background samples . . . 57
5.1.2 Signal samples . . . 59
5.2 Event selection . . . 59
5.3 Multivariate selection using an ANN . . . 67
5.3.1 Feature selection . . . 67
5.3.2 ANN architecture . . . 71
5.3.3 First training (BCE) . . . 72
5.3.4 Second training (Punzi) . . . 73
5.3.5 Training results . . . 76
5.4 Remaining background events . . . 82
5.5 Statistical evaluation . . . 86
5.5.1 Likelihood specification . . . 86
5.5.2 The Python library pyhf . . . 88
5.5.3 Binning . . . 88

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

5.5.4 Testing the validity of the asymptotic formulae 91
5.5.5 Procedure on simulated data 98
5.5.6 Calculating the global significance 104
5.5.7 The procedure for increased luminosity 107

6 Discussion of the results **109**
6.1 Expected sensitivity 109
6.2 Conclusion and outlook 110
6.2.1 Particle identification studies 110
6.2.2 Search for an invisibly decaying Z' boson 112

List of Tables **114**

List of Figures **117**

References **124**

A Tables **128**

B PID summary plots **135**

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Abbreviations

ΛCDM Lambda cold dark matter 6

ANN artificial neural network iii, 20, 22, 23, 61, 67, 69, 71–78, 80–85, 112–115, 120, 121

ARICH Aerogel Ring Imaging Cherenkov counter 29, 31, 32, 38

basf2 Belle II analysis software framework 36, 37, 40, 46

BCE binary cross-entropy 23, 67, 71–74, 76, 78, 80, 120

BDT boosted decision tree 67, 69, 71, 114

BF branching fraction 8, 9, 117

BSM beyond the Standard Model 1, 37

CDC Central Drift Chamber 27, 29–31, 33, 35, 38, 41, 48, 53, 55, 56, 64, 118

CDF cumulative distribution function 14, 92

CL confidence level 14, 16–18, 101, 103, 109–111, 122, 123

CMB cosmic microwave background 6

CMS centre-of-mass system 10, 25–27, 42, 44, 45, 59, 61–66, 69, 114, 118, 119

DAQ data acquisition 34, 35

ECL Electromagnetic Calorimeter 27, 29, 31–33, 38, 41, 48, 53, 55, 56, 61, 62, 64–66, 111, 118, 119

FOM figure of merit 17, 18, 42, 62, 64, 67, 73, 75, 77, 78, 80, 81, 83, 120

FPGA Field Programmable Gate Array 30

FSR final state radiation 64–66, 119

FWHM full width at half maximum 88–90, 115, 121, 133

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

HEP high energy physics 2, 10, 11

HER high energy storage ring 25, 26, 34

HLT high level trigger 34

HPD highest posterior density 19

IP interaction point 27, 28, 30, 32, 60, 61, 117

ISR initial state radiation 58, 89

KLM K_{L} - Muon Detector 27, 29, 33–35, 38, 52–54, 56, 62, 112, 118

LEE look elsewhere effect 16, 17, 101, 104, 107

LER low energy storage ring 25, 26, 34

LR learning rate 22, 72, 75, 76

MC Monte Carlo 12, 37, 38, 40–42, 45–49, 51–53, 56–58, 60, 62, 72, 75, 83, 87, 89, 91, 98–101, 106, 108, 110–112, 114, 118, 122

ML maximum likelihood 11–13, 15, 88, 98, 115, 133

MLE maximum likelihood estimator 12, 18

MVA multivariate analysis 69, 110

PDF probability density function 11, 13, 40, 86, 89, 92, 98

PDG Particle Data Group 57

PID particle identification 31, 40, 45, 47, 48, 51–55, 64, 110, 118, 119

PMT photomultiplier tube 32

POI parameter of interest 14–16

PSD pulse shape discrimination 33

PXD Pixel Detector 27, 29, 30, 34, 35, 117

QED quantum electrodynamics 5, 34

RPC Resistive Plate Chamber 33

SM Standard Model 1, 3, 4, 6–8, 11, 14, 27, 37, 65, 67, 70, 117, 120

SVD Silicon Vertex Detector 27, 29, 30, 38, 117

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

TOP Time-of-propagation counter 27, 31, 32, 38, 53, 118

VXD Vertex Detector 29, 30, 35, 117

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Chapter 1

Introduction

The Standard Model (SM) of elementary particle physics is a very successful theory. It can explain many experimental observations with very high accuracy. It gives an insight into how the most fundamental interactions arise from a mathematical description that, as most physicists would say, has a sense of beauty. Nevertheless, there are also reasons to be dissatisfied with the SM. Some discrepancies between experiments and the predicted outcome based on the SM exist, like the measurement of the anomalous magnetic dipole moment of the muon or anomalies involving the decay of B mesons. However, these are not the main reasons why we should be concerned. Some phenomena that we observe do not appear in the theory, such as gravity and dark matter. While Einstein's general relativity provides a well-established theory for gravity, no fundamental explanation for dark matter is accepted. This is why in recent years physicists started to search for a new particle which could extend the SM and account for the existence of dark matter. However, even if this is successful, the SM is still a theory based on many seemingly arbitrary choices of nature. Physicists always seek more fundamental theories. In the past, these theories always provided a simpler (in the sense of more compact[1]) description of our universe. It can only be assumed that the SM is not an exception and eventually we will arrive at a new theory which might also include the currently missing parts. The steadiest path towards this is to find experimental evidence for phenomena not described by the current theory (even if they should). This is why physicists want to disprove the SM and find 'new' physics beyond the Standard Model (BSM).

This thesis will describe the search for a new particle in electron-positron collisions at the Belle II experiment. The new particle is a Z' gauge boson which could act as a mediator of an unknown force. It is based on theoretical models that could solve some of the issues described above, most notably the dark matter puzzle. An overview of the theoretical concepts relevant to this thesis is given in Chapter 2. The experimental setup utilised for this analysis is described in Chapter 3, and Chapter 5 gives the details of the search procedure, particularly the selection of events and the statistical evaluation of the results. Additionally, in Chapter 4, the performance of particle identification at Belle II is studied and compared to simulations. The accurate identification of particles is an essential requirement for every analysis. Chapter 6 summarises the

[1] A mathematical representation of the standard model is already so compact that it can fit on a coffee mug.

main findings of this thesis and discusses how one can proceed and improve this search in the future.

It is worth to stress that all the analysis regarding the Z' has only been performed and optimised on simulated data. This gives the foundation to perform the search using actual data from collisions in 2019 and 2020. It is also worth noting that the same search has already been done at Belle II with a small dataset collected in 2018 [1]. While the overall concept is the same, many aspects of the experiment have changed, resulting in almost completely new conditions. Some findings of the previous measurement are still used for the current analysis, but they will be referenced accordingly. This thesis deals with the implementation of entirely new methods that constitute an improvement to the previous analysis. Certainly this present work could also benefit from the expertise of the analysts from the first search, most of which are still involved in the upcoming measurement.

Natural units

Throughout this thesis, the natural unit system, commonly used in high energy physics (HEP), will be employed. In these units, the constants ε_0, $\hbar$, and c take the value 1. Useful conversions are $\hbar c \approx 197$ MeVfm and $(\hbar c)^2 \approx 0.39$ GeV^{2}mb.
Nevertheless, as it is custom at Belle II, the units will sometimes be shown with additional factors of c corresponding to the respective quantity's conversion.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

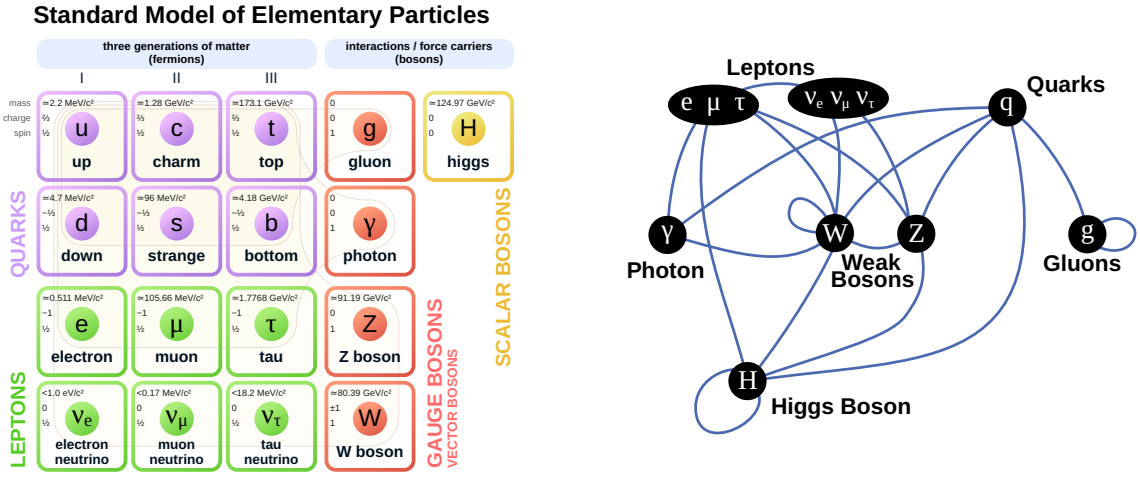
Chapter 2

Theoretical overview

2.1 The Standard Model of particle physics

The SM of elementary particle physics describes our current best understanding of how nature works at the most fundamental level. Mathematically it can be described by 37 quantum fields and up to 18 free parameters (26 when including neutrino masses). This already shows how complex this model is and, although based on Lorentz and gauge symmetry principles, also arbitrary. We also know that the SM is incomplete. Not only does it not explain dark matter, but it also completely fails to describe gravity. Also, while there is CP violation embedded in the SM it is not enough to explain the asymmetry of matter over antimatter in the universe through baryogenesis. There are many approaches to complete the SM and arrive at a 'theory of everything', but so far none have been successful.

The SM describes the interactions of particles, which are excitations in the quantum fields mentioned above. These particles are divided in fermions which follow the Fermi-Dirac statistics and bosons which follow Bose-Einstein statistics. Thus, based on the spin-statistics theorem, fermions carry half odd integer spin and bosons integer spin. In the SM, fermions are the particles that compose matter and bosons are the force carriers responsible for interactions. The three forces described by the SM are the electromagnetic force carried by the photon (γ), the strong force carried by gluons (g) and the weak force carried by the Z boson and the $W^{\pm}$ bosons.
These are the four vector bosons (bosons with spin 1) in the standard model. Besides, there is a scalar boson with spin 0, called the Higgs boson. It is interactions with the corresponding field what gives particles their mass. This is called the Higgs mechanism, which also describes how the weak and the electromagnetic interactions arise from the more fundamental electroweak interaction.
The fermions can be further divided into quarks and leptons. Quarks engage in all interactions and are always found in bound states (mesons or baryons) which is a consequence of their charge under the strong force. Leptons, however, do not experience the strong force. Both quarks and leptons can be classified into three generations of doublets. This results in six distinct flavours of quarks (up, down, charm, strange, top, bottom) and three types of leptons (electron, muon and tau) each with an accompanying neutral particle, called the neutrino. All the particle types in the



(a) The particles which constitute the SM and their most important properties. [2]

(b) The possible interactions between particles. [3]

Figure 2.1: An overview of the particles and interactions described by the Standard Model.

SM and their interactions are depicted in Figure 2.1. In addition to what is shown in Figure 2.1 every particle also has an antiparticle with the same mass but opposite charge.

Gauge theories

One of the essential concepts in the theoretical development of the SM is *gauge invariance*. This means requiring invariance of the Lagrangian under local transformations from certain Lie groups. New vector fields, so-called *gauge fields*, need to be defined to ensure the system's invariance under these transformations. The beauty of the SM is that these gauge fields are real, measurable quantum fields that give rise to the force carrier bosons (gauge bosons) mentioned before.

The gluons originate from requiring local SU(3) gauge invariance of the quark fields. For every quark flavour, there are three colour fields ('red', 'green', 'blue') that form a triplet on which the SU(3) acts. This results in 8 gluon fields (one for each generator of the SU(3)). By requiring invariance under $SU(2)_L \otimes U(1)_Y$ gauge transformations of the quark and lepton doublets four new fields arise, resulting in the electroweak sector of the SM. The 'L' in $SU(2)_L$ denotes that only left-chiral[1] fermions transform under the SU(2) group. Right-chiral fermions form singlets under SU(2) and do not undergo weak interactions. The 'Y' in $U(1)_Y$ stands for weak hypercharge, which is the quantum number corresponding to the U(1) gauge symmetry. The associated quantum number to the $SU(2)_L$ gauge symmetry is called weak isospin.

Both the weak hypercharge and the weak isospin are conserved by all interactions, except for interactions with the Higgs field. Only the electric charge, which is a combination of weak isospin and hypercharge, is conserved. This is because the $SU(2)_L \otimes U(1)_Y$ symmetry is broken by the

[1] Chirality is a concept arising from the representation theory of the Lorentz group.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table 2.1: The approximate values of the fine structure constant of the electromagnetic, the strong and the weak force for different energy scales.

coupling	$q^2 \approx 0$	$q^2 = \mathrm{m}_Z^2$
α	$\frac{1}{137}$	$\frac{1}{128}$
α_S	$O(1)$	0.12
α_W	$\frac{1}{30}$	

Higgs mechanism into $\mathrm{U}(1)_\mathrm{em}$. It is this spontaneous symmetry breaking which separates the electromagnetic and weak forces. The gauge bosons which mediate the weak force acquire mass, while the photon remains massless.

Interaction strength

The strength of the interaction is given by the couplings g_S (for the strong force), g_W (for the weak force) and e (for the electromagnetic force). It is found from experiments that the weak coupling is the same for all lepton flavours. This is called lepton flavour universality. In the quark sector the flavours mix, which introduces differences in the observed couplings and leads to phenomenons studied in flavour physics. The couplings are often expressed in terms of the fine-structure constants

$$\alpha_\mathrm{S} = \frac{g_\mathrm{S}^2}{4\pi}, \quad \alpha_\mathrm{W} = \frac{g_\mathrm{W}^2}{4\pi} \quad \text{and} \quad \alpha_\mathrm{em} = \frac{e^2}{4\pi}.$$

Contrary to what the name might suggest, their values are not constant. So-called virtual processes renormalise the bare coupling and make it dependent on the energy scale q^2 of the interaction. This is called the 'running of the coupling' and is described by renormalisation group theory. What is experimentally measured is only the effective strength of the coupling. The approximate values of the measured couplings in the atomic limit ($q^2 \approx 0$) and around $q^2 = \mathrm{m}_Z^2$ are given in Table 2.1. [4, Section 10.5]
For couplings much less than one, perturbation theory can be used to expand the interaction in powers of g and perform practical calculations. For a larger coupling, non-perturbative methods have to be used (e.g. calculations on a discrete lattice).

Because the weak and strong gauge bosons can interact with themselves, the couplings decrease with higher energy. Quantum electrodynamics (QED) on the other hand is an abelian gauge theory (the symmetry group $\mathrm{U}(1)_\mathrm{em}$ is commutative), where no self-interactions are possible. This increases the electromagnetic coupling for higher energies, suggesting that the couplings unify at some higher energy scale. This energy dependence is sketched in Figure 2.2.

2.2 Dark matter

Already since the beginning of the 20th century, there are hints for the existence of additional 'dark' matter in our universe. The term dark refers to the fact that this new matter is not charged

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

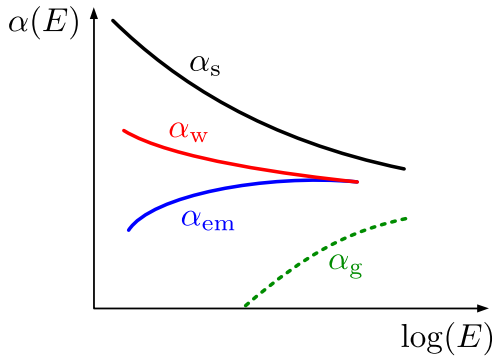


Figure 2.2: A sketch of coupling constants' energy dependence for the strong, weak and electromagnetic interaction. A hypothetical increase of the gravitational interaction (not included in the SM) is also shown, suggesting that all forces unify at some scale.

under the electromagnetic or the strong force. It only interacts gravitationally with ordinary, baryonic matter. Therefore most of the evidence originates from astronomical observations. Most strikingly, the rotation curves of spiral galaxies deviate from the expected velocities based on the distribution of visible (luminous) matter. Using the virial theorem, the predicted velocities of an object at a distance r around a massive body M are $v(r) = \sqrt{\frac{GM}{r}}$, where G is the gravitational constant. So at large distances from the centre of the galaxy the rotation velocities are expected to decrease, but they are found to be roughly constant. Assuming our understanding of gravity is correct, this can be explained by a spherical halo of dark matter around the galaxy. In Figure 2.3 the observed rotation curves for the galaxy NGC 2403 as measured by Begeman [5] are shown. The solid lines show the fitted contribution of luminous matter and the dark halo to the overall observed velocity.

Other convincing arguments for the existence of dark matter come from the study of galaxy clusters and their mass distribution through gravitational lensing. In many cases, there is much more mass measured than what is expected from calculations based on luminous matter. Similarly, the mass distribution in a colliding galaxy cluster (the 'Bullet-Cluster') is displaced to what is expected for baryonic matter alone.

Other arguments can be made at the cosmological scale. Because of the distinct time evolution and interactions of dark and ordinary matter, they affect the cosmic microwave background (CMB) in different ways. By precise measurement of anisotropies in the CMB cosmological models can be tested [see 7, Section 2.3]. The Lambda cold dark matter (ΛCDM) model can explain the CMB quite well. Within this model, dark matter represents 26.5 % of the universe's energy density. Only 4.9 % account for all the ordinary matter which is described by the SM (and only 10 % of that is luminous). The remaining and largest fraction of the energy density is attributed to dark energy.

Another hint for dark matter comes from the structure formation of the universe. Structure emerges from density perturbations in the early universe. Ordinary matter alone would be prevented to collapse by radiation and take a much longer time to cluster. Dark matter could have

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

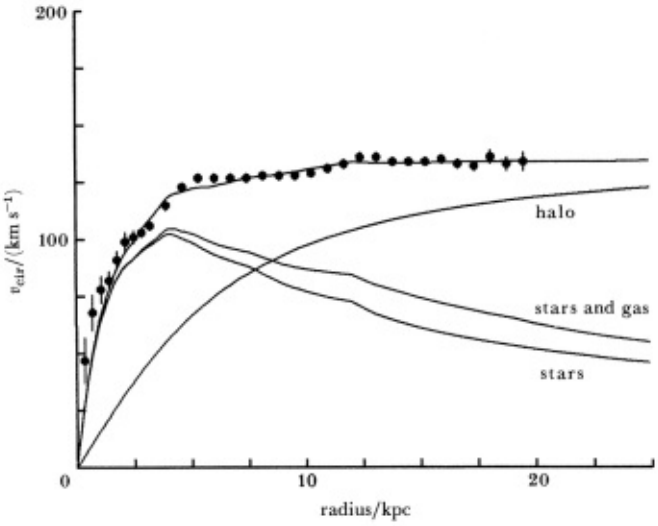


Figure 2.3: The observed rotation curve for NGC 2403 (dots with error bars) together with the fitted contribution of luminous matter and the dark halo to the overall observed velocity (solid lines). [from 6, Figure 4]

acted as a seed for ordinary matter to collapse and form the large condensed structures (stars, black holes, galaxies) that we observe today.

Although there is so much evidence for dark matter from astronomical observations, so far it could not be observed as a particle. Many theoretical models exist that provide candidates for such dark matter particles. The cosmological model often constrains these models, but there is still a lot of room to search for such dark matter candidates with the experimental methods available in particle physics.

2.3 The $L_\mu - L_\tau$ gauge boson Z'

The Z' boson is a hypothetical massive vector boson without any electric or colour charge. Numerous extensions of the SM predict such a particle [see 8, p. 786]. The simplest models are where the Z' originates as the gauge boson of a new abelian $\mathrm{U}(1)'$ gauge symmetry. One group of such models considers global U(1) symmetries already present in the SM, for example the lepton flavour numbers L_e, L_μ and L_τ. It is possible to make combinations of these symmetries such as $L_e - L_\mu$, $L_e - L_\tau$ and $L_\mu - L_\tau$ to build a model which extends the SM gauge group in an anomaly-free way. [9–11]

In particular, the $L_\mu - L_\tau$ models are attractive since they could explain open questions such as the small masses of the neutrinos [12] and provide mechanisms for dark matter production [13, 14]. They could also explain some anomalies observed in B decays [13, 15] and the discrepancies observed concerning the anomalous magnetic moment of the muon [14, 16]. The effects of this new gauge symmetry would be well hidden from observation since it couples primarily to the second and third generations of leptons.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

2.3.1 The invisibly decaying Z'

This thesis concerns a light (MeV- GeV) leptophilic (it interacts only with leptons) Z' belonging to the $L_\mu - L_\tau$ symmetry [see 14].
The Lagrangian describing the interactions of the new gauge boson can be written as

$$\mathcal{L} = -g'\bar{\mu}\gamma^\alpha Z'_\alpha \mu + g'\bar{\tau}\gamma^\alpha Z'_\alpha \tau - g'\bar{\nu}^{\mathrm{L}}_\mu \gamma^\alpha Z'_\alpha \nu^{\mathrm{L}}_\mu + g'\bar{\nu}^{\mathrm{L}}_\tau \gamma^\alpha Z'_\alpha \nu^{\mathrm{L}}_\tau, \tag{2.1}$$

where the superscript L denotes the left-chiral component of the Dirac spinor, and g' is the coupling constant associated with U(1)$'$. The partial decay widths are

$$\Gamma\left(Z' \to \ell^+\ell^-\right) = \frac{(g')^2 M_{Z'}}{12\pi}\left(1 + \frac{2M_\ell^2}{M_{Z'}^2}\right)\sqrt{1 - \frac{4M_\ell^2}{M_{Z'}^2}}\,\theta(M_{Z'} - 2M_\ell) \quad \text{and} \tag{2.2}$$

$$\Gamma\left(Z' \to \nu_\ell \bar{\nu}_\ell\right) = \frac{(g')^2 M_{Z'}}{24\pi}, \tag{2.3}$$

where ℓ denotes either μ or τ and M is the mass of the respective particle [derived from 16, Equation 2.12].
If it is kinematically allowed for the Z' to decay into dark fermions χ and $\bar{\chi}$ this is expected to happen with a branching fraction (BF) of 1 (because it is expected to couple only weakly to the SM particles compared to dark matter). If no such light dark matter particle exists the Z' can still decay invisibly (i.e. not seen by our detectors) via neutrinos. The BF for this decay is given by

$$\mathrm{BF}\left(Z' \to \nu_\ell \bar{\nu}_\ell\right) = \frac{2\,\Gamma\left(Z' \to \nu_\ell \bar{\nu}_\ell\right)}{2\,\Gamma\left(Z' \to \nu_\ell \bar{\nu}_\ell\right) + \Gamma\left(Z' \to \mu^+\mu^-\right) + \Gamma\left(Z' \to \tau^+\tau^-\right)}$$

and would depend on the mass of the Z'. Considering the available decay modes and using Equation 2.2 the resulting BF are

$$\begin{aligned} \mathrm{BF}\left(Z' \to \nu_\ell \bar{\nu}_\ell\right) &= 1 && \text{for} && M_{Z'} < 2M_\mu, \\ \mathrm{BF}\left(Z' \to \nu_\ell \bar{\nu}_\ell\right) &\approx 1/2 && \text{for} && 2M_\mu < M_{Z'} < 2M_\tau \quad \text{and} \\ \mathrm{BF}\left(Z' \to \nu_\ell \bar{\nu}_\ell\right) &\approx 1/3 && \text{for} && M_{Z'} > 2M_\tau. \end{aligned} \tag{2.4}$$

So when considering an invisibly decaying Z' there are two possibilities:

1. It can decay into dark matter particles χ with $M_\chi < 0.5 M_{Z'}$ resulting in a BF of $Z' \to$ invisible close to 1. This case is currently most constrained by a measurement from Belle II [1] (see also the next Section).

2. It only decays in SM particles resulting in a BF of $Z' \to$ invisible according to 2.4. This case is currently most constrained by a measurement from BaBar for a Z' decaying in a pair of muons [17]. Because this is a fully visible final state they could only exclude a Z' with a mass greater than two times the muon mass.

This thesis will focus on the first case, but a generalisation to the second is straightforward.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

2.3.2 Detection at Belle II

To find such a particle at Belle II, we need to search for its signature in e^+e^- collisions. One possibility is to consider the process $e^+e^- \to \mu^+\mu^-$ ($Z' \to$ invisible). A Feynman diagram of this process where the Z' decays to neutrinos ($\nu\bar{\nu}$) or dark leptons ($\chi\bar{\chi}$) is shown in Figure 2.4.

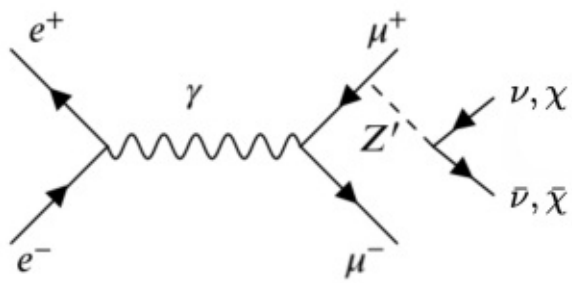


Figure 2.4: A Feynman diagram showing the process $e^+e^- \to \mu^+\mu^-$ ($Z' \to$ invisible) where invisible is either $\nu\bar{\nu}$ or $\chi\bar{\chi}$.

The result would be two muons in the final state plus missing energy, which can be detected with Belle II. This search was already done with a relatively small dataset, resulting in the limits on the coupling constant shown in Figure 2.5.

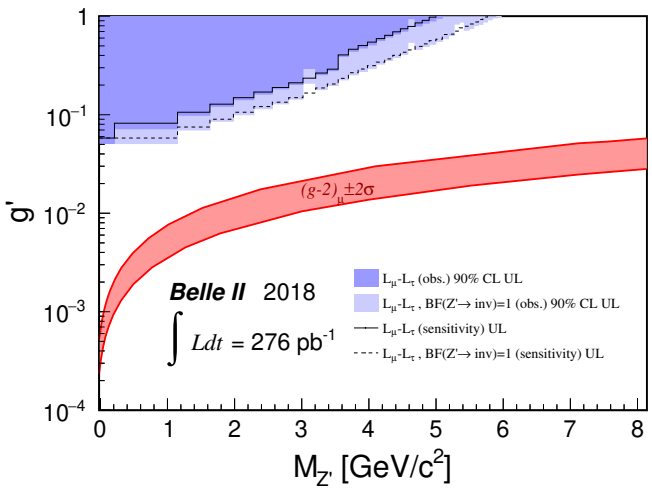


Figure 2.5: The current 90 % upper limits on the g' coupling constant for different cases of the $Z' \to$ invisible BF. The region which could explain the muon anomalous magnetic moment is highlighted $(g-2)$. [1]

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Recoil variables

The most important quantities to perform this search are the Z' mass and momentum. Since there is no direct access to this (the Z' decays invisibly), they need to be inferred from the information provided by the visible final state particles. Quantities like this are called recoil variables because they describe the particle recoiling against the $\mu^+\mu^-$ system. It is also necessary to know the initial energy and momentum of the colliding electron beams. Then one can use the Equation

$$p_{e^+} + p_{e^-} = p_{\mu^+} + p_{\mu^-} + p_{Z'}, \tag{2.5}$$

where p denotes the particles four-vector, to calculate the desired properties. The squared recoil mass is simply

$$\mathrm{m}^2_{\mathrm{rec}} := p^2_{Z'}, \tag{2.6}$$

where $p_{Z'}$ is calculated from Equation 2.5. At Belle II the initial state four-vector is given by

$$p_{e^+} + p_{e^-} = \begin{pmatrix} 11.006 \\ 0.4566 \\ 0 \\ 2.9994 \end{pmatrix} \text{GeV}, \tag{2.7}$$

which results in a maximum value of $\mathrm{m}_{\mathrm{rec}}$ corresponding to the centre-of-mass system (CMS) energy of the collision $\sqrt{s} \equiv \sqrt{(p_{e^+} + p_{e^-})^2} = 10.58$ GeV. This value defines the upper bound of the mass range in which the search is possible.

The strategy is to hunt for a peak or 'bump' in the recoil mass spectrum, which would appear at the mass of the hypothetical Z'. This is a very distinct feature, but only a small number of signal events can be expected (given that a Z' exists within our sensitivity). The expected and observed events in a given recoil mass region have to be statistically evaluated to provide results. With more data, the sensitivity of this search will increase until systematic effects dominate.

The last thing to mention here is that this search, although based on one single Z' model, is general enough to constrain a wide range of possible models or even find new physics that is not predicted by any model.

2.4 Statistical methods

This Section describes the basic statistical methods commonly used in HEP, which are also relevant for this analysis. First, a general overview of possible approaches is given, then a frequentist method for hypothesis testing and obtaining confidence intervals is described in more detail. This approach will be used in Section 5.5 to perform the statistical evaluation of the analysis.

2.4.1 Inferential statistics

Contrary to descriptive statistics, where one extracts and analyses information from a given data sample, an inferential statistical analysis is performed when we want to deduce properties of the

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

assumed underlying distribution of a given sample. In a typical high energy physics analysis, this underlying distribution represents the currently accepted model (the SM) or a new model that can include new physics.

This statistical model, $P_{\mathrm{m}}(\boldsymbol{x}|H)$, is the assumed probability of observing the data $\boldsymbol{x}$ given a hypothesis H. This hypothesis may depend on one or more parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)$, so the model can be written as $P_{\mathrm{m}}(\boldsymbol{x}|\boldsymbol{\theta})$. To perform statistical inference the likelihood function

$$L(\boldsymbol{\theta}) = P_{\mathrm{m}}(\boldsymbol{x}_0|\boldsymbol{\theta}), \tag{2.8}$$

which is the model evaluated at the observed data $\boldsymbol{x}_0$, is used as a basis to perform tests on the parameters $\boldsymbol{\theta}$.

2.4.2 Parameter estimation

To get a point estimate for any parameter θ of the model, an *estimator* $\widehat{\theta}(\boldsymbol{x})$, which is a function of the data, is used. When $\widehat{\theta}$ is written, this refers to the value of the estimator at the given data, sometimes called the estimate. Because the data are a random variable, the estimator is also a random variable with an associated probability distribution. There are no restrictions on how an estimator is constructed, but one can study its properties and compare them with other estimators. The most important of these properties are *consistency*, *bias*, *efficiency* and *robustness* [8, p. 527].

The most common estimators in HEP are the least squares and the maximum likelihood (ML) estimator. Here only the latter will be described, details on the least squares estimator can be found elsewhere (e.g. [18]). The ML estimate of the parameters can be found by maximising the likelihood function with respect to $\boldsymbol{\theta}$, so solving the equations

$$\frac{\partial \ln L}{\partial \theta_i} = 0, \quad i = 1, \ldots, N. \tag{2.9}$$

While sometimes this can be done analytically, usually numerical methods are used to maximise the likelihood function. If the data $\boldsymbol{x} = (x_1, \ldots, x_n)$ are independent and follow the same probability density function (PDF) $f(x; \boldsymbol{\theta})$, as it is often the case, the likelihood function can be written as the product of the individual likelihoods

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f\left(x_i; \boldsymbol{\theta}\right). \tag{2.10}$$

If the length of the data vector itself is not fixed, and the probability of observing n events also depends on the model's parameters, this needs to be included in the likelihood function. In HEP n usually follows a Poisson distribution with mean $\mu(\boldsymbol{\theta})$ and the likelihood function becomes what is called the *extended likelihood*

$$L(\boldsymbol{\theta}) = \frac{\mu^n}{n!} e^{-\mu} \prod_{i=1}^{n} f\left(x_i; \boldsymbol{\theta}\right). \tag{2.11}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Asymptotic properties of the ML estimator

The maximum likelihood estimator (MLE) has some desired properties in the limit of very large data samples, which is called the asymptotic limit. The estimator is said to be

- *consistent*: The estimate $\widehat{\theta}$ converges to the true value.
- *efficient*: The variance of the considered estimator is the smallest possible variance for any estimator, which is given by the Rao-Cramér-Fréchet bound.
- *normal*: The estimates $\widehat{\theta}$ are normally (Gaussian) distributed.

As a consequence, if the range of the allowed data is independent of $\boldsymbol{\theta}$, the variance of the MLE can be obtained as the curvature of the natural logarithm of the likelihood function at its minimum

$$\widehat{V}_{ij} = -\left(\frac{\partial^2 \ln L}{\partial\theta_i \partial\theta_j}\right)^{-1}\Bigg|_{\widehat{\boldsymbol{\theta}}}. \tag{2.12}$$

These asymptotic properties are usually not met with finite sample size, but often approximately valid for large samples. The number of observations required to enter the asymptotic regime depends in general on the model and therefore has to be always checked if one wants to make use of these properties. Equation 2.12 is used extensively to calculate the uncertainty of the point estimates, for example by the HESSE method of the `MINUIT` minimiser [19].

The binned ML method

A special case of the ML method occurs if the observed data are binned in one or more variables with a total number of N bins. Then a new data vector is obtained as the number of entries in each bin, $\boldsymbol{n} = (n_1, \ldots, n_N)$. If the total number of events is not fixed and each of the n_i follows a Poisson distribution with expectation value $\mu_i(\boldsymbol{\theta})$ the likelihood function can be written as

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} \frac{\mu_i^{n_i}}{n_i!} e^{-\mu_i}. \tag{2.13}$$

In the limit of infinitely many bins the ML estimate obtained from Equation 2.13 is equivalent to the one obtained from the extended likelihood when using the unbinned method (Equation 2.11). While there is a loss of information when the data are binned, this method has some advantages. It is usually faster to perform a numerical optimisation compared to the unbinned ML method, where the CPU consumption increases linearly with the total number of events. Also, by performing a binned fit, it is possible to use templates for the underlying distributions (for example from Monte Carlo (MC) simulations) if no analytical form is available. Compared to the least squares method, the binned ML method is more efficient, less biased, and less sensitive to binning choices [18].

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

2.4.3 Goodness-of-fit

Using the binned approach of the ML method, it is possible to quantify how well the model describes the data using a p-value [20]. This can be done with the *likelihood ratio*

$$\lambda(\boldsymbol{\theta}) = \frac{L(\boldsymbol{\theta})}{L(\widehat{\boldsymbol{\mu}})}, \tag{2.14}$$

where the likelihood function in the denominator is the same as in Equation 2.13 but instead of μ depending on some parameters $\boldsymbol{\theta}$, the $\boldsymbol{\mu}$ themselves are the free parameters. In this case, $\widehat{\mu}$ will correspond to the number of events n in the bin. According to Wilks' theorem [21], the distribution of the quantity

$$-2\ln\lambda(\widehat{\boldsymbol{\theta}}) \tag{2.15}$$

asymptotically (for large μ) approaches a χ^2 distribution. The number of degrees of freedom is equal to the difference in the number of free parameters of the numerator and the denominator in the likelihood ratio. Therefore, the p-value for the model describing the data is

$$p = \int_{-2\ln\lambda(\widehat{\boldsymbol{\theta}})}^{\infty} f_{\chi^2}(z; n_\mathrm{d})\, dz, \tag{2.16}$$

where f_{χ^2} is the PDF of the χ^2 distribution and n_d is the associated number of degrees of freedom.

2.4.4 Nuisance parameters

Even if the assumed model correctly describes the data's underlying distribution, systematic effects influence the observed distribution of the data. These systematic effects can be included by introducing additional free parameters in the model. These other parameters are called nuisance parameters $\boldsymbol{\nu}$ and the likelihood function will now depend on both $\boldsymbol{\theta}$ and $\boldsymbol{\nu}$.

If there is an additional measurement (called an auxiliary measurement), that can help constrain the values of the nuisance parameters this information can be included by multiplying the likelihood $L(\boldsymbol{\theta}, \boldsymbol{\nu})$ with

$$L_\nu(\nu) = P_\nu(\boldsymbol{y}|\nu), \tag{2.17}$$

where P_ν is the model of the parameter ν and $\boldsymbol{y}$ is the observed data of the auxiliary measurement.

2.4.5 Hypothesis tests and confidence intervals

A test of the statistical model $P_\mathrm{m}(\boldsymbol{x}|H)$ with a specific hypothesis H can be performed using a *test statistic* t. The value of this test statistic is calculated based on the observed data. Therefore it is a random variable that follows some probability distribution. If this distribution is known (under the assumption that H is true), the level of agreement of the observed data with the hypothesis can be quantified by computing a p-value as a measure of incompatibility with H. Usually, this p-value, the statistical significance, is expressed in standard deviations of a unit Gaussian where

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

the one-sided tail probability is equal to the p-value. This quantity, called Z, can be calculated as

$$Z = \Phi^{-1}(1 - p), \tag{2.18}$$

where Φ is the cumulative distribution function (CDF) of the standard normal distribution. When looking for new particle physics processes, the hypothesis that is tested usually corresponds to the SM (often called the background-only hypothesis). Before performing the measurement, one has to decide on a threshold α_0, called the *significance level*, that defines a 'critical region'. If the observed p-value is smaller than α_0, the background-only hypothesis is rejected, and a discovery can be claimed. In particle physics it has been established to set the significance level to $\alpha_0 = 2.87 \times 10^{-7}$ ($Z = 5$), meaning that if H is true, there is only about a 1 in 3 500 000 chance that it will be rejected (type I error).

When rejecting a signal hypothesis, usually a threshold of 0.05 or 0.1 is used instead. In some cases, a family of signal hypothesis $H(\boldsymbol{\mu})$, depending on one or more parameters $\boldsymbol{\mu}$, can be tested. If the hypothesis is parametrised with the signal's strength, then even the background-only hypothesis can be expressed in the same framework as the particular case of zero signal strength. In the following, we will assume such a hypothesis $H(\mu)$ where μ is the signal strength. This parameter will also be called the parameter of interest (POI). The *power* of a test is defined as the probability that a discovery will be claimed (rejecting $H(0)$ at a significance level α_0) when $H(\mu)$ with $\mu > 0$ is true.

When testing such a parametrised hypothesis $H(\mu)$, an interval for the parameter can be obtained by performing a test for many parameter values (also referred to as a scan of the parameter). The boundary of this interval will be given by the values of μ where the p-value of this hypothesis is equal to some threshold value α. This interval will depend on the observed data, and it will either contain the true value of the parameter μ or not. If the measurement is repeated many times, a fraction of the obtained intervals will contain (cover) the true value of μ. This fraction of intervals is also called the *coverage probability*, and it is a statement about a family of intervals, rather than one single interval. By construction, this coverage probability should be equal to $1 - \alpha$, independent of the actual value of μ. Sometimes this is not possible, for example, when one deals with discrete parameters or additional nuisance parameters. In that case, the intervals are constructed so that the coverage probability is greater than $1 - \alpha$ (they are allowed to *overcover*). These intervals are said to have a confidence level (CL) of $1 - \alpha$.

This procedure, known as the *test-inversion method*, is equivalent to constructing intervals with a construction by Neyman [22]. The core of Neyman's construction, the ordering rule, is given by the choice of the test statistic. Since different ordering rules will lead to different intervals, the same is true for the test statistic. Therefore, it is essential to state what ordering rule or test statistic was used to construct the intervals. [23]

When dealing with nuisance parameters, the proper way to construct the intervals would be to construct a higher dimensional confidence region in the same manner as described above. This region can then be projected on the parameter μ, resulting in intervals with coverage probability of at least CL, independent of the true value of μ *and* $\boldsymbol{\nu}$. Because this is cumbersome, it is desired to eliminate the nuisance parameters before constructing the intervals. In the following, I will describe one possible way to do this.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

The $\tilde{t}_\mu$ test statistic

Some test statistics and their properties are described by Cowan, Cranmer, Gross and Vitells [24]. I will focus on the $\tilde{t}_\mu$ test statistic since I found it the best suitable for the analysis.

Many test statistics use a ratio of likelihoods. Neyman and Pearson showed that a test based on this ratio gives the most powerful test among all possible statistical tests [25]. In particular, many test statistics commonly used are based on the *profile likelihood ratio*

$$\lambda_p(\mu) = \frac{L(\mu, \widehat{\widehat{\boldsymbol{\nu}}}(\mu))}{L(\widehat{\mu}, \widehat{\boldsymbol{\nu}})}, \tag{2.19}$$

where $\widehat{\widehat{\boldsymbol{\nu}}}$ denotes the ML estimate of $\boldsymbol{\nu}$ for a fixed value of μ, thus eliminating the dependency of L on $\boldsymbol{\nu}$. The quantity in the numerator is also called the profile likelihood[2], and the denominator is the global maximum of the likelihood for unconstrained values of μ. As mentioned in Section 2.4.3 the quantity $-2\ln\lambda_p(\mu)$ will asymptotically follow a χ^2 distribution, here with one degree of freedom. Using the profile likelihood ratio, the test statistic $\tilde{t}_\mu$ is defined as

$$\tilde{t}_\mu = \begin{cases} -2\ln\left(L\big(\mu, \widehat{\widehat{\boldsymbol{\nu}}}(\mu)\big)/L\big(0, \widehat{\widehat{\boldsymbol{\nu}}}(0)\big)\right) & \hat{\mu} < 0 \\ -2\ln\lambda_p(\mu) & \hat{\mu} \geq 0 \end{cases}. \tag{2.20}$$

This test statistic is useful if the POI is constrained by a boundary, here $\mu = 0$. In the case of signal strength, this is sensible since negative values would be not physical. If the estimated value of μ would yield a negative number, the best-fit value in the denominator of λ_p is set to 0 and $\widehat{\boldsymbol{\nu}}$ is replaced by $\widehat{\widehat{\boldsymbol{\nu}}}(0)$.

As shown by Cowan et al. [24], in the asymptotic limit the cumulative distribution of $\tilde{t}_\mu$ can be written as

$$F(\tilde{t}_\mu|\mu') = \Phi\left(\sqrt{\tilde{t}_\mu} + \frac{\mu-\mu'}{\sigma}\right) + \begin{cases} \Phi\left(\sqrt{\tilde{t}_\mu} - \frac{\mu-\mu'}{\sigma}\right) - 1 & \tilde{t}_\mu \leq \mu^2/\sigma^2 \\ \Phi\left(\frac{\tilde{t}_\mu - (\mu^2 - 2\mu\mu')/\sigma^2}{2\mu/\sigma}\right) - 1 & \tilde{t}_\mu > \mu^2/\sigma^2 \end{cases}. \tag{2.21}$$

where σ is the standard deviation of $\widehat{\mu}$ and μ' is the true value of the parameter μ. For the special case where $\mu = \mu'$ this becomes

$$F(\tilde{t}_\mu|\mu) = \begin{cases} 2\Phi\left(\sqrt{\tilde{t}_\mu}\right) - 1 & \tilde{t}_\mu \leq \mu^2/\sigma^2 \\ \Phi\left(\sqrt{\tilde{t}_\mu}\right) + \Phi\left(\frac{\tilde{t}_\mu + \mu^2/\sigma^2}{2\mu/\sigma}\right) - 1 & \tilde{t}_\mu > \mu^2/\sigma^2 \end{cases}. \tag{2.22}$$

The standard deviation σ can be estimated either from Equation 2.12 or with a special dataset called the *Asimov data*, defined as the values of $\boldsymbol{x}$ that, if used to estimate the parameters of the model, given the true parameter values [24, Section 3.2].

[2] The profile likelihood is actually not a likelihood and has not the same mathematical properties. In many cases however, it behaves similarly.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

A considerable benefit of using this test statistic is that it avoids a common problem, often referred to as 'flip-flopping'. This refers to the issue when analysts quote upper limits if there is no observed excess in data, but instead switch to quote two-sided confidence regions if there is an excess. If this switch is done by changing the ordering principle or the used test statistic, this can lead to an undesired undercoverage of the produced intervals [see 26, Figure 4]. By using the $\tilde{t}_\mu$ test statistic, this switch is done automatically. One either obtains upper limits or two-sided intervals, depending on the observed data, while still maintaining the correct coverage. This also makes it possible to quote a hypothetical discovery's statistical significance by simply testing the background-only hypothesis using $\tilde{t}_\mu$. In the case of no nuisance parameters, using the test statistic $\tilde{t}_\mu$ is equivalent to the method described by Feldman and Cousins [26].

Another advantage of this test statistic is that it never yields empty intervals or intervals only containing one point of the POI, which is an undesired property.

It is important to note that in general the asymptotic formulae are not valid and only approximately correct for some problems. Also, the usage of the profile likelihood can mess up the coverage of the obtained intervals. Therefore it is always necessary to check their coverage explicitly with simulated data. When dealing with nuisance parameters, one has to generate data for different values of μ and $\boldsymbol{\nu}$. The fraction of the obtained intervals covering the true value has to be at least CL, for all reasonable values of $\boldsymbol{\nu}$.

If the asymptotic formulae do not hold, it is always possible to construct the intervals directly based on simulated data, guaranteeing the coverage by construction. With nuisance parameters, the correct way to do this is called the *supremum method*. Here one constructs intervals for all possible values of $\boldsymbol{\nu}$ for each tested μ and then selects the values of $\boldsymbol{\nu}$ which yield the largest interval. This is computationally very expensive and rarely done in practice. An alternative is to restrict the values of $\boldsymbol{\nu}$ to the best fit values observed in data, $\widehat{\boldsymbol{\nu}}$ or even $\widehat{\widehat{\boldsymbol{\nu}}}$. This is called the *plugin method* and assumes that the nuisance parameters' true values are equal to the estimated ones. The method can be improved by considering values of $\boldsymbol{\nu}$ around $\widehat{\boldsymbol{\nu}}$, resulting effectively in a mixture of the plugin and the supremum method.

2.4.6 Look elsewhere effect

If one does not know all the signal model parameters, it is possible to search in different, often disjoint, phase space regions. This unknown parameter could be the mass of a new particle that determines the shape and kinematics of the events one is looking for. This can result in many disjoint search regions, like looking for a peak in a distribution. In such a case, it is expected that an excess in some of these regions due to statistical fluctuations is observed. This is called the look elsewhere effect (LEE), and it has to be accounted for when quoting the significance for a signal. While the *local significance* is the significance obtained for a fixed parameter of the signal model, the *global significance* considers the probability of observing an excess anywhere in the search region. For a test statistic $t_0(m)$ which is calculated for different values of the parameter m the global significance can be obtained from the distribution of $t_0(\widehat{m})$, where $\widehat{m}$ is the value of m that gives the largest t_0 (assuming that larger values of t_0 indicate a increasing disagreement with the data). One way to get this distribution is to simulate many datasets corresponding to the

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

background-only hypotheses and then to calculate $t_0(\widehat{m})$ for each of them. This is computationally very expensive and in some cases not feasible, since to correct a 5σ effect, one would have to repeat this procedure about 10^7 times. Unfortunately, there is no simple formula for this distribution, but approximations exist.

The most notable is a result described by Gross and Vitells [27]. It gives an estimation on the global p-value for large observed values of $t_0(\widehat{m})$ (which is also the case one usually is most interested in). For lower values of $t_0(\widehat{m})$ the result still gives an upper bound on the global p-value. It states that if t_0 follows a χ^2 distribution with s degrees of freedom

$$P\big(t_0(\widehat{m}) > c\big) \approx P\big(\chi^2_s > c\big) + \langle N(c)\rangle, \tag{2.23}$$

where c is the observed value of $t_0(\widehat{m})$ and $\langle N(c)\rangle$ is the mean number of 'upcrossings' of $t_0(m)$ above c in the range of m. The value of $\langle N(c)\rangle$ can be estimated from a reference value c_0, using a result from Davies [28]. In the typical case of $s = 1$

$$\langle N(c)\rangle = \mathcal{N} e^{-c/2}, \tag{2.24}$$

where $\mathcal{N}$ is a constant independent of c which can be intuitively interpreted as the average number of independent search regions over m. Using this result one can write

$$\langle N(c)\rangle = \langle N(c_0)\rangle e^{-(c-c_0)/2}. \tag{2.25}$$

The quantity $\langle N(c_0)\rangle$ can be estimated from simulated data since now much fewer iterations are needed, depending on the chosen value of c_0.

In the context of the LEE sometimes the *trial factor* is quoted, which is defined as the ratio

$$\frac{p_{\text{global}}}{p_{\text{local}}} = \frac{P(t_0(\widehat{m}) > c)}{P(t_0(m) > c)}. \tag{2.26}$$

2.4.7 Optimising the search region

When comparing different experiments or analysis selections, it is useful to have a simple quantity representative of the resulting sensitivity. This quantity is called a figure of merit (FOM). If the signal strength is not known a priori, it is common to use a FOM introduced by Punzi [29]. This FOM is related to the values of the signal strength μ for which the power of the corresponding test is greater or equal to the CL one wishes to use to construct confidence intervals. Punzi shows that for a simplified model the quantity

$$\sigma_{\min} = \frac{\frac{b^2}{2} + a\sqrt{B(t)} + \frac{b}{2}\sqrt{b^2 + 4a\sqrt{B(t)} + 4B(t)}}{\epsilon(t)\cdot L} \quad \text{with} \quad \begin{array}{l} a = \Phi^{-1}(1-\alpha_0) \\ b = \Phi^{-1}(CL) \end{array} \tag{2.27}$$

can be interpreted as the 'minimum detectable cross section' and will be referred to as the *Punzi sensitivity region*. Here $B(t)$ is the total number of expected background events and $\epsilon(t)$ is the signal efficiency, both depending on an additional parameter t which could be for example the

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

value of a cut. L is the total integrated luminosity, and α_0 is the chosen significance level as defined in Section 2.4.5.

This quantity has a few properties that make it well suited to use for comparison. For example, if the true cross section of the signal is larger than σ_{min}, the probability that performing this experiment will lead to discovery is at least CL. If there is no signal observed, at least all the values greater than σ_{min} can be excluded at the given CL (independently of the true value of σ). This means that for $\sigma > \sigma_{\text{min}}$ the experiment is guaranteed to provide a result (either exclusion or discovery).

Since a FOM is usually a quantity that should be maximised, taking the inverse of 2.27 and removing constant factors yields

$$\text{f.o.m.}_{\text{Punzi}} = \frac{\epsilon(t)}{b^2 + 2a\sqrt{B(t)} + b\sqrt{b^2 + 4a\sqrt{B(t)} + 4B(t)}}, \tag{2.28}$$

which is referred to as the Punzi figure of merit.

Another approach is to maximise a FOM directly related to the expected significance [see 24, Equation 97]. When instead of discovering a signal, it is desired to measure the yield of a process with known signal strength, a possible figure of merit is

$$\text{f.o.m.} = \frac{s(t)}{\sqrt{s(t) + B(t)}}, \tag{2.29}$$

where $s(t)$ is the expected number of signal events. This FOM originates from the variance of the MLE of the signal yield when considering a Poisson counting experiment. The MLE is $\widehat{s} = (n - B)/s$ (with n being the number of observed events) and its variance $\frac{s+B}{s^2}$. Hence maximising 2.29 is equivalent to minimising the standard deviation $\sigma_{\widehat{s}} = \sqrt{s + B}/s$ which results in a greater precision of the measurement [8, p. 541].

2.4.8 Bayesian statistics

Everything discussed previously was based on frequentist statistics. In this Section I want to briefly mention the Bayesian method, which gives an alternative approach and is often considered a different 'philosophy' of statistics. While the underlying concepts of probability are certainly different, both methods have their advantages and disadvantages. In the end, the problem at hand should determine which method is suited best.

The core difference is that, in Bayesian statistics, a probability distribution is assigned on the hypothesis (or the parameters of the model). This is not possible from a frequentist view, since a parameter has to have a fixed value (even if it is unknown).

Bayesian inference is based on a very fundamental law about probability, called *Bayes' theorem*:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \tag{2.30}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

This can be written as

$$P(H|\boldsymbol{x}) = \frac{P\left(\boldsymbol{x}|H\right)\pi\left(H\right)}{\int P\left(\boldsymbol{x}|H'\right)\pi\left(H'\right)dH'}, \tag{2.31}$$

where $\pi(H)$ is the *prior probability* for the hypothesis H which represents the degree of believe one has about H before performing the measurement, $P(\boldsymbol{x}|H)$ is the likelihood and $P(H|\boldsymbol{x})$ is the probability of H given the data $\boldsymbol{x}$, called the *posterior probability*. The integral in the denominator is simply a constant so that the posterior probability is normalised to unity. With access to the posterior distribution, statistical inference becomes as 'easy' as calculating integrals (usually numerically). The analogue of frequentist confidence intervals in a Bayesian framework is called a *credible interval*. One often quotes the regions of the posterior distribution with the highest probability (resulting in the narrowest interval), called the highest posterior density (HPD) interval.

So while a frequentist will make a statement about the probability that the collected data produces an interval that contains the true value of a parameter (when considering all possible outcomes of the experiment), a Bayesian talks about the probability that, based on the observed data and the prior knowledge about the parameter, the true value *is* in the interval.

One critique of the frequentist method is that it produces accurate results, but not correct ones. This is because there are situations where the intervals with the correct coverage give results that we perceive as incorrect. To study this, it helps to refer to a Bayesian analysis of the problem. [see 30, Section 3]

In some cases, the Bayesian approach feels more natural and resembles better how we think and estimate probabilities ourselves. If there is prior knowledge about a hypothesis, for example from another experiment, a Bayesian interpretation of the data is straightforward. It should be preferred if one wants to include this additional information in the final result. However, if this prior information is not available, an argument can be made that this procedure is not objective since the result depends on the prior choice. This introduces some complications. In many cases, to claim objectivity, it is shown that the choice of the prior does not influence results or a special 'objective prior' like *Jeffreys prior* is used, and frequentist properties of the result are studied. While this approach is certainly valid, it differs from the original Bayesian view where the prior is supposed to be subjective.

Another interesting variation of the Bayesian method is *Empirical Bayes*, where the prior distribution is estimated from data. This approach can be seen as a bridge to the frequentist world, and it has been shown that this method produces desirable results and solves known problems with the frequentist method.

In my opinion, the frequentist and Bayesian methods should be used in synergy, and it can only be beneficial to compare and study results obtained with complementary approaches. It is however important always to state how the results were obtained to avoid misinterpretation. There is no clear advantage of either method in terms of complexity and computation since to validate results checks with simulated data have to be done in either case.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

2.5 Artificial neural networks

In this Section I will briefly describe the basic principles of feedforward ANNs. ANNs are inspired by the neurons in our brain. Of course, a brain is much more complicated, but some similarities exist between an artificial and a biological neural network. In both cases, we speak of neurons to refer to some unit connected with other units of the same type. If a neuron receives a signal above some threshold (which is different for each neuron), it will send signals to other neurons. Based on the connections between the neurons and these activation thresholds, the network can serve many purposes. The process of tweaking these 'parameters' of the network is called learning.

One of the simplest ways to realise something like this on a computer is by grouping some of these neurons into layers and then sequentially connecting neurons from one layer to another. These connections have a direction starting from the *input layer* towards the *output layer*. Between the input and output layer, there can be any number of additional layers called *hidden layers*. When every neuron of one layer is connected to every neuron of the consequent layer, it is called *fully connected*. The connections itself have a *weight* associated with them which can be any real number. Additionally, each neuron also has a number assigned which is called the *bias*. Using these parameters, we can define the output of any neuron to be

$$a_j^l = \sigma\left(\sum_k w_{jk}^l a_k^{l-1} + b_j^l\right), \tag{2.32}$$

where a_j^l defines the output of the jth neuron in layer l, w_{jk}^l is the weight connecting the kth neuron from layer $l-1$ to the jth neuron of layer l, and b_j^l is the bias of the jth neuron in layer l. σ is a called the *activation function* which is usually some kind of continuous version of a step function, returning 0 for large negative and 1 for large positive inputs. One commonly used example for this is the Fermi function,

$$\sigma(z) = \frac{1}{1+e^{-z}}, \tag{2.33}$$

which in this context is often called the sigmoid function. A graph of this function is depicted in Figure 2.7. Figure 2.6 shows how the relations defined by Equation 2.32 can be drawn for a network with three fully connected layers (two neurons in the input layer, three neurons in the hidden layer and one neuron in the output layer). Using Equation 2.32 any input to the network will get mapped to some well-defined output. With enough parameters ($\boldsymbol{w}$ and $\boldsymbol{b}$) the network can be used to approximate any function. The task at hand is now to choose these parameters so that the network behaves as desired. In *supervised learning*, this is done using training data that match pairs of input $\vec{x}$ and anticipated output $\vec{y}(\vec{x})$. When these inputs are fed to the network, the network's output can be compared to the output given by the training data. To measure how well the network is performing on this data set a *loss function* is defined, returning larger values when the network performs worse and smaller values for a good match between actual output $\vec{a}(\boldsymbol{w}, \boldsymbol{b})$ and desired output $\vec{y}$. One of the simplest loss functions is the quadratic loss,

$$C(\boldsymbol{w}, \boldsymbol{b}) \equiv \frac{1}{2n} \sum_{\boldsymbol{x}} \|\vec{y}(\vec{x}) - \vec{a}(\boldsymbol{w}, \boldsymbol{b})\|^2, \tag{2.34}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

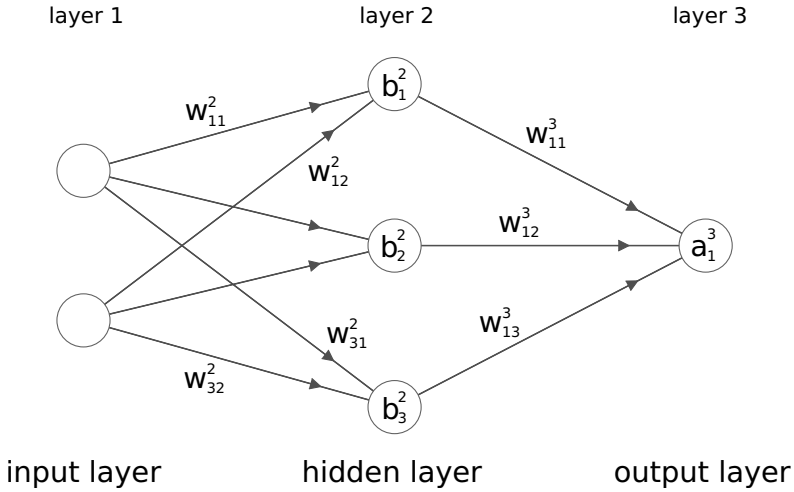


Figure 2.6: The layout of a simple neural network with three fully connected layers (two neurons in the input layer, three neurons in the hidden layer and one neuron in the output layer). The connection between the neurons have the weight w^l_{jk}, and each neuron has some bias b^l_j associated.

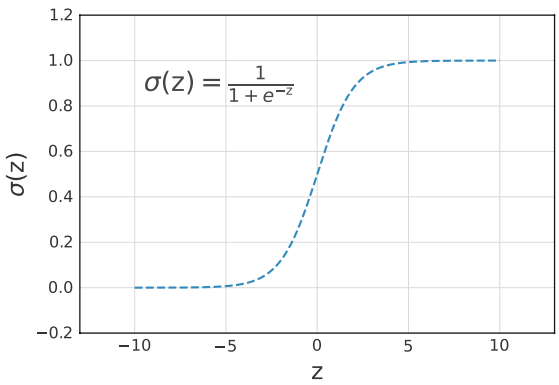


Figure 2.7: A graph of the sigmoid (Fermi) activation function.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

where the sum is over all training inputs $\boldsymbol{x}$ and n denotes the total number of training inputs. This gives simply the mean of the squared error between $\vec{a}$ and $\vec{x}$.

Using this quantity, choosing proper parameters for the network becomes an optimisation problem: to minimise the loss. In the context of ANNs, this process is called 'learning'. In practice, this can be done using an iterative algorithm like gradient descent. This means that in each iteration, the parameters are adjusted in the opposite direction of the gradient of the loss function, evaluated at the parameters' current values. This is where an infinitesimal change of the parameters will lead to the largest decrease in the loss. A parameter η gives the size of the step in this direction, called the *learning rate (LR)*. This is a so-called hyperparameter that needs to be set before training the neural network[3], which can significantly influence the training procedure. Finding a good choice for hyperparameters is a tedious task that often involves trial and error, though more methodical approaches exist. Expressed as equations the update rules for the parameters $\boldsymbol{w}$ and $\boldsymbol{b}$ are

$$w^l_{jk} \quad \rightarrow \quad \tilde{w}^l_{jk} = w^l_{jk} - \eta \frac{\partial C}{\partial w^l_{jk}} \quad \text{and} \tag{2.35}$$

$$b^l_j \quad \rightarrow \quad \tilde{b}^l_j = b^l_j - \eta \frac{\partial C}{\partial b^l_j}. \tag{2.36}$$

The derivatives $\frac{\partial C}{\partial w^l_{jk}}$ and $\frac{\partial C}{\partial b^l_j}$ can be calculated efficiently, using the backpropagation algorithm. This algorithm is necessary to make gradient descent feasible for ANNs with many parameters. It works by calculating the derivatives using the chain rule starting from the last layer, incrementally stepping through the network until reaching the first layer. Thereby redundant calculations of intermediate terms are avoided, and all the derivatives can be obtained by performing one 'backward pass'. The backpropagation algorithm is generalised by automatic differentiation which is the backbone of all modern machine learning libraries.

When training neural networks, one usually deals with lots of training data. Thus it is not feasible to calculate the gradient of the loss function for the whole training set at once. Therefore, a modification to the gradient descent algorithm, called *stochastic gradient descent*, is often used. Here the training data is split randomly into smaller chunks of data called *mini-batches*. These are then used to update the weights and biases by calculating the gradient of the loss function only for the given subset. The gradient will not be equal to what would be obtained using the full sample, but it will give a good approximation for large enough subsets. Similar to a simulated annealing algorithm, introducing this randomness can help escape from a local minimum to reach a better minimum. The size of the mini-batches is called the batch size and is another hyperparameter one has to optimise, which in turn influences the optimal choice of the LR again. One extreme would be to choose a batch size of 1, called online learning. However, to utilise the existing highly optimised linear algebra packages, it is worth considering larger batch sizes since many calculations can be performed simultaneously by writing them in matrix form. After stepping through all mini-batches and using every training sample, one *epoch* is complete. This

[3]The LR can also be varied while training the network. In general, one wants to decrease the LR with time to converge at a minimum and avoid overshooting. Another possibility is to include 'momentum', a method of varying the step size depending on the previous step.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

process will then be repeated with a new random partition of the training data until a maximum number of epochs is reached and the training is complete. When to stop the training is another hyperparameter. In general, all the hyperparameters need to be optimised in such a way to find a balance between execution speed and performance.

One additional issue to consider is overtraining. The ANN can pick up some particular features of the training data (which might be statistical and do not describe the underlying function we want to approximate). Because ANNs give are a very flexible model, they can learn these features very efficiently. This results in a bad generalisation to samples that were not used in training. To monitor this effect, it helps to split all available data into two or more independent sets and evaluate the performance with data not used for training. Overtraining can be dealt with in many ways: by using a simpler model, using more training data, stopping the training before overtraining occurs (early stopping) or implementing regularisation techniques. Overtraining can also occur when optimising the network's hyperparameters, even if done by trial and error. This is why the final performance of the ANN should ideally be evaluated on completely independent data that was not involved in the training at all.

The last thing I want to address here is an issue encountered when using the quadratic loss (Equation 2.34) and the sigmoid activation function (Equation 2.33). If the neurons initially have values very close to 0 or 1, the learning process will be slowed down (it is said that the neurons 'saturate'). This is because the derivative of the loss function is proportional to the derivative of the activation function $\sigma'(z)$, evaluated at the individual neurons (this is a simple consequence of the chain rule). The derivative of the sigmoid function is almost 0 for very large and very small values (see Figure 2.7), thus leading to tiny steps when updating the network's parameters even if the error is large. This problem can be resolved by using a different loss function called the binary cross-entropy (BCE) function and given by

$$C_{\mathrm{BCE}} = -\frac{1}{n}\sum_{\boldsymbol{x}}[\vec{y}\ln\vec{a} + (1-\vec{y})\ln(1-\vec{a})]. \tag{2.37}$$

The BCE is designed so that in combination with the sigmoid activation it effectively leads to cancellation of all the $\sigma'(z)$ terms. It still has all the necessary properties of a loss function and compared to the quadratic loss, it almost always results in a faster learning process.

The ANN described here is one of the simplest types of networks but certainly sufficient for classifying signal against some background. It is called a feedforward network because the input is passed only in one direction from the first layer to the last and no information is fed back in loops like in recurrent neural networks.

Many other interesting types of networks for supervised and also unsupervised learning exist. Different learning algorithms and techniques to improve learning exist and are continually being developed since this is an active field of research.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Chapter 3

Experimental setup

The Belle II experiment is the successor of the Belle experiment, which ran from 1999 to 2010. Its primary purpose is to continue the success story of Belle by performing high precision measurements in the heavy flavour sector while also being able to search for new physics. The goal is to collect a large data set corresponding to an integrated luminosity of 50 ab^{-1} at the $\Upsilon(4S)$ resonance. This makes Belle II the forwardmost experiment at the intensity frontier (compared to the LHC energy frontier). Belle II should achieve the high luminosity with the asymmetric electron-positron accelerator SuperKEKB, the successor of KEKB. In this Chapter, I will give some details about SuperKEKB and Belle II.

3.1 Luminosity

The notion of luminosity will be used frequently throughout this thesis. The expected number of events N for a given process are calculated by the corresponding cross section σ times the total integrated luminosity:

$$N = \sigma \times \int L(t)dt. \tag{3.1}$$

Here $L(t)$ is the *instantaneous* luminosity which is integrated over some time. For a collider with bunched beams and head-on collisions, an expression for the instantaneous luminosity is

$$L = f_{\text{coll}} \frac{n_1 n_2}{4\pi\sigma_x^* \sigma_y^*} \mathcal{F}, \tag{3.2}$$

where f_{coll} is the average collision frequency, n_1 and n_2 are the number of particles in each of the colliding bunches, σ_x^* and σ_y^* give the horizontal and vertical beam size at the interaction point, and $\mathcal{F}$ is a factor which incorporates additional effects. [8, p. 433]

When these quantities are unknown or change over time, the luminosity can be estimated using a reference process with known cross section σ_{ref}. By reconstructing this process with a

known efficiency ε_{rec} (e.g. based on simulations) the luminosity can be calculated from the number of observed events N_{obs}, using

$$\int L(t)dt = \frac{N_{\text{obs}}}{\sigma_{\text{ref}} \cdot \varepsilon_{\text{rec}}}.$$

The instantaneous luminosity is usually given in units of $\text{cm}^{-2}\text{s}^{-1}$. The integrated luminosity is commonly given in inverse barn, the unit used for cross sections ($1\,\text{barn} \equiv 1 \times 10^{-24}\,\text{cm}^2$).

3.2 SuperKEKB

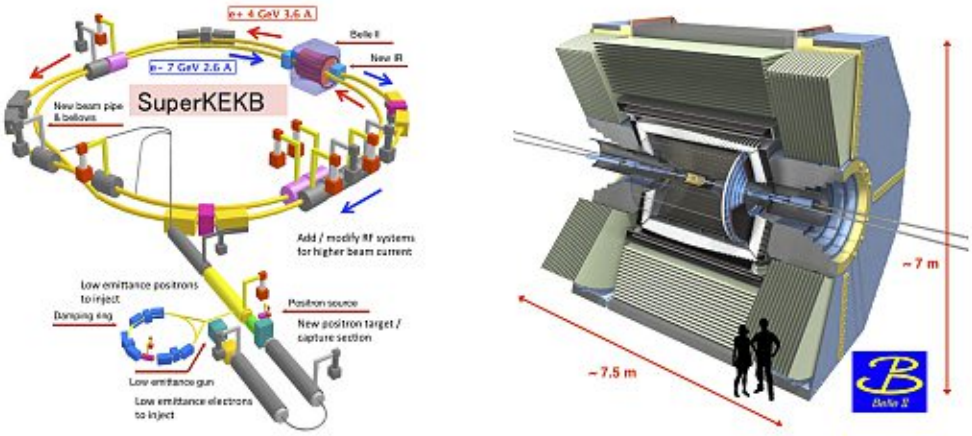


Figure 3.1: The SuperKEKB accelerator and the Belle II detector. [31]

SuperKEKB is a 3 km circumference particle collider located in Tsukuba, Japan. It accelerates e^- (e^+) beams to 7 GeV (4 GeV) and delivers collisions to the Belle II detector at $\sqrt{s} = 10.58$ GeV. The asymmetric beam energies lead to collisions where the CMS is boosted with respect to the laboratory system with a Lorentz factor $\beta\gamma = 0.28$. Currently, SuperKEKB holds the world record for the highest luminosity recorded at a particle collider. This record was set in 2020 with an instantaneous luminosity of $2.4 \times 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$. The design luminosity of $8 \times 10^{35}\,\text{cm}^{-2}\text{s}^{-1}$ (40 times of what was achieved with KEKB) is only expected to be reached in a few years.

A sketch of the SuperKEKB accelerator is shown on the left-hand side in Figure 3.1. It consists of a 600 m linear accelerator (lower side of the Figure) to produce and accelerate electron and positrons which are then injected into the high energy storage ring (HER) and low energy storage ring (LER). In order to achieve high luminosities, the particles need to be injected with low emittance[1]. This is achieved with a photo-cathode high-current radio-frequency gun [see 32, p. 13] for the electrons. The gun produces electrons with a bunch charge of 4 nC and an emittance of 10 mm · mrad.
For positron production, a higher charge (10 nC) electron beam is created with a thermionic gun which then interacts with a 14 mm thick tungsten target. The positrons are created from the

[1] A measure for the spread of the particle beam in position-momentum phase space.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

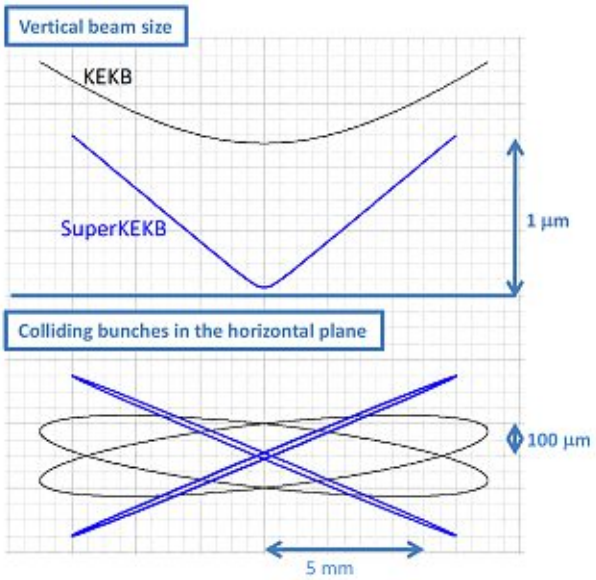


Figure 3.2: A comparison of the crab crossing scheme, which was used in KEKB from 2007 to 2010, and the crab waist scheme deployed in SuperKEKB. [32, p. 3]

resulting electromagnetic showers in the material. High emittance for the positrons is achieved through a damping ring with an energy of 1.1 GeV. [32]

The particles are then injected in the main storage ring with a circumference of 3 km. Each ring can hold around 1500 bunches, circulating with a frequency at the order of 100 kHz. Every bunch has a length of approximately 0.5 cm and contains around 3×10^{10} particles [8, p. 442]. The storage current target is 3.6 A for the LER and 2.6 A for the HER, about twice as high as in KEKB. This increase in beam current is one reason why SuperKEKB is expected to provide such a high luminosity. The other reason for this is a novel *nano-beam* scheme (or 'crab waist' scheme) for the interaction region. This scheme is based on an idea initially proposed for the SuperB[2] project [see 33, p. 131]. It utilises a large crossing angle (4.8 ° at Belle II) and highly squeezed beams with superconducting magnets at the interaction region. The bunches should only intersect in a focused region with a target width of 0.3 mm and a height of 100 nm. A comparison to KEKB, where a crab crossing scheme was used, is outlined in Figure 3.2.

3.3 The Belle II experiment

Belle II is a multi-purpose detector. Because it is operated at a CMS energy of 10.58 GeV which corresponds to the rest mass of the $\Upsilon(4S)$ the process $e^{\pm} \to \Upsilon(4S)$ is favoured. The $\Upsilon(4S)$ is an excited bound state of two b quarks ($b\bar{b}$) which decays with a probability of about 51.4 % in B^+B^- and with 48.6 % in $B^0\overline{B}^0$. This is why Belle II is also called a B factory since these B

[2]The SuperB project was another high luminosity e^+e^- collider proposed to be built in Italy but eventually cancelled in 2012.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table 3.1: The most dominant processes at Belle II with a CMS energy of 10.58 GeV. The cross sections are taken from 'The Belle II Physics Book' [36, p. 48].

physics process		cross section [nb]
$e^+e^- \rightarrow$	$\Upsilon(4S)$	1.11
	$u\bar{u}(\gamma)$	1.61
	$d\bar{d}(\gamma)$	0.40
	$s\bar{s}(\gamma)$	0.38
	$c\bar{c}(\gamma)$	1.30
	$\tau^+\tau^-(\gamma)$	0.919
	$\mu^+\mu^-(\gamma)$	1.148
	$e^+e^-(\gamma)$	~ 300

mesons can be used to study the SM with high precision. In particular, the study of CP violation and the measurement of the CKM parameters is a high priority for Belle II. The Belle experiment successfully established (together with BaBar at SLAC) the existence of CP violation in the B meson system by measuring a time-dependent asymmetry in the golden channel $B^0 \rightarrow J/\psi K_S^0$ [34, 35]. This type of measurement is also why in Belle and Belle II the CMS does not coincide with the laboratory frame. To be able to do time-dependent measurements an excellent vertex resolution is required and by boosting the B mesons, one can increase the average decay length up to a point where the detector can resolve it.

However, the study of B mesons is not the only physics that can be done with Belle II. It can also be used to study D mesons and τ leptons extensively. This becomes clear when looking at the cross sections of the most dominant processes in Belle II (see Table 3.1). In its lifetime Belle II will collect data containing information from a vast number of D mesons and τ leptons which can result in world-leading measurements in these areas.

Because of the detector's hermetic design and the precise knowledge of the initial energy and momentum, it is also possible to perform missing mass and recoil measurements. With the efficient reconstruction of low multiplicity events, Belle II is also well suited to search for new physics processes involving 'invisible' particles such as dark matter.

A detailed horizontal cross-section of the Belle II detector is given in Figure 3.3. The electron beam (going from left to right) and the positron beam (going from right to left) cross at the centre of the detector at the interaction point (IP). The coordinate system of the detector has its origin at the IP and is such that the z-axis is pointing to the right-hand side in Figure 3.3 (this is also called the forward region of the detector). The x-axis is pointing up, and the y-axis is pointing perpendicular to the plane, resulting in a right-handed coordinate system. The polar angle θ and the azimuthal angle φ are defined in the usual way.

Looking from the IP radially outward there are six subdetectors: the Pixel Detector (PXD), the Silicon Vertex Detector (SVD), the Central Drift Chamber (CDC), the Time-of-propagation counter (TOP), the Electromagnetic Calorimeter (ECL) and the K_L - Muon Detector (KLM). Because of the boosted CMS the detector has no perfect forward-backwards symmetry since

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

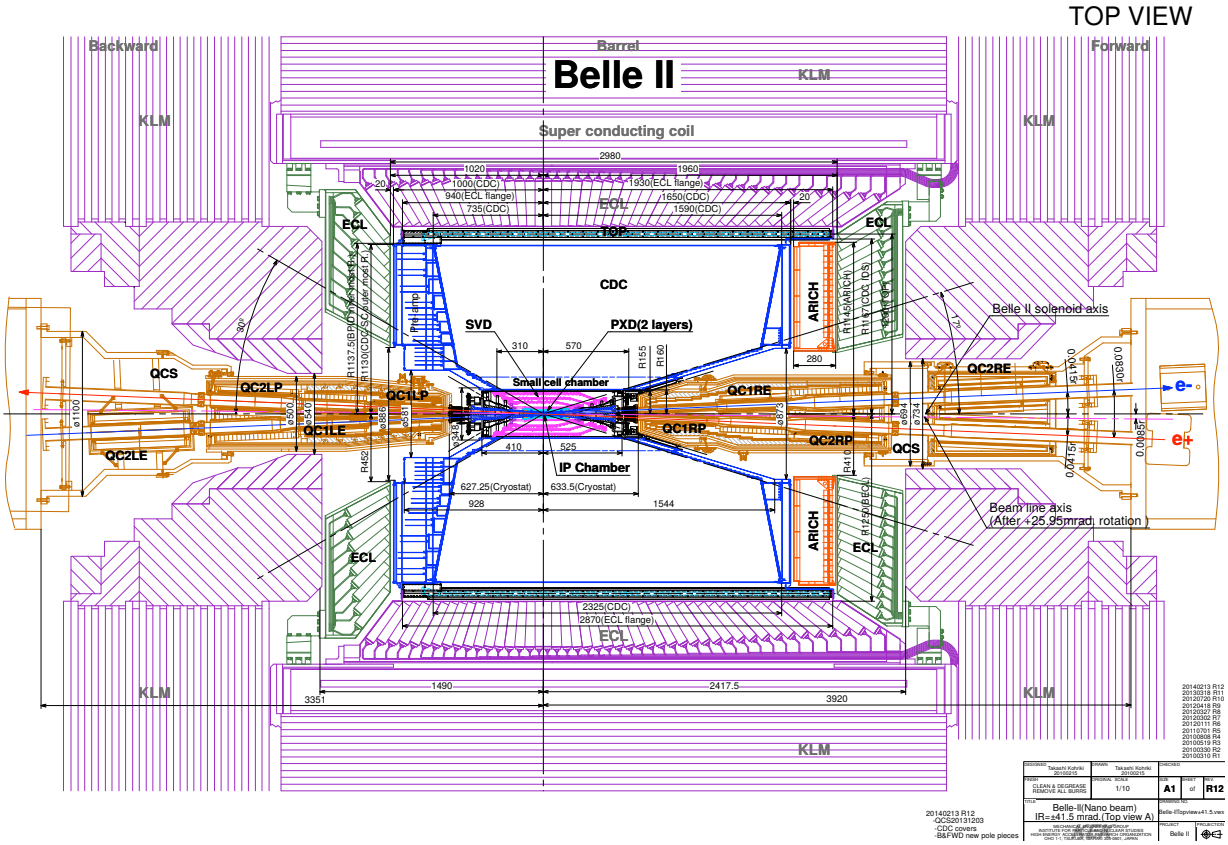


Figure 3.3: A cross-section of the Belle II detector, showing the interaction point and all the subdetectors from above. [36]

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

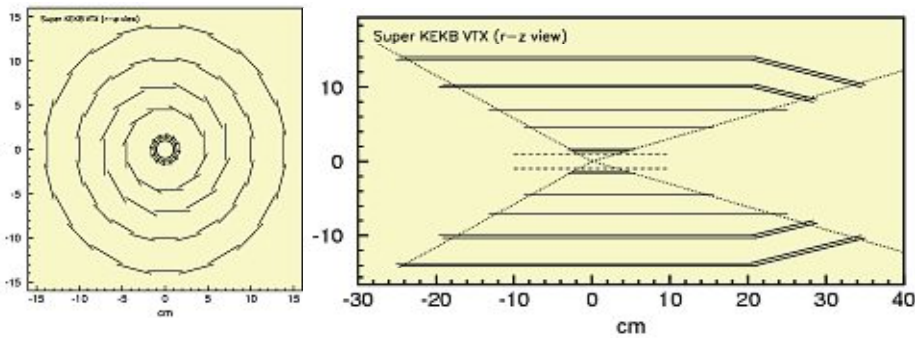


Figure 3.4: The VXD viewed from the front (r-φ view) and the side (r-z view). The four outer layers are the SVD and the two inner layers the PXD. The innermost circle on the left and the dashed lines on the right correspond to the beam pipe. [36]

more particles are expected in the forward region. This is also why an additional detector is only located in the forward part: the Aerogel Ring Imaging Cherenkov counter (ARICH).

In the detector's inner regions is a homogeneous 1.5 T magnetic field along the z direction, generated by a superconducting solenoid between the ECL and the KLM. Because the Belle II detector will face higher event rates and increased backgrounds, almost all the detector components were considerably upgraded from the predecessor experiment. In the following, I will give a short description of each subdetector's operating principle and purpose. More details can be found in the 'Belle II Technical Design Report' [37].

3.3.1 Subdetectors

The Vertex Detector

The Vertex Detector (VXD) consists of PXD and SVD which are the innermost detectors right around the 10 mm beryllium beam pipe. The PXD consists of two layers at $r = 14$ mm and $r = 22$ mm (currently the second layer is only partially installed). It uses DEPFET (Depleted p-Channel Field Effect Transistor) sensors which can be made very thin ($\sim 50\,\mu$m) to minimise multiple scattering. In total, the PXD will have around 8 million pixels.
The SVD consists of four layers located at $r = 38$ mm, $r = 80$ mm, $r = 115$ mm and $r = 140$ mm. Each layer is equipped with double-sided silicon strip sensors to avoid a large number of readout channels, as they are present for pixel detectors. Figure 3.4 illustrates the position of the VXD modules in the r-φ and the r-z plane. In Figure 3.5 a 3D rendering of the full VXD with the PXD in the centre is shown.

The primary purpose of the vertex detectors is to measure and reconstruct decay vertices with high resolution. The particle tracks are reconstructed in the CDC and then extrapolated to the PXD with information from the SVD. The SVD is also able to reconstruct particles that decay outside of the PXD such as K_S^0 and also low momentum tracks (down to $\mathrm{p_t} \sim 10$ MeV) that do not reach CDC.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

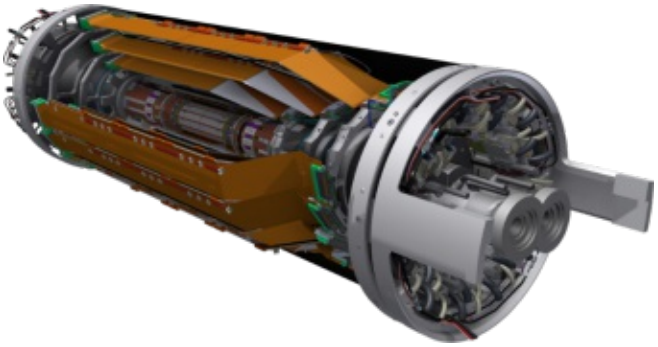


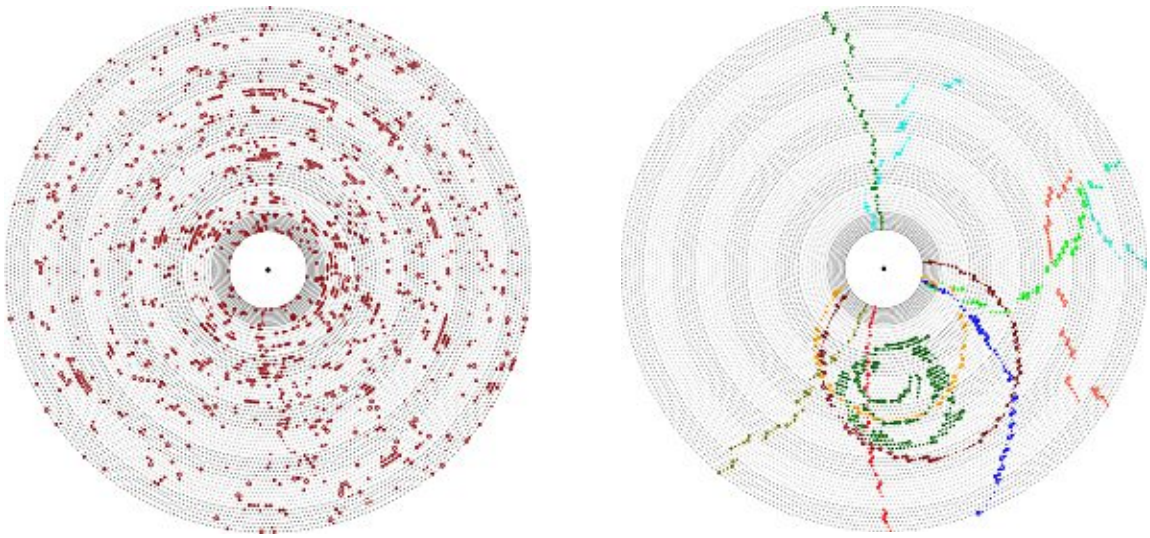
Figure 3.5: A 3D rendering of the VXD, showing the PXD in the centre close to the beam pipe and the SVD layers arranged around it. [38]

The Central Drift Chamber

The CDC is the main tracking device of Belle II, detecting traces of charged particles. It is a large volume drift chamber consisting out of 14 336 wires (30 µm diameter) arranged in 56 layers from $r = 160\,\text{mm}$ up to $r = 1130\,\text{mm}$. The layers are grouped in super-layers which consist of layers either in axial orientation (aligned with the magnetic field) or in stereo orientation (skewed to the magnetic field). This gives access to the z position of the tracks, enabling a reconstruction in three dimensions. The gas in the CDC is a mixture of He and C_2H_6 (50:50). The largest drift times are about 350 ns for a 17 mm drift cell. Figure 3.6 shows the wires of the CDC in the r-φ plane with the strength of the measured current indicated as circles ('hits') around the wires. The image in 3.6a is based on a simulation of the expected beam background for the target instantaneous luminosity. The image in 3.6b is a simulated event where the hits are categorised to belong to the same track. This shows the challenges one faces when the task is to accurately and reliably find and reconstruct tracks using the CDC measurements in a high background environment.

In Belle II, the CDC has multiple purposes. It reconstructs tracks of charged particles which gives information about the location but also their momenta. This is because the curvature of the track in the magnetic field depends on its momentum. With the curvature and the exact value of the magnetic field inside the CDC the particle's momentum can be inferred. The CDC also provides particle identification about the tracks, using the characteristic energy loss $\frac{dE}{dx}$ at a given momentum, which can be measured in the gas volume. This is especially important for low momentum tracks that do not reach the outer detectors designed for particle identification. The CDC is also used to trigger relevant events based on charged particles (the trigger system is described in Section 3.3.3). It can particularly provide the z coordinate of the tracks at $r = 0$, which is crucial to identify background events that do not originate from the IP. Field Programmable Gate Arrays (FPGAs) can be used to implement a fast reconstruction of these three-dimensional tracks. With this information, background events can be rejected while keeping relevant physics events and maintaining an acceptable trigger rate.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



(a) The hits expected from beam background for the target luminosity. [38]

(b) A physics event and the resulting signal in the CDC. The colour indicates which hits were caused by the same particle. [39, p. 57]

Figure 3.6: Simulated CDC measurements shown in the r-φ plane.

Particle identification detectors

Although all detectors contribute to the identification of particles, the TOP and ARICH are usually referred to as the particle identification (PID) detectors. Their main task is to improve the discrimination of kaons from pions in a momentum range of 0.4 GeV/c up to about 4 GeV/c and to separate pions, muons and electrons at low momentum (below 1 GeV/c). Both detectors are Cherenkov detectors, which are based on the fact that a charged track with a velocity β that exceeds the speed of light in a medium with refractive index n emits Cherenkov radiation at a characteristic angle given by

$$\cos\theta_c = \frac{1}{n(\lambda)\beta}. \tag{3.3}$$

Particles with the same momentum but distinct masses will have different velocities, thus emitting photons at different Cherenkov angles. Additionally, there is a velocity threshold below which the particles will not emit any Cherenkov radiation, so there will be a momentum range (depending on the refractive index) where pions radiate, but kaons do not, providing a good separation between the two. The method of how the TOP and the ARICH measure the Cherenkov angles differs quite significantly.

The TOP is located in the barrel region. One of the requirements for this detector was to make it as thin and homogeneous as possible to improve the ECL response to electromagnetic particles. This was achieved with 16 quartz bars (2.6 m long and 2 cm), arranged at $r = 120$ mm to cover the whole barrel.

When particles traverse the bar and emit Cherenkov photons, they get reflected inside the

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

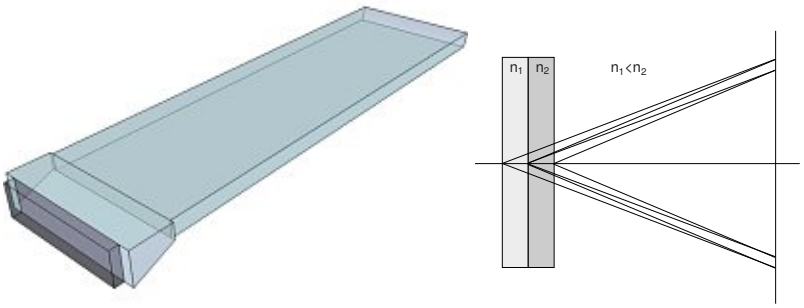


Figure 3.7: Left side: One of the quartz bars of the TOP detector. The Cherenkov photons get reflected inside the bar until they reach the detector located at one side of the bar. Right side: A sketch of how two aerogel layers with different refractive indices can be chosen to give overlapping rings from the Cherenkov photons at the detector plane. [36]

material (with total internal reflection) until they reach one end of the detector where a fast position-sensitive micro-channel plate photomultiplier tube (PMT) is located. A different θ_c will lead to different path lengths of the photons and hence to a different time of propagation. By measuring this time of propagation together with the Cherenkov photon's impact position at the photodetector, the angle θ_c can be reconstructed. A graphic of one of the quartz bars is shown in Figure 3.7 on the left.

For the ARICH aerogel tiles are arranged in a disk at the forward endcap region of Belle II. These tiles serve as a radiator for Cherenkov photons which then propagate through 20 cm of an expansion volume after which single-photon sensitive detectors are located. The Cherenkov angle can be reconstructed directly through the radius of the ring formed by the photons at the detector plane. With a thicker aerogel layer, the number of detected photons increases but so does the uncertainty of the emission point, which determines how well the Cherenkov angle can be resolved. By placing two aerogel layers with different refractive indices behind each other, the number of emitted photons can be increased without spreading out. If the refractive indices are well-chosen the two rings produced by the Cherenkov photons will overlap at the photon detectors. This situation is sketched on the right side in Figure 3.7. At Belle II it was found that two 2 cm thick layers with refractive indices of $n = 1.045$ and $n = 1.055$ give the best results.

The Electromagnetic Calorimeter

The ECL consists of 8736 thallium-doped caesium iodide CsI(Tl) crystals distributed around the barrel (ranging from $r = 125$ cm to $r = 162$ cm) and at the endcaps (at $z = -102$ cm and $z = 196$ cm). Each crystal is about 6 cm $\times$ 6 cm wide and 30 cm long. The shape is a truncated pyramid, and the crystals are oriented in such a way that the long side points towards the IP (see Figure 3.3).

The crystals were reused from Belle and chosen because CsI(Tl) is one of the brightest scintillators, has a short radiation length and good mechanical properties. For scintillation light readout two photodiodes are glued to the back of the crystal. The readout electronics were upgraded in

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Belle II with waveform sampling, which allows measuring the signal waveform. Since the scintillation response is distinct for particles with different ionisation properties, this allows separating the signals after recording the data (pulse shape discrimination (PSD)).

The ECL's main purpose is to detect photons and electrons (or positrons), which will deposit all their energy in the crystals. The efficient detection of photons (energy, position and angle) in an energy range from 20 MeV to 4 GeV is essential for many physics programs at Belle II. At Belle, the energy resolution varied from 4 % at 100 MeV with an angular resolution of 13 mrad to 1.6 % at 8 GeV with an angular resolution of 3 mrad [40, p. 31].
The performance is very similar in Belle II, but it is expected to worsen with increasing background levels. The reconstruction of π^0 from these detected photons gives a good indication of the ECL's performance. At Belle II the mass resolution is currently around 5.3 MeV/c (4.5 MeV/c at Belle). Other tasks of the ECL are to provide information for particle identification and the trigger system.

The K_L - Muon Detector

The KLM consists of 14-15 alternating layers of iron plates and active detector material located around the barrel and at the endcaps (see Figure 3.3). At the barrel, the KLM extends from around 200 cm to 250 cm. The iron plates are 4.7 cm thick, corresponding to 3.9 interaction lengths λ_0, to detect particles that pass through the inner detectors. They also serve as a flux return of the magnetic field since they are located outside of the solenoid.
The detector layers are either Resistive Plate Chambers (RPCs) or scintillator strips. RPCs are a special type of proportional counter based on an ionisation chamber principle and were already used in Belle. At Belle II, they are only used for the outer barrel layers because the dead time of the RPCs is too large to handle the large number of neutrons produced in electromagnetic showers due to the increased background rates. Therefore scintillator strips with silicon photomultipliers are used instead for the two inner barrel layers and the endcaps.

Together with the ECL which provides 0.9 λ_0 the KLM is used to detect K_L^0 which shower hadronically. Muons do not shower hadronically and can penetrate many KLM layers depending on their momentum and electromagnetic energy deposition. Most muons below ~ 0.7 GeV/c do not reach the KLM at all and muons with more than 1.5 GeV/c usually traverse through all layers and escape the Belle II detector. Using extrapolated tracks from the CDC with a muon hypothesis the expected patters in the KLM are compared to the actually observed ones, which gives the likelihood for the track to be a muon.

3.3.2 Beam induced backgrounds

As already mentioned before, Belle II has to deal with higher background levels than its predecessor because of the high luminosity goals. The background rates depend on many factors like luminosity, beam current, beam size, number of bunches or vacuum conditions and can be reduced by a well-designed interaction region and shielding. The most dominant sources for what is considered to be beam induced background are mentioned in the following. [described in 36, p. 50]

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

- Touschek scattering: Two particles Coulomb scatter inside the bunch and hit the beam pipe. The subsequent particle shower can reach the Belle II detector. The scattering rate is inversely proportional to the beam energy cubed, thus much higher for the LER.

- Beam-gas scattering: A beam particle interacts with a residual gas molecule in the vacuum tube and hits the beam pipe, creating shower particles. This can happen via Coulomb scattering (change of momentum) or bremsstrahlung (energy loss).

- Synchrotron radiation: When the charged beams are accelerated in focusing magnets' magnetic field, they can emit synchrotron radiation. The emitted photons have energy at the keV scale and can damage the inner detectors. The inner surface of the beam pipes near Belle II is specially designed to prevent this. Synchrotron radiation is proportional to the beam energy squared, thus more considerable for the HER.

- Radiative Bhabha scattering: Photons from the large cross section process $e^+e^- \to e^+e^-\gamma$ propagate along the beam axis and interact with the magnets' iron. This results in many neutrons produced via the giant resonance effect, which is the main background for the KLM. Additionally, the electron and positron involved in the scattering loose energy and can hit the wall of the magnets and produce electromagnetic showers.

- Two-photon process: Similarly to Bhabha scattering the QED process[3] $e^+e^- \to e^+e^-e^+e^-$ has a large cross section and the primary electron-positron pair can hit parts of the detector when scattered with a large angle or when losing enough energy. The other electron-positron pair will have a low momentum and can spiral in the magnetic field, leaving many hits in the innermost detectors.

3.3.3 Trigger system

The trigger system is responsible for identifying events for physics interest during data taking. The hardware-based trigger system ("L1") needs to provide a decision whether to keep an event or not in less than 5 µs while still being able to separate two events within 200 ns. This is because not all events can be processed by the data acquisition (DAQ) system, which has a limited rate of around 30 kHz. At the target luminosity the event rate for a process with a cross section of 1 nb is 800 Hz (Bhabha events alone have a cross section of around 300 nb - see Table 3.1). Most of these events are from luminosity dependent beam backgrounds such as (radiative) Bhabha or two-photon processes with two or less charged particles in the detector. This results in a challenge, especially for low-multiplicity physics studies where the topologies can be very similar. After passing the low-level trigger L1, events are again evaluated with the software-based high level trigger (HLT). The HLT has access to more advanced reconstruction algorithms (the same used to perform an analysis) with information from all subdetectors except the PXD[4]. The HLT must reduce the event rates to around 10 kHz in order to make data storage possible. [36, Section 3.7]

[3] It is called a two-photon process because there are two photons involved when drawing a Feynman diagram.

[4] This is because the data rates from the PXD are too high to be processed. Only some regions of interest are read out, based on tracks reconstructed at the HLT level.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table 3.2: The names and activation conditions for a few selected trigger lines.

`fff`	number of full 2D tracks ≥ 3
`ffo`	number of full 2D tracks ≥ 2 CDC r-φ opening angle $> 90\,^\circ$ no ECL Bhabha veto signal
`ff30`	number of full 2D tracks ≥ 2 CDC r-φ opening angle $> 30\,^\circ$ no ECL Bhabha veto signal
`hie`	ECL total energy > 1 GeV no ECL Bhabha veto signal

Many trigger lines exist, but only a few of them will be used explicitly in this thesis. Their names and trigger conditions are listed in Table 3.2. A 'full 2D track' is a track that fully traverses all CDC layers, reconstructed only with 2D information from the projection to the r-φ plane. In addition, all the trigger lines have the requirement that no injection veto is present from SuperKEKB to prevent the DAQ when there is larger noise right after beams are injected. Only these trigger lines were used for now because they are currently best studied and well understood. Additionally, they were fully activated for the whole data taking period (except `ff30` which was not present before 2020). They provide a reasonable efficiency of around 90 % for low multiplicity events with tracks in the barrel region. Unfortunately, these triggers do not perform well in the endcaps. The requirement on the opening angle is currently also a restriction for the high recoil mass region in the Z' analysis (i.e. large Z' mass), which is why the `ff30` trigger was introduced.

Other more recent triggers that are especially interesting to be used for future studies are triggers based on the KLM, short track triggers for tracks that do not fully traverse the CDC (e.g. tracks pointing in the endcap region) and triggers based on 3D track information obtained with neural networks, called neuro triggers.

3.3.4 Data taking periods

The first data taken with Belle II was during a detector commissioning period called *Phase 2* (February 2018 - July 2018). In Phase 2 all subdetectors except for the vertex detectors were fully present, and its main aim was to start the operation of SuperKEKB and make sure that it was safe to install the VXD [36, p. 44].
Additionally to these commissioning tasks, it was possible to collect around 510 pb^{-1} of data for calibration and first physics analyses without the need for the VXD.

First collisions with the full Belle II detector started on March 25th 2019. The phase of the experiment since then is dubbed *early Phase 3*. This name is used to refer to the current geometry of the detector, e.g. in simulation. The 'early' is because currently a part of the PXD is still missing and the full PXD will only be installed during a shutdown in 2022. The data taking is further divided into periods called 'experiments'. These experiments are numbered consecutively.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

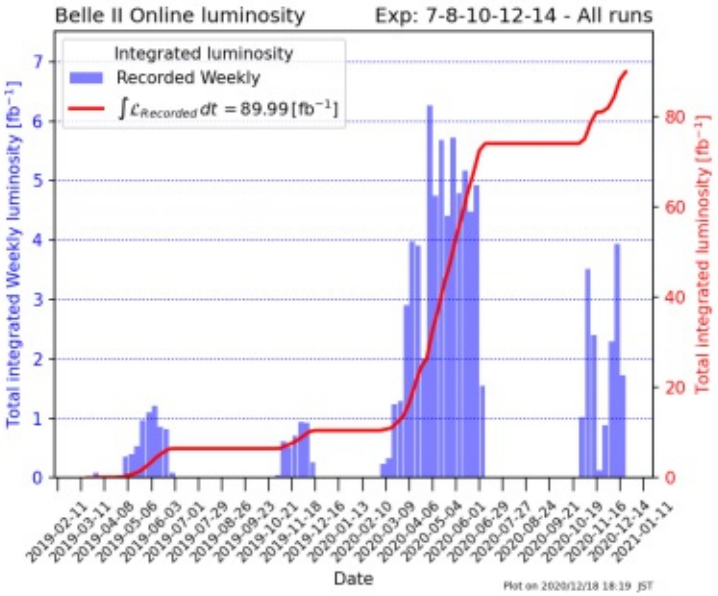


Figure 3.8: The weekly recorded luminosity at Belle II for the years 2019 and 2020 overlayed with the total integrated luminosity. Out of the 90 fb^{-1} only around 63 fb^{-1} pass the quality requirements for physics analyses. [41]

For example, in the fall of 2020, data for experiment 14 was taken. The for physics relevant data taking periods (experiment 7, 8, 10, 12 and 14) and the amount of data taken in terms of luminosity can be seen in Figure 3.8. This is shown in periods of one week together with the overall recorded luminosity. Currently, around 90 fb^{-1} of data are collected in early Phase 3, from which only $\sim$ 63 fb^{-1} pass the requirements to be used in physics analyses. Why events are excluded from an analysis can have many reasons, but most commonly runs are flagged as 'bad' if one of the detector's subsystems is not working as expected.

3.4 Software

At the core of Belle II's software is the Belle II analysis software framework (basf2) [42]. The framework works by processing modules in sequential order. These modules can be written either in Python or C++ and perform a task using information stored in a common object store. The order and specific configuration of the modules is defined using Python scripts called steering files.

The modules contain the reconstruction algorithms which are used to process the data. Initially, the raw data with detector information like track hits and calorimeter clusters is used to reconstruct higher-level objects, like charged tracks. On an even higher level, these objects can then be used to calculate properties of hypothesised particles (like the four-momenta) and information about the full event (like the event shape, missing mass or recoil quantities). The final objects are stored in mDST (mini Data Summary Tables) files. [see 36, Section 5]

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

The basf2 is updated continuously with better algorithms, new functionalities and bug fixes. If not stated otherwise, the software release used in this thesis is 05-01-05.

3.5 Simulation

Simulation is essential to understand the detector response for different physics events. It is usually done in three steps: event generation, simulation of detector interaction and digitisation of the detector response.

In event generation, the production points, four-momenta and decays of particles are generated for a specific process based on the SM or a new model that could describe BSM physics. Many different generators are used to perform this task. Typically the generation of different processes requires different generators. For example, taupair events are generated with `KKMC` while the subsequent decay of the τ leptons is handled by `TAUOLA`, `EvtGen` is used for decaying B and D mesons, `BABAYAGA.NLO` for generating $e^+e^- \to e^+e^-(\gamma)$ events and `MadGraph` for BSM processes [more information about generators in 36, Section 4.3].

The simulation of the particle interactions with the detector is based on `Geant4` [43], which uses Monte Carlo methods to generate realistic trajectories, secondary particles and hits (energy deposits) in the detector. It uses the exact geometry and composition of the Belle II detector.

The next step is to acquire the detector response for each subdetector. This is done by simulating the physics processes in the detectors under consideration of their geometry, electronic effects, and random noise. The output of this simulation is a detector signal. This translation from the `Geant4` hits to the detector digits is done by individual modules for each detector. [37, Section 14.4.6]

This simulated data can be analysed and processed in the same way as real (raw) data with the benefit of knowing which process or particle corresponds to the signal in the detector. A large amount of this *simulated data* or MC are required to approximate the distributions from the underlying model. The simulation will never be perfect, always resulting in some discrepancies between real data and simulated data, which must be carefully evaluated.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Chapter 4

Particle identification studies

In this chapter, I will describe how taupair events ($e^+e^- \to \tau^+\tau^-$) can be used to study particle identification performance. In particular, the pion identification efficiencies as well as $\pi^\pm \to \ell$ misidentification ('mis-id') rates ($\ell \in \{e^\pm, \mu^\pm\}$) are calculated and compared between data and MC. This is done for different cases in bins of momentum and polar angle to provide results in a channel independent way. The obtained data/MC ratios and estimated systematic uncertainties can be used in physics analyses to correct discrepancies between data and simulation.

4.1 Methodology

To calculate the mentioned efficiencies and mis-id rates, a clean pion sample is needed. In taupair events, this can be achieved by selecting pions from the 3-prong decay[1] of one of the tau leptons. The other τ is required to decay via a 1-prong channel, resulting in the 3-prong 1-prong topology shown in Figure 4.1 for the case that the τ decays to a muon on the 1-prong side. The reason for this is that because of the distinctive event shape, it is possible to select this topology from the data without much additional background.

The tracks on the 3-prong side are expected to be mostly pions since the branching fraction of $\tau \to \pi\pi\pi + n\pi^0$ constitutes about 95% of all 3-prong decays (see Figure 4.2). The rest consists mostly of decays with one or two pions and additional kaons. Counting only the expected tracks on the 3-prong side, this results in about 98% pions [8].

4.2 Particle identification variables

The particle identification variables at Belle II are based on likelihood ratios. Each subdetector (det $\in$ {SVD, CDC, TOP, ARICH, ECL, KLM}) provides information that can be used to identify a particle. For every subdetector a likelihood

$$\mathcal{L}_{\text{det}}(i) \coloneqq L_{\text{det}}(\boldsymbol{x}|i) \quad \text{with} \quad i \in \{e, \mu, \pi, K, p, d\} \tag{4.1}$$

[1] The term n-prong means that the τ decays to n charged particles.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

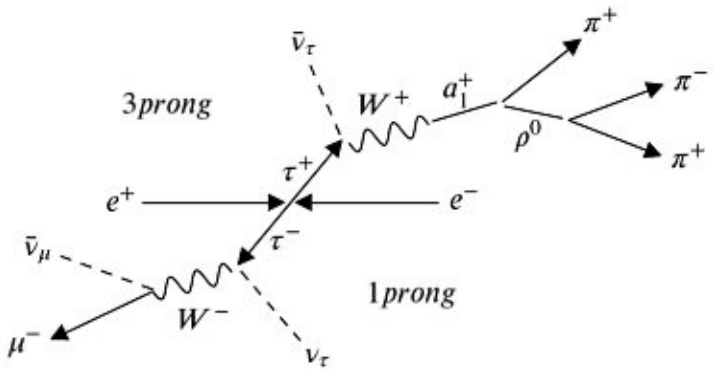
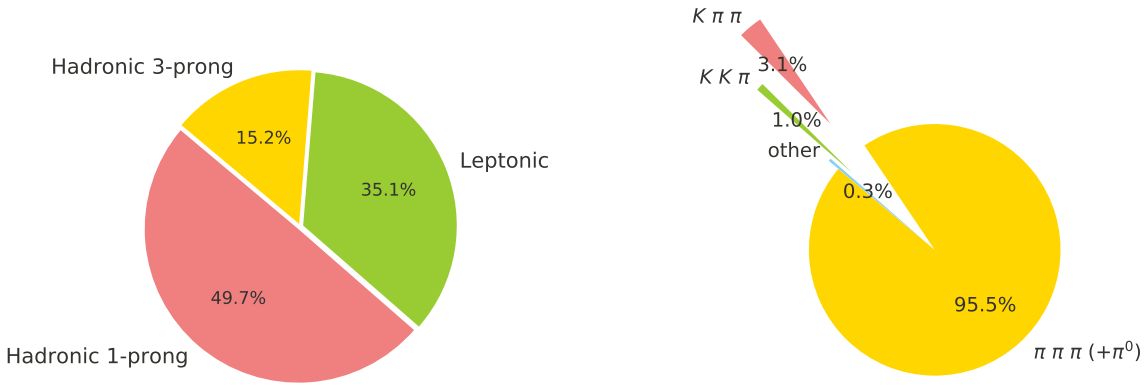


Figure 4.1: An example of a taupair 3-prong 1-prong decay.



(a) Relative branching fractions of all tau decays. About 15.2 % of tau leptons decay into three charged particles.

(b) Relative branching fractions of only 3-prong tau decays. About 98 % of all decay products are charged pions.

Figure 4.2: A pie chart showing the relevant branching fractions of tau leptons.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

is defined with a PDF either from simulation, analytical calculation or directly from data. The considered particle hypothesis i correspond to the long-lived charged particles produced in Belle II: electrons (e), muons (μ), pions (π), kaons (K), protons (p) and deuterons (d). If the individual likelihoods from the subdetectors are independent, the overall likelihood for hypothesis i can be expressed as the product $\prod_{\text{det}} \mathcal{L}_{\text{det}}(i)$. Using this product a likelihood ratio can be defined as

$$\text{globalID}\,(i) = \frac{\prod_{\text{det}} \mathcal{L}_{\text{det}}(i)}{\sum_{i'} \prod_{\text{det}} \mathcal{L}_{\text{det}}(i')}, \tag{4.2}$$

where the sum is over all particle hypothesis, and i is the hypothesis one wants to test. This ratio can be interpreted as a probability using Bayes' theorem (see Section 2.4.8):

$$P(i|\boldsymbol{x}) = \frac{P\,(\boldsymbol{x}|i)\,\pi\,(i)}{\sum_{i'} P\,(\boldsymbol{x}|i')\,\pi\,(i')}. \tag{4.3}$$

With $P(\boldsymbol{x}|i) = \prod_{\text{det}} \mathcal{L}_{\text{det}}(i)$ and assuming an equal prior probability $\pi(i)$ for each hypothesis this directly yields 4.2.

The ratio in Equation 4.2 is what defines the *global PID* variables in Belle II. These variables are named according to the test hypothesis (e.g. muonID, electronID, pionID) and have a value between 0 and 1 for each track. If no information is available for a particle hypothesis from a subdetector, this particular likelihood is set to 1. Additionally, there are *binary PID* variables, defined as

$$\text{binaryID}\,(i,j) = \frac{\prod_{\text{det}} \mathcal{L}_{\text{det}}(i)}{\prod_{\text{det}} \mathcal{L}_{\text{det}}(i) + \prod_{\text{det}} \mathcal{L}_{\text{det}}(j)}, \tag{4.4}$$

that discriminate only between two particle hypothesis i and j. Often binary PID variables perform better when only two particle types need to be separated, at the cost of higher mis-id rates for other particle types.

4.3 Monte Carlo and data samples

The MC samples used in this study were produced in an official Monte Carlo campaign by the data production group of Belle II. These samples are periodically reproduced with new software releases. The latest campaign currently available is called MC13a and was produced with basf2 release 04-00-03. Only certain processes that could constitute a background for this analysis were considered. The cross section used to generate these samples and the corresponding luminosity is shown for each process in Table 4.1.

The collision data used for this study is all the data currently available from experiment 7, 8, 10, 12 and 14. Selecting only high-quality data collected at $\sqrt{s} = 10.58\,\text{GeV}$ this corresponds to $62.8\ \text{fb}^{-1}$.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table 4.1: The simulated processes with their cross section and the corresponding luminosity of the generated samples (MC13a). If not suggested otherwise all particles are to be understood as charged particles such that the overall charge sums to zero.

Process	cross section [nb]	MC luminosity [fb^{-1}]
$ee \to \tau\tau$	0.919	100
$ee \to ee(\gamma)$	294	100
$ee \to eeee$	39.7	1000
$ee \to ee\mu\mu$	18.9	1000
$ee \to \mu\mu(\gamma)$	1.15	1000
$ee \to d\bar{d}$	0.401	1000
$ee \to c\bar{c}$	1.33	1000
$ee \to s\bar{s}$	0.383	1000
$ee \to u\bar{u}$	1.61	1000
$ee \to B^+B^-$	0.54	1000
$ee \to B^0\overline{B}^0$	0.51	1000

4.4 Event selection

Initial selections

To select the desired events only tracks for which the distance of the point of closest approach to the interaction point in the r-phi plane (dr) is less than 1 cm and the relative distance in the z-direction (dz) is between −3 cm and 3 cm are considered. Furthermore, three charged tracks are required to be in one hemisphere while only one is in the other. The hemispheres are defined by the thrust vector $\hat{t}$ obtained by maximising the thrust value

$$V_{\mathrm{thrust}} \overset{\max}{=} \sum_i \frac{|\vec{\mathrm{p}}_i \cdot \hat{t}|}{\sum_j |\vec{\mathrm{p}}_j|}, \tag{4.5}$$

where $\vec{\mathrm{p}}$ denotes the three-momentum, and the sum is over all reconstructed particles. Additionally, the sum of all charged tracks is required to be zero, and the track on the 1-prong side has to have associated CDC hits, an ECL cluster and $\mathrm{p_t} > 0.1$ GeV/c. Photons are reconstruction from ECL clusters consistent with a photon hypothesis and minimum energy of 200 MeV. Only clusters within the CDC acceptance region in θ from 17° until 150° were considered. Neutral pions are reconstructed from photons passing the same selections except with an energy threshold of only 100 MeV. The invariant mass of the two photons is required to be between 0.115 MeV and 0.152 MeV to match the mass of the π^0.

Multivariate selections

After these initial selection of the events, further background reduction is achieved with two multivariate selections, described in the following.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

The first is applied in the space spanned by the two possible combinations of the $\pi^+\pi^-$ invariant mass from the 3-prong τ decay products, $m_{\pi\pi(1)}$ and $m_{\pi\pi(2)}$. The distribution of events in this combination of variables is also called a Dalitz plot. This selection targets events where the invariant mass of one of the two pion pairs is consistent with the ρ^0 resonance at around 770 MeV. Initially, only events where both invariant masses are larger than 0.31 GeV/c^2 are selected.

The multivariate selection is computed by minimising a slightly altered form of the FOM described in Section 2.4.7 (Equation 2.29). The FOM can be changed to give more weight to background events to optimise for a purer signal selection. The defining equation for this altered formula is

$$\mathrm{FOM}_{\mathrm{pure}} = \frac{-N_{\mathrm{sig}}}{\sqrt{N_{\mathrm{sig}} + s \cdot N_{\mathrm{bkg}}}}. \tag{4.6}$$

Here N_{sig} is the number of signal events which in this case is chosen to be all taupair events with a 1-prong 3-prong topology where the 3-prong tau decays as $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau + n\pi^0$ (or the charge conjugated decay). N_{bkg} is the number of background events which are all events that are not defined as signal events. This number is scaled by a factor s, which was chosen in the present case to be 100.

The FOM optimisation is done in 50×50 dimensions, corresponding to the binning in the 2D plane. Each dimension has two possible states (1 and 0), corresponding to the bin's inclusion in calculating the overall FOM. The optimal configuration of all the possible states to minimise the FOM is found with a simulated annealing algorithm. Additionally, a small positive coupling between neighbouring bins is introduced to obtain a connected region in the plane as an optimised result. The border of this region is then fitted with polynomials obeying special symmetry requirements, namely the reflection symmetry at the diagonal $m_{\pi\pi(1)} = m_{\pi\pi(2)}$. For the optimisation a larger, fully independent sample of taupair MC corresponding to 500 fb^{-1} was used in addition to the other processes in Table 4.1. The contribution of the different processes to N_{sig} and N_{bkg} was scaled according to their luminosity.

Figure 4.3a shows the signal distribution, Figure 4.3b the background distribution, and Figure 4.3c the purity

$$p = \frac{N_{\mathrm{sig}}}{N_{\mathrm{sig}} + N_{\mathrm{bkg}}}, \tag{4.7}$$

for each bin in the plane where the first selection is applied. The ρ^0 resonance is visible in the signal distribution. The white lines are the polynomials enclosing the selected region. The initial value of the temperature parameter T in the simulated annealing was set to 150. Each iteration the temperature was decreased by a factor of $\frac{1}{1.01}$, stopping after 1000 iterations. The value of the FOM after each iteration can be found in Figure 4.3d.

The second multivariate selection is applied in the 2D space spanned by the thrust value and the total visible energy in the CMS. Here initially only events with a thrust value between 0.91 and 0.986 and a visible energy between 2.5 GeV and 11 GeV are selected. The subsequent optimisation procedure is the same as for the first selection. The results are shown in Figure 4.4.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

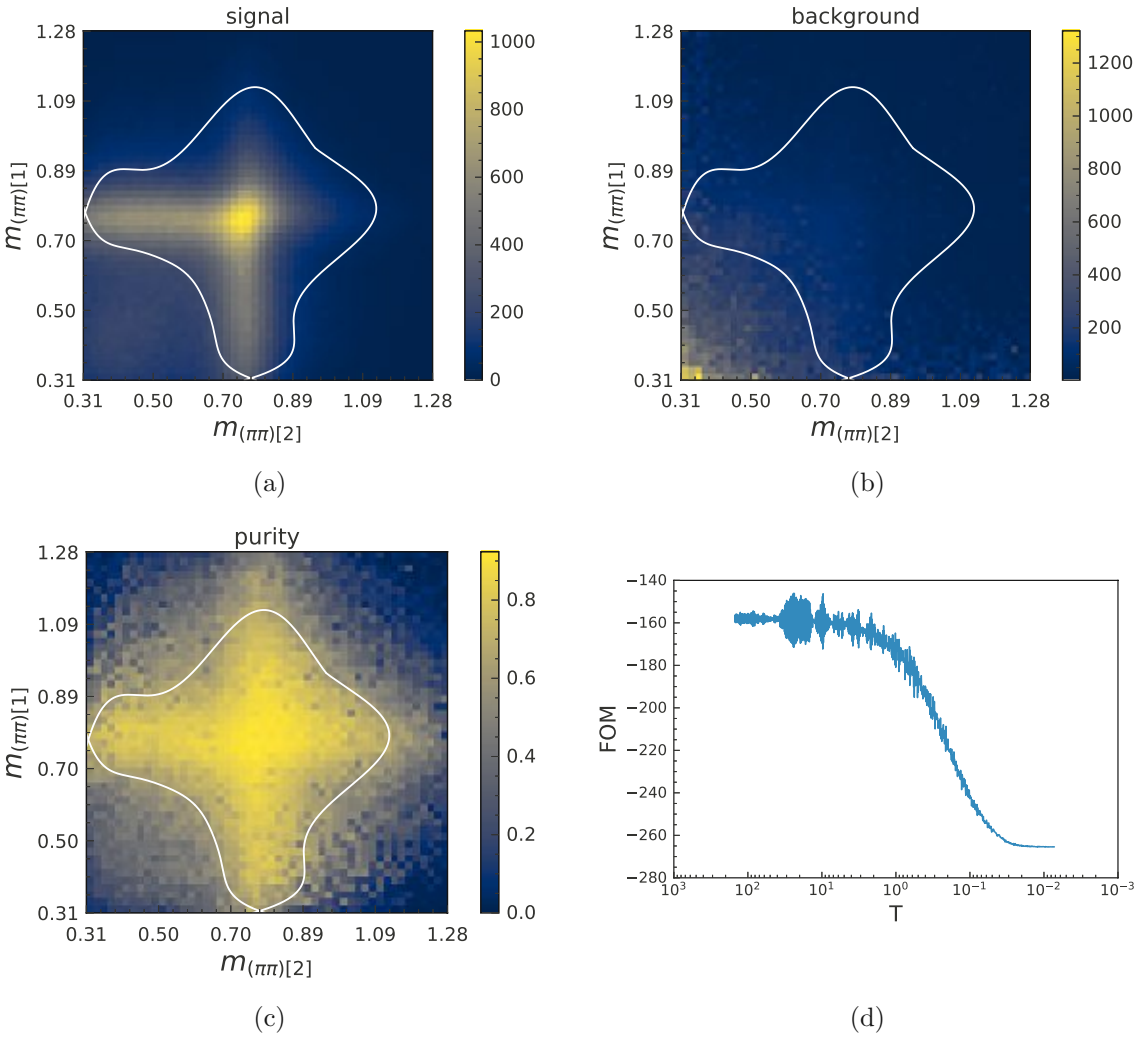


Figure 4.3: Optimisation of the first selection in the 2D plane spanned by the two possible combinations of the $\pi^+\pi^-$ invariant mass from the 3-prong τ decay products.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

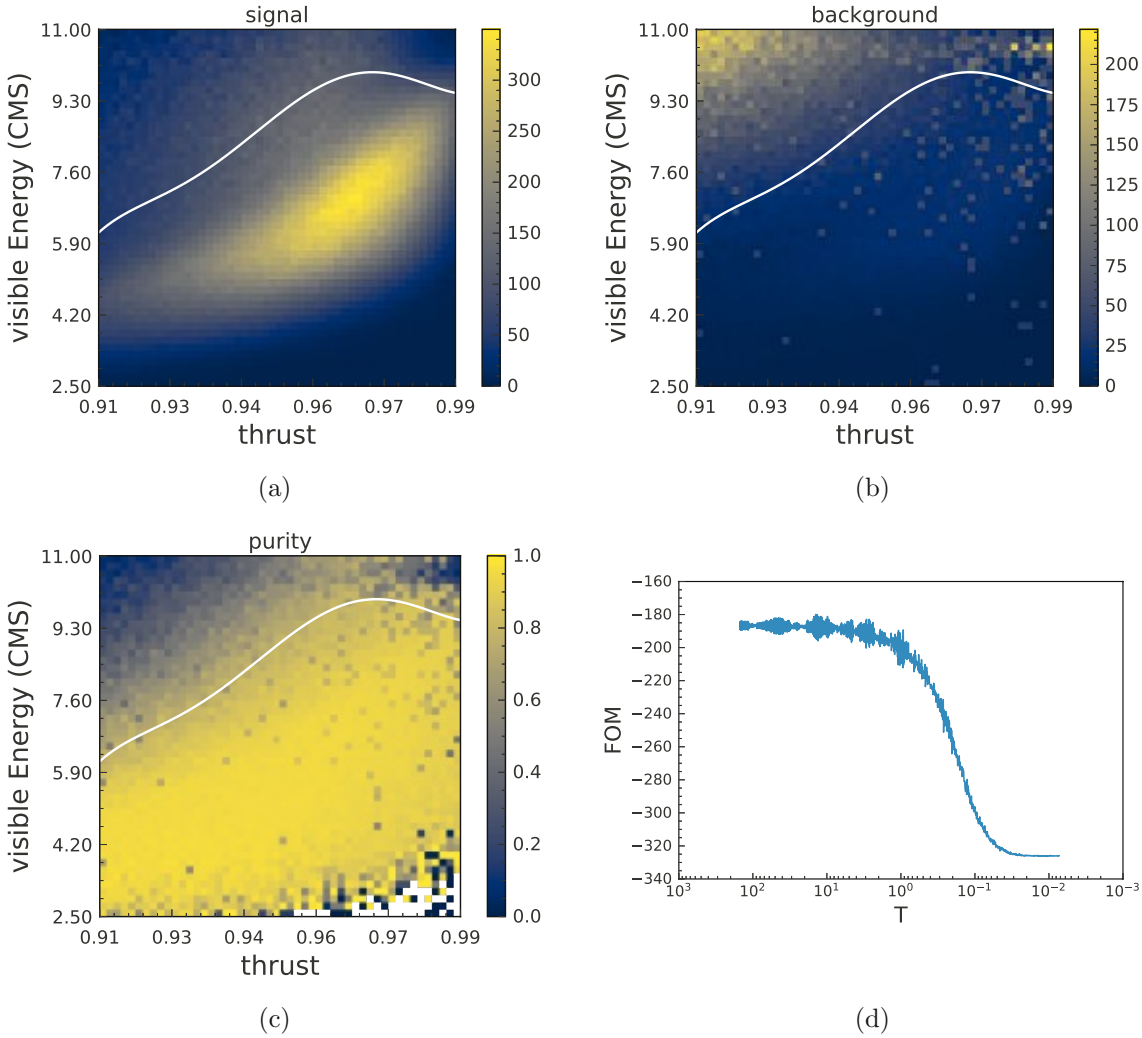
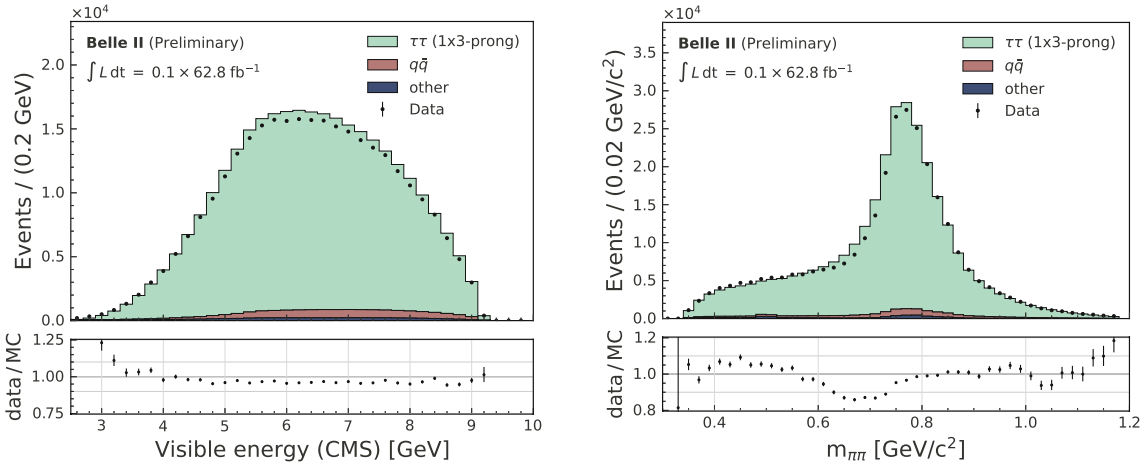


Figure 4.4: Optimisation of the second selection in the 2D plane spanned by the thrust value and the total visible energy in the CMS.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



(a) (b)

Figure 4.5: The distributions of visible energy in the CMS and the invariant mass of the pion pairs in data and MC after background suppression.

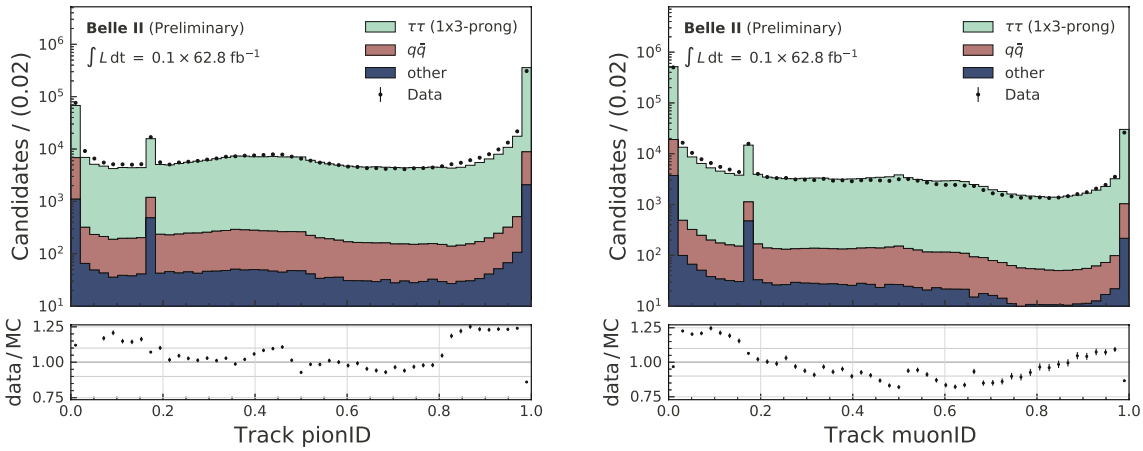
Further background suppression

To further suppress the background, a vertex fit on the 3-prong side is performed using `TreeFitter` [44]. The fit is required to converge, and a loose cut on the fit quality is applied. Furthermore, no two tracks are allowed to be within a 5° cone. Further background suppression could be achieved by applying a PID requirement for the tracks. Since we are interested in the effect of these variables, it can not be applied for the track we want to study. It would be possible to require that two of the three 3-prong tracks are very likely pions (i.e. apply a pionID cut) and then study the third one (this method is also called 'tag and probe'). However, this cut would introduce data/MC discrepancies (just like the ones we would like to estimate) that need to be accounted for either by assigning a systematic error or correcting this effect with results from other studies. For the present study, this additional tag selection is not applied since the systematic uncertainty introduced would be larger than the uncertainty due to the sample's impurity. Nevertheless, it is already possible to obtain a very clean pion sample through the described selections based mostly on topology and kinematics.

Selection results

The distribution for the visible energy in the CMS and the pion pairs' invariant mass is shown in Figure 4.5 for data and MC. The MC samples were scaled to match the data luminosity. The overall data/MC ratio is 97.66 %. Figure 4.6 shows the pionID and muonID variables for a combined sample of all the 3-prong tracks. Most of the tracks have a value very close to 0 or 1 (note the logarithmic scale). While the overall shape agrees between data and simulation, some discrepancies are introduced when a cut is applied on this variable. This is what will be quantified in this study.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



(a) (b)

Figure 4.6: PionID and muonID distributions in data and MC after background suppression.

Table 4.2: A comparison of the different tracks on the 3-prong side.

track	fraction of π (MC13a)	events in data (62.8 fb^{-1})
1	96.4 %	2 298 091
2	97.72 %	2 370 915
3	97.26 %	2 297 824
0 (1-3)	97.13 %	6 966 830

The signal efficiency after all selections, using $\tau_{\mathsf{3prong}} \to 3\pi(+n\pi^0)$ as the signal, is

$$\varepsilon_{\mathrm{sig}} = \frac{N_{\mathrm{rec}}}{2 \cdot \mathcal{B}_{1p} \cdot \mathcal{B}_{\mathrm{sig}} \cdot N_\tau}, \tag{4.8}$$

with $\mathcal{B}_{1p}$ being the branching ratio of 1-prong tau decays, $\mathcal{B}_{\mathrm{sig}}$ the branching ratio of the signal decays on the 3-prong side, N_{rec} the total number of reconstructed signal events and N_τ the total number of generated taupair events. For MC13a this results in a signal efficiency of 15.75 %. The purity for the signal as defined in Equation 4.7 is 94.1%.

The fraction of pions in MC and the total number of events in data for each track are shown in Table 4.2. Track number 1 and number 3 have the same charge, whereas track number 2 is the track with the different charge. The difference between the 1st and 3rd track originates from the ordering of particle lists in basf2 (this ordering determines what will be track number 1 and 3). The second track has a higher purity because in this sample the two same charged tracks are tagged. This reduces more background of the kind $\tau \to K\pi^+\pi^-\nu_\tau$. Track number 0 is assigned to the combined sample of all three tracks. The particle types that constitute this combined sample in MC are shown in Table 4.3.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table 4.3: The composition according to particle types in the MC sample after applying all selections.

particle type	MC13a
π	97.13 %
K	1.45 %
'none'	0.75 %
μ	0.32 %
p	0.22 %
e	0.13 %

4.5 Calculation of efficiencies

The efficiencies and mis-id rates in data are calculated with

$$\epsilon_{\text{data}} = \frac{N_{\text{pid}} - \sum_j n^j_{\text{pid}} \cdot r_j}{N - \sum_j n^j} \quad \text{with} \quad j \in \{e, \mu, K, p, d\}, \tag{4.9}$$

where N is the total number of events remaining after all selections and N_{pid} is the number of events remaining after applying an additional cut on a PID variable. n^i is the number of MC background events per particle hypothesis i and n^i_{pid} the same after the PID selection. r_i is the misidentification rate ratio between data and MC. Using Equation 4.9 the efficiencies will be calculated in bins of momentum and polar angle. The MC events n are scaled such that the relation

$$N \equiv \sum_i n^i \quad \text{with} \quad i \in \{\pi, e, \mu, K, p, d\}$$

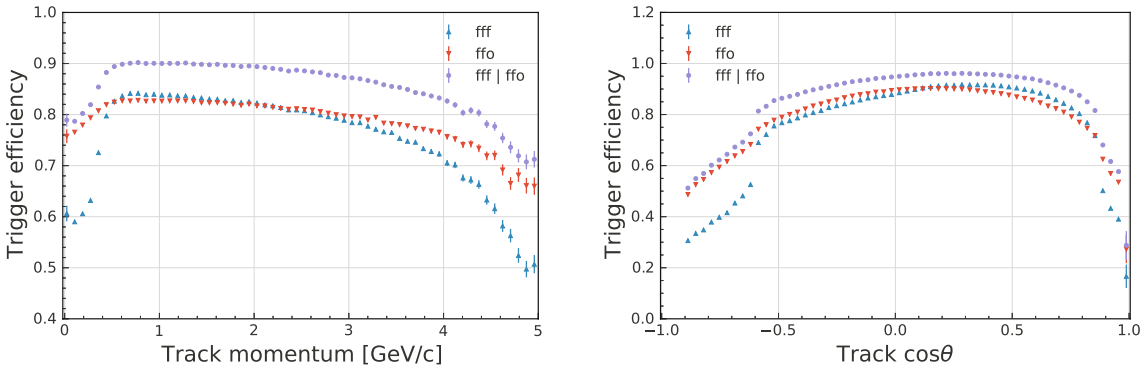
holds for each bin. If this results in a scaling that differs more than 50 % to luminosity based scaling, no results are reported for the respective bin.
The efficiencies and mis-id rates in the simulation are calculated with

$$\epsilon_{\text{MC}} = \frac{n^\pi_{\text{pid}}}{n^\pi}, \tag{4.10}$$

where the n^π and n^π_{pid} are evaluated based on true pions in MC.

The systematic error is calculated by varying n and r in Equation 4.9. A mismatch between simulated backgrounds and the actual data is estimated by varying n in one direction up to the value obtained when the MC events are scaled based on luminosity, but at least for 5 %. In the other direction, n is always varied by 5 %, which is the dimension of the overall observed data/MC mismatch. The statistical error on n is also propagated as a systematic error. The mis-id rate corrections r can in principle be taken from other performance measurements, but not all of them are available. Thus, this ratio is set to a nominal value of 1 and varied from 0.5 to 2, which corresponds to the typical magnitude of the mis-id rate corrections.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



(a) (b)

Figure 4.7: The momentum and polar angle dependency of the trigger efficiency for the `fff` trigger, the `ffo` trigger, and their combination.

Trigger correction

To correct for a possible bias introduced by the L1 trigger, a correction has to be applied. This correction will be calculated for events triggered by the `fff` or the `ffo` trigger (for the definition of the trigger lines see Table 3.2 in Section 3.3.3). Thus only events in the available data that were triggered by one of these triggers can be used. In the currently available data, this corresponds to 85.7 %. The event topology already meets the trigger conditions, which is why no additional cut needs to be applied in MC. However, because the trigger is not simulated in MC, we need to apply a correction based on the measured trigger efficiency.

To measure the trigger efficiency in data, we use an orthogonal (independent) reference trigger. Here we select the `hie` trigger for this purpose since this trigger is based on the ECL while the `fff` and `ffo` triggers are based on the CDC. Thus we can assume that they are independent. The `hie` trigger is fired by 84 % of the selected data, which ensures a large enough sample on which the calculation can be performed. To calculate the trigger efficiency we use

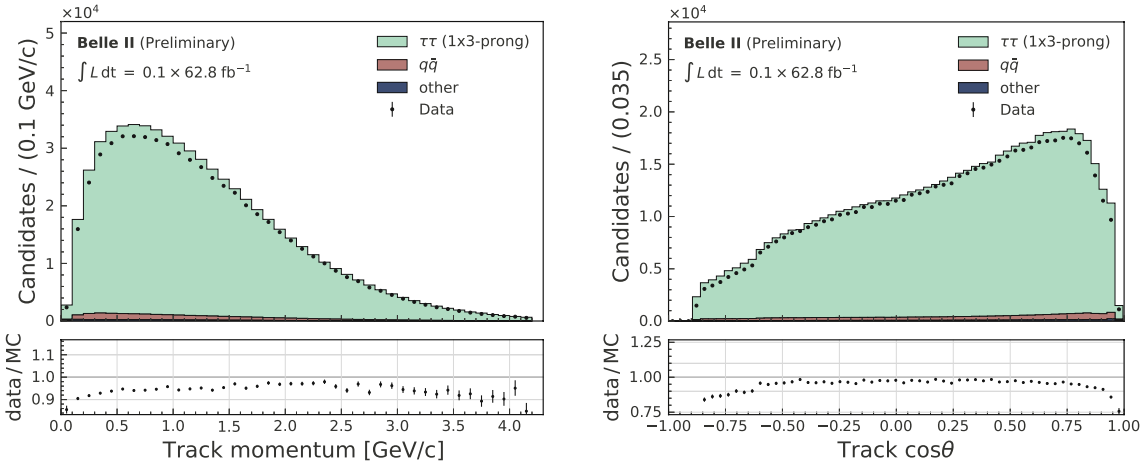
$$\varepsilon_{\text{trig}} = \frac{N_{(fff \vee ffo) \wedge hie}}{N_{hie}} \tag{4.11}$$

where $N_{(fff \vee ffo) \wedge hie}$ denotes the number of events where either the `fff` or the `ffo` trigger was active together with the `hie` trigger, and N_{hie} denotes the number of events the `hie` trigger was activated. The total efficiency for the present data sample calculated in this way is 88.48 %. Figure 4.7 shows the momentum and θ dependency of the trigger efficiency for the `fff` trigger and the `ffo` trigger and their combination. The efficiency decreases in the endcap regions and for lower momentum tracks, so short track triggers could also be considered for the future.

The trigger efficiency will be calculated in each bin for the selection without the PID cut (ε_{tag}) as well as the selection including the PID cut (ε_{probe}). The measured efficiencies correspond to an expression that can be written as

$$\epsilon \sim \frac{\mathcal{N}_{probe} \cdot \varepsilon_{probe}}{\mathcal{N}_{tag} \cdot \varepsilon_{tag}}, \tag{4.12}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



(a) (b)

Figure 4.8: The distributions of the combined sample containing all tracks on the 3-prong side. In data only events where the `fff` or `ffo` trigger was activated are shown and the MC distribution is scaled according to the calculated trigger efficiency.

with $\mathcal{N}_{probe}$ and $\mathcal{N}_{tag}$ being the actual number of probe and tag events without any trigger effects. What we want to measure is $\frac{\mathcal{N}_{probe}}{\mathcal{N}_{tag}}$, thus the trigger correction factor we apply on efficiencies measured in data amounts to $\frac{\varepsilon_{tag}}{\varepsilon_{probe}}$. If this correction is below 0.5 or above 1.5, no results are reported in that bin. The statistical uncertainty on this correction is also propagated as a systematic uncertainty for the final result.

In Figure 4.8 the data and MC distributions can be seen for the momentum and the cosine of the polar angle for all the 3-prong tracks combined. Only events where the `fff` or `ffo` trigger were activated are shown and the MC distribution is scaled according to the calculated trigger efficiency. The overall ratio between data and MC is now 94.6 %.

Binning

The binning used for this study can be found in Table 4.4 for the $\pi \to e$ mis-id rates ('electron related binning') and in Table 4.5 for the $\pi \to \mu$ mis-id rates and pion efficiencies ('muon related binning'). The bin edges are chosen based on critical regions (like gaps or instrumentation) and the division in barrel and endcap components of the relevant subdetectors. A sketch of the Belle II detector from the side with the regions corresponding to these bins is shown in Figure 4.9.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table 4.4: The binning used for calculating the $\pi \to e$ mis-id rates.

momentum	polar angle			
p [GeV/c]	θ [rad]	θ [deg]	$\cos\theta$	correspondence
(0.4, 0.5)	(0.22, 0.56)	(12.61°, 32.09°)	(0.98, 0.85)	1
(0.5, 1.0)	(0.56, 1.13)	(32.09°, 64.74°)	(0.85, 0.43)	2
(1.0, 1.5)	(1.13, 1.57)	(64.74°, 89.95°)	(0.43, 0.00)	3
(1.5, 2.0)	(1.57, 1.88)	(89.95°, 107.72°)	(0.00, −0.30)	4
(2.0, 2.5)	(1.88, 2.23)	(107.72°, 127.77°)	(−0.30, −0.61)	5
(2.5, 3.0)	(2.23, 2.71)	(127.77°, 155.27°)	(−0.61, −0.91)	6

Table 4.5: The binning used for calculating the $\pi \to \mu$ mis-id rates and pion efficiencies.

momentum	polar angle			
p [GeV/c]	θ [rad]	θ [deg]	$\cos\theta$	correspondence
(0.4, 0.5)	(0.40, 0.64)	(22.92°, 36.67°)	(0.92, 0.80)	1
(0.5, 0.7)	(0.64, 0.82)	(36.67°, 46.98°)	(0.80, 0.68)	2
(0.7, 1.0)	(0.82, 1.16)	(46.98°, 66.46°)	(0.68, 0.40)	3
(1.0, 1.5)	(1.16, 1.46)	(66.46°, 83.65°)	(0.40, 0.11)	4
(1.5, 2.0)	(1.46, 1.78)	(83.65°, 101.99°)	(0.11, −0.21)	5
(2.0, 2.5)	(1.78, 2.13)	(101.99°, 122.04°)	(−0.21, −0.53)	6
(2.5, 3.0)	(2.13, 2.22)	(122.04°, 127.20°)	(−0.53, −0.60)	7
	(2.22, 2.60)	(127.20°, 148.97°)	(−0.60, −0.86)	8

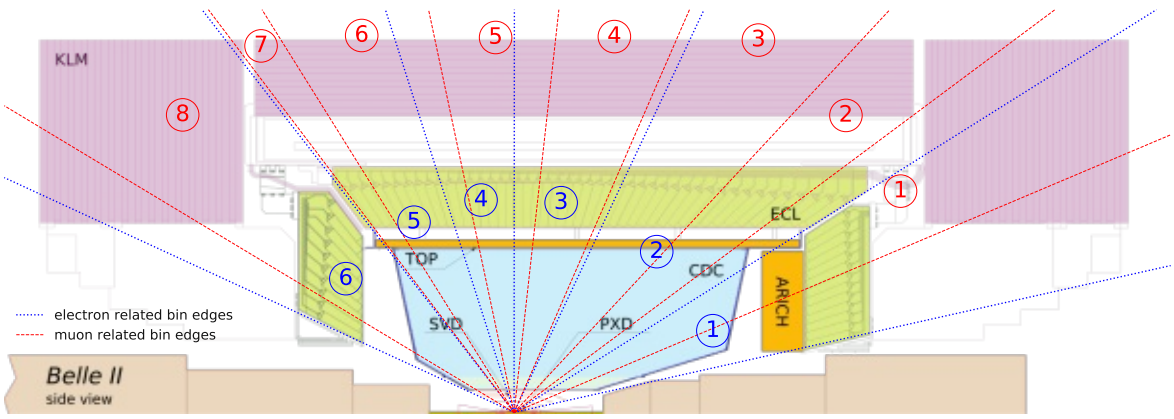


Figure 4.9: The polar angle bins in a side view of the Belle II detector. The blue markers correspond to the bins defined in Table 4.4, and the red markers correspond to the bins defined in Table 4.5.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

4.6 Results

The efficiencies and mis-id rates were compared for the global PID variables as well as the binary PID variables with

$$\begin{aligned} \text{binary pion ID} &= \text{binaryID}\,(\pi, \mu)\,, \\ \text{binary muon ID} &= \text{binaryID}\,(\mu, \pi) \text{ and} \\ \text{binary electron ID} &= \text{binaryID}\,(e, \pi)\,. \end{aligned}$$

In addition to that the subdetector performance was evaluated independently, using the (binary) likelihood ratio restricted to only one subdetector:

$$\text{binaryID}_{\text{det}}\,(i, j) = \frac{\mathcal{L}_{\text{det}}(i)}{\mathcal{L}_{\text{det}}(i) + \mathcal{L}_{\text{det}}(j)}. \tag{4.13}$$

If there was no subdetector information available (`nan` in one of the likelihoods), the event was excluded in the calculation. Multiple comparisons were made using these PID variables:

1. A comparison of data and Monte Carlo and the ratio thereof. The efficiencies are either calculated for the combined sample of all three tracks or only the sample with the different charged track (track 2 in Table 4.2), depending on where the total uncertainty (statistical + systematic) is smaller.

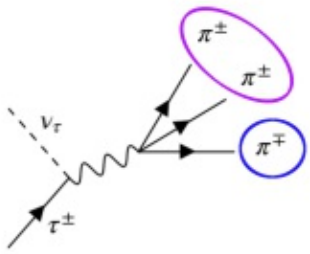
2. A comparison of positive and negative charged tracks, in data and MC together with the charge asymmetry, defined as $\mathcal{A}^{+-} = \frac{N^+ - N^-}{N^+ + N^-}$.

3. A comparison of the two tracks with the same charge ('1' and '3') to the remaining track ('2') (see Figure 4.10a). Both samples include the respective charge conjugation. Except for a slight shift in the momentum distribution, which should not be visible with fine enough binning, the two samples should give almost the same results. This was initially thought as a consistency check for the systematic error estimation since the two-track sample is expected to contain a higher fraction of non-pion tracks (see Table 4.2). However, after observing significant differences, it was found that the particle identification performance depends on track isolation (i.e. the distance to the closest track). This explains the disagreement between the samples because the different charged track will most of the time be more isolated than the other two tracks since it bends differently in the magnetic field (see Figure 4.10b).

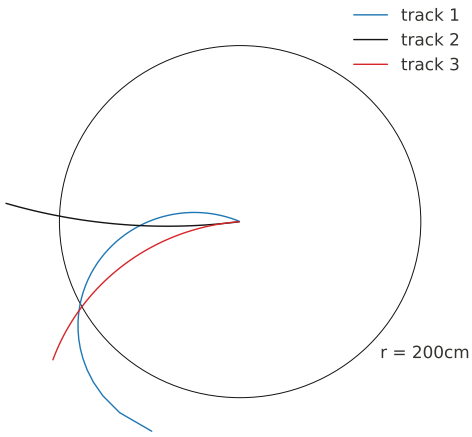
The results are provided in tables (used for corrections) and 1D histograms, binned in either momentum or polar angle and restricted to one bin in the other variable (based on the binning scheme defined in the previous Section). Only a few notable examples of these results are shown since the total amount of plots with all the considered combinations amounts to 1176. In Appendix B a summary of the PID performance in each of the bins can be found for the $\pi \to \mu$ mis-id rates (B.1), the $\pi \to e$ mis-id rates (B.2) and the pion identification efficiencies (B.3).

In the following Figures, pion efficiencies, $\pi \to \mu$ mis-id rates and $\pi \to e$ mis-id rates are denoted with $\epsilon_\pi(x)$, $\epsilon_\mu(x)$ and $\epsilon_e(x)$ respectively, where x is the cut value on the associated PID

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



(a) A sketch showing the 'track' samples: the two tracks with the same charge and the track with the opposite charge.



(b) A typical 3-prong tau decay in the r-φ plane. The same charged tracks bend due to the magnetic field in the in the same direction.

Figure 4.10: The three tracks from the 3-prong tau decay. Track number 1 and 3 have the same charge, while track number two has the charge with the opposite sign.

variable. The solid error bar indicates the statistical error only, while the transparent bar is systematic and statistical error combined. In the subplot, this division into errors is denoted by two error bars. All Figures show the results for data and simulation in bins of either polar angle or momentum. The title shows the restricted region of the other variable. For the combined sample restricted to subdetectors, the fraction of events with no subdetector information (`nan`'s) in data and MC is shown. For the charge comparison, the subplot shows charge asymmetry in data, and for the track comparison, the subplot shows the ratio of efficiencies between the two samples in data.

Figure 4.11 shows pion efficiencies with information from all subdetectors. In 4.11a the efficiencies for the global PID is shown in bins of the polar angle. The pion identification efficiency decreases in the backward endcap region. This can be explained by the lack of a PID detector in that region. In 4.11b the efficiencies for the binary PID is shown in bins of momentum. Lower momentum pions can not be identified as well as higher momentum pions. A good data/MC agreement is evident for the binary PID while in the global PID there are discrepancies up to 10 %. This is because less information which can introduce a mismatch between data and MC is used to construct the binary variables. The efficiencies themselves can not be compared directly between binary and global variables, because the cut at 0.5 on the variables will correspond to different working points. One useful feature of the binary PID variable is that the mis-id rates can be inferred from efficiencies and vice versa. In this case, with a cut at 0.5, the binary $\pi \to \mu$ mis-id rates are simply $1 - \epsilon_\pi(0.5)$.

Figure 4.12 shows the $\pi \to \mu$ mis-id rates using only KLM information. In 4.12a the track samples are compared. The discrepancies between the samples stem from different track isolation. This effect is present in data and MC, but more pronounced for the lower momentum tracks shown

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

here. In 4.12b the full sample is shown. For higher momenta the mis-id rates are below 4 %, and the data/MC ratio is almost consistent with one. For lower momentum tracks the mis-id rates increase up to 10 % and differ notably between data and MC. Only tracks with p > 0.7 GeV/c reach the KLM and have a likelihood defined.

Figure 4.13 shows the $\pi \to e$ mis-id rates for different subdetectors. In 4.13a the charge-separated samples are shown with only ECL information. A difference between the samples is evident, resulting from charge-dependent interactions of pions in the detector material. This is much more pronounced for lower momentum tracks. Interestingly, for a track momentum below 0.6 GeV/c, the negative pions in MC show larger mis-id rates than the positive pions, while in data it is exactly the opposite. In 4.13b only CDC information is used. Here a very large discrepancy is observed with mis-id rates in data more than ten times larger than in MC. Furthermore, the mis-id rates in data show a peak at $\theta = 90\,°$. The bad agreement between data and simulation for the CDC also shows itself in the global electron PID, where the data/MC ratio is usually between 2 and 3.

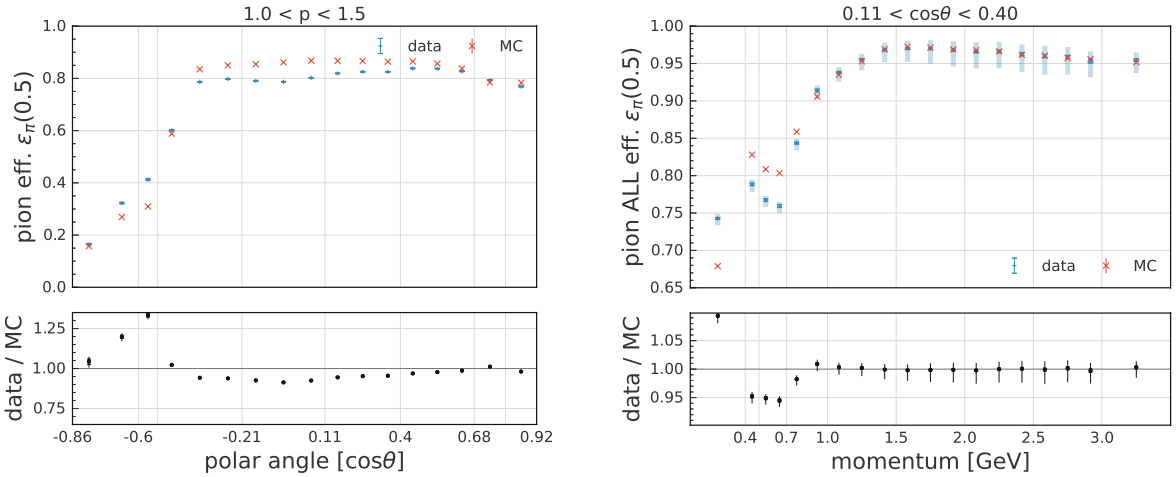
Figure 4.14 shows the dependence of mis-id rates on the threshold value of the related global PID variable, compared for the different track samples. In 4.14a the $\pi \to e$ mis-id rates are shown. The two-track sample has overall higher mis-id rates. This can be related to the fact that if two pions are close together at the ECL they are more likely to be identified as an electron (see 4.15b). The mis-id rates in data range from 2 % up to 10 %, while in MC they are less than half of that. In 4.14b the $\pi \to \mu$ mis-id rates are shown. Data and MC agree within a few percent, and although there is also some distinction between the track samples, it is much less pronounced than for the electronID variable. A tighter cut on the muonID variable seems to reduce the effect. The mis-id rates range from 3.5 % up to 7 %.

Track isolation

As already mentioned before, it was found that the particle identification performance depends on track isolation. This can be visualised by calculating the mis-id rates in bins of the distance to the closest track, at a given radius r. To take into account the effects of the magnetic field, the tracks need to be extrapolated up to r first. By repeating this process for multiple values of r (in the present case this was done every 5 cm), the dependence can be studied at the location of the different subdetectors. In order to see the effect at all subdetectors at once, 2D histograms were used where the x-axis gives the distance to the closest track, and the y-axis gives the radial distance r.

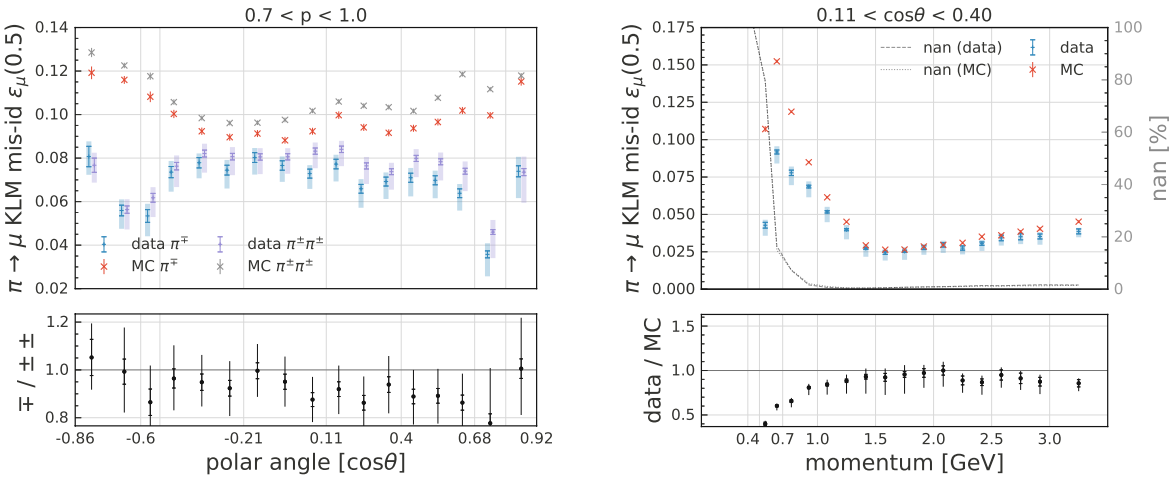
Some of these histograms are shown in Figure 4.15. In 4.15a the global $\pi \to e$ mis-id rates are shown, calculated in data. The dependence on the distance to the closest track is clearly visible and especially pronounced at the ECL (r = 125 - 162 cm) the TOP ($r \sim 120$ cm) and the CDC (r = 16 - 113 cm). Tracks that are closer together show an increase in mis-id rates. For the CDC the effect is more substantial when the tracks are not well separated at the first layers. For the ECL this effect could be explained by a larger energy deposited in the crystals and assigned to one or both of the tracks. In 4.15b the binary PID is shown using only ECL information. The black dotted lines mark the area of the ECL. This illustrates the effect of the ECL and how it

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



(a) (b)

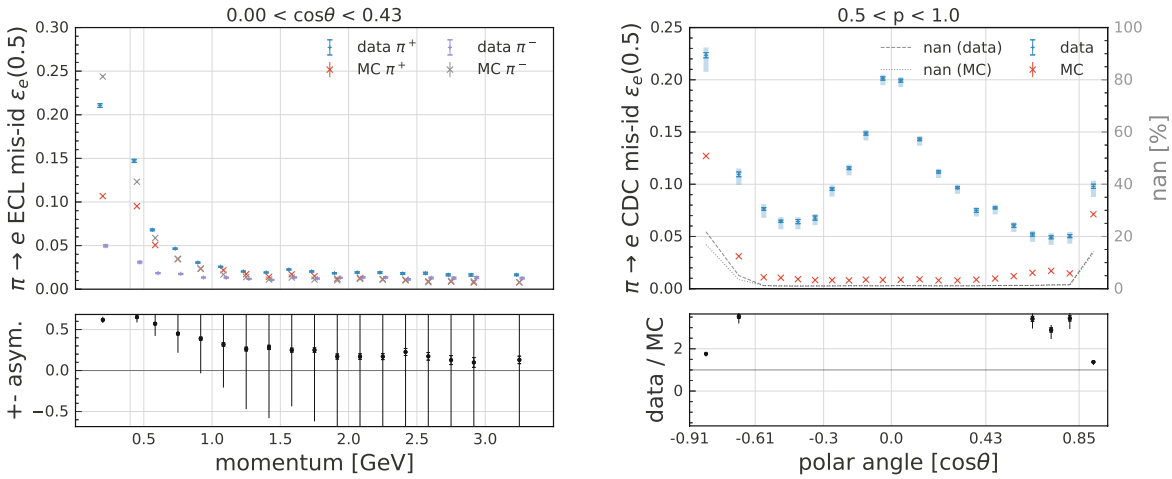
Figure 4.11: The pion efficiencies using the global PID (a) and the binary PID (b). A better data/MC agreement is evident for the binary PID.



(a) (b)

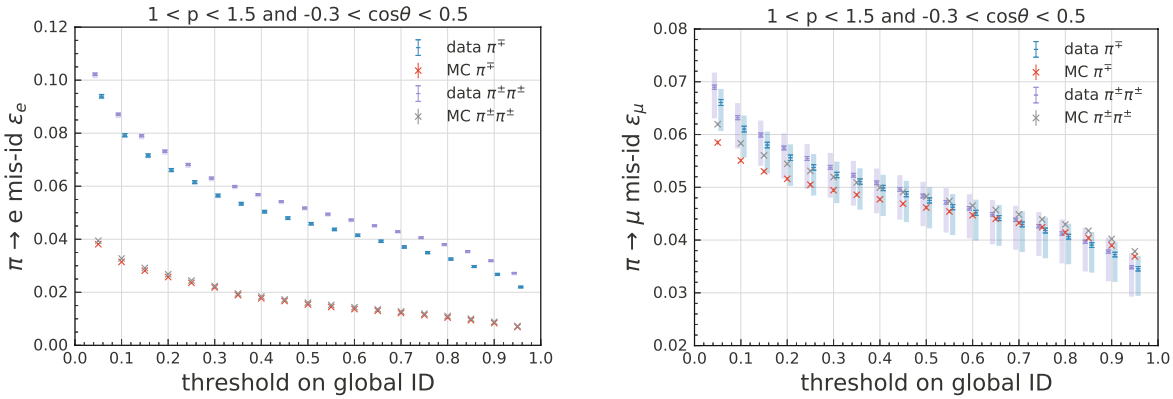
Figure 4.12: The $\pi \to \mu$ mis-id rates using only KLM information, in (a) for the track samples and in (b) for the full sample. Only tracks with p > 0.7 GeV reach the KLM and have a likelihood defined.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



(a) (b)

Figure 4.13: The $\pi \rightarrow e$ mis-id rates for different subdetectors. In (a) the charge-separated samples are shown with only ECL information. A difference between the samples is evident. In (b) only CDC information is used. Here a large data/MC discrepancy is observed.



(a) (b)

Figure 4.14: The dependence of the $\pi \rightarrow e$ mis-id rates (a) and the $\pi \rightarrow \mu$ mis-id rates (b) on the threshold value of the related global PID variable. The mis-id rates are displayed for the different track samples.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

only affects only the corresponding region. In 4.15c the global $\pi \to e$ mis-id rates in MC are shown. While a similar dependence can be observed for the ECL, it is no longer present for the CDC. This becomes also evident when the data/MC ratio is calculated (shown in 4.15d). Here the data/MC ratio for close tracks is as high as 4, while for the isolated tracks it is around 2. This is because the CDC simulation is done per-track rather than per-event, thus not considering any correlation effects.

For the KLM a similar dependence on track isolation can be observed when looking at $\pi \to \mu$ mis-id rates. This effect was found to be much more relevant for low momentum tracks and overall well simulated.

These results suggest that track isolation is a non-negligible effect and should be considered when dealing with different event topologies.

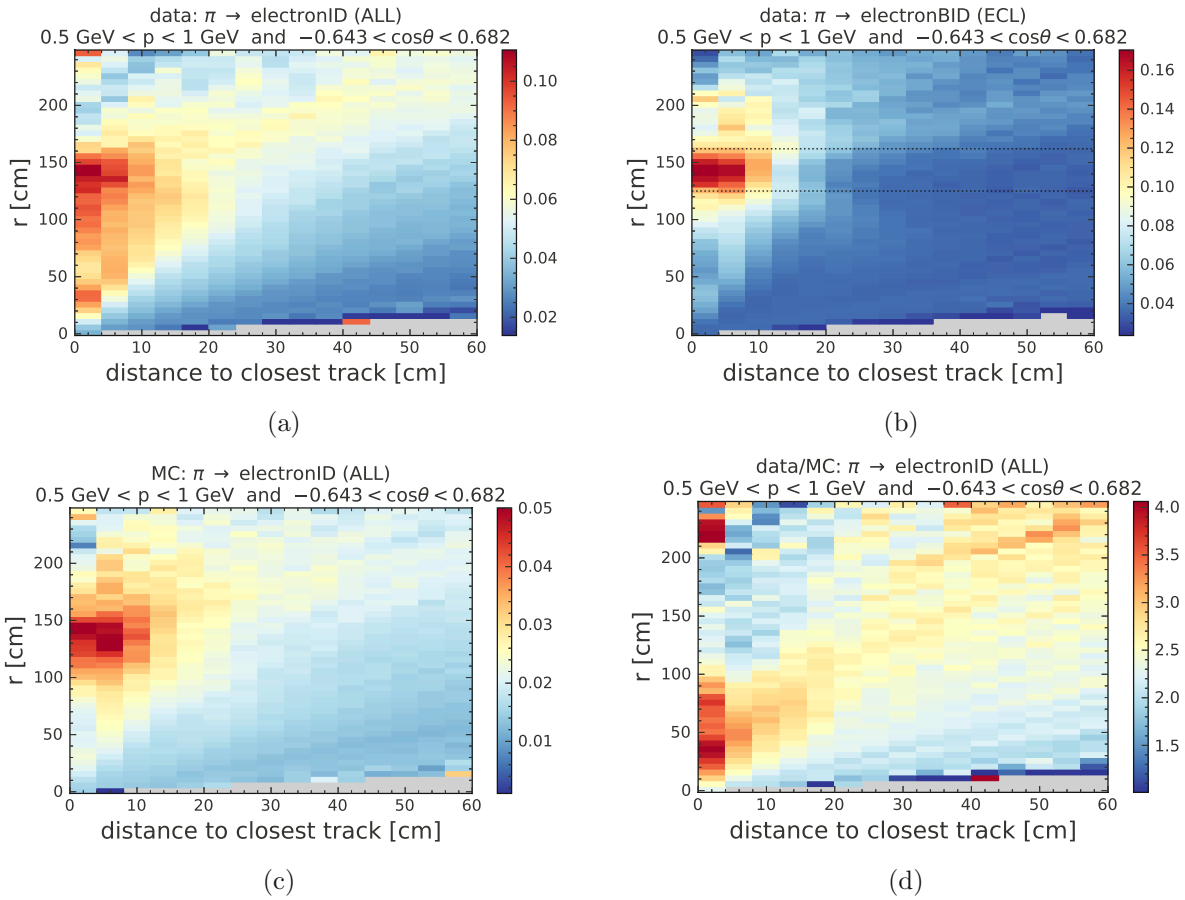


Figure 4.15: 2D histograms showing how the mis-id rates depend on the distance to the closest track at varying radius. Here the $\pi \to e$ mis-id rates are shown in a given momentum range and angular acceptance.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Chapter 5

Analysis procedure

This chapter describes the analysis procedure. To not introduce a subjective bias[1] it is desirable to perform what is called a *blind analysis*. This refers to an analysis where the whole procedure is first performed and tested only on simulated data. To compare data and MC, it is possible to define *control regions* in the data that do not coincide with the actual data sample one wants to use for the analysis (the *signal region*), but show similar properties. After the analysis procedure is fixed, the whole process is repeated on the data without further modifications. The result should be reported as-is, even it is not what was expected or desired (except a significant mistake is found in the analysis). This prevents the often subconscious decisions we make to 'shape' the result to give us the answers we want.

Only the first part of the analysis is presented in this work, which is merely performed on simulation. Before moving on and looking at the data, this analysis procedure must be internally reviewed by the collaboration. Only then the final results can be obtained. Nevertheless, what is shown in the following chapter is the full analysis procedure that will also be applied to data. The statistical evaluation is tested on a pseudo dataset based on simulation but has real data properties. This analysis's sensitivity to discover a Z' and the parameter space expected to be excluded when there is no evidence is also estimated.

5.1 Monte Carlo samples

5.1.1 Background samples

This Section will describe the background MC samples used for this analysis. Table 5.1 shows all the processes that were considered to contribute as a background. For each of the processes the corresponding production cross section, generated events and the resulting equivalent luminosity is shown. In general, the MC sample should be as large as possible to minimise the statistical

[1] What a subjective experimenter expectancy can lead to can be seen in many wrong discoveries such as the 'double peak of the A_2 meson' [45] or even simply in the history of measurements of Standard Model parameters (Particle Data Group (PDG) history plots [8, p. 20]).

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table 5.1: Generated Monte Carlo samples for different processes considered as a possible background for the analysis with corresponding luminosities. If not suggested otherwise all particles are to be understood as charged particles such that the overall charge sums to zero.

process	cross section [nb]	MC luminosity [fb^{-1}]	generated events
$ee \to \tau\tau$	0.919	3×10^3	2.76×10^9
$ee \to ee\mu\mu$	18.9	1×10^3	1.89×10^{10}
$ee \to \mu\mu$	1.15	1×10^3	1.15×10^9
$ee \to u\bar{u}$	1.6	100	1.60×10^8
$ee \to \pi\pi\gamma$	0.153	65.3	1×10^7
$ee \to ee\pi\pi$	1.9	1×10^3	1.9×10^9
$ee \to c\bar{c}$	1.33	100	1.33×10^8
$ee \to d\bar{d}$	0.401	100	4.01×10^7
$ee \to s\bar{s}$	0.383	100	3.83×10^7
$ee \to B^0\overline{B}^0$	0.51	100	5.1×10^7
$ee \to eeKK$	0.0798	1×10^3	7.98×10^7
$ee \to eeee$	39.7	500	1.99×10^{10}
$ee \to ee$	294	20	5.88×10^9
$ee \to eepp$	0.0117	1×10^3	1.17×10^7
$ee \to B^+B^-$	0.54	100	5.4×10^7

error that will eventually enter as a systematic error in the final result. For some processes, the luminosity is below 100 fb^{-1}, but they will not contribute to the background after performing all the selections as will be shown in Section 5.4.

All the relevant MC samples are taken from the official MC production, centrally produced for all Belle II members. Internally this production is referred to as `MC13a`, where the 'a' refers to a run independent production. The basf2 release used for this production is `release-04-00-03` and the detector geometry corresponds to 'Early Phase 3'.

The most important processes will be $ee \to \tau\tau$ ("taupair"), $ee \to \mu\mu$ ("$\mu\mu$") and $ee \to ee\mu\mu$ ("$ee\mu\mu$"). For the taupair sample, the relevant process is where both tau leptons decay to a muon or to a pion which is misidentified as a muon. These events will also have missing energy because of the neutrinos (ν_τ and ν_μ). This will result in the same final state as the signal events. Section 5.3 will show how these events can be suppressed with a multivariate selection using kinematic variables. The taupair background is distributed mainly between a recoil mass of 1 GeV and 8 GeV. For the $\mu\mu$ sample, the final state is equivalent to signal events when there is initial state radiation (ISR)[2] and the photon is not detected. These events will contribute mostly as background for very low Z' masses, corresponding to a low recoil mass (peaking around 0 GeV). The $ee\mu\mu$ sample can have the same signature in the detector if both electrons are outside of the detector acceptance or not detected. These events will cluster at very large recoil masses (above 6 GeV). In Figure 5.1 the recoil mass distributions for these processes are shown after the

[2]Even though not shown explicitly, the generated processes for $\mu\mu$ and taupair events include radiative effects, so there can be one or more photons in the final state.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

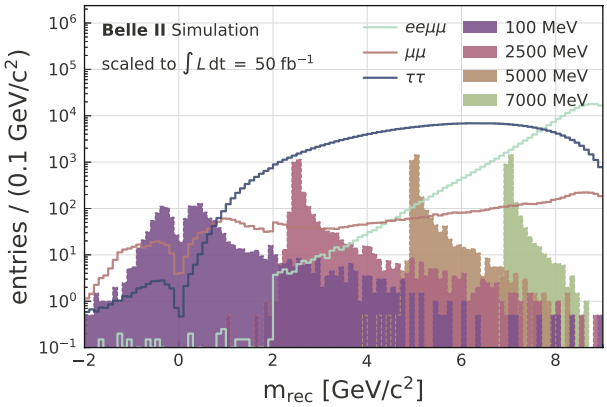


Figure 5.1: Signal and background distributions of the recoil mass after all univariate selections, showing only the main backgrounds and signal distributions for 100 MeV, 2.5 GeV, 5 GeV and 7 GeV.

univariate selections described in Section 5.2. The signal distributions for some generated signal samples are overlayed. As expected, they are peaking at the value corresponding to the Z' mass. The width of the peak depends on the recoil mass resolution.

5.1.2 Signal samples

The signal samples were generated using `MadGraph 5` [46] for Z' masses ranging from 100 MeV to 10 GeV in steps of 100 MeV. For each mass point, 20 000 events were generated. A region for each generated mass corresponding to $\pm 2\sigma$ was defined, where σ denotes the standard deviation of the signal peak in the recoil mass distribution. The finite width of the signal is caused in this case by resolution effects. These regions will be referred to as 'mass windows', and they will be used to estimate the number of backgrounds falling inside the signal region of the respective Z'. The Punzi figure of merit, which is used to optimise the selections, will also be calculated individually for each mass point in the corresponding window.

5.2 Event selection

In this Section, I will describe the initial selection of the events to reduce the background while retaining as many signal events as possible. Most of the selections (also referred to as 'cuts') applied here are either obvious to select the signal topology or imposed because of constraints from trigger considerations. Some more special cuts I will describe in more detail.

An overview of all the cuts applied before the multivariate selection can be found in Table 5.2 with their respective cut number. This number will be referred to throughout the text. Note that the order does not matter and is chosen to reflect the similar origin of the respective selections. A star (*) denotes that the respective variable's reference frame is given by the CMS. If not explicitly

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table 5.2: Overview of the event selections applied before the multivariate selection. A bold number indicates a cut which was applied at a later stage of the event processing.

1	$dr < 2\,\text{cm}$ and $\|dz\| < 5\,\text{cm}$	2 clean tracks
2	at least one CDC hit	
3	$\Delta\varphi_{\text{lab}} > 30°$ (trigger)	dimuon system
4	3D opening angle (CMS) $< 179.5°$	
5	no photon within 15° of $\vec{p}_{\text{rec}}$	rest of event
6	no extra tracks with CDC hits	
7	extra energy < 0.5 GeV	
8	ECL cluster match	additional track selections
9	$(E_{\text{cluster}}/p)_\mu < 1\,c$	
10	$37° < \theta_\mu < 120°$	
11	muonID > 0.5	
12	$\theta_{\text{rec}} < 123.19°$	
13	$33.8° < \theta_{\text{rec}} < 88.85°$ or $\theta_{\text{rec}} > 91.15°$ or $m_{\text{rec}} > 2$ GeV/c^2	
14	$m_{\text{rec}} < 9$ GeV/c^2	
15	$p_t^*(\mu^+\mu^-) > 1$ GeV/c	

mentioned, the laboratory frame is used. A bold number indicates a cut that was only applied after preprocessing the events. The cuts with number $\{1, 2, 3, 5, 6, 7, 10, 15\}$ will be referred to as preselections and follow largely what was used in the phase 2 measurement [see 1, p. 5].

The cuts with number 1 and 2 ensure that only events with exactly two 'clean' tracks are selected. Here a clean track is defined as a track whose distance of the point of closest approach to the IP in the r-phi plane (dr) is less than 2 cm, and the relative distance in the z-direction (dz) is between −5 cm and 5 cm. Additionally, at least one CDC hit has to be present. These requirements ensure that the track's origin is at the IP and it reaches beyond the vertex detectors. This is mainly to not count tracks that arise from beam background or cosmic radiation that could also be present in a signal event.

The cuts with number 3 and 10 are connected to the trigger. Since for the present analysis the `ffo` trigger is used the requirement for the trigger to be activated ($\Delta\varphi_{\text{lab}} > 30°$) has also to be imposed on the MC samples. The distribution of $\Delta\varphi_{\text{lab}}$ for the dominating background processes and selected Z' masses after all preselections were applied can be seen in Figure 5.2. Cut number 10 is applied because in a separate trigger study [Belle II internal document: 47, p. 7] it was found that the trigger efficiency drops significantly in the endcap regions. In the range of $37° < \theta_\mu < 120°$ it is stable around a value of 89%.

The cuts with number 5, 6 and 7 are connected to the rest of the event, which is the set of clusters or tracks not used for the clean track reconstruction. The requirements are that there is no extra track with CDC hits (cut 6) and the sum of all the extra clusters in the ECL (extra energy) does not exceed 0.5 GeV (cut 7). The latter rejects events with high energetic photons

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

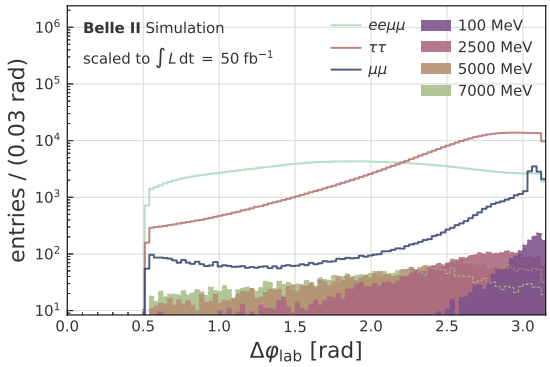


Figure 5.2: Signal and background distributions of the 2D opening angle between μ^+ and μ^- in the r-φ plane after preselections.

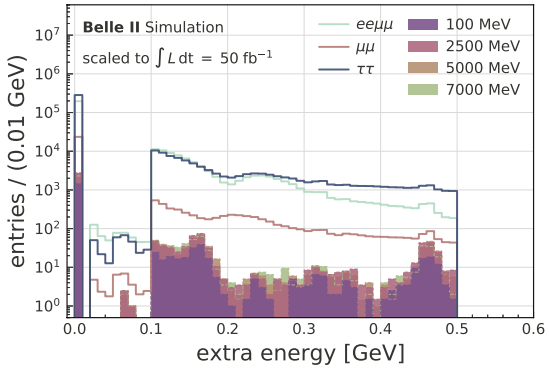


Figure 5.3: Signal and background distributions of the extra energy deposited in the ECL after preselections.

or increased ECL activity and is shown in Figure 5.3 after all preselections. Additionally, there is the requirement that there is no photon within 15° of the direction of the recoil momentum (cut 5). This ensures that $ee \to \mu\mu\gamma$ events do not pass the selections.

The cut with number 15 rejects events where the transverse momentum of the dimuon ($\mu^+\mu^-$) system in the CMS is less than 1 GeV/c. This is equivalent to the transverse recoil momentum in the CMS. This cut rejects a large amount of background and keeps the processed files at a reasonable size. The distribution of $\mathrm{p}_\mathrm{t}^*(\mu^+\mu^-)$ after all univariate selections is shown in Figure 5.4. While this variable is a good discriminator between signal and background, cutting around 1 GeV also removes a significant amount of signal events. This value was chosen because the cut is superseded by the multivariate selection described in the next chapter. This can be seen in Figure 5.5 where the remaining signal events for all generated Z' masses are shown after the multivariate selection. The multivariate selection is based on an ANN minimising the Punzi figure of merit, thus providing an optimised selection of the signal region and justifying the removal of the signal events. Since it was consistently found for different luminosities that the ANN removes these events, the cut was already applied at preselection level. This reduction of event number is also desirable since it speeds up the training of the ANN. An additional purpose of this cut is that it rejects cosmic muons. Suppose a muon (positive or negative) transverses the Belle II detector and passes the IP. In that case, it will be indistinguishable from an event where two muons with opposite charge originate from the IP. Because of the asymmetric beam energies, it will appear as if there is energy missing in these events. Of course, these cosmic events have other striking features. For example, the dimuon momentum in the laboratory frame $\mathrm{p}(\mu^+\mu^-)$, defined as the absolute value of the vector sum of the two muons' momenta, is expected to be very close to 0. Since these events can not survive the $\mathrm{p}_\mathrm{t}^*(\mu^+\mu^-)$ cut, no other cosmic veto needs to be considered.

Cut number 4 requires that the opening angle between μ^+ and μ^- in the CMS is less than 179.5°. This rejects some $ee \to \mu\mu$ events for which the muons are expected to be back to back in the CMS. In Figure 5.6 the peak at 180° is well visible. The cut corresponds to removing the rightmost bin. A tighter cut is unnecessary since the multivariate selection will be performed on

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

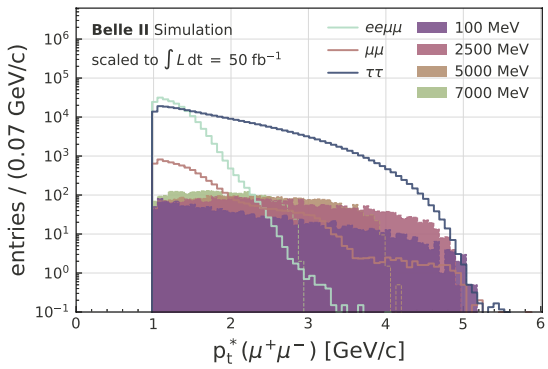


Figure 5.4: Signal and background distributions of the transverse momentum of the dimuon system in the CMS after all univariate selections.

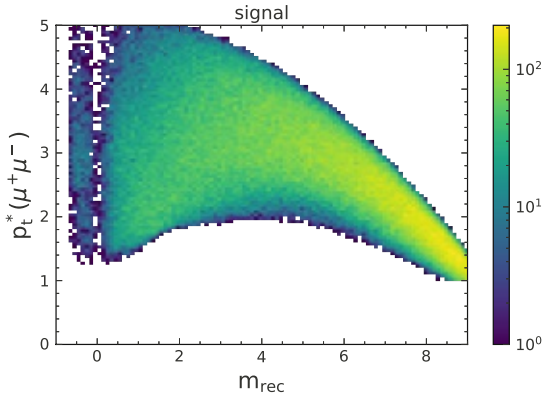


Figure 5.5: A two dimensional histogram showing signal events for different values of m_{rec} and $p_t^*(\mu^+\mu^-)$ after the multivariate selection was applied.

variables correlated with this opening angle and therefore optimise the selection.

Cut number 12 and 13 concern the recoil momentum, in particular, it is a veto of detector regions with missing instrumentation. The ECL has a gap between forward ($\theta = [12.4°; 31.4°]$), barrel ($\theta = [32.2°; 128.7°]$) and backward region ($\theta = [130.7°; 155.1°]$) as well as at $\theta = 90°$ to provide space for cables and detector supports. This corresponds to a range with polar angles from 0.22 rad to 2.71 rad with gaps at 0.56 rad, 1.57 rad and 2.26 rad. The ECL is necessary to detect photons, in particular for $\mu\mu\gamma$ events. If the photon goes outside of detector acceptance or inside a gap, it will escape the detection, and the event will be indistinguishable from a signal event. This can be seen in Figure 5.7 where the background after preselections is shown for different values of the recoil momentum polar angle. The background peaks at the regions not covered by the ECL. One solution would be to veto these events using the KLM to detect the photons that pass through the ECL. Another solution is to exclude those regions in our search, which is the approach chosen for the present work, leading to cuts number 12 and 13. The corresponding regions are highlighted in Figure 5.8, showing only the $\mu\mu$ background in 5.8a and the combined signal samples in 5.8b. In 5.8a the peaking background is well visible, both in θ_{rec} and m_{rec}. The reason why large regions at the low and high θ_{rec} are removed can again be found when looking at the selected events after the multivariate selection was applied. This is Shown in Figure 5.9, for the $\mu\mu$ background in 5.9a and for the combined signal samples in 5.9b. The final selection will exceed or at least match the rectangular cuts applied here, thus justifying the rough choice of the veto region.

Cut number 11 requires that the muonID (defined in Section 4.2) of both charged tracks is greater than 0.5. This uses information from all subdetector systems to determine how likely the observed track is a muon. The largest contamination will be from pions that still have some probability to pass this selection (see the detailed study in Chapter 4). In principle, the cut on the muonID variable can be performed on any value between 0 and 1 and should be optimised using the Punzi FOM. However, because of large discrepancies between data and MC, correction

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

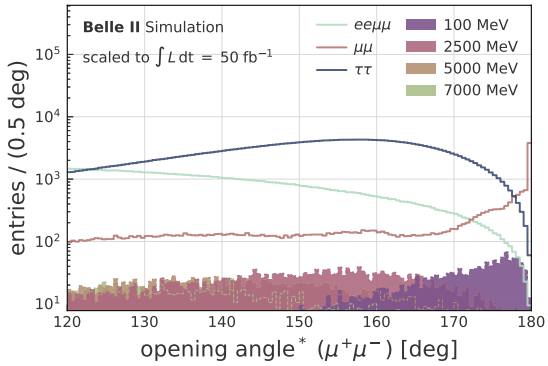


Figure 5.6: Signal and background distributions of the opening angle between μ^+ and μ^- in the CMS after preselections.

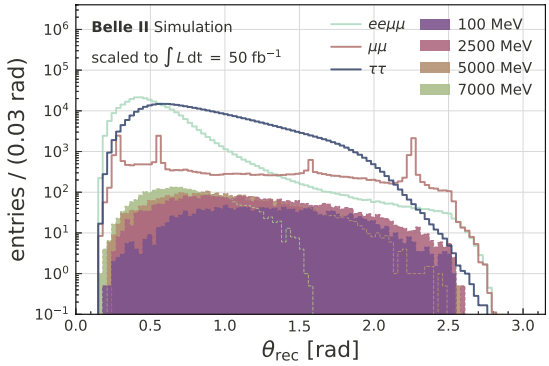
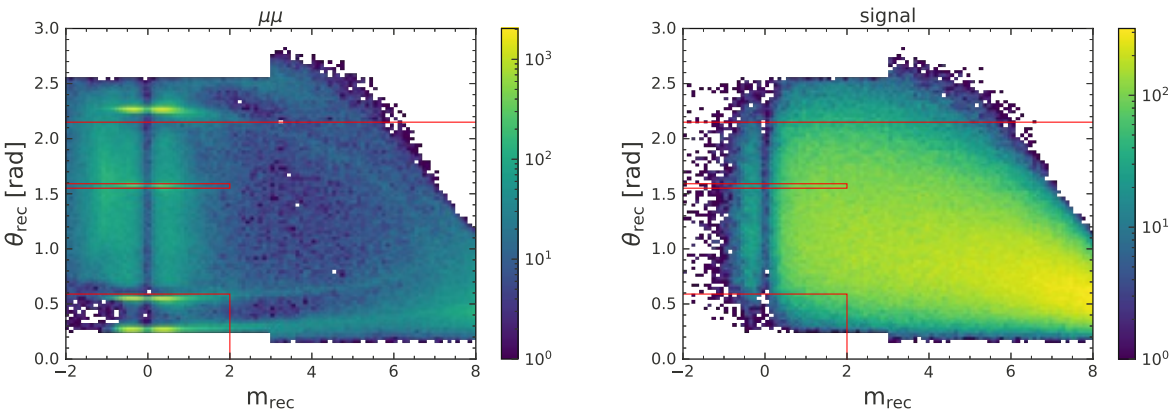


Figure 5.7: Signal and background distributions of θ_{rec} after preselections.

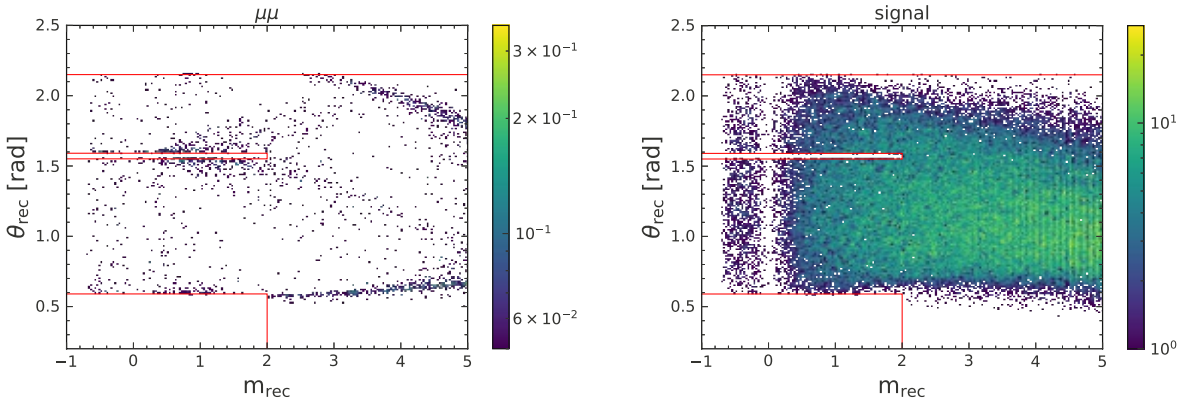


(a) $\mu\mu$ background

(b) signal events

Figure 5.8: Two dimensional histograms showing background and signal events for different values of m_{rec} and θ_{rec} after preselection.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



(a) $\mu\mu$ background

(b) signal events

Figure 5.9: Two dimensional histograms showing background and signal events for different values of m_{rec} and θ_{rec} after the multivariate selection was applied.

factors have to be applied and systematic uncertainties assigned. These are currently provided from different studies (including the study on taupair events performed for this thesis) for the whole collaboration, but only at working points of muonID $> 0.5, 0.9$ and 0.95. Because most tracks are distributed very close to the extreme values[3] there was not a large difference found when comparing these values, but muonID > 0.5 resulted in the best value of the FOM.

Cut number 8 guarantees that an ECL cluster is associated with each track, ensuring that the tracks reach at least the ECL. This improves particle identification and rejects events where a low momentum particle curls inside the CDC, leading to complications with tracking and an overall less clean event environment. Figure 5.10 shows the distribution of the cluster energy for the positively charged muon candidate after preselections. Typically, muons only deposit a small amount of energy in the ECL. Cut number 9 rejects events where the ratio of the associated ECL cluster energy over the track's momentum is larger than 1. For muons, this ratio is almost always less than 0.8, which is valid for all the signal samples, as shown in Figure 5.11.
However, there appears a large amount of $\mu\mu$ background for values greater than 1.2. Two coinciding circumstances can explain this: Firstly, a muon can emit final state radiation (FSR) or Bremsstrahlung. The momentum of the photon will point in a direction similar to the muon momentum. Sometimes, the muon's track and the photon overlap exactly at the ECL detector (the muon bends in the magnetic field, the photon not) and the photon energy will be assigned to the muon track. This leads to a situation to be seen in Figure 5.12 where only tracks with $E_{cluster}/p > 1$ are shown. The histogram shows the CMS energy of the lower energetic muon on the vertical axis and the deposited ECL energy on the horizontal axis. Since the energy is inferred from the particle momentum, this corresponds to the actual muon energy, while the cluster energy is much larger. Also apparent in the Figure is the correlation between the two variables. This can be explained by noting that the muon CMS energy in $ee \to \mu\mu$ events is expected to peak around $\frac{\sqrt{s}}{2}$ which at Belle II is 5.29 GeV. This can be seen in Figure 5.13 where the CMS energy

[3]This is a common feature of PID variables, see for example Figure 4.6.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

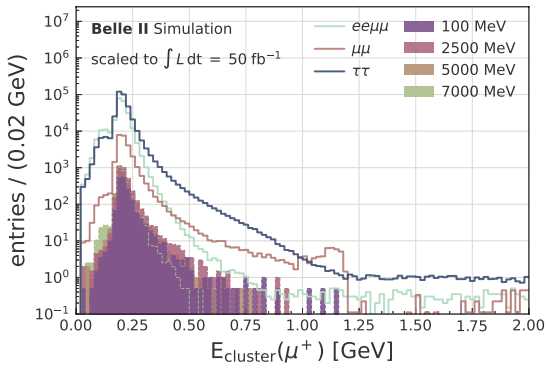


Figure 5.10: Signal and background distributions of $E_{cluster}$ for the positively charged muon after preselections.

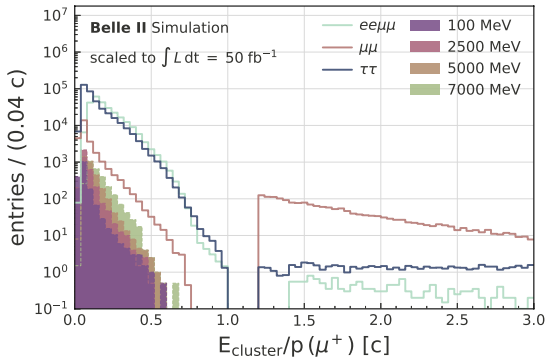


Figure 5.11: Signal and background distributions of the ratio $E_{cluster}/p$ from the positively charged muon after preselections.

distribution of the higher energetic muon is shown. For the $\mu\mu$ background and very low Z' masses this variable peaks around 5.29 GeV. The energy difference to this value, in Figure 5.12 around 3 GeV, is taken by the FSR photon and then deposited in the ECL. For such events, a lower CMS energy would lead to larger cluster energy, which is also evident in Figure 5.12. The other factor that leads to the sharp increase of background events seen in Figure 5.11 is that the ECL likelihood functions used for particle identification are not defined above a value of $E_{cluster}/p > 1.2$. For muons these should be 0 for such large $E_{cluster}/p$ values. Then the events in question would not pass the muonID cut. Until this issue is fixed, we can simply apply the $E_{cluster}/p < 1$ cut to remove all such events.

The last cut in Table 5.2 not mentioned so far is $m_{rec} < 9$ GeV/c^2 with cut number 9. The whole purpose of this cut is to reduce file size since a substantial amount of $ee\mu\mu$ events pass the selections for large recoil masses. By performing this cut, we accept that all signal events with a Z' mass greater than 9 GeV will be removed. This is not relevant since we have no sensitivity to discover a Z' with such high mass because of dominating SM backgrounds.

After all these selections, the resulting distribution of events over the recoil mass can be found in Figure 5.14. Here all backgrounds with remaining events are shown. Because of resolution effects, the squared recoil mass may be negative. To calculate the nominal recoil mass, we define

$$m_{rec} = \text{sign}(m_{rec}^2) \cdot \sqrt{|m_{rec}^2|},$$

resulting in the double peak structure around $m_{rec} = 0$. The backgrounds will still be significantly reduced by the multivariate selection described in the next Section.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

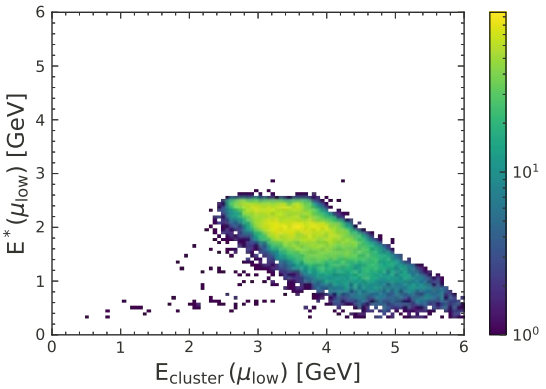


Figure 5.12: The CMS energy of the lower energetic muon and the deposited ECL energy for tracks with $E_{cluster}/p > 1$. For lower CMS energies a larger cluster energy is observed, consistent with the FSR hypothesis.

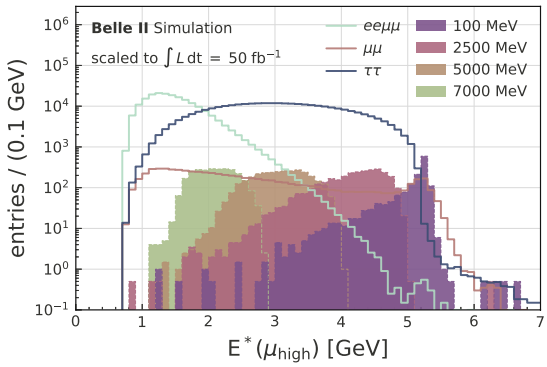


Figure 5.13: Signal and background distributions of the CMS energy from the higher energetic muon after all univariate selections. Lower values correspond to higher recoil masses.

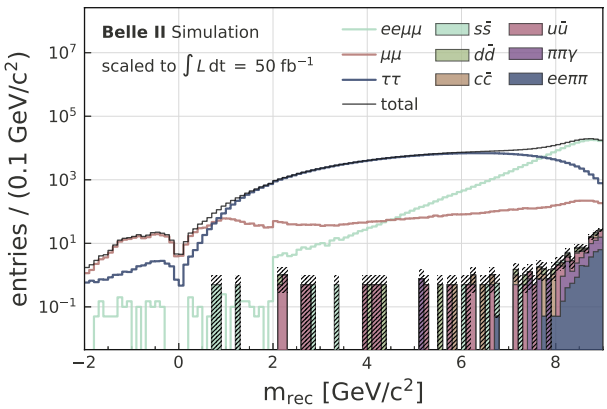


Figure 5.14: Signal and background distributions of the recoil mass after all univariate selections, showing all backgrounds with remaining events. The filled histograms are stacked.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

5.3 Multivariate selection using an artificial neural network

To further suppress the background, a simple feedforward ANN is deployed. This ANN is also used to optimise the selections by minimising directly a quantity related to the Punzi FOM. We will see how this simplifies the utilisation of the ANN for the usage with different mass regions that should in principle be optimised independently. It will be possible to apply a single ANN for all mass regions at once, which directly yields the optimal selection for each of them. Additionally, the final selection will be improved compared to commonly used loss functions. For this training to converge, it was found that the parameters of the network should already be initialised in a way that defines some separation between signal and background. This can simply be achieved by training the network first using a BCE loss function and then use this trained network as a starting point for the training with the FOM loss function. More details about this will be given in Section 5.3.4.

5.3.1 Feature selection

To select the most discriminating features for separating signal and background around 50 variables were considered initially. These initial variables were mostly related to kinematic aspects of the finals state particles and the overall event topology. Multiple combinations of energy and momentum were formed in different reference frames to find specific compositions that are most efficient at distinguishing the various processes. Variables related to lower-level detector signals were disregarded because this could introduce additional systematic effects if they are not perfectly simulated. With the use of fast boosted decision trees (BDTs) [48] (implemented in basf2) and a forest of trees (using scikit-learn [49]) feature importances based on the mean decrease in impurity as well as the permutation importance [50] were calculated. This was done by randomly selecting a subset of around 10 features out of all variables and training a BDT for each subset. The procedure was repeated 10 000 times[4] and based on the feature importance in an independent test set the most promising variables were selected. The correlation between the variables is shown in Figure 5.15. Out of these variables between 4 and 8 were selected randomly and used again to train 1000 BDTs, while also varying the hyper-parameters randomly within a given range. One of the resulting feature importance rankings is shown in Figure 5.16. Depending on what variables enter the training and how they are correlated, the result of these feature rankings could differ quite significantly. However, a clear trend towards some variables could be observed. It was found that, when taking into account the correlations, 4 features are sufficient to achieve the best discrimination power. The eventually selected features are listed and described in Table 5.3.

The distribution and pairwise correlations for these variables are shown in Figure 5.17 for different Z' masses and SM backgrounds. Only events within the respective mass windows of the Z' are considered. Especially for the taupair background (in 5.17c), there seems to be a good separation from the signal with almost 0 background remaining. For the $\mu\mu$ (5.17a and 5.17b) and $ee\mu\mu$ (5.17d) background such a clean separation is not obvious. It can only be noted that

[4] The training was done on a smaller, representative dataset which resulted in a training time of less than one minute for each BDT.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

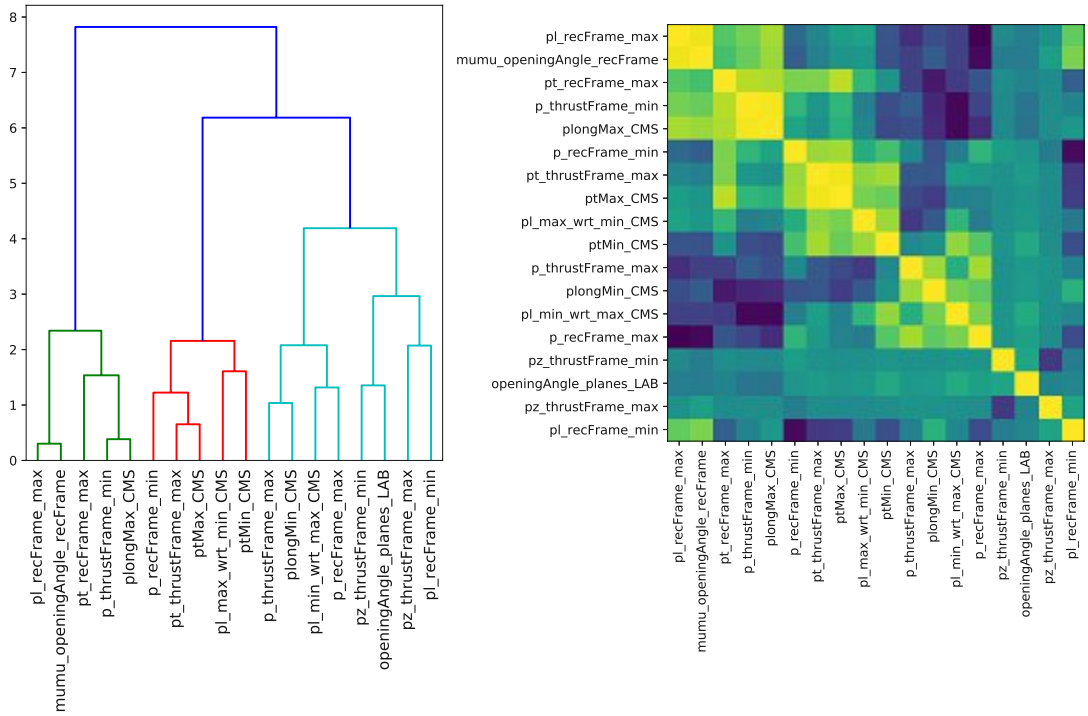


Figure 5.15: The most promising features and their correlations, selected from the 50 initial variables. They are different combinations of kinematic and topology related variables in different reference frames.

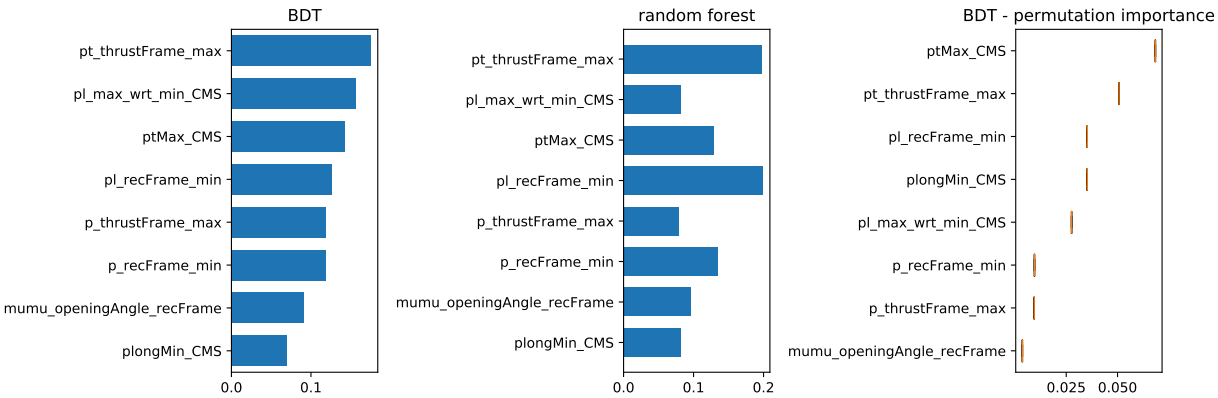


Figure 5.16: The feature importance ranking for one combination of variables, obtained with three different methods. Depending on what variables enter the training, the result of these rankings can differ quite significantly.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table 5.3: The most important features found after training a large amount of BDTs. These variables are used for training the ANN.

variable	description
$p^*_{t,\mathrm{thrust}}(\mu)$	The transverse momentum component of the muons with respect to the thrust axis in the CMS.
$p^*_{t,\mu_{\mathrm{min}}}(\mu_{\mathrm{max}})$	The transverse momentum component of the higher energetic muon with respect to the lower energetic muon in the CMS.
$p^*_{l,\mu_{\mathrm{min}}}(\mu_{\mathrm{max}})$	The longitudinal momentum component of the higher energetic muon with respect to the lower energetic muon in the CMS.
$p^*_t(\mu^+\mu^-)$	The transverse momentum of the dimuon system in the CMS.

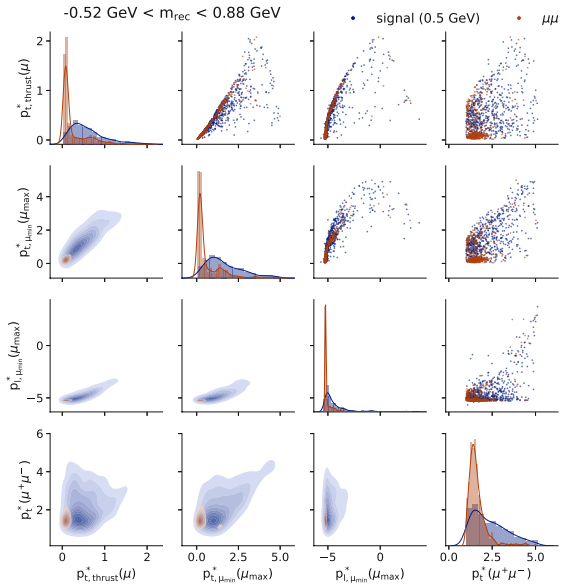
for both backgrounds in particular $p^*_t(\mu^+\mu^-)$ is important.
However, this is just a glimpse into the four-dimensional hyperspace that can be utilised when performing a multivariate analysis (MVA).

When comparing the fast BDT to a simple feedforward ANN an almost identical performance in terms of separation power inside the individual mass windows was found with the BDT being faster to train in general. However, a crucial factor when training a classifier for multiple signal samples at once is whether the results can also be applied to a signal mass that was not used for training. The individual signal samples have a peak in the recoil mass variable. When using kinematic variables (which will always be correlated in some way with the recoil mass) as input for the classifier, these peaks will correspond to separated clusters in the multidimensional hyperspace spanned by the training variables. This gives rise to the risk that the classifier excludes regions in between clusters, even if it would be occupied by another Z' mass (not used in training). So the classifier needs to be able to generalise the selection to these regions even without any training data. This can be checked by excluding some signal samples from the training process and studying the classifier's effect on the excluded samples. It would also manifest itself in the recoil mass distribution after applying a selection defined by the classifier. If the input variables are correlated to the recoil mass, and the classifier does not generalise there will be increased background ('bumps') in the recoil mass spectrum at the location of the signal samples used for training. This is referred to as *background sculpting*. It was found that the BDT output was biased towards some recoil masses while the ANN did not exhibit this behaviour at all. This could be related to the fact that the ANN describes a continuous function while the BDT is based on a series of rectangular cuts.

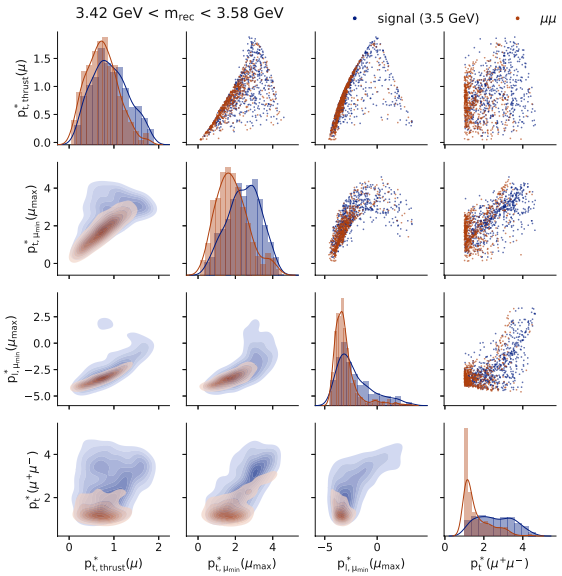
Another reason why the ANN was preferred over the BDT is that with Python libraries like `PyTorch` [51] it is straightforward to adapt any aspect of the algorithm, like designing a new custom loss function. This freedom of experimenting with different implementations while maintaining high performance was why the final ANN was realised with `PyTorch`.

Multiple combinations of different features were also checked with the ANN architecture described in the next ection. The performance was found to be worse or at best equivalent to a network using the four variables listed in Table 5.3. Also, almost no change in performance was observed when additional features were added. On the other hand, when one of the variables was

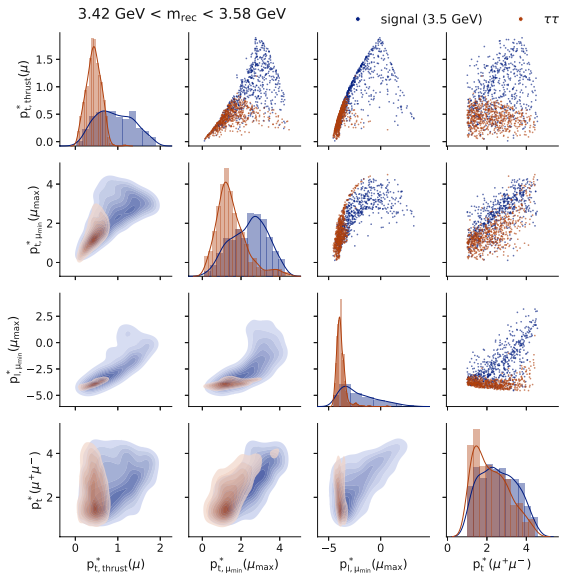
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



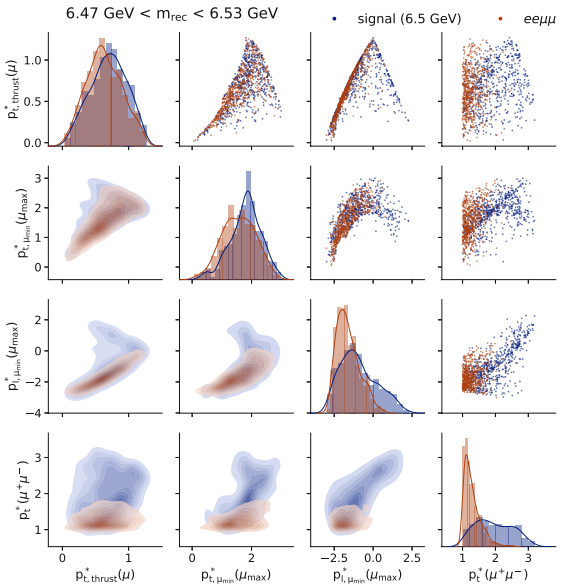
(a) A 500 MeV Z' and $\mu\mu$ background.



(b) A 3500 MeV Z' and $\mu\mu$ background.



(c) A 3500 MeV Z' and taupair background.



(d) A 6500 MeV Z' and $ee\mu\mu$ background.

Figure 5.17: The distribution and pairwise correlations for the variables used in the multivariate selection. The shown signal events and SM backgrounds are selected from the respective Z' mass window.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

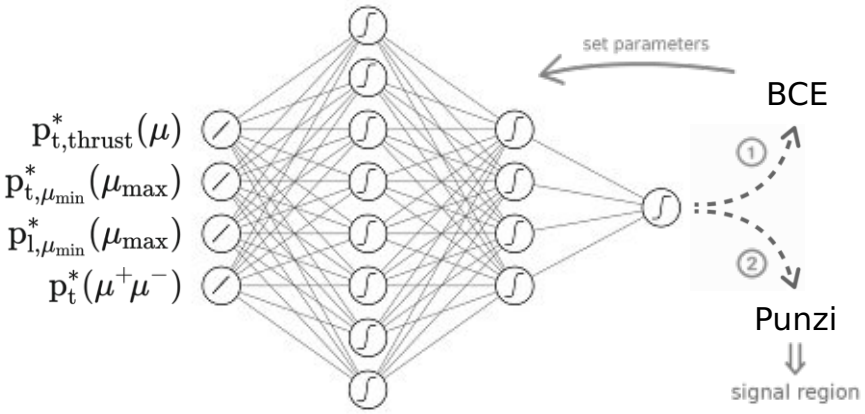


Figure 5.18: A sketch of the ANN architecture. The first training with the BCE loss function was used to set the weights and biases of the net for the second training with the custom loss function based on the Punzi figure of merit.

Table 5.4: The minimum and maximum values of the input variables, used for scaling the distributions to lie between 0 and 1.

input	minimum	maximum
$p^*_{t,thrust}(\mu)$	0	2.2
$p^*_{t,\mu_{min}}(\mu_{max})$	0	5.2
$p^*_{l,\mu_{min}}(\mu_{max})$	-5.5	2.2
$p^*_t(\mu^+\mu^-)$	1	5.2

dropped, the separation power worsened, confirming the checks performed on the BDTs.

5.3.2 ANN architecture

A rough sketch of the ANN architecture is given in Figure 5.18. Besides the four input neurons and one output neuron, two hidden layers with 8 and 4 neurons respectively are used. All layers are fully connected. The size of the net was determined empirically to give good results while keeping the network relatively small. This is to avoid overtraining and achieve a good generalisation to all possible Z' masses. It was found that a deeper net or additional neurons per layer can not improve the results, leading to the conclusion that this network already has enough flexibility to find the optimal separation for the chosen input variables.

For the activation function of the neurons in the hidden layers, a hyperbolic tangent is used while the output neuron uses a sigmoid function. Before training, the input variables were scaled to lie between 0 and 1. The minimum and maximum values of the input variables used to perform this scaling are listed in Table 5.4.

Before training the weights and biases were randomly initialised with values drawn from a normal distribution. For the training itself, only every second generated Z' mass was used,

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table 5.5: The number of training events and weight factors used to train the ANN.

	training events	scaling (50 fb^{-1})	scaling (normed)
signal	313 507	0.488	13.65
$\mu\mu$	56 465	0.05	1.4
taupair	7 774 186	0.0167	0.47
$ee\mu\mu$	413 667	0.05	1.4
total	8 557 825		
mean		0.036	1

starting from 100 MeV up to 8700 MeV. Additionally, all signal and background events that do not lie in any of the $\pm 2\sigma$ mass windows of the used signal samples were removed for the training.

5.3.3 First training (BCE)

After preparing the data as described in the previous Section about 8.5×10^6 events were remaining to use for the training, resulting in 1045 mini-batches. The exact number of events for the individual samples are listed in Table 5.5. Because there is a large difference in available events between the samples the ANN would give more importance to the larger samples (in this case this would be taupair which represents 90% of all events). This would bias the results since the number of training events does not represent the actual number of expected events. This issue can be solved by replacing the mean in Equation 2.37 with a weighted mean:

$$C_{\mathrm{BCE}}(\boldsymbol{w}, \boldsymbol{b}) = -\frac{1}{n} \sum_i [y_i(\boldsymbol{w}, \boldsymbol{b}) \ln a_i + (1 - y_i(\boldsymbol{w}, \boldsymbol{b})) \ln(1 - a_i)] \cdot s_i. \tag{5.1}$$

Here the sum is over all events in a mini-batch with size n and s_i denotes the weight factor. The desired output y_i is set to 1 for all signal events and 0 for all background events. The scaling for the background events is done according to the MC luminosities L_{MC} of the respective processes. When scaling to 50 fb^{-1} this factor is $50\ \mathrm{fb}^{-1}/L_{\mathrm{MC}}$. The weight factor for the signal events is then calculated so that the relative weight of all signal samples corresponds to 50%. To get the same weights for different target luminosities, the resulting factors are then scaled by a common factor such that the mean weight of all events is equal to 1. The resulting weights for the different samples are listed in Table 5.5.

Using the BCE loss function the network was trained with a batch size of 8192 and a LR of 1. When the loss did not decrease for 10 epochs, the LR was reduced by a factor of 0.5. The training was stopped after 300 epochs. The evolution of the loss can be seen in Figure 5.19. The output of the ANN after this first training for different values of the recoil mass $\mathrm{m_{rec}}$ is shown in Figure 5.20 for the signal events (5.20a), the $\mu\mu$ sample (5.20b), the taupair sample (5.20c) and the $ee\mu\mu$ sample (5.20b). To better visualise the output distributions the inverse of the sigmoid activation function,

$$z(a) := \sigma^{-1}(a) = \ln \frac{a}{1-a},$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

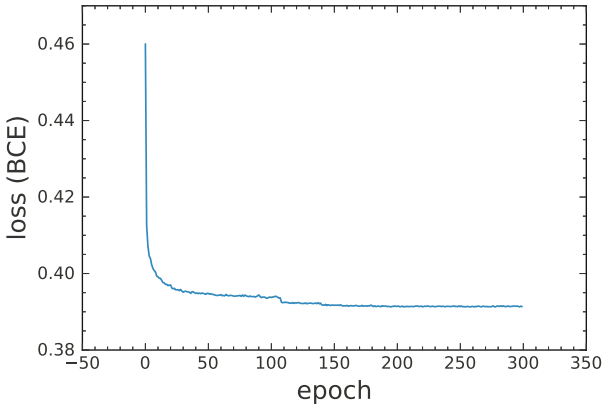


Figure 5.19: The evolution of the BCE loss for 300 epochs in the first training.

was applied on the network output, undoing the last calculation of the ANN. Note that $\sigma(5) = 0.993$ and $\sigma(-5) = 0.0067$. Above 8 GeV/c^2 the ANN can not distinguish the signal events from the $ee\mu\mu$ sample. Below 7 GeV/c^2 on the other hand, many signal events are well classified (note the logarithmic scale on the colour bar), though some $\mu\mu$ background remains. The taupair events can almost be separated entirely.

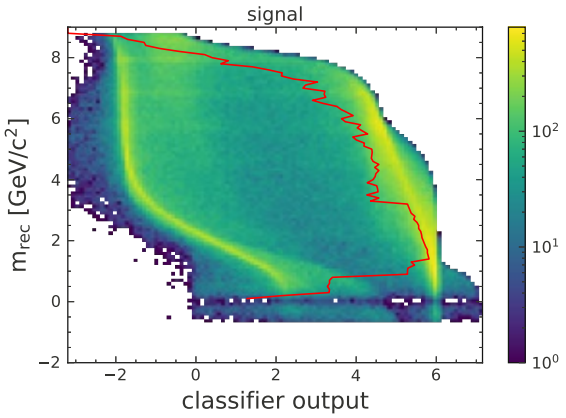
The red line in Figure 5.20 indicates the cut for the individual recoil masses one would have to apply to maximise the Punzi figure of merit in the respective mass windows. In principle, one could use this cut to define a selection. One problem with this is that the optimisation of the cut in the different windows is not very stable. This can lead to increased systematic effects if the data distribution does not precisely follow the simulated events (which is rarely the case). Especially between the generated mass points, the linear interpolation shown in the Figure is not optimal. In the next ection, we will see that we can actually do better without this additional step and it will be possible to apply a single straight cut instead.

5.3.4 Second training (Punzi)

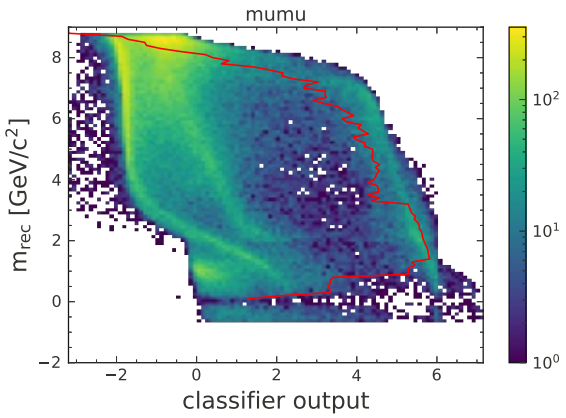
The second training utilises a custom loss function based on the Punzi figure of merit (described in Section 2.4.7). In particular, since the loss function is always minimised, the Equation for the Punzi sensitivity region (2.27) will be used. This corresponds to the inverse of the Punzi FOM. Alternatively, the negative Punzi FOM could be used for minimisation. It was found that the inverse Punzi results in slightly better performance at the lower and higher end of the recoil mass spectrum where the minimum detectable cross section is larger (corresponding to a worse sensitivity). This can be explained by the fact that these regions will get a larger weight when calculating the mean of this quantity over all the signal samples.

However, Equation 2.27 can not be used directly because the number of background events B and the signal efficiency ϵ are discrete functions of the network parameters for any fixed cut on the classifier output. The loss function should at least be differentiable. We can build a differentiable function by replacing the fixed cut on the output with a sum over all events weighted with the

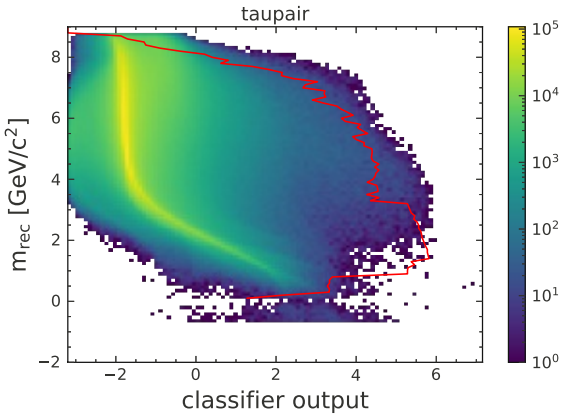
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



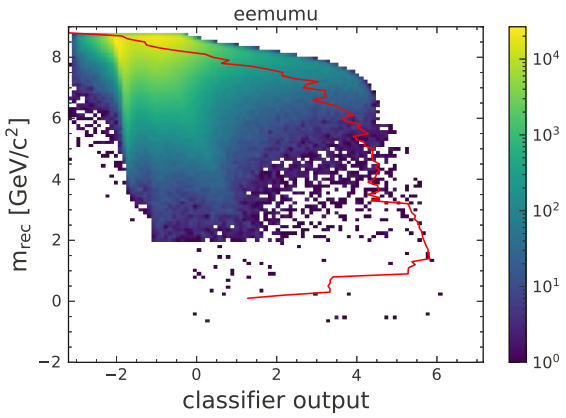
(a) The output distribution for signal events.



(b) The output distribution for $\mu\mu$ events.



(c) The output distribution for taupair events.



(d) The output distribution for $ee\mu\mu$ events.

Figure 5.20: The output of the ANN after the first training (BCE) for different values of the recoil mass. The red line indicates the cut which yields the largest Punzi figure of merit in the respective mass windows. The output distributions are shown before applying the sigmoid activation function on the output layer.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

respective value of the output. Since events classified as signal are expected to cluster around 1 and events classified as background around 0 this quantity will approximate the original function very closely. In the equation this weighting can be captured by performing the replacements

$$\begin{aligned} \epsilon(t) \to \epsilon(\boldsymbol{w}, \boldsymbol{b}) &= \sum_{i \in y_{\{1\}}} \frac{a_i(\boldsymbol{w}, \boldsymbol{b}) \cdot \varepsilon_{\text{trig}}}{N_{\text{gen}}} \quad \text{and} \\ B(t) \to B(\boldsymbol{w}, \boldsymbol{b}) &= \sum_{i \in y_{\{0\}}} a_i(\boldsymbol{w}, \boldsymbol{b}) \cdot l_i \cdot \varepsilon_{\text{trig}} \end{aligned} \qquad \text{with} \quad y_{\{j\}} := \{i \mid y_i = j\}\,. \tag{5.2}$$

The sum for the signal efficiency is over all signal events used in training denoted by the set of indices $y_{\{1\}}$ and the sum for the total number of background events is over all background events used in training with indices $y_{\{0\}}$. N_{gen} is the total number of generated signal events, $\varepsilon_{\text{trig}}$ is the trigger efficiency and l_i is a weight factor to scale the MC luminosity for the background samples to the target luminosity. This factor is different for each process and given for 50 fb^{-1} by the values in the second column of Table 5.5.

The final loss function is then given by the arithmetic mean of this continuous Punzi sensitivity calculated for all Z' masses that are used in training

$$C_{\text{Punzi}}(\boldsymbol{w}, \boldsymbol{b}) = \frac{1}{N_{Z'}} \sum_{m_{Z'}} \frac{\frac{b^2}{2} + a\sqrt{B(\boldsymbol{w}, \boldsymbol{b})} + \frac{b}{2}\sqrt{b^2 + 4a\sqrt{B(\boldsymbol{w}, \boldsymbol{b})} + 4B(\boldsymbol{w}, \boldsymbol{b})}}{\epsilon(\boldsymbol{w}, \boldsymbol{b})}, \tag{5.3}$$

where with respect to 2.27 the constant luminosity factor L was dropped.

For the training with the Punzi FOM the data was not divided into batches since the loss function is no longer based on a single event but instead on a large set of events. To still achieve a reasonable training time, Equation 5.3 is implemented in an efficient way by using sparse matrices so that $B(\boldsymbol{w}, \boldsymbol{b})$ and $\epsilon(\boldsymbol{w}, \boldsymbol{b})$ can be calculated with a single matrix multiplication.

The learning rate was initially set to 5×10^{-4} and when the loss did not decrease for 20 epochs the LR was reduced by a factor of 0.5. The training was stopped after 2500 epochs. Figure 5.21 shows the loss after every epoch. The oscillatory feature up to epoch 400 is because the learning rate was large enough that a step in the opposite direction of the gradient could increase the loss. After epoch 400, the learning rate was automatically reduced, and the algorithm could converge to a minimum.

The output of the ANN after this second training for different values of the recoil mass m_{rec} is shown in Figure 5.22 for the signal events (5.22a), the $\mu\mu$ sample (5.22b), the taupair sample (5.22c) and the $ee\mu\mu$ sample (5.22d). As before, what is shown is the network output before the last sigmoid activation, and the red line indicates the cut for the individual recoil masses one would have to apply to maximise the Punzi figure of merit in the respective mass windows. Compared to the output distribution after the first training, the classified signal and background events are now much better separated. It is possible to apply a cut at $z = 0$, corresponding to $\sigma(0) = 0.5$, and achieve a good Punzi FOM everywhere. Alternatively, it would be possible to weight the events according to the neural network output in the same way it was done for the

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

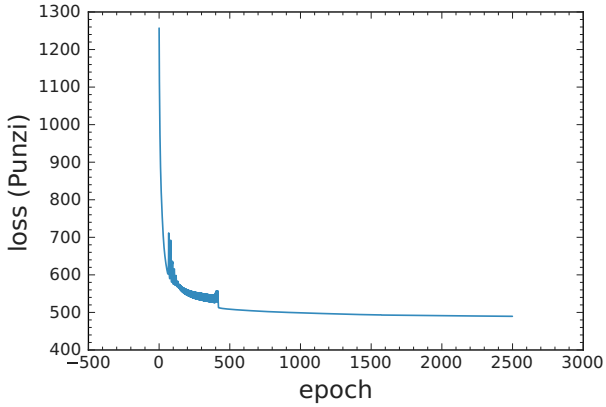


Figure 5.21: The evolution of the Punzi based loss for 2500 epochs in the second training.

Table 5.6: Hyperparameters for the first and second training of the ANN based on different loss functions.

	1. training	2. training
loss	BCE	'inverse Punzi'
batch size	8192	-
epochs	300	2500
learning rate	1	5×10^{-4}
reduce LR on plateau by	0.5	0.5
patience	10	20
training time	$\sim 30\,\text{min}$	$\sim 1\,\text{h}$

loss function. In this case, a cut was chosen since this simplifies the statistical treatment and systematic error estimation, and almost no differences were observed between the two options.

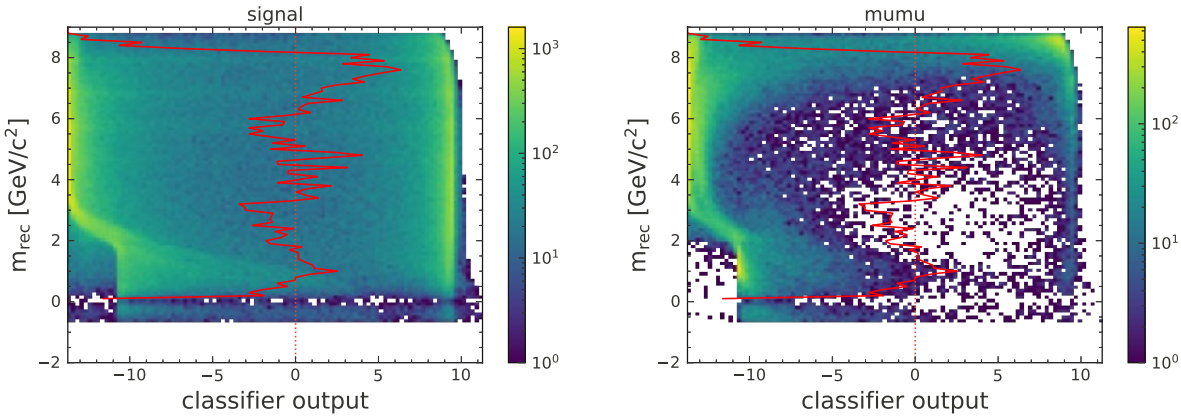
To summarise the training processes, all the hyperparameters are listed in Table 5.6. The training time is only an approximate estimate and measured on a local machine when running on an Intel(R) i7-8550U CPU @1.80 GHz. The code was also tested on a GPU where a significant speedup was observed.

5.3.5 Training results

In this Section, the performance of the ANN will be discussed. This has to be done for the different Z' masses individually since, depending on the recoil mass, the overall event shape is quite diverse and different background are dominating. All the quantities (signal efficiencies and background events) will be calculated in the $\pm 2\sigma$ windows of the respective Z' mass and always include a trigger efficiency of 89%.

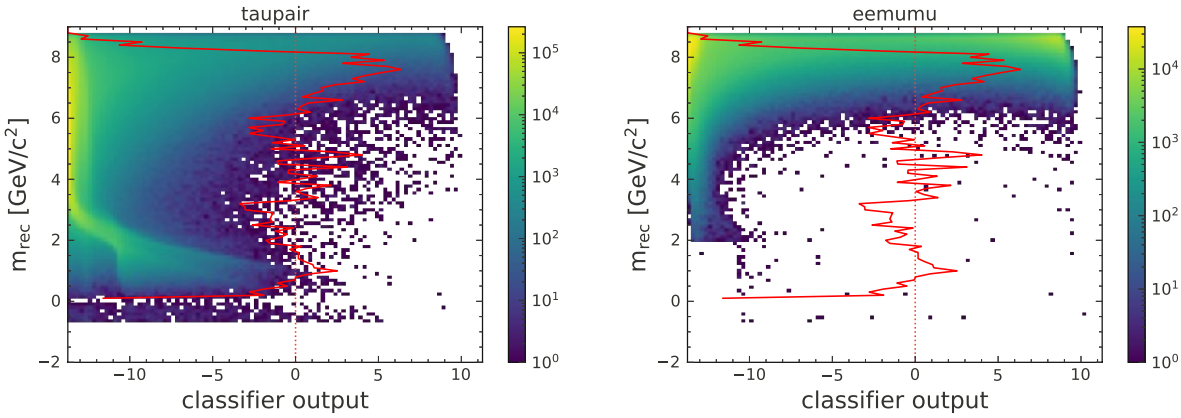
Figure 5.23 shows how the overall signal efficiency and the number of background events change for different cut values on the classifier output. The individual background contributions,

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



(a) The output distribution for signal events.

(b) The output distribution for $\mu\mu$ events.



(c) The output distribution for taupair events.

(d) The output distribution for $ee\mu\mu$ events.

Figure 5.22: The output of the ANN after the second training (FOM) for different values of the recoil mass. The red line indicates the cut which yields the largest Punzi figure of merit in the respective mass windows. The output distributions are shown before applying the sigmoid activation function on the output layer.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

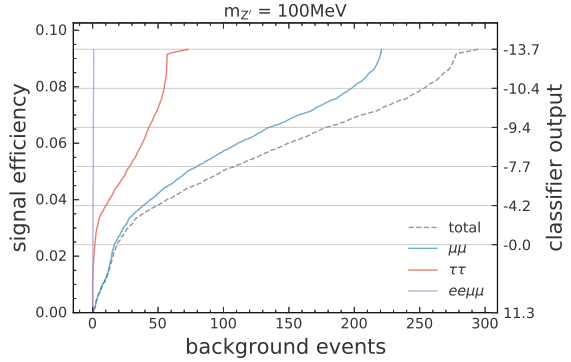
as well as the total number of background events, are shown. As shown previously in Figure 5.22 a cut around a classifier output of 0 gives the highest value of the Punzi FOM. This cut will be used to define the final selection, and we are particularly interested in signal efficiency and remaining background events at this value.
In 5.23a a Z' mass of 100 MeV is considered. The dominating background is $\mu\mu$, and while the background can be heavily reduced, some events will still remain. This also comes at the cost of a low signal efficiency of around 2%.
In 5.23b a Z' mass of 1000 MeV is considered. Here $\mu\mu$, as well as taupair backgrounds, are relevant. The ANN manages to almost completely remove all taupair events while some $\mu\mu$ background remains. The signal efficiency is with 6% now much larger than for 100 MeV.
In 5.23c a Z' mass of 3000 MeV is considered. In this region of the recoil mass, the dominating background is taupair. As before the network can find a good separation between signal and the taupair events, leading to less than one expected background event ("0 background regime"). The signal efficiency is slightly larger at 7.5%.
In 5.23d a Z' mass of 7000 MeV is considered. Here initially the taupair events still give the largest contribution to the background, but $ee\mu\mu$ processes start to become relevant. At the selected working point, the taupair background is again well separated, and $ee\mu\mu$ events compose most of the remaining background. The signal efficiency is still around 7.5%.
For larger Z' masses the remaining background increases steadily due to $ee\mu\mu$ processes that can not be separated effectively.

Similar information is given in Figure 5.24 where the true positive rate is shown together with the false positive rate for different cuts on the classifier output. The crosses mark the mentioned working points. Note that such a plot can be misleading since it does not portray information about the selections applied before the classifier. The rates are only with respect to these selections, and the shape of the curves can change when they are altered.

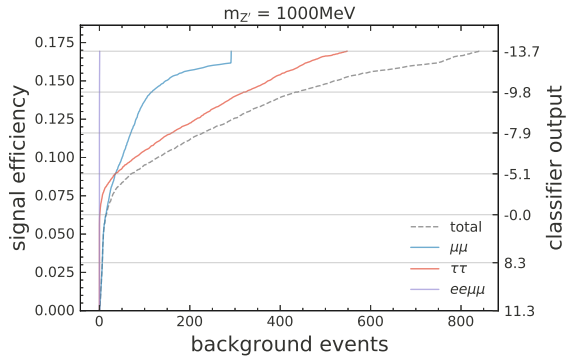
In Figure 5.25 the Punzi sensitivity region as defined in Equation 2.27 is shown for different Z' masses before applying the ANN and after. This can be interpreted as the minimum detectable cross section. More details about the properties of this quantity are described by Punzi [see 29, p. 2].

Figure 5.26 shows the Punzi FOM as defined in Equation 2.28. For this quantity, higher values correspond to better sensitivity. It is compared for three different selections. One after the first training with the BCE for the optimised values with respect to the Punzi FOM (red line in Figure 5.20). The other two after the second training (FOM), once again for the optimised values with respect to the Punzi FOM (red line in Figure 5.22) and then also for the straight cut at 0 (dotted line in Figure 5.22). Almost no difference can be observed comparing these two selections, showing that the Punzi loss function training worked as expected. It is justified to apply a single cut on the ANN output without further optimisation. However when comparing this selection to the best possible selection after the first training an increased FOM between 1 GeV and 5 GeV is apparent while performing similarly for the rest of the mass range. By changing the loss function, the neural network is able to define a selection that was not possible before. This should not come as a surprise since it is trained explicitly to maximise the FOM rather than minimise the BCE. It seems that this improvement is most pronounced for recoil masses with a low number of expected backgrounds.

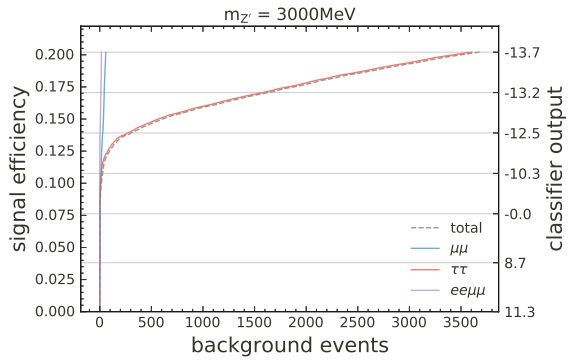
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



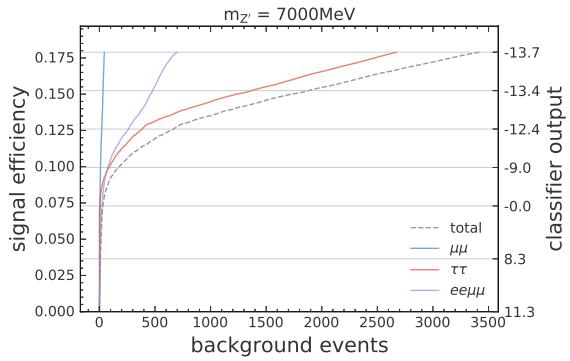
(a) For a Z' mass of 100 MeV.



(b) For a Z' mass of 1 GeV.



(c) For a Z' mass of 3 GeV.



(d) For a Z' mass of 7 GeV.

Figure 5.23: The signal efficiency and the number of remaining background events for different working points. The background events are counted in the respective mass windows and a trigger efficiency of 89 % is applied on both signal and background. The classifier output before applying the sigmoid activation function is used. A cut at 0 will be chosen for the final selection.

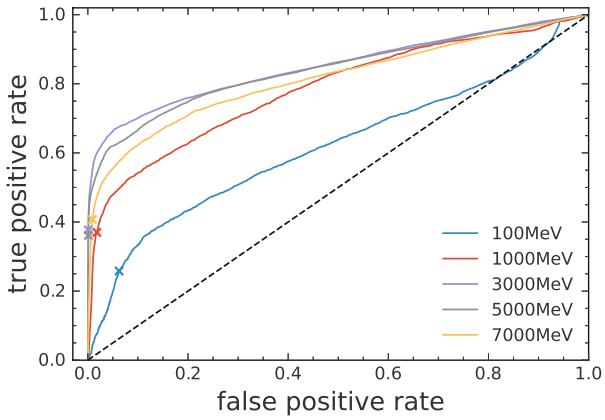


Figure 5.24: The true positive and false positive rates of the classifier, shown for different Z' masses. Only events within the mass windows are used. The crosses mark the final selection.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

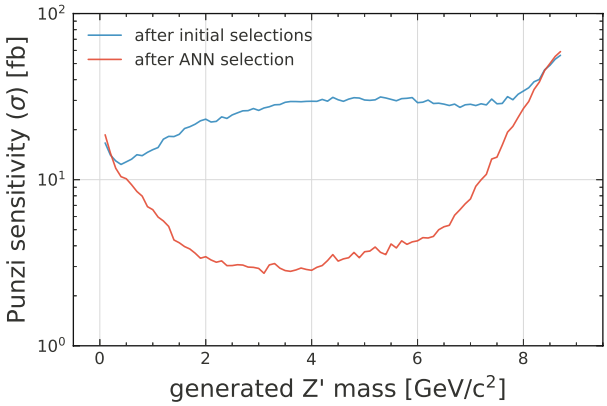


Figure 5.25: The Punzi sensitivity region ('minimum detectable cross section') for different Z' masses before and after applying the ANN.

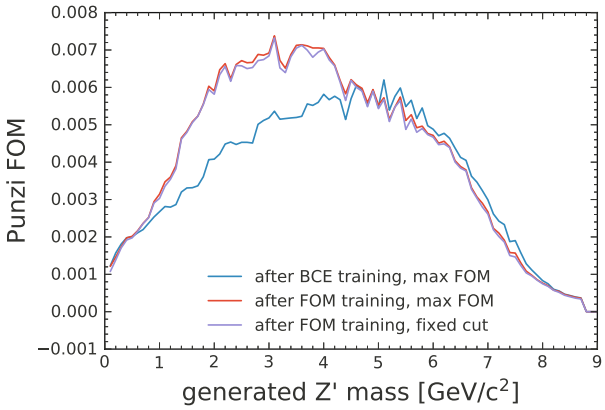
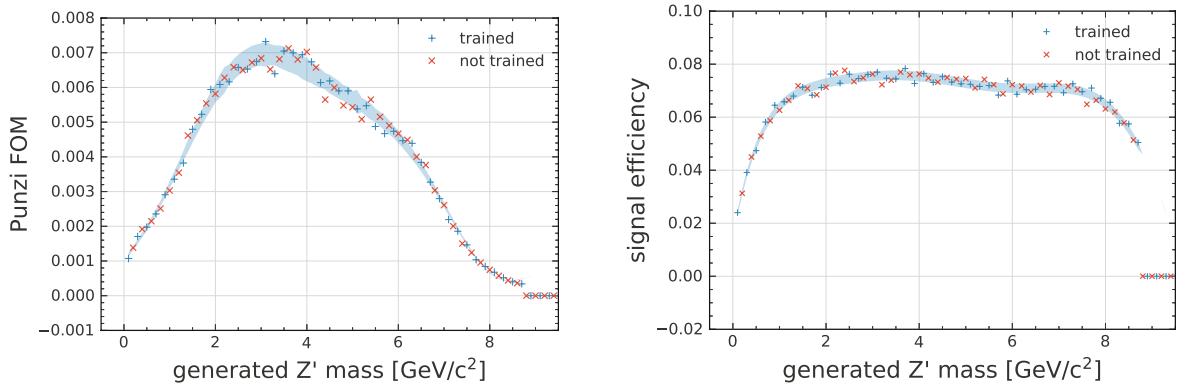


Figure 5.26: The Punzi FOM compared for three different selections: The optimised selection after the first training (BCE), the optimised selection after the second training (FOM) and the cut at a fixed value of the ANN output after the second training.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



(a) The Punzi FOM for all generated Z' masses. It gives the best results between 2 GeV and 5 GeV.

(b) The signal efficiency for all generated Z' masses. It is around 7% for a Z' mass between 1 and 8 GeV.

Figure 5.27: A comparison of the ANN performance for the masses used in training and those left out. No difference can be observed between the samples. The shaded area indicates the expected $\pm 1\sigma$ statistical error.

The figures in this chapter showed all generated Z' masses (in steps of 100 MeV) but bear in mind that for the training of the ANN only every second mass was used and all events in between were excluded from training. If the network is biased towards the used masses, this would be apparent in the distributions, yet all the figures show a very smooth change when varying the Z' mass. This is shown in Figure 5.27 explicitly, where the masses used in training are shown as `'+'` and those left out as `'x'`. Both the Punzi FOM (5.27a) and the signal efficiency (5.27b) after the final selection do not exhibit any obvious difference between these samples. The signal efficiency is for a large part of the mass range stable around $7 - 7.5\,\%$. The shaded area denotes the $\pm 1\sigma$ statistical error which is consistent with the variation in between the samples.

This shows that the ANN can generalise even to Z' masses that were never used to train the network. This guarantees that no signal will be lost when applying the classifier, no matter what mass it might have. Furthermore, this shows that there is no issue with overtraining of the network since the unused Z' masses can be seen as a test sample. In the current training, no division in training and test data is done for the used masses, but in initial studies the dataset was separated using 60% for training and 40% for testing. Since no difference between the output distributions of these samples could be observed, it was decided to deploy all data for training but still use only every second mass so the rest can serve as a test sample. The reason why overtraining is not an issue at all could be that the number of training events is very large, and the model is relatively simple with 'only' 81 parameters.

An argument as to why the ANN is able to generalise so well and why it is possible to only use one network for the whole recoil mass range can be made based on Figure 5.28. Here a 3D scatter plot is shown for the (scaled) variables $\mathrm{p}^*_{\mathrm{t,thrust}}(\mu)$, $\mathrm{p}^*_{\mathrm{t},\mu_{\mathrm{min}}}(\mu_{\mathrm{max}})$ and $\mathrm{p}^*_{\mathrm{l},\mu_{\mathrm{min}}}(\mu_{\mathrm{max}})$, using only events in a region of 0.5 GeV around a value of $\mathrm{p}^*_{\mathrm{t}}(\mu^+\mu^-) = 2.2$ GeV/c. This makes it possible to show the separation boundary (the green sheet in Figure 5.28) defined by the final selection, where the input of $\mathrm{p}^*_{\mathrm{t}}(\mu^+\mu^-)$ was set to 2.2 GeV/c. To see where different recoil masses

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

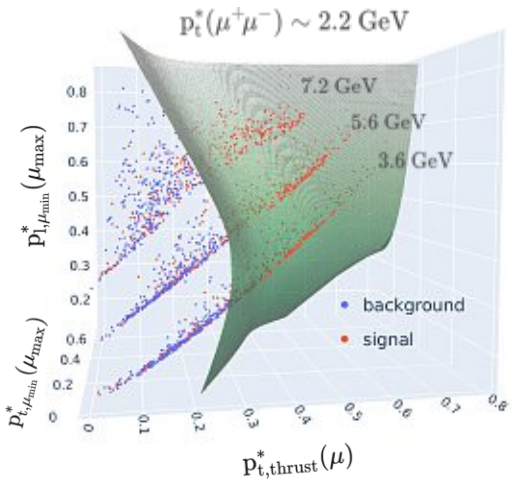


Figure 5.28: A 3D scatter plot showing the input space of the ANN with $p_t^*(\mu^+\mu^-)$ set to 2.2 GeV/c. The separation boundary defined by the final selection (green sheet) separates the planes corresponding to different recoil masses in a way that optimises the selection for all Z' masses.

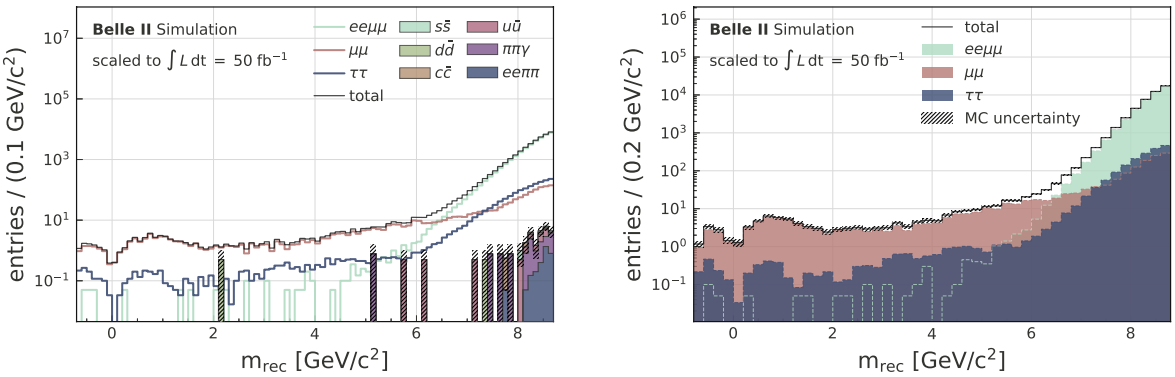
are distributed only events in the mass windows of a 3.6 GeV, a 5.6 GeV and a 7.2 GeV Z' are shown and labelled as signal and background.

It is visible that the respective masses correspond to planes in this three-dimensional space and occupy distinct regions. This partition enables the network to adapt for differences in the recoil mass regions, and there is no need to train separate classifiers. This partition also leads to the desired generalisation property. The separation boundary slices all the mass planes and defines an optimised signal region for each of them. The network interpolates the selection between the planes, defining a natural extension to all the Z' masses. However, Figure 5.28 does not reflect the actual situation since only three masses are shown. In reality, there are no such big gaps in this space since the mass windows can overlap, and the distribution is spread because of resolution effects. Nevertheless, the same argument applies, and it is possible that just this smearing enables the desired generalisation. Moreover, there is one additional dimension that the network has available. For different inputs of $p_t^*(\mu^+\mu^-)$ the picture is qualitatively similar, but the separation boundary (and the distribution of the events) will vary depending on its value. This also highlights the power of adding one additional dimension when defining a selection.

5.4 Remaining background events

After applying the multivariate selection by cutting on the ANN output, the remaining background events in bins of the recoil mass are shown in Figure 5.29. In 5.29a all processes with remaining events are shown. The processes $u\bar{u}$, $d\bar{d}$, $s\bar{s}$, $c\bar{c}$, $\pi\pi\gamma$ and $ee\pi\pi$ contribute only for $m_{rec} > 7$ GeV and considering the dominating $ee\mu\mu$ background in that region they will be neg-

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



(a) Showing all backgrounds with remaining events after the multivariate selection. The filled histograms are stacked.

(b) Showing only the relevant backgrounds: $ee\mu\mu$, $\mu\mu$ and taupair.

Figure 5.29: Background distributions of the recoil mass after applying all selections.

lected in the following. Figure 5.29b shows only the relevant background processes. The taupair events which were dominant in a large region before (see Figure 5.14 on page 66) contribute now the least. The $\mu\mu$ background is most relevant up to 6 GeV. Above that value, the $ee\mu\mu$ events start to take over.

In Table 5.7 the total number of remaining background events for all generated processes are listed before and after applying the ANN for a luminosity of 50 fb^{-1}. The taupair background got reduced by 99.3%, the $\mu\mu$ background by 77% and the $ee\mu\mu$ background by 67.6%. The first column shows the total number of events that are expected to occur for this luminosity.
Table 5.8 lists the remaining background events after all selections separated in recoil mass regions. The error is the statistical error of the MC samples.

Table 5.9 and 5.10 show the same for a target luminosity of 500 fb^{-1}. The initial selections are the equivalent, but the ANN is trained using the increased luminosity. This affects the second training (the Punzi FOM depends on luminosity) and the resulting signal region differs, leading in this case to a tighter selection.

More detailed listings of the signal efficiencies and the expected number of background events in the mass windows of all the generated Z' masses can be found in the appendix (page 129 for 50 fb^{-1} and page 131 for 500 fb^{-1}).

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table 5.7: The expected number of events for each of the simulated processes scaled to 50 fb^{-1}. The first column gives the total number of events expected with this luminosity, the second column gives the number of events after all univariate selections and the third column the number of events after applying a cut on the ANN output.

process	expected events	after univariate selections	after ANN
$ee \to \tau\tau$	4.6×10^7	$(3.1059 \pm 0.0007) \times 10^5$	2202 ± 6
$ee \to ee\mu\mu$	9.44×10^8	$(1.9635 \pm 0.0010) \times 10^5$	$(6.368 \pm 0.006) \times 10^4$
$ee \to \mu\mu$	5.74×10^7	7261 ± 19	1676 ± 9
$ee \to u\bar{u}$	8.02×10^7	38 ± 4	22.5 ± 3.4
$ee \to \pi\pi\gamma$	7.65×10^6	35 ± 5	34 ± 5
$ee \to ee\pi\pi$	9.48×10^7	21.5 ± 1.0	12.8 ± 0.8
$ee \to c\bar{c}$	6.64×10^7	12.0 ± 2.4	4.5 ± 1.5
$ee \to d\bar{d}$	2.00×10^7	12.0 ± 2.4	6.5 ± 1.8
$ee \to s\bar{s}$	1.92×10^7	7.0 ± 1.9	1.0 ± 0.7
$ee \to B^0\overline{B}^0$	2.55×10^7	2.5 ± 1.1	0
$ee \to eeKK$	3.99×10^6	0.15 ± 0.09	0.05 ± 0.05
$ee \to eeee$	1.99×10^9	0	0
$ee \to ee$	1.47×10^{10}	0	0
$ee \to eepp$	5.85×10^5	0	0
$ee \to B^+B^-$	2.7×10^7	0	0

Table 5.8: The expected number of events for three different mass regions after applying all the selections. The numbers are scaled to correspond to 50 fb^{-1}.

mass range [GeV/c^2]:	$\mathrm{m_{rec}} < 3$	$3 < \mathrm{m_{rec}} < 6$	$6 < \mathrm{m_{rec}} < 9$
$ee \to \tau\tau$	4.38 ± 0.27	12.6 ± 0.5	2185 ± 6
$ee \to ee\mu\mu$	0.50 ± 0.16	9.2 ± 0.7	$(6.367 \pm 0.006) \times 10^4$
$ee \to \mu\mu$	58.6 ± 1.7	113.5 ± 2.4	1504 ± 9
$ee \to u\bar{u}$	0	0.5 ± 0.5	22.0 ± 3.3
$ee \to \pi\pi\gamma$	0	0.8 ± 0.8	33 ± 5
$ee \to ee\pi\pi$	0	0	12.8 ± 0.8
$ee \to c\bar{c}$	0	0	4.5 ± 1.5
$ee \to d\bar{d}$	0.5 ± 0.5	0	6.0 ± 1.7
$ee \to s\bar{s}$	0	0	1.0 ± 0.7
$ee \to B^0\overline{B}^0$	0	0	0
$ee \to eeKK$	0	0	0.05 ± 0.05
$ee \to eeee$	0	0	0
$ee \to ee$	0	0	0
$ee \to eepp$	0	0	0
$ee \to B^+B^-$	0	0	0

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table 5.9: The expected number of events for each of the simulated processes scaled to 500 fb^{-1}. The first column gives the total number of events expected with this luminosity, the second column gives the number of events after all univariate selections and the third column the number of events after applying a cut on the ANN output.

process	expected events	after univariate selections	after ANN
$ee \to \tau\tau$	4.6×10^8	$(3.1059 \pm 0.0007) \times 10^6$	$(3.150 \pm 0.007) \times 10^4$
$ee \to ee\mu\mu$	9.44×10^9	$(1.9635 \pm 0.0010) \times 10^6$	$(7.255 \pm 0.006) \times 10^5$
$ee \to \mu\mu$	5.74×10^8	$(7.261 \pm 0.019) \times 10^4$	$(1.711 \pm 0.009) \times 10^4$
$ee \to u\bar{u}$	8.02×10^8	$(3.8 \pm 0.4) \times 10^2$	240 ± 35
$ee \to \pi\pi\gamma$	7.65×10^7	$(3.5 \pm 0.5) \times 10^2$	$(3.5 \pm 0.5) \times 10^2$
$ee \to ee\pi\pi$	9.48×10^8	215 ± 10	144 ± 8
$ee \to c\bar{c}$	6.64×10^8	120 ± 24	55 ± 17
$ee \to d\bar{d}$	2.00×10^8	120 ± 24	65 ± 18
$ee \to s\bar{s}$	1.92×10^8	70 ± 19	10 ± 7
$ee \to B^0\overline{B}^0$	2.55×10^8	25 ± 11	0
$ee \to eeKK$	3.99×10^7	1.5 ± 0.9	1.0 ± 0.7
$ee \to eeee$	1.99×10^{10}	0	0
$ee \to ee$	1.47×10^{11}	0	0
$ee \to eepp$	5.85×10^6	0	0
$ee \to B^+B^-$	2.7×10^8	0	0

Table 5.10: The expected number of events for three different mass regions after applying all the selections. The numbers are scaled to correspond to 500 fb^{-1}.

mass range [GeV/c^2]:	$\mathrm{m_{rec}} < 3$	$3 < \mathrm{m_{rec}} < 6$	$6 < \mathrm{m_{rec}} < 9$
$ee \to \tau\tau$	19.0 ± 1.8	80 ± 4	$(3.140 \pm 0.007) \times 10^4$
$ee \to ee\mu\mu$	5.5 ± 1.7	58 ± 5	$(7.254 \pm 0.006) \times 10^5$
$ee \to \mu\mu$	521 ± 16	1005 ± 22	$(1.559 \pm 0.009) \times 10^4$
$ee \to u\bar{u}$	0	0	240 ± 35
$ee \to \pi\pi\gamma$	0	8 ± 8	$(3.4 \pm 0.5) \times 10^2$
$ee \to ee\pi\pi$	0	0	144 ± 8
$ee \to c\bar{c}$	0	0	55 ± 17
$ee \to d\bar{d}$	0	0	65 ± 18
$ee \to s\bar{s}$	0	0	10 ± 7
$ee \to B^0\overline{B}^0$	0	0	0
$ee \to eeKK$	0	0	1.0 ± 0.7
$ee \to eeee$	0	0	0
$ee \to ee$	0	0	0
$ee \to eepp$	0	0	0
$ee \to B^+B^-$	0	0	0

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

5.5 Statistical evaluation

After selecting the signal region, a statistical analysis was performed to obtain the expected limits (when assuming no signal in the data) and quantify how well the Standard Model describes the observed data. If a significant deviation is found, this would be a hint for new physics. All the studies shown here are based on simulated data only, even when the term 'observed data' is used.

5.5.1 Likelihood specification

A binned likelihood method was used and systematic uncertainties are included with nuisance parameters as described in Section 2.4. The form of the likelihood is

$$L(\mu, \boldsymbol{\nu}) = \prod_{i=1}^{N} \frac{(\mu s_i(\boldsymbol{\nu}) + B_i(\boldsymbol{\nu}))^{n_i}}{n_i!} e^{-(\mu s_i(\boldsymbol{\nu}) + b_i(\boldsymbol{\nu}))} \prod_{\nu \in \boldsymbol{\nu}} \mathrm{Gaus}(y_\nu; \nu, \sigma_\nu), \tag{5.4}$$

where

N	is the total number of bins,
μ	is the signal strength (will correspond to the signal cross section in fb),
$\boldsymbol{\nu}$	are the nuisance parameters,
s_i	is the number of expected signal events in bin i, depending on $\boldsymbol{\nu}$,
B_i	is the total number of expected background events in bin i, depending on $\boldsymbol{\nu}$,
n_i	is the observed number of events in bin i,
Gaus	is the PDF of a normal distribution with mean ν and standard deviation σ_ν,
y_ν	is the measured value for ν (auxiliary measurement) and
σ_ν	is the expected value of the uncertainty on ν.

Furthermore, the total number of expected signal and background events can be decomposed in

$$s_i(\boldsymbol{\nu}) = \prod_{\nu \in \boldsymbol{\kappa}_i^s(\boldsymbol{\nu})} \nu \cdot s_i(\boldsymbol{\nu})^0 \qquad \text{and} \qquad B_i(\boldsymbol{\nu}) = \sum_{b \in \mathcal{B}} \left(\prod_{\nu \in \boldsymbol{\kappa}_i^b(\boldsymbol{\nu})} \nu \cdot b_i(\boldsymbol{\nu})^0 \right), \tag{5.5}$$

where

$\mathcal{B}$	are independent background samples that contribute to the total background,
$\boldsymbol{\kappa}_i^b$	are multiplicative modifiers for the background sample b in bin i,
$\boldsymbol{\kappa}_i^s$	are multiplicative modifiers for the signal in bin i,
$s_i^0(\boldsymbol{\nu})$	is the nominal number of expected signal events per fb^{-1} and
$b_i^0(\boldsymbol{\nu})$	is the nominal number of expected background events per fb^{-1}.

If the signal or background PDF is available in analytical form, s_i^0 and b_i^0 can be calculated with

$$\eta \int_a^b f(x|\boldsymbol{\theta})\, dx, \tag{5.6}$$

where $f(x|\boldsymbol{\theta})$ is the respective PDF, a and b are the upper and lower edge of bin i and η is the total number of expected events from the sample. If the distribution of the samples is not

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

known or too complicated to describe analytically, s_i^0 and b_i^0 can be estimated directly from MC simulations. This is called the *template fit* method and the approach used for the present analysis.

The multiplicative modifiers used for this analysis are concerning only the normalisation of the individual samples. In general, this framework can be extended to include additional systematic uncertainties, like the statistical error from MC simulation or the shape of the signal or background distribution. While the modifiers for the normalisation share a single nuisance parameter for all the bins in a specific sample, additional modifiers might introduce one nuisance parameter per bin. When performing this analysis on actual data, such additional uncertainties should be included in the model. For now, they were not introduced since the high number of additional parameter increases the fit time significantly without affecting the results. This is expected since data was generated from MC without changing its shape and the statistical error from MC is negligible for the luminosities considered. However, for real data, it can not be expected that the shape of the distribution follows what is obtained from MC. This has to be accounted for with additional free parameters.

In Table 5.11 the multiplicative modifiers included in the model are listed together with the auxiliary measurement y and standard deviation σ used in the constraint term $\mathcal{N}(y|\nu,\sigma)$. The luminosity $\mathcal{L}$ corresponds to the total integrated luminosity of the observed data in fb^{-1} and will be provided by a separate study based on counting Bhabha and digamma events [see 52]. The uncertainty on the luminosity (1%) is the systematic error given by measurements of the luminosity in 2020. The auxiliary measurement for the normalisation modifiers is set to 1 since our best knowledge of the expected number of events comes from MC simulation and is already included in s_i^0 and b_i^0. Here the MC simulation itself can be seen as the auxiliary measurement. Alternatively, a different normalisation could be obtained from a control channel. The uncertainty on the normalisation is based on the uncertainty of additional systematic effects that influence the yields. For example, this could be the trigger efficiency, particle identification, tracking or data/MC disagreement in control channels. This uncertainty can be estimated from dedicated systematic studies (e.g. the particle identification study described in Chapter 4). Since some of these studies are common for multiple analyses, they are often provided by the performance group for the whole collaboration. Other, more channel-specific studies, can be performed on control channels or directly on the analysis dataset (without unblinding). These studies are not subject to this thesis since at the time of writing the processed data was not yet fully available. However, to still perform the statistical evaluation, a conservative estimate on the normalisation uncertainty based on the systematic studies conducted for the phase 2 measurement [1] was used. Since both the software and the detector have improved significantly since then, it is not expected to exceed these values.

Taking the logarithm of the likelihood 5.4 yields

$$\ln L(\mu,\boldsymbol{\nu}) = \sum_{i=1}^{N}\Big(-\mu s_i(\boldsymbol{\nu}) - B_i(\boldsymbol{\nu}) + n_i \ln\big(\mu s_i(\boldsymbol{\nu}) + B_i(\boldsymbol{\nu})\big)\Big) - \frac{1}{2}\sum_{\nu\in\boldsymbol{\nu}}\frac{(y_\nu-\nu)^2}{\sigma_\nu^2} + C, \tag{5.7}$$

where C is a constant factor, not depending on μ or $\boldsymbol{\nu}$. The logarithm of the profile likelihood

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table 5.11: The multiplicative modifiers considered, together with their auxiliary measurements.

	y_ν	σ_ν	validity
signal normalisation	1	0.1	signal
taupair normalisation	1	0.2	background sample ($\tau\tau$)
$\mu\mu(\gamma)$ normalisation	1	0.2	background sample ($\mu\mu\gamma$)
$ee\mu\mu$ normalisation	1	0.2	background sample ($ee\mu\mu$)
luminosity	$\mathcal{L}$ [fb^{-1}]	$0.01 \cdot \mathcal{L}$ [fb^{-1}]	all samples

ratio used to construct the test statistic $\tilde{t}_\mu$ (see Equation 2.20) is

$$\ln \lambda_p(\mu) = \ln L(\mu, \widehat{\widehat{\boldsymbol{\nu}}}) - \ln L(\widehat{\mu}, \widehat{\boldsymbol{\nu}}). \tag{5.8}$$

5.5.2 The Python library pyhf

In this Section, I will briefly describe the Python library `pyhf` [53], which was used to perform the statistical analysis.

`Pyhf` is a Python implementation of `HistFactory` [54] and can be used to perform fits and statistical tests based on [24] with the binned likelihood method. The likelihood is defined in a declarative way using the JSON data format. This fully specifies the statistical model of the analysis in a single file that can easily be preserved, published and used to reproduce the analysis results with various statistical methods. Published likelihoods would also make it easier to combine results from different measurements.

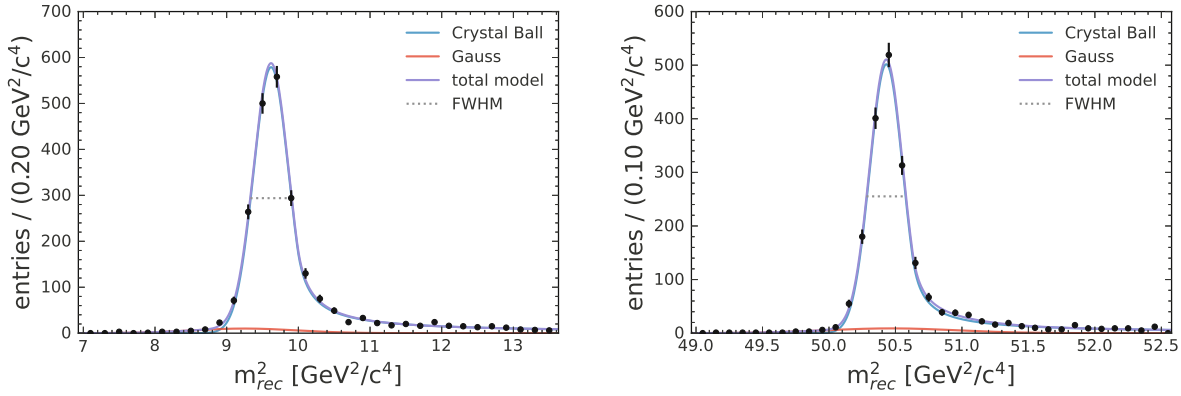
The optimisation in `pyhf` is performed either with `MINUIT` or `scipy.optimize`. As a backend it uses either `numpy`, `JAX`, `TensorFlow` or `PyTorch`, which are optimised to efficiently perform calculations based on vectors and tensors. The latter three are also designed for fast machine learning and inherently support multiprocessing and GPU acceleration. This makes `pyhf` much faster than the C++ based `HistFactory`. An additional feature enabled by using these machine learning libraries is autodifferentiation.

In the current version of `pyhf` it is not possible to perform statistical inference based on the test statistic $\tilde{t}_\mu$, described in Section 2.4. However, I could implement this feature in a private fork of the library, which was then used to obtain the following results.

5.5.3 Binning

In order to use the binned ML method, a binning scheme needs to be defined. The obvious variable to divide in bins is the squared recoil mass. The bin width should not be too large compared to the width of the signal, because this would lead to a loss of statistical power. A bin width equal to the full width at half maximum (FWHM) of the signal peak was found to be small enough to not lose any information while keeping the total number of bins to a reasonable number. Because no high number of signal events is expected, it is not an issue that the exact

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



(a) The fit for $m_{Z'} = 3100$ MeV.

(b) The fit for $m_{Z'} = 7100$ MeV.

Figure 5.30: Fits of the signal distribution (after all selections) in order to compute the FWHM. The signal is modelled with the sum of a Gaussian and a Crystal Ball function.

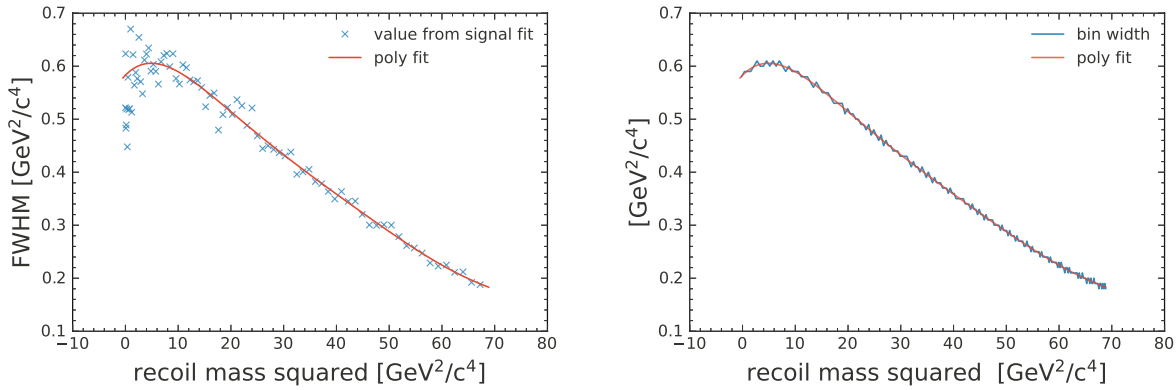
signal shape is not resolved. The FWHM is easy to compute by fitting the signal distribution (after all selections) as is shown in Figure 5.30 for two mass points.

In this case, an unbinned maximum likelihood fit was performed where the model was the sum of a Gaussian and a Crystal Ball function. The relative weight and all parameters for the PDFs, were free-floating during the fit. The Crystal Ball function was chosen to describe the asymmetric tail of the signal distribution, which is a result of ISR processes. The distribution can be described quite well by a Crystal Ball function alone, but it was found that including a Gaussian helps to model the tails, resulting in a much better fit. Since this signal model is only used to obtain the bin sizes, no thorough validation of this model is necessary.

In Figure 5.31a the FWHM obtained from the fit is shown for each generated mass point. The width of the signal peak decreases for larger recoil masses because of a better momentum resolution for the (low momentum) muons in the event. To approximate the FWHM analytically, these values were fit with a polynomial of degree six, resulting in the red curve shown in Figure 5.31a.

The bin widths were calculated so that the value of the polynomial at the centre of each bin is equal to the respective bin width. This resulted in a total of 196 bins. Figure 5.32 shows the bin edges (vertical lines) together with the generated mass points (crosses). The bin edges were rounded to two decimal places and are listed in Table A.5. In Figure 5.31b the resulting bin widths are shown together with the polynomial. When searching for a signal in data, the statistical procedure has to be repeated with many signal models, corresponding to different mass hypotheses of the Z'. This should be done approximately in steps of the recoil mass resolution. Otherwise, a signal might be unnoticed by the procedure. With the binning defined above, this is achieved with testing at least one signal model per bin. As apparent in Figure 5.32, at large values of the squared recoil mass, not every bin is covered by a generated signal. For the present work, this has been neglected since the procedure is only tested with simulated data. When performing the statistical treatment on the actual data, the same procedure can be used, only with more signal models. These models can be obtained by generating a new signal MC sample per bin

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



(a) The fit results together with the polynomial used to capture the dependence of the width on the Z' mass.

(b) The width resulting width of the bins, calculated to follow the shape of the polynomial.

Figure 5.31: The FWHM for each generated mass point and the thereof resulting width of the bins in the squared recoil mass distribution.

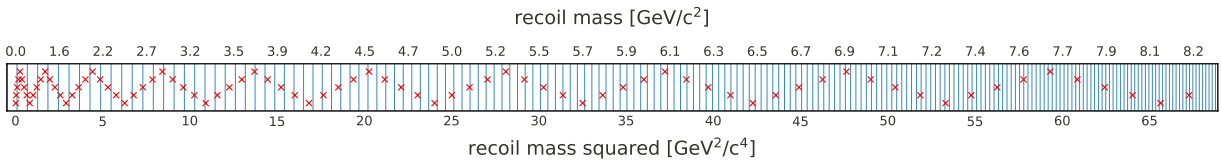


Figure 5.32: The bins defined for the recoil mass to be used for the fit. The bin width is calculated based on the FWHM of the signal. The vertical lines show the bin edges and the crosses indicate the mass points where a signal was generated. For visual clarity, the individual mass points are shifted with respect to the adjacent points.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

and using this as a template. Alternatively, the analytical model from the fits of the available signal distributions can be used to describe the signal at arbitrary mass points by interpolating the parameters of the fit. This procedure is not guaranteed to work and needs to be validated carefully. In the present case, since the number of bins is not too large, the template method is preferential.

5.5.4 Testing the validity of the asymptotic formulae

This Section will compare the asymptotic formulae to the empirical distributions obtained from many simulated samples of the statistical model. This is also called the *toy MC* method or it is said to 'throw toys'. If not stated otherwise, the sample size is 5000. I will compare the distribution of $\tilde{t}_\mu$ for different assumed values of the Z' mass m and the cross section μ. The distribution will differ depending on the true value of the cross section, denoted μ'. Here two cases are of interest: $\mu' = 0$ and $\mu' = \mu$. The distribution for $\mu' = 0$ will be called the *background like* distribution and it is used to obtain the expected value of the intervals, assuming no signal in the data. The distribution for $\mu' = \mu$ will be called the *signal like* distribution and it is used to construct the confidence intervals for μ. The special case $\mu' = \mu = 0$ is important for testing the background-only hypothesis and calculating the significance of a hypothetical signal.

Before comparing the distribution of $\tilde{t}_\mu$ (as defined in Equation 2.20) and its asymptotic approximation (see Equation 2.21 for the asymptotic formula of the cumulative distribution) it is instructive to study the estimator $\widehat{\mu}$. The asymptotic formulae are based on a result formulated by Wald [55], which states

$$-2\ln\lambda(\mu) = \frac{(\mu-\widehat{\mu})^2}{\sigma^2} + \mathcal{O}(1/\sqrt{N}). \tag{5.9}$$

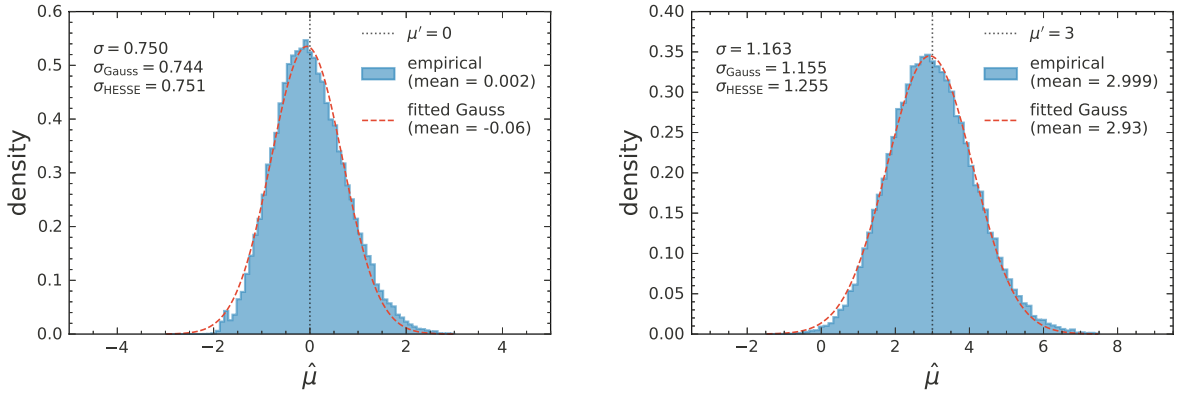
This equation holds if $\widehat{\mu}$ is normal distributed with mean μ' and standard deviation σ. The additional term can be neglected for large sample size N, leading to the results discussed by Cowan et al. [24]. The distribution of $\widehat{\mu}$ is shown in Figure 5.33 explicitly for a Z' mass hypothesis of 2 GeV, using 50 000 samples. In Figure 5.33a it is shown for $\mu' = 0$ and in Figure 5.33b for $\mu' = 3$. A normal distribution was fitted to the sample and overlayed in the Figure. While the estimator in this specific case is unbiased (the empirical mean is equal to the true mean), it is evident that the distribution is not perfectly Gaussian. In particular, there is an asymmetry present in the empirical distribution.

Also shown in Figure 5.33 is the standard deviation σ obtained empirically from the distribution together with the value for the shown Gaussian and the value given by the HESSE method when performing the fit to the Asimov data. In both cases, the HESSE method gives a good estimate of σ.

To check Equation 5.9 explicitly for $\tilde{t}_\mu$ one can plot

$$\tilde{t}_\mu = \begin{cases} \frac{\mu^2}{\sigma^2} - \frac{2\mu\widehat{\mu}}{\sigma^2} & \widehat{\mu} < 0 \\ \frac{(\mu-\widehat{\mu})^2}{\sigma^2} & \widehat{\mu} \geq 0 \end{cases} \tag{5.10}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



(a) (b)

Figure 5.33: The distribution of the estimated value $\widehat{\mu}$ around the true value μ' obtained with 50 000 toys. The hypothesised Z' mass is 2 GeV, and the chosen value for μ' is indicated by the vertical dotted line.

together with the empirical values of $\tilde{t}_\mu$ for different values of $\widehat{\mu}$. This is done in Figure 5.34 for four different cases. In Figure 5.34a, 5.34c and 5.34d the true value is 0 and the assumed value is 0, 0.5 and 3 respectively. The red dotted line shows Equation 5.10 with σ obtained from the HESSE method. In 5.34c and 5.34d the same Equation is also shown with

$$\sigma^2_{\text{Asimov}} = \frac{(\mu - \mu')^2}{\tilde{t}_{\mu,\text{A}}}, \tag{5.11}$$

where $\tilde{t}_{\mu,\text{A}}$ is the value of the test statistic evaluated with the Asimov data [see 24, Section 3.2]. Using σ_{Asimov} gives a better agreement with the empirical values. One explanation for this could be that when using the Asimov data, higher order terms in Equation 5.9 are taken into account. For all future calculations, this method will be used. Figure 5.34b shows $\tilde{t}_\mu$ for $\mu' = \mu = 3$, which corresponds to $\widehat{\mu}$ shown in Figure 5.33b. Again the asymmetry is evident but the overall agreement in this case is reasonable.

Checking the validity of the assumptions made to derive the asymptotic formulae is useful. Still, in the end, the only thing that matters is how well the asymptotic PDF of $\tilde{t}_\mu$ agrees with the empirical one. Figure 5.35 shows the signal like distribution for three different values of m and μ. The distributions agree quite well. Naturally, the empirical distribution shows statistical fluctuations for extreme values of $\tilde{t}_\mu$. In Figure 5.36 some of the background like distributions are shown. Here the asymptotic formulae fail to describe the distribution for increasing values of μ. The sample size is 5000 for all the cases, except for the background like distribution of $m = 3100\,\text{MeV}$, where 50 000 samples were produced.

To quantify the agreement of the empirical and asymptotic distributions it is instructive to calculate the p-value $= 1 - \text{CDF}(\tilde{t}^{\text{obs}}_\mu)$ for different values of $\tilde{t}^{\text{obs}}_\mu$. The p-value as a function of $\tilde{t}^{\text{obs}}_\mu$ for the signal like distribution is shown in Figure 5.37 for some combinations of m and μ. The asymptotic formulae seem to describe the empirical to a good extent. Since a sample size

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

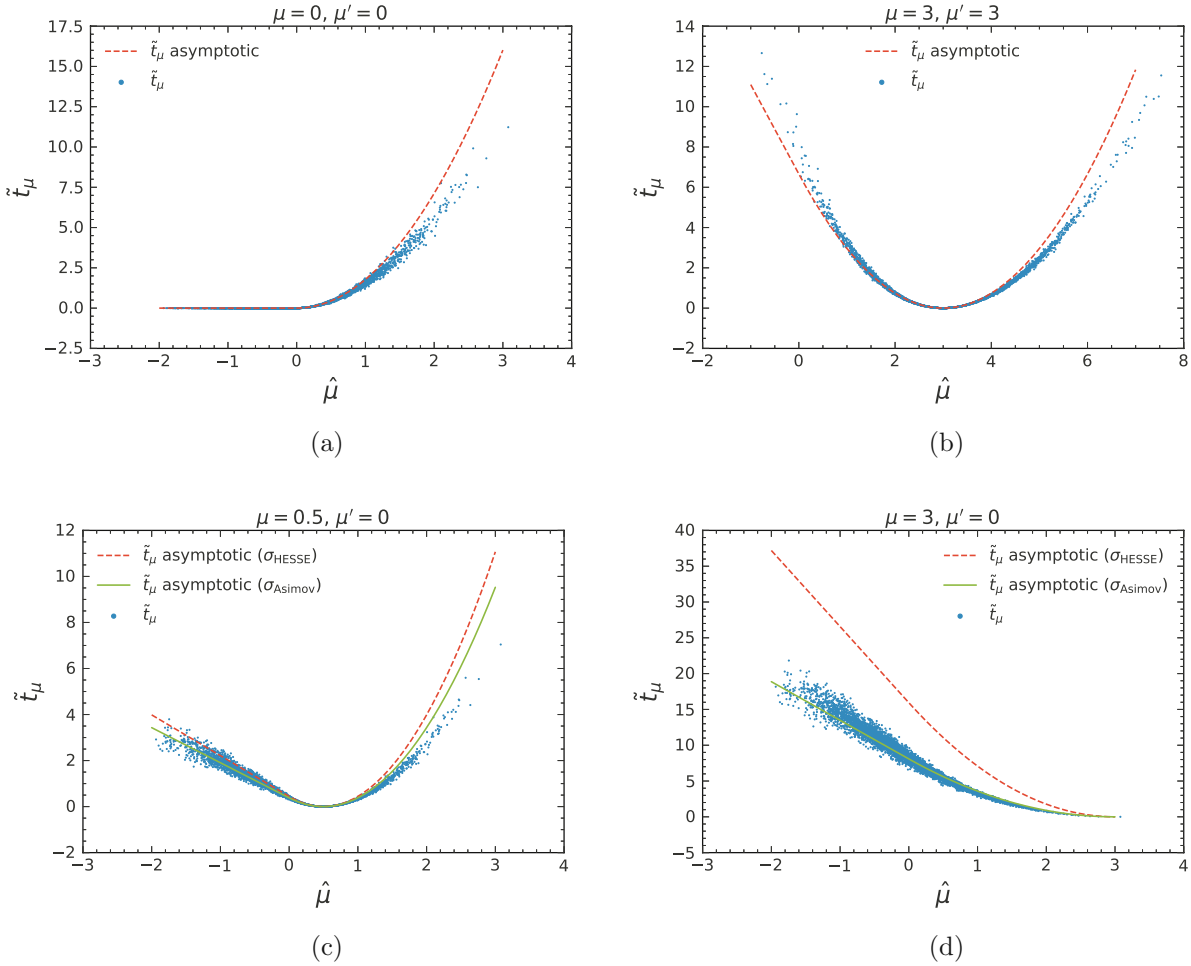


Figure 5.34: The test statistic $\tilde{t}_\mu$ depending on the estimated value $\widehat{\mu}$ for different combinations of the assumed value μ and the true value μ'. The asymptotic approximation based on Wald's formula is overlayed over the empirical values of $\tilde{t}_\mu$ using toy samples. σ is obtained from the HESSE method of MINUIT. For $\mu \neq \mu'$ also the approximation where σ is calculated from the Asimov data is shown.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

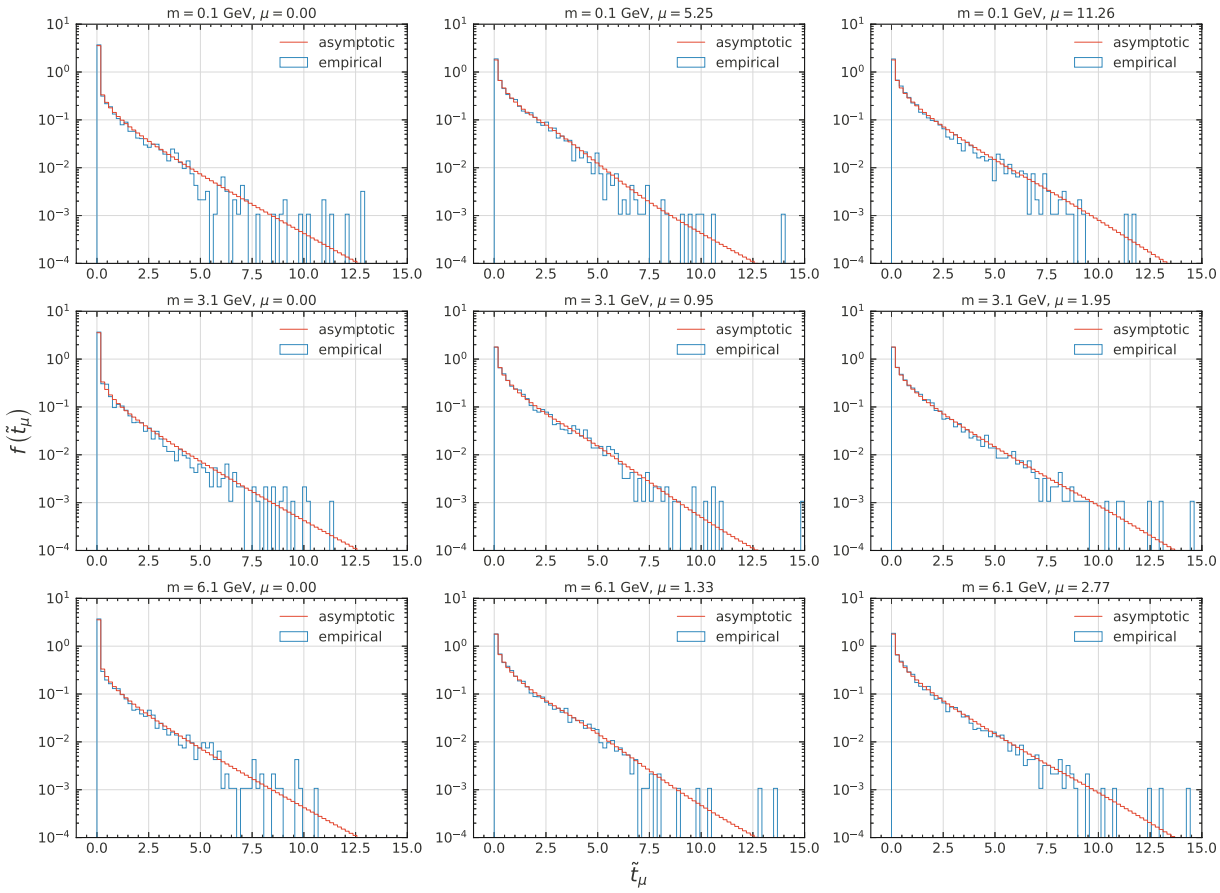


Figure 5.35: The signal like distribution for three different values of m and μ, comparing toy samples and the asymptotic approximations, showing a good agreement between the distributions.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

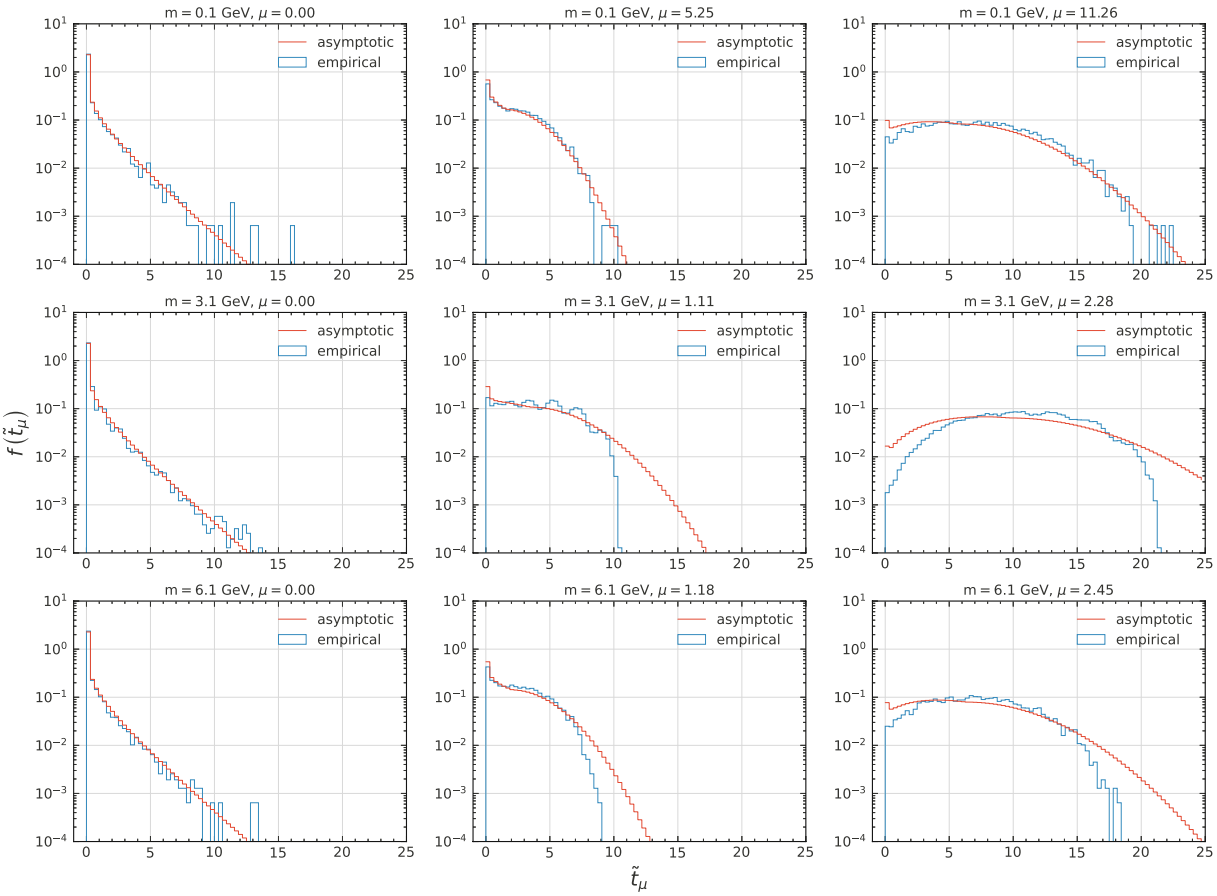


Figure 5.36: The background like distributions for three different values of m and μ, comparing toy samples and the asymptotic approximations. Here the asymptotic formulae fail to describe the distribution for increasing values of μ.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

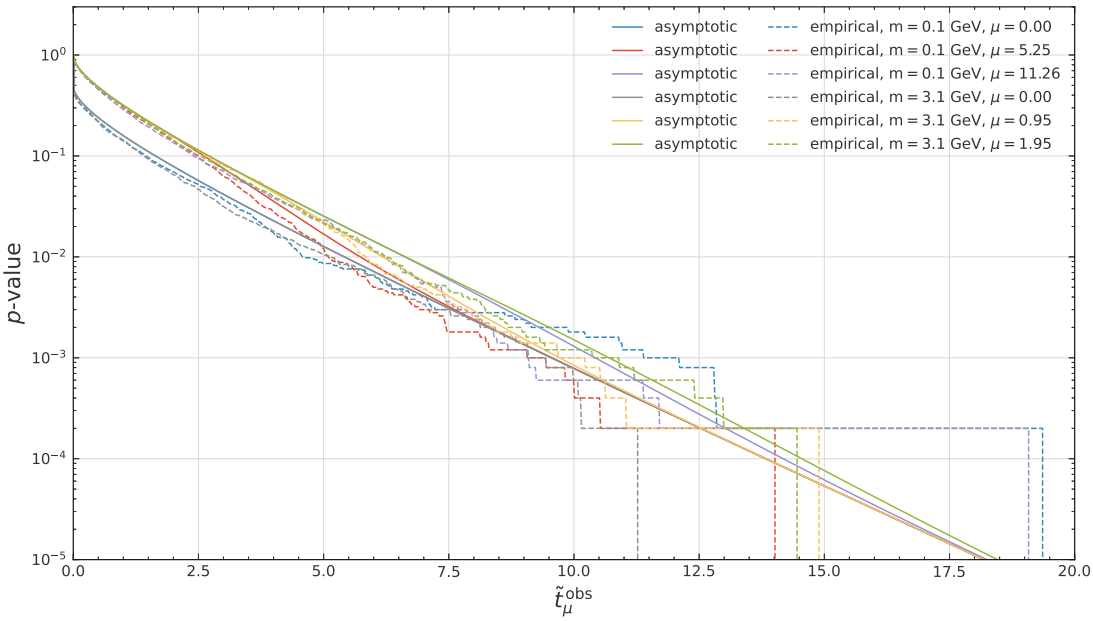


Figure 5.37: The p-value as a function of $\tilde{t}_\mu^{\text{obs}}$ for the signal like distribution for some combinations of m and μ. The asymptotic and the empirical distributions agree reasonably well.

of 5000 was used, the statistical error of the empirical distribution is small for p-values greater than $\sim 1/100$. In Figure 5.38 the same is shown for the background like distribution. A significant deviation of the asymptotic p-values from the empirical ones is observed for most $\tilde{t}_\mu^{\text{obs}}$ (except for the special case $\mu = 0$ where the background like and signal like distributions are the same).

To visualise the disagreement between the distributions for more than just some selected values of m and μ, the value of the p-value obtained with the asymptotic formula was calculated at the 90% quantile of the empirical distribution (corresponding to $\tilde{t}_\mu^{\text{obs}}$ with yields a p-value of 0.1). This value is shown in Figure 5.39 for the signal like (5.39a) and the background like (5.39b) case. The y-axis shows different values of m and the x-axis shows the index of the value of μ that was tested.

The index 0 always corresponds to the special case $\mu = 0$ and the index 1 to $\mu = 0.1$. For indices greater than one, values of μ are linearly increasing. The exact values are different for each mass point and depend on the expected sensitivity. Some of the values are given in Table A.6 for the signal like case and in Table A.7 for the background like case. The colour of each bin shows the p-value obtained with the asymptotic formula, where red corresponds to a lower and blue to a larger p-value with respect to the empirical one. White signifies a value around 0.1 and therefore a good agreement at the considered value of $\tilde{t}_\mu^{\text{obs}}$. For the signal like distribution, only three mass points were checked, and overall good agreement was found. There is a slight bias towards larger p-values when using the asymptotic distributions, which is consistent with what is shown in Figure 5.37. This will result in slightly more conservative intervals and significances, but it is expected to be only a small effect. For that reason, the asymptotic formulae will be used for the signal like case to save computation time. The background like distribution shows large

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

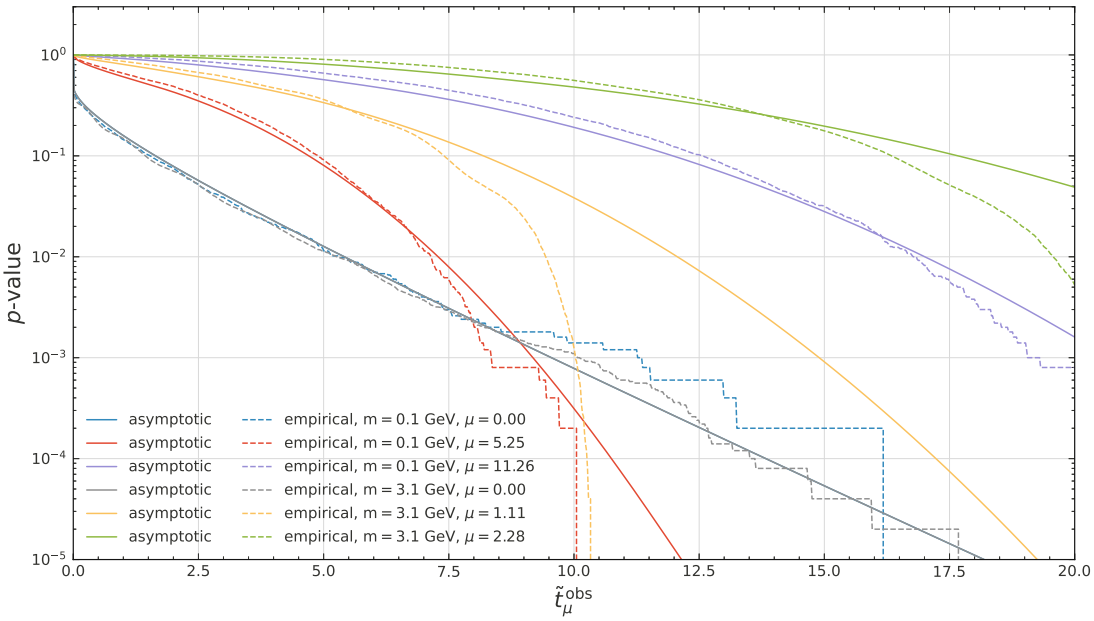
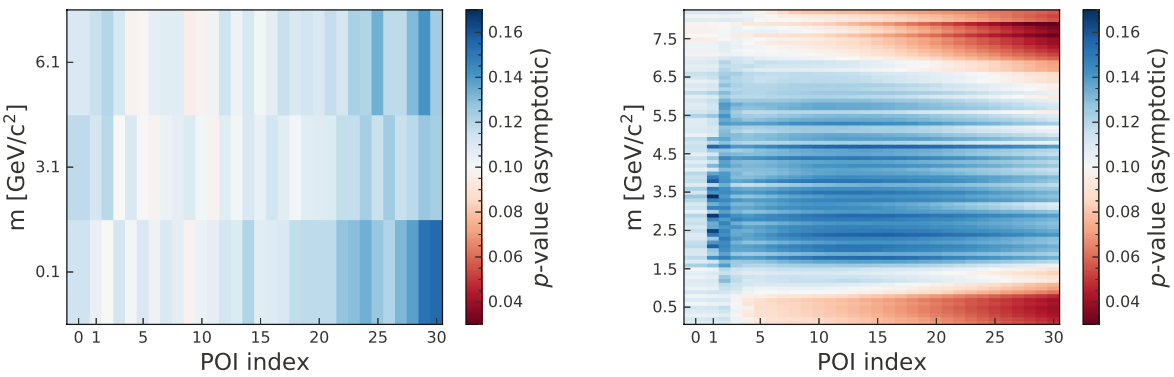


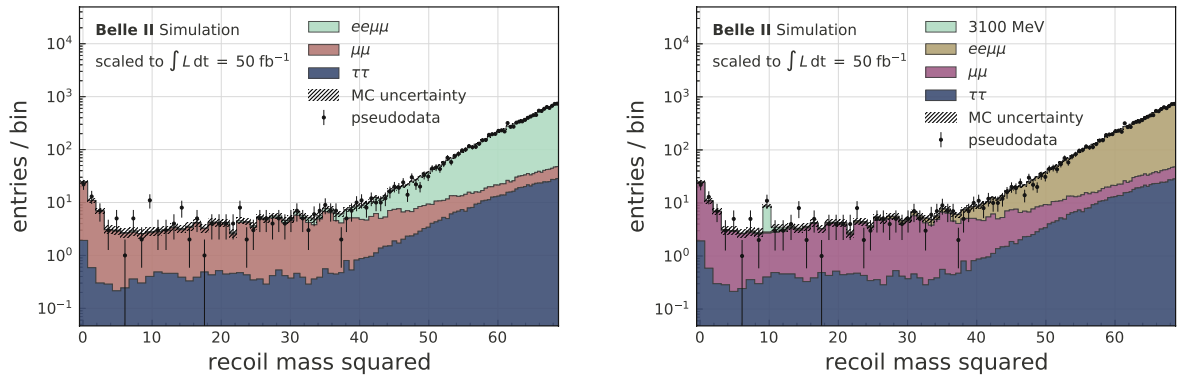
Figure 5.38: The p-value as a function of $\tilde{t}_\mu^{\text{obs}}$ for the background like distribution for some combinations of m and μ. The asymptotic and empirical distributions divert for larger $\tilde{t}_\mu^{\text{obs}}$ except for $\mu = 0$.



(a) (b)

Figure 5.39: The value of the p-value obtained with the asymptotic formula calculated at the 90% quantile of the empirical distribution for different values of m and μ. For the signal like distribution (a) the agreement is reasonably well. For the background like distribution (b) significant deviations are observed for $\mu > 0$. The mapping of the POI index to μ is given in Table A.6 and Table A.7.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



(a) (b)

Figure 5.40: The simulated dataset together with the nominal MC expectation for a total integrated luminosity of 50 fb^{-1}. Additionally to the background contributions a 3100 MeV Z' was injected with a cross section of 1.5 fb. In (b) also the signal MC which was used to sample the dataset is overlayed.

disagreement in some bins, hence the asymptotic formulae will not be used in this case. Instead, an empirical distribution was calculated for each considered value of m and μ.

5.5.5 Procedure on simulated data

A simulated pseudo dataset was sampled from the PDF of the statistical model to test the procedure. Additionally to the background, a 3100 MeV Z' was injected with a cross section of 1.5 fb. The resulting number of 'observed' and expected events are shown in Figure 5.40. Figure 5.40a shows only the background events from MC, in Figure 5.40b also the signal MC is shown. Note that the width of the bins shown here is two times larger than what is used in the construction of the model (for better visualisation).

The first check one can perform with this simulated dataset is to perform the ML fit and compare the estimates to the true values used to generate the sample. This is shown in Figure 5.41 for each of the nuisance parameters with a Z' hypothesis of 3100 MeV. The Figure shows on the y-axis the number of standard deviations the estimated value deviates from the value of the auxiliary measurement (the expected value from MC), where the standard deviation is given by the Gaussian constraint term used in the likelihood (see Equation 5.4). This is also called a *pull plot*[5].

The marker indicates the point estimate after the fit, and the error bars give the uncertainty obtained from the HESSE method of `MINUIT`. Since, in this case, the expected value is equal to the true value, the estimated value should be consistent with the expected value. In Figure 5.41 it can be seen that the data do not constrain the nuisance parameters for the luminosity and the signal normalisation. This is expected since it is always 'cheaper' to vary the unconstrained

[5]The estimated value 'pulls' the mean of the Gaussian constraint away from the value given by the auxiliary measurement.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

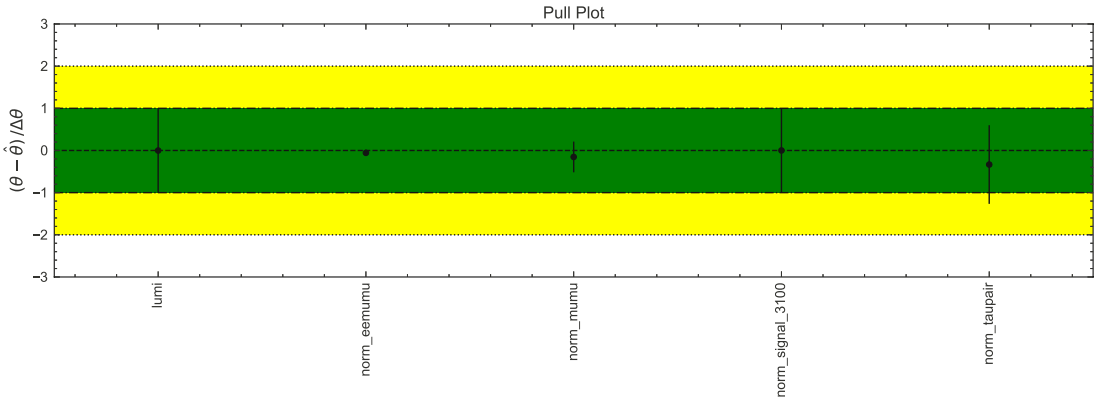


Figure 5.41: A pull plot showing on the y-axis the number of standard deviations the estimated nuisance parameters deviate from the value of the auxiliary measurement, which in this case also corresponds to the true value.

estimate of the signal cross section μ instead of the signal normalisation. However, the signal normalisation still influences the results since the value of μ is fixed in the profile likelihood. The luminosity modifier acts as an overall normalisation, and since the luminosity has a very tight constraint, it is cheaper to vary all background normalisations instead. The normalisation of the $ee\mu\mu$ background has a very precise estimate since there are a lot of bins with a large number of $ee\mu\mu$ events (see Figure 5.40). On the contrary, the taupair normalisation can not be estimated very precisely because the low expected yield is dominated by other backgrounds.

Introducing discrepancies

Previously the true value of the nuisance parameters was set to the expected value. This is very rarely the case since our simulation is not perfect. If the model is flexible enough, it should fit to the data and compensate for any discrepancies between data and MC. This is shown for a very simple case where a new simulated dataset was sampled with a variation on the normalisation of the three background contributions (the signal normalisation was not altered since this would simply correspond to a change in cross section). The variation and the corresponding number of standard deviations of the Gaussian constraint term are shown in Table 5.12.

Table 5.12: The variation on the normalisation artificially introduced on a new simulated dataset and the equivalent standard deviations of the Gaussian constraint term.

$ee\mu\mu$	-40%	-2σ
$\mu\mu$	$+30\%$	$+1.5\sigma$
taupair	$+20\%$	$+1\sigma$

The simulated dataset together with the MC expectation is shown in Figure 5.42. The discrepancy is clearly visible. In Figure 5.43 the estimated values of $\boldsymbol{\nu}$ after performing the fit are

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

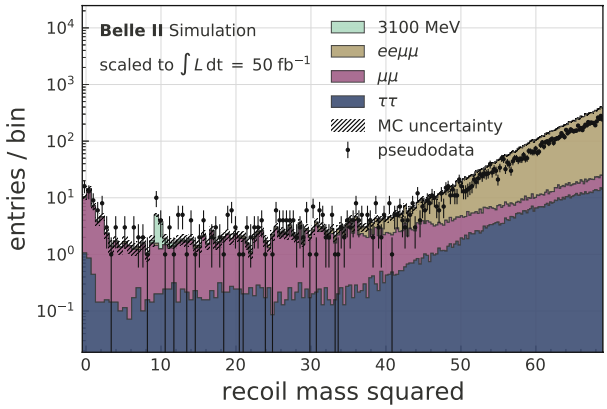


Figure 5.42: The simulated dataset with artificially introduced discrepancies together with the nominal MC expectation for a total integrated luminosity of 50 fb^{-1}. Additionally to the background contributions a 3100 MeV Z' was injected with a cross section of 1.5 fb.

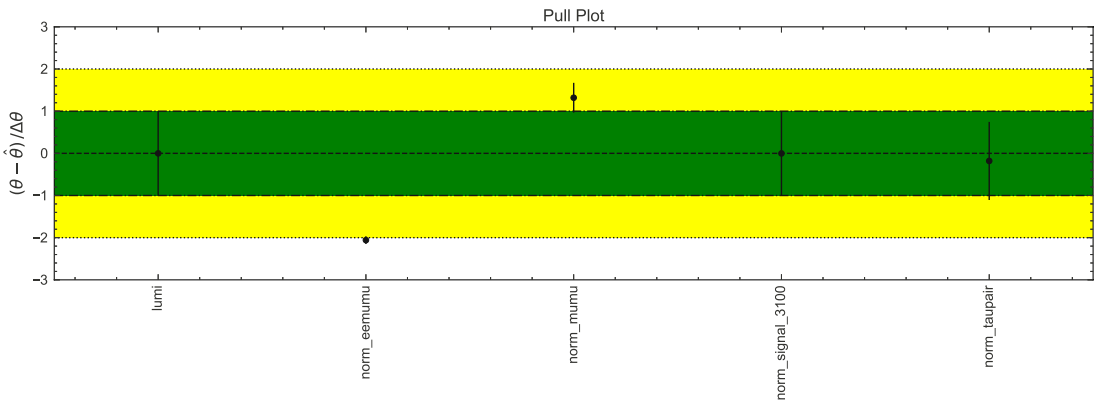


Figure 5.43: The estimated values of $\boldsymbol{\nu}$ after performing the fit for the simulated dataset with discrepancies. The true value of $\boldsymbol{\nu}$ is correctly estimated within the errors.

shown. As mentioned before, also in this case the taupair normalisation can not be estimated very precisely. Additionally, if the taupair sample in data is much smaller or larger than expected, the estimator will not pick it up. Because of the constraint, it is cheaper to include the discrepancy in the $ee\mu\mu$ and $\mu\mu$ normalisation. This is not an issue as long as the overall yield is estimated correctly. In this case, however, the estimated values are still consistent with the true values.

p-values and intervals

Using the test statistic $\tilde{t}_\mu$ and the distributions shown in Section 5.5.4 a local p-value can be calculated for the 'observed' value of $\tilde{t}_\mu$, in this case obtained from the simulated pseudo dataset in Figure 5.40. For the generation of the toy samples, the true value of the nuisance parameters is fixed to the best estimate $\widehat{\boldsymbol{\nu}}$ for the simulated data shown in Figure 5.41 (the plugin method).

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

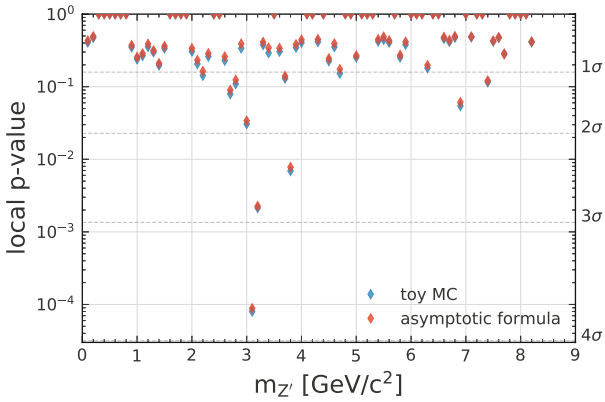


Figure 5.44: The local p-values for different mass hypothesis for the simulated pseudo dataset corresponding to a total integrated luminosity of 50 fb^{-1}. p-values are obtained from toy samples as well as from the asymptotic formulae.

When testing the background-only hypothesis ($\mu = 0$) this results in the p-values shown in Figure 5.44. The distributions shown in Section 5.5.4 showed good agreement for $\mu = 0$ for the asymptotic formula and the toy MC method. This can again be seen in Figure 5.44, where both methods are compared. As already mentioned before, the asymptotic formula gives slightly larger (more conservative) p-values. For local p-values below 10^{-2}, to decrease the statistical error, 50 000 toys were used instead of 5000.

At 3100 MeV, where the signal was injected, a significant deviation from the background-only hypothesis of nearly 4σ is observed. This gives the probability to find data as extreme as the one observed if the mass would be known before performing the measurement. Since the search is performed over a range of possible signal masses, the LEE (see Section 2.4.6) has to be accounted for to give the global significance of this excess. The trial factor will be calculated in Section 5.5.6.

The $\tilde{t}_\mu$ test-statistic can also be used to construct intervals as described in 2.4.5. The interval construction is shown for $m = 3100$ MeV in Figure 5.45. It shows the p-value for different values of μ being probed. The two curves correspond to what is obtained with the simulated data (red) and what would be expected if there is no signal in the data (blue). The values of μ where the curves cross 0.1 define the borders of the 90% CL intervals. The curve has a maximum (a p-value of 1) at the point estimate $\widehat{\mu}$ given by the fit (the fit result was 2.11 ± 0.85). For the simulated pseudo-data these intervals, calculated for all mass points, are shown in Figure 5.46 as a shaded blue region. The colour bar indicates the number of signal events that would be expected for a given cross section μ. Since the test statistic $\tilde{t}_\mu$ is used, the obtained intervals automatically change from upper limits to two-sided intervals. As expected, this happens around 3100 MeV, whereas in this case the obtained interval also contains the true value of $\mu' = 1.5$.

In Figure 5.47 the p-values when testing against the signal plus background hypothesis are shown for all considered values of m and μ. The overlayed contours denote a p-value of 0.1, defining the borders of the 90% CL intervals. Figure 5.47a shows the observed values with the simulated data, Figure 5.47b the expected values and Figure 5.47c and 5.47d the expected values in case of a -2σ or $+2\sigma$ fluctuation.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

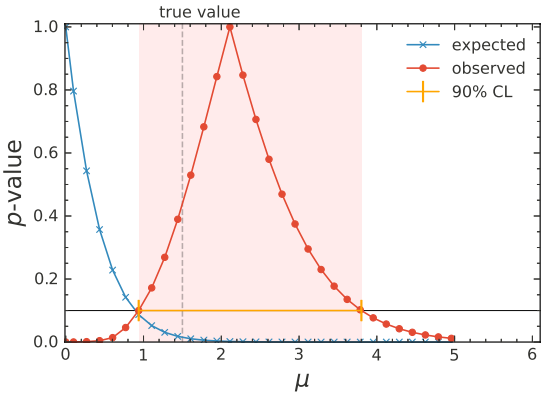


Figure 5.45: The interval construction for $m = 3100\,\text{MeV}$. The shaded area corresponds to the observed two-sided confidence interval. The expected upper limit is given by the intersection of the blue line with a p-value of 0.1.

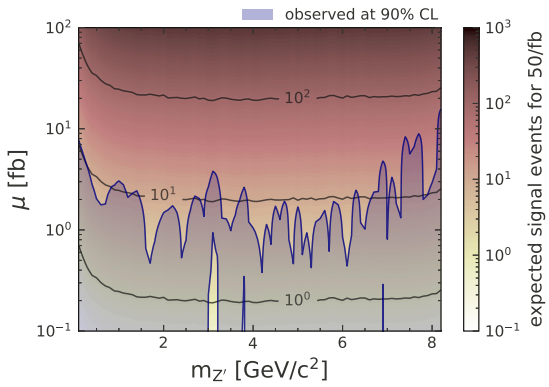


Figure 5.46: The confidence intervals for all mass points calculated with the simulated pseudo-data (shaded blue region). The colour bar indicates the number of signal events that would be expected for a given cross section μ.

The expected p-values are calculated with a value $\tilde{t}_\mu^{\text{exp}}$ obtained from the background like distribution. This corresponds to the value of $\tilde{t}_\mu$ given by the median value of the $\widehat{\mu}$ sample. The $\pm 1\sigma$ and $\pm 2\sigma$ variations thereof correspond to the value of $\tilde{t}_\mu$ given by the quantiles of the $\widehat{\mu}$ sample which match ± 1 or ± 2 standard deviations of a standard normal distribution. These values are listed in Table 5.13. If the test statistic and $\widehat{\mu}$ are related monotonically one can simply

Table 5.13: The quantiles corresponding to $\pm 1\sigma$ and $\pm 2\sigma$ deviations from the mean of a standard normal distribution.

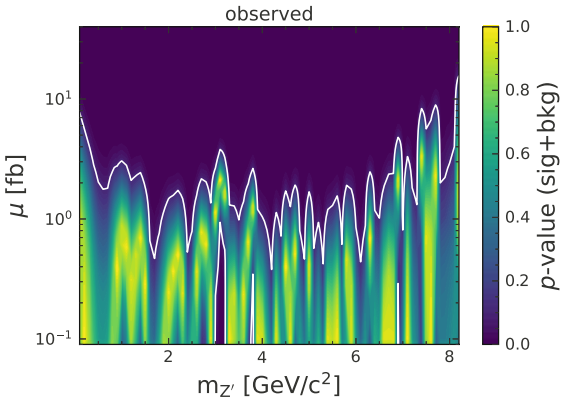
	-2σ	-1σ	expected	$+1\sigma$	$+2\sigma$
quantile	0.02	0.16	0.5	0.84	0.98

calculate the quantile of the $\tilde{t}_\mu$ sample instead of $\widehat{\mu}$. Unfortunately for $\tilde{t}_\mu$ this is only the case for $\mu = 0$ or $\mu \gg \mu'$. In general the relation is resemblant to the one shown in Figure 5.34c. However by noting that monotonicity holds separately for $\widehat{\mu} > \mu$ and $\widehat{\mu} < \mu$ the quantile can be obtained similarly. It is possible to define a new quantity,

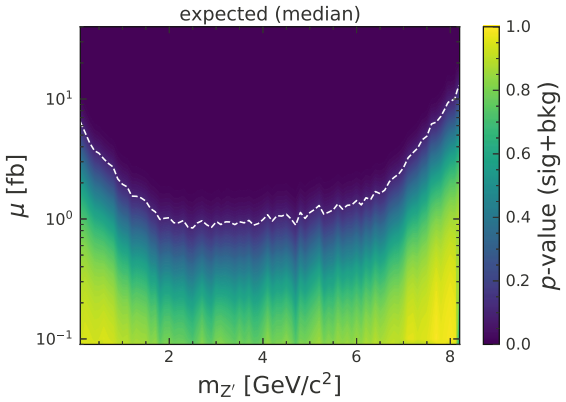
$$\tilde{t}_\mu^{\pm} = \begin{cases} \tilde{t}_\mu & \widehat{\mu} > \mu \\ -\tilde{t}_\mu & \widehat{\mu} < \mu \end{cases}, \tag{5.12}$$

which is again monotonically related to $\widehat{\mu}$. Now the quantile can be calculated directly on the $\tilde{t}_\mu^{\pm}$ sample. The absolute value of this quantile will correspond to the desired $\tilde{t}_\mu$.

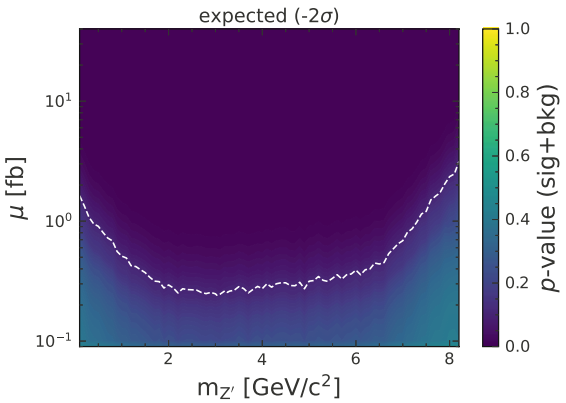
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



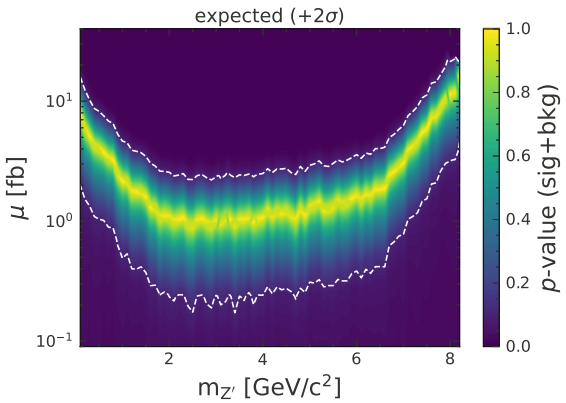
(a) The observed values using the simulated dataset.



(b) The expected values (median).



(c) The expected values in case of a -2σ fluctuation.



(d) The expected values in case of a $+2\sigma$ fluctuation.

Figure 5.47: The p-values when testing against the signal plus background hypothesis for all considered values of m and μ. The overlayed contours denote a p-value of 0.1, defining the borders of the 90% CL intervals.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

5.5.6 Calculating the global significance

As mentioned before, when many signal hypotheses are possible at different values of an additional parameter (in this case m), the LEE (see Section 2.4.6) has to be taken into account.

First the parameter $\mathcal{N}$ has to be estimated using Equation 2.24. To do this, the mean number of upcrossings $\langle N(c_0)\rangle$ of $\tilde{t}_\mu^{\text{obs}}$ above some reference value c_0 has to be calculated. The only reliable way to achieve this is with simulation. For the present case, a reasonable number was 5000 toy samples which allows for a reference value as high as ~ 7.

The procedure mentioned in Section 2.4.6 requires the distribution of the test statistic to follow a χ^2 distribution (with 0 degrees of freedom). However, when testing the background-only hypothesis, the test statistic $\tilde{t}_{\mu=0}$ follows asymptotically

$$f\left(\tilde{t}_0 \mid \mu' = 0\right) = \frac{1}{2}\delta\left(\tilde{t}_0\right) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\tilde{t}_0}}e^{-\tilde{t}_0/2}, \tag{5.13}$$

which is a mixture of a delta distribution at $\tilde{t}_\mu = 0$ and a χ^2 distribution, each with a weight of ½. This is because for the background like case half of the estimates $\widehat{\mu}$ are expected to fall below 0, resulting in $\tilde{t}_0 = 0$ (see Equation 2.20). Therefore in the asymptotic limit, one would expect half as many upcrossings. Similarly, the distribution of $\tilde{t}_0(\widehat{m})$ for large values is expected to be half of the distribution of $t_0(\widehat{m})$ if t_0 follows a χ^2 distribution. This is because half of the very extreme values for $\tilde{t}_0$ will end up at 0, resulting in a shift to lower values of $\tilde{t}_0(\widehat{m})$. If the asymptotic formula holds, the same is true for the distribution of $\tilde{t}_0$ and χ^2.

This results in effectively introducing a factor ½ for every term in Equation 2.23, which can then be written as

$$P\big(\tilde{t}_0(\widehat{m}) > c\big) \approx P\big(\tilde{t}_0 > c\big) + \langle N(c)\rangle. \tag{5.14}$$

In Figure 5.48 the (empirical) distribution of $\tilde{t}_0(\widehat{m})$ is shown together with the distribution of $t_0(\widehat{m})$ divided by two. The right tail of the distributions seems to agree well, confirming the expectations from the argument above.

In Figure 5.49 the values of the test statistic $\tilde{t}_0$ are shown together with t_0 for the same sample. The number of upcrossings above the threshold $c_0 = 2$, in this case, is 5 and 10 respectively. Figure 5.50 shows the mean number of upcrossings together with the value of $\mathcal{N}$, calculated for different values of c_0 using

$$\mathcal{N} = \langle N(c_0)\rangle e^{c_0/2}. \tag{5.15}$$

The mean number of upcrossings is predicted to follow $\langle N(c)\rangle = \mathcal{N}e^{-c/2}$, which would result in a flat distribution of $\mathcal{N}$ in Figure 5.50. While the exponential form of $\langle N(c)\rangle$ is evident, only around $c_0 = 8$ the curve follows the predicted form. Above that, the statistical error becomes too large to estimate $\mathcal{N}$ reliably. One reason why in this case $\mathcal{N}$ is not constant could be that Equation 5.13 does not describe the actual distribution of $\tilde{t}_0$ perfectly. To obtain an approximate estimate for the trial factor, it is sufficient to evaluate $\mathcal{N}$ for a high reference value. If $\mathcal{N}$ stays constant after the reference value, the global p-value can be estimated correctly. If $\mathcal{N}$ decreases for higher values of c this will result at least in an upper bound for the global p-value. From Figure 5.50 it is not evident what happens after $c \approx 10$, but the fact that for $7 < c < 10$ the

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

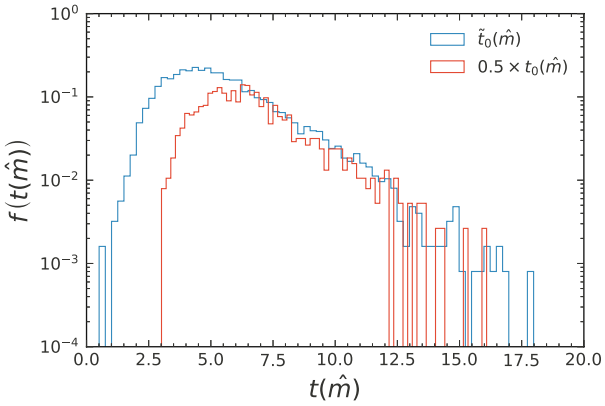


Figure 5.48: The empirical distribution of $\tilde{t}_0(\widehat{m})$ overlayed with the distribution of $t_0(\widehat{m})$ divided by two.

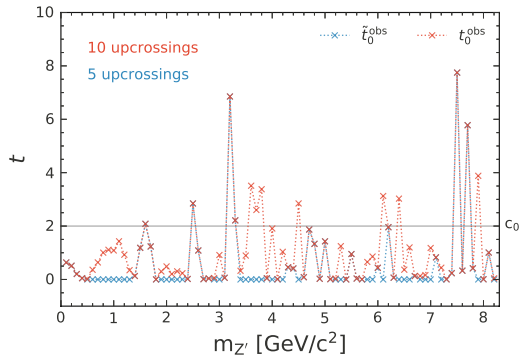


Figure 5.49: The values of the test statistic $\tilde{t}_0$ overlayed with t_0 for different mass hypothesis, calculated for the same simulated sample.

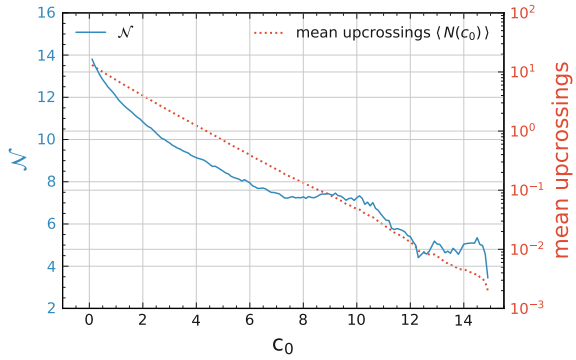


Figure 5.50: The mean number of upcrossings and the value of $\mathcal{N}$ as a function of c_0. Only around $c_0 = 8$ is $\mathcal{N}$ independend of c_0.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

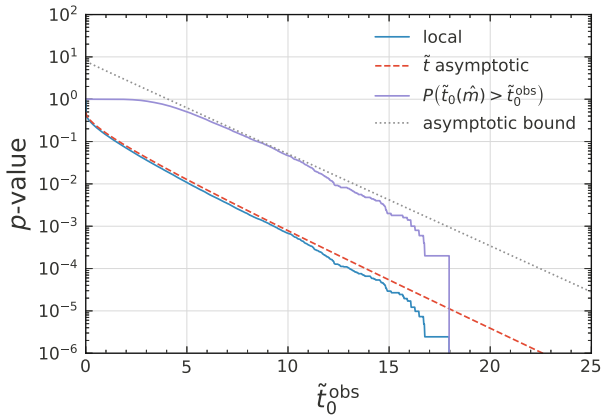


Figure 5.51: A comparison of the p-values obtained empirically and using the asymptotic approximations as a function of $\tilde{t}_\mu^{\text{obs}}$.

Table 5.14: Values for $\langle N(c_0)\rangle$ and $\mathcal{N}$ compared for the test statistic $\tilde{t}_0$ and t_0, evaluated at $c_0 = 7$.

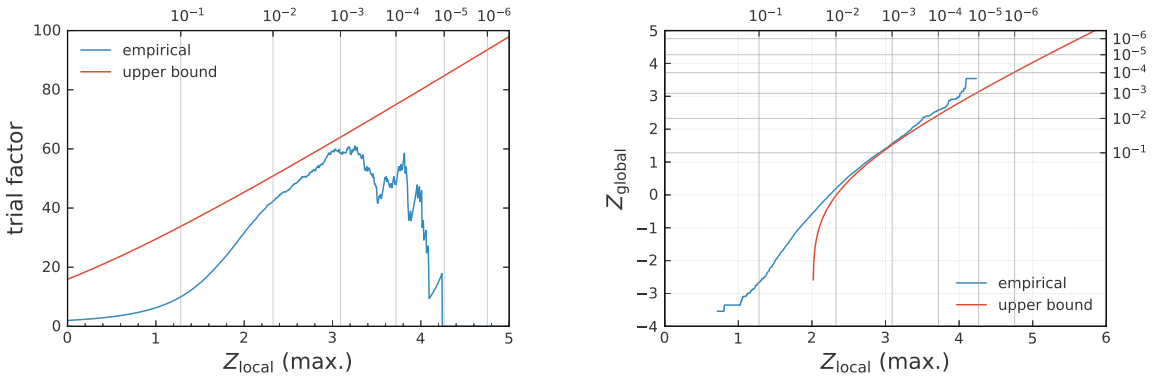
	$\tilde{t}_0$	t_0
$\langle N(c_0)\rangle$	0.225	0.432
$\mathcal{N}$	7.458	14.306

value of $\mathcal{N}$ is approximately constant could mean that for high values of the test statistic the asymptotic approximation holds. For the present case a reference value of $c_0 = 7$ was chosen to evaluate $\mathcal{N}$, resulting in a value of 7.458. Values for both $\tilde{t}_0$ and t_0 at $c_0 = 7$ are compared in Table 5.14, showing good agreement with the expectation of twice as large values for t_0.

Figure 5.51 shows a comparison of the p-values obtained empirically and using the asymptotic approximations for different values of $\tilde{t}_\mu^{\text{obs}}$. The blue line gives the local p-value from the distribution of $\tilde{t}_0$, and the red dashed line the prediction from Equation 5.13. The purple line gives the global p-value from the distribution of $\tilde{t}_0(\widehat{m})$ and the grey dotted line the prediction from Equation 5.14. It seems that the asymptotic approximations give a good continuation of the p-values for high values of $\tilde{t}_\mu^{\text{obs}}$ where the statistics of the toy MC sample is not sufficient any more. If a significant excess in data is found, the validity of this approach can always be checked with more simulated samples.

The trial factor (ratio of global to local p-value) can be calculated with these results. This is shown in Figure 5.52a. In Figure 5.52b a direct conversion from local to global significance is given. $Z_{\text{local}}(\text{max.})$ gives the largest local significance and is equal to $\sqrt{\tilde{t}_0(\widehat{m})}$. Up to $Z_{\text{local}} \sim 3$ the empirically obtained curve gives the correct values. Above this value, the statistical error becomes large, but the upper bound is expected to asymptotically give the correct values for high Z. This makes it possible to translate the local p-value of an observed excess to a global one. For example the excess 'observed' in the simulated pseudodata sample (see Figure 5.44) with a p-value of 10^{-4} would only correspond to $Z_{\text{global}} \sim 2.5$. For a discovery ($Z_{\text{global}} > 5$) at least a

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



(a) The ratio of global to local p-value.

(b) A direct translation from local to global significance.

Figure 5.52: The relation between local and global significance. The empirical relation is due to limited sample size only usable up to $Z_{\text{local}} \sim 3$, while the asymptotic bound is valid for high significances.

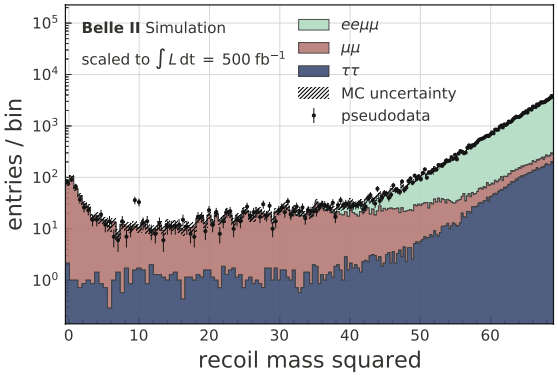
value of $Z_{\text{local}} = 6$ would be required. Since the trial factor depends on how many mass points were checked, this study needs to be repeated when more signal samples are used for additional mass points.

5.5.7 The procedure for increased luminosity

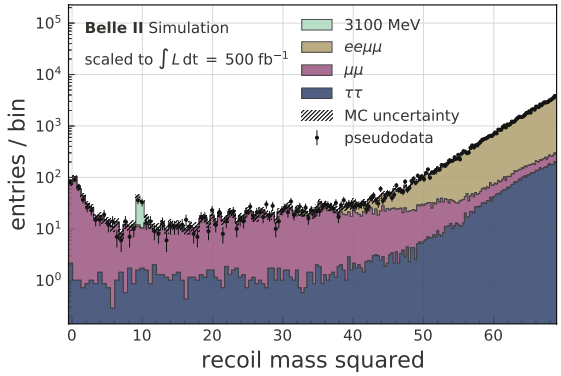
The statistical procedure described in this chapter was also repeated for the selections which were optimised for a data sample corresponding to 500 fb^{-1}. Everything mentioned so far still applies, and the procedure can be repeated on the new sample. Of course the expected number of background events per bin increases, together with the number of signal events for a fixed cross section. This will also improve the validity of the asymptotic formulae. The results differ in the sense that in case of no discovery a tighter limit on the cross section μ can be set. A comparison of the expected limits given by two considered luminosities can be found in the next chapter. With more and more data, the limits will decrease until systematic uncertainties dominate the measurement or eventually a Z' is found. It is instructive to see how the hypothetical 3100 MeV Z' injected with a cross section of 1.5 fb would appear in the larger dataset of 500 fb^{-1}. For this purpose, a new pseudo dataset was produced analogously to Section 5.5.5. The squared recoil mass distribution can be seen in Figure 5.53. Now the peak at 3100 MeV is clearly visible (compared to Figure 5.40), and even by eye we can say that this is not merely a statistical fluctuation.

When performing the calculation for this new pseudo-data, it is found that the local p-value at 3100 MeV is less than 1×10^{-14}, reaching almost 8σ significance. This can be seen in Figure 5.54 together with the local p-values for the other Z' mass hypotheses. Even considering the LEE this corresponds to a significance well beyond 5σ and a discovery could be claimed.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



(a) The pseudodata overlayed with the background MC.



(b) The pseudodata overlayed with background and signal MC which was used to sample the dataset.

Figure 5.53: The simulated dataset together with the nominal MC expectation for a total integrated luminosity of 500 fb^{-1}. Additionally to the background contributions a 3100 MeV Z' was injected with a cross section of 1.5 fb.

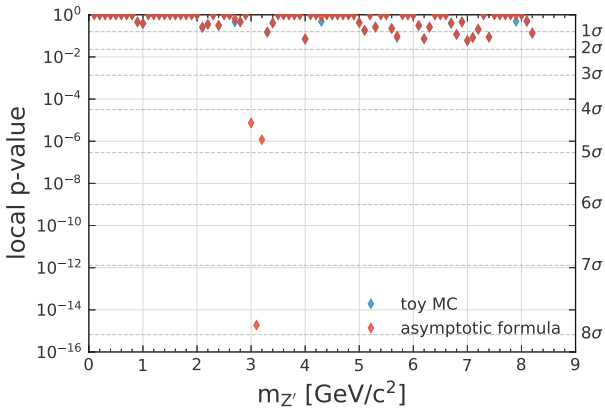


Figure 5.54: The local p-values for different mass hypothesis for the simulated pseudo dataset corresponding to a total integrated luminosity of 500 fb^{-1}. p-values are obtained from toy samples as well as from the asymptotic formulae.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Chapter 6

Discussion of the results

6.1 Expected sensitivity

In this Section I want to show the expected sensitivity for the cross section $e^+e^- \to \mu^+\mu^-$ *invisible* as well as the Z' coupling constant g'. These results are obtained with the methods described in Chapter 5. In Figure 6.1 the expected limits at 90 % CL on the cross section are shown for a

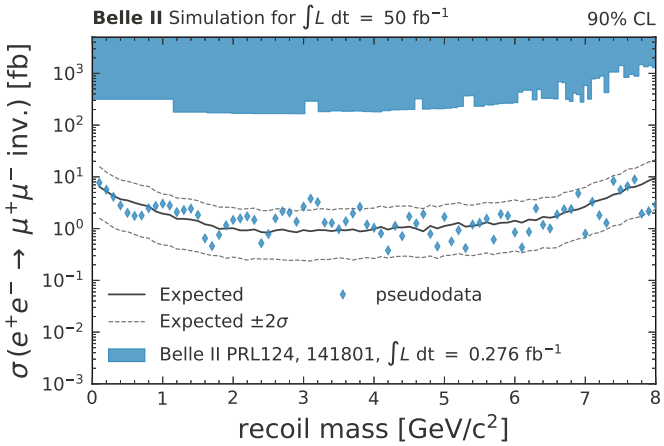


Figure 6.1: The expected limits at 90 % CL on the cross section for a luminosity of 50 fb^{-1}. The blue diamonds mark the results of the statistical procedure applied on simulated pseudo-data, whereas a 3.1 GeV Z' was injected with a cross section of 1.5 fb.

luminosity of 50 fb^{-1}. The dashed lines show the expected $\pm 2\sigma$ variations of the limits. The blue diamonds mark the results of the statistical procedure applied to the simulated pseudo-data from Section 5.5.5. The limits obtained from the simulated dataset are within the $\pm 2\sigma$ variation of the expected upper limit except for an excess around 3.1 GeV. This is the result of the Z' signal injected with a cross section of 1.5 fb. The filled region in the Figure shows the already excluded limits from the phase 2 measurement with 0.275 fb^{-1} [1]. In Figure 6.2 the same information is

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

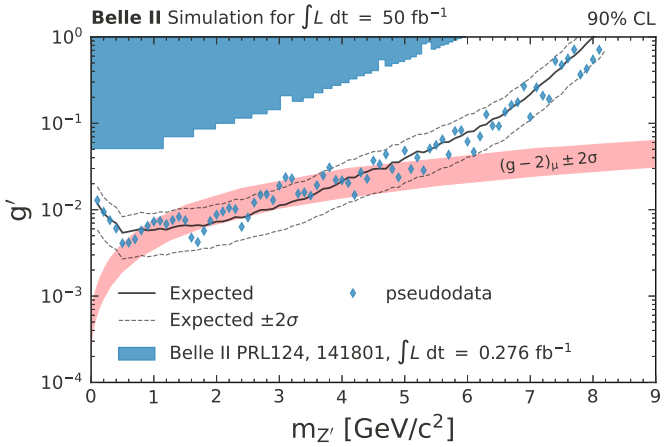


Figure 6.2: The expected limits at 90 % CL on g' for a luminosity of 50 fb^{-1}. The blue diamonds mark the results of the statistical procedure applied on simulated pseudo-data, whereas a 3.1 GeV Z' was injected with a cross section of 1.5 fb.

shown with the cross section translated to the g' coupling constant. The values of g' that could explain the current discrepancies concerning the muon's anomalous magnetic dipole moment are indicated with a band labelled $g-2$.

In Figure 6.3 the expected limits on the cross section are compared for 50 fb^{-1} and 500 fb^{-1}. Figure 6.4 shows again these limits translated to g'. The limits for the increased luminosity were obtained from separately optimised selections. This shows that with a 10-fold increase in luminosity, a significantly larger parameter space can be excluded. Also, while for 50 fb^{-1} the $g-2$ band can barely be covered, for 500 fb^{-1} the band is well covered by the expected limits for a Z' mass between 1 GeV and 4.5 GeV. What would be a hint for an excess at 50 fb^{-1} could already be an observation at 500 fb^{-1}.

6.2 Conclusion and outlook

6.2.1 Particle identification studies

A pion sample originating from 3-prong τ decays was selected to study particle identification performance in Belle II. High purity was achieved with multivariate selections based on a simulated annealing optimisation. The pion identification efficiencies, $\pi \to \mu$ and $\pi \to e$ misidentification probabilities were studied in data and simulation. Corrections for the observed differences in data and MC were provided to analysts in bins of polar angle and momentum. While there are currently some large discrepancies between data and MC, the simulation is continuously improved based on the found disagreements. This is why it is important to continue these PID studies and monitor effects originating from the different subdetectors. Additionally there are multiple efforts to improve particle identification with MVA tools, which needs to be validated.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

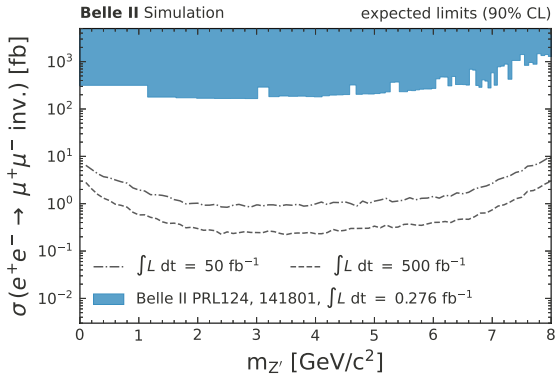


Figure 6.3: The expected limits at 90 % CL on the cross section, compared for a luminosity of 50 fb^{-1} and 500 fb^{-1}.

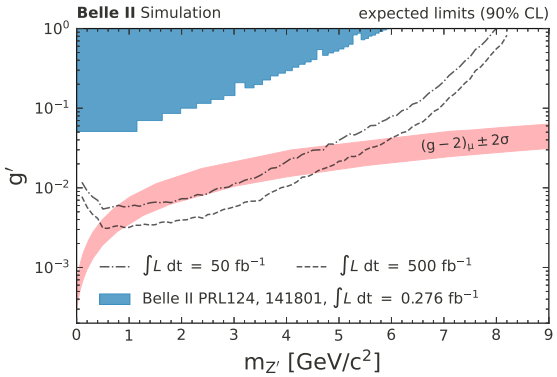


Figure 6.4: The expected limits at 90 % CL on g', compared for a luminosity of 50 fb^{-1} and 500 fb^{-1}.

It was also found that the efficiencies depend on track isolation which is a non-negligible effect that has to be considered. This could explain some discrepancies in particle identification studies performed with different event topologies. The discrepancies between different channels are assigned as a systematic error, but an additional separation in bins of isolation could be considered for the future. Since the three 3-prong tracks are naturally close together, this effect can be observed very well with this sample and studied even further. One next step would be to also look at the angle of the extrapolated tracks at their point of closest approach. Another idea is to take a closer look at the subdetectors to understand better where these effects originate. For example, for the ECL it would be interesting to see what track the clusters get associated to and if there is some charge dependence.

Something else to consider in the future for this study is the usage of more dedicated trigger lines. Especially in the forward and backward regions of the detector, the particle identification is not well understood yet because the trigger efficiency decreases for low multiplicity events. However, with the taupair 3-prong 1-prong topology, it should also be possible to do a more dedicated study in the endcaps, especially with the utilisation of short track or neuro triggers. In general, the considered taupair channel can provide a high statistics pion sample. With 63 fb^{-1} the current selections yielded 7 million pions. In the future, when even more data is collected, it would be possible to obtain an even purer sample on the cost of some efficiency. Additionally, the mis-id rates from other studies could be used to correct the remaining backgrounds that are estimated from MC. This would decrease the systematic error for the calculated efficiencies.
Another option would be to shift from the current 'cut and count' approach to a fitting method. The ρ^0 peak could be fitted in the invariant mass distribution of the pion pairs on the 3-prong side (the ρ^0 decays 100 % into two pions). With an appropriate fit model, the number of pions could be estimated in data (and in MC) based on the fit result. This could give more accurate results and provide a better way of calculating systematic effects, for example, with nuisance parameters.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

6.2.2 Search for an invisibly decaying Z' boson

In Chapter 5 the analysis procedure to search for a new Z' vector boson was laid out. The selection was optimised based on simulated data with a dedicated ANN. A detailed description of the statistical methods for the interpretation of the results was given in Section 5.5. This provides the basis to perform this analysis on actual data.

One of the remaining tasks is to evaluate systematic effects on control channels in data. The effect of the ANN has to be checked since it trained on simulated data. Because of the relatively simple model with only four kinematic input variables and a large amount of training data, the network is not expected to give rise to large data/MC discrepancies. In particular, since it was shown that the classification boundary in 3D forms a smooth sheet, one could argue that the selection defined by the ANN is not much different from a univariate cut performed on any other variable. Moreover, after evaluating the systematic effects, the statistical procedure has to be verified again with a detailed toy study. It is also worthwhile to consider cross-checks using a Bayesian technique. Other ways to improve this analysis are with the utilisation of new trigger lines. This would help to extend this study also using tracks in the endcap region of the detector. Especially the KLM trigger looks very promising since two muons in the final state could be triggered on.

Table 6.1: Values for the expected limits on the cross section $\sigma(e^+e^- \to \mu^+\mu^-$ inv.) and the coupling constant g' for different Z' masses. They were obtained for a dataset corresponding to an integrated luminosity of 50 fb^{-1} and 500 fb^{-1}.

	50 fb^{-1}		500 fb^{-1}	
mass [GeV]	σ [fb]	g'	σ [fb]	g'
0.1	6.44	1.17×10^{-2}	2.61	7.48×10^{-3}
0.5	3.51	5.40×10^{-3}	1.07	2.99×10^{-3}
1.0	1.94	5.83×10^{-3}	5.42×10^{-1}	3.08×10^{-3}
1.5	1.41	6.61×10^{-3}	3.71×10^{-1}	3.39×10^{-3}
2.0	1.02	7.28×10^{-3}	2.88×10^{-1}	3.87×10^{-3}
2.5	8.38×10^{-1}	8.42×10^{-3}	2.42×10^{-1}	4.52×10^{-3}
3.0	9.37×10^{-1}	1.12×10^{-2}	2.30×10^{-1}	5.58×10^{-3}
3.5	9.43×10^{-1}	1.43×10^{-2}	2.10×10^{-1}	6.75×10^{-3}
4.0	9.82×10^{-1}	2.13×10^{-2}	2.23×10^{-1}	1.02×10^{-2}
4.5	1.09	2.96×10^{-2}	2.61×10^{-1}	1.45×10^{-2}
5.0	1.12	3.95×10^{-2}	3.13×10^{-1}	2.08×10^{-2}
5.5	1.18	5.38×10^{-2}	3.48×10^{-1}	2.91×10^{-2}
6.0	1.42	7.96×10^{-2}	3.85×10^{-1}	4.14×10^{-2}
6.5	1.62	1.18×10^{-1}	4.60×10^{-1}	6.27×10^{-2}
7.0	2.67	2.17×10^{-1}	7.07×10^{-1}	1.12×10^{-1}
7.5	4.88	4.42×10^{-1}	1.31	2.29×10^{-1}
8.0	9.51	9.92×10^{-1}	2.91	5.48×10^{-1}

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

The obtained expected limits for an integrated luminosity of 50 fb^{-1} are at the order of 1 fb for Z' masses between 1.5 GeV and 6.5 GeV. For 500 fb^{-1} they are around 0.3 fb for the same mass range. For lower and higher masses, the sensitivity decreases slightly. The exact values for the limits on the cross section and the coupling constant are shown in Table 6.1. Above 8 GeV the analysis loses sensitivity to a Z' within the $L_\mu - L_\tau$ model, but it is worthwhile also to include this region in the analysis. However, the multivariate selection currently removes all events for very high recoil masses since the number of background events is too high. Some efforts were made to increase the sensitivity for low and high masses. For the low masses, this was achieved with some dedicated selections targeting the $\mu\mu$ background. It was not possible to get a better performance out of the ANN, even with more input variables and a deeper architecture.
More advanced machine learning techniques can be tried for the future, but with the currently available high-level information, I believe that further separation is not feasible. Nevertheless, when other detector information and signals from a deeper level are considered, this could better identify the signal events. Such an approach brings new complications and requires an excellent understanding and simulation of the detector. The experiment is still in an early stage, and while it is already performing very well, many improvements will yet be made on both the software and the hardware side in the future, leading to higher precision. New techniques can also be used to achieve better particle identification and reconstruction algorithms, which would improve all Belle II results.

The expected limits for an invisibly decaying Z' projected in this analysis will be reached in the near future. Already data corresponding to more than 50 fb^{-1} of integrated luminosity are recorded. If there is a Z' in the accessible mass range it could be observed soon and provide a solution to the dark matter puzzle. In any case, with the large dataset that will be collected in the coming years, many exciting physics results will come out of Belle II.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

List of Tables

2.1 The approximate values of the fine structure constant of the electromagnetic, the strong and the weak force for different energy scales. 5

3.1 The most dominant processes at Belle II with a CMS energy of 10.58 GeV. The cross sections are taken from 'The Belle II Physics Book' [36, p. 48]. 27

3.2 The names and activation conditions for a few selected trigger lines. 35

4.1 The simulated processes with their cross section and the corresponding luminosity of the generated samples (MC13a). If not suggested otherwise all particles are to be understood as charged particles such that the overall charge sums to zero. . . . 41

4.2 A comparison of the different tracks on the 3-prong side. 46

4.3 The composition according to particle types in the MC sample after applying all selections. 47

4.4 The binning used for calculating the $\pi \to e$ mis-id rates. 50

4.5 The binning used for calculating the $\pi \to \mu$ mis-id rates and pion efficiencies. . . . 50

5.1 Generated Monte Carlo samples for different processes considered as a possible background for the analysis with corresponding luminosities. If not suggested otherwise all particles are to be understood as charged particles such that the overall charge sums to zero. 58

5.2 Overview of the event selections applied before the multivariate selection. A bold number indicates a cut which was applied at a later stage of the event processing. 60

5.3 The most important features found after training a large amount of BDTs. These variables are used for training the ANN. 69

5.4 The minimum and maximum values of the input variables, used for scaling the distributions to lie between 0 and 1. 71

5.5 The number of training events and weight factors used to train the ANN. 72

5.6 Hyperparameters for the first and second training of the ANN based on different loss functions. 76

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

5.7 The expected number of events for each of the simulated processes scaled to 50 fb^{-1}. The first column gives the total number of events expected with this luminosity, the second column gives the number of events after all univariate selections and the third column the number of events after applying a cut on the ANN output. 84

5.8 The expected number of events for three different mass regions after applying all the selections. The numbers are scaled to correspond to 50 fb^{-1}. 84

5.9 The expected number of events for each of the simulated processes scaled to 500 fb^{-1}. The first column gives the total number of events expected with this luminosity, the second column gives the number of events after all univariate selections and the third column the number of events after applying a cut on the ANN output. 85

5.10 The expected number of events for three different mass regions after applying all the selections. The numbers are scaled to correspond to 500 fb^{-1}. 85

5.11 The multiplicative modifiers considered, together with their auxiliary measurements. 88

5.12 The variation on the normalisation artificially introduced on a new simulated dataset and the equivalent standard deviations of the Gaussian constraint term. 99

5.13 The quantiles corresponding to $\pm 1\sigma$ and $\pm 2\sigma$ deviations from the mean of a standard normal distribution. 102

5.14 Values for $\langle N(c_0) \rangle$ and $\mathcal{N}$ compared for the test statistic $\tilde{t}_0$ and t_0, evaluated at $c_0 = 7$. 106

6.1 Values for the expected limits on the cross section $\sigma(e^+e^- \to \mu^+\mu^-$ inv.) and the coupling constant g' for different Z' masses. They were obtained for a dataset corresponding to an integrated luminosity of 50 fb^{-1} and 500 fb^{-1}. 112

A.1 Recoil mass windows around each generated Z' mass corresponding to ± 2 times the standard deviation of the respective signal peak. For each window the signal efficiency and the total number of expected background events are shown for a integrated luminosity of 50 fb^{-1}. 129

A.2 Continuation of Table A.1. 130

A.3 Recoil mass windows around each generated Z' mass corresponding to ± 2 times the standard deviation of the respective signal peak. For each window the signal efficiency and the total number of expected background events are shown for a integrated luminosity of 500 fb^{-1}. 131

A.4 Continuation of Table A.3. 132

A.5 The bin edges defined over the squared recoil mass to perform the binned ML fit. The bin widths were calculated in such a way that they correspond to the FWHM of the signal distribution, which resulted in a total of 196 bins. (Section 5.5.3) . . . 133

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

A.6 The mapping of the POI index to the assumed signal strength μ for different values of m. The respective μ was used for calculating the p-values of the signal like distribution in Figure 5.39a. 133

A.7 The mapping of the POI index to the assumed signal strength μ for different values of m. The respective μ was used for calculating the p-values of the background like distribution in Figure 5.39b. 134

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

List of Figures

2.1 An overview of the particles and interactions described by the Standard Model. . 4

2.2 A sketch of coupling constants' energy dependence for the strong, weak and electromagnetic interaction. A hypothetical increase of the gravitational interaction (not included in the SM) is also shown, suggesting that all forces unify at some scale. 6

2.3 The observed rotation curve for NGC 2403 (dots with error bars) together with the fitted contribution of luminous matter and the dark halo to the overall observed velocity (solid lines). [from 6, Figure 4] 7

2.4 A Feynman diagram showing the process $e^+e^- \to \mu^+\mu^-$ ($Z' \to$ invisible) where invisible is either $\nu\bar{\nu}$ or $\chi\bar{\chi}$. 9

2.5 The current 90 % upper limits on the g' coupling constant for different cases of the $Z' \to$ invisible BF. The region which could explain the muon anomalous magnetic moment is highlighted $(g-2)$. [1] 9

2.6 The layout of a simple neural network with three fully connected layers (two neurons in the input layer, three neurons in the hidden layer and one neuron in the output layer). The connection between the neurons have the weight w^l_{jk}, and each neuron has some bias b^l_j associated. 21

2.7 A graph of the sigmoid (Fermi) activation function. 21

3.1 The SuperKEKB accelerator and the Belle II detector. [31] 25

3.2 A comparison of the crab crossing scheme, which was used in KEKB from 2007 to 2010, and the crab waist scheme deployed in SuperKEKB. [32, p. 3] 26

3.3 A cross-section of the Belle II detector, showing the interaction point and all the subdetectors from above. [36] 28

3.4 The VXD viewed from the front (r-φ view) and the side (r-z view). The four outer layers are the SVD and the two inner layers the PXD. The innermost circle on the left and the dashed lines on the right correspond to the beam pipe. [36] 29

3.5 A 3D rendering of the VXD, showing the PXD in the centre close to the beam pipe and the SVD layers arranged around it. [38] 30

3.6 Simulated CDC measurements shown in the r-φ plane. 31

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

3.7 Left side: One of the quartz bars of the TOP detector. The Cherenkov photons get reflected inside the bar until they reach the detector located at one side of the bar. Right side: A sketch of how two aerogel layers with different refractive indices can be chosen to give overlapping rings from the Cherenkov photons at the detector plane. [36] . . . 32

3.8 The weekly recorded luminosity at Belle II for the years 2019 and 2020 overlayed with the total integrated luminosity. Out of the 90 fb^{-1} only around 63 fb^{-1} pass the quality requirements for physics analyses. [41] . . . 36

4.1 An example of a taupair 3-prong 1-prong decay. . . . 39

4.2 A pie chart showing the relevant branching fractions of tau leptons. . . . 39

4.3 Optimisation of the first selection in the 2D plane spanned by the two possible combinations of the $\pi^+\pi^-$ invariant mass from the 3-prong τ decay products. . . . 43

4.4 Optimisation of the second selection in the 2D plane spanned by the thrust value and the total visible energy in the CMS. . . . 44

4.5 The distributions of visible energy in the CMS and the invariant mass of the pion pairs in data and MC after background suppression. . . . 45

4.6 PionID and muonID distributions in data and MC after background suppression. . . . 46

4.7 The momentum and polar angle dependency of the trigger efficiency for the `fff` trigger, the `ffo` trigger, and their combination. . . . 48

4.8 The distributions of the combined sample containing all tracks on the 3-prong side. In data only events where the `fff` or `ffo` trigger was activated are shown and the MC distribution is scaled according to the calculated trigger efficiency. . . . 49

4.9 The polar angle bins in a side view of the Belle II detector. The blue markers correspond to the bins defined in Table 4.4, and the red markers correspond to the bins defined in Table 4.5. . . . 50

4.10 The three tracks from the 3-prong tau decay. Track number 1 and 3 have the same charge, while track number two has the charge with the opposite sign. . . . 52

4.11 The pion efficiencies using the global PID (a) and the binary PID (b). A better data/MC agreement is evident for the binary PID. . . . 54

4.12 The $\pi \to \mu$ mis-id rates using only KLM information, in (a) for the track samples and in (b) for the full sample. Only tracks with $p > 0.7$ GeV reach the KLM and have a likelihood defined. . . . 54

4.13 The $\pi \to e$ mis-id rates for different subdetectors. In (a) the charge-separated samples are shown with only ECL information. A difference between the samples is evident. In (b) only CDC information is used. Here a large data/MC discrepancy is observed. . . . 55

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

4.14 The dependence of the $\pi \to e$ mis-id rates (a) and the $\pi \to \mu$ mis-id rates (b) on the threshold value of the related global PID variable. The mis-id rates are displayed for the different track samples. 55

4.15 2D histograms showing how the mis-id rates depend on the distance to the closest track at varying radius. Here the $\pi \to e$ mis-id rates are shown in a given momentum range and angular acceptance. 56

5.1 Signal and background distributions of the recoil mass after all univariate selections, showing only the main backgrounds and signal distributions for 100 MeV, 2.5 GeV, 5 GeV and 7 GeV. 59

5.2 Signal and background distributions of the 2D opening angle between μ^+ and μ^- in the r-φ plane after preselections. 61

5.3 Signal and background distributions of the extra energy deposited in the ECL after preselections. 61

5.4 Signal and background distributions of the transverse momentum of the dimuon system in the CMS after all univariate selections. 62

5.5 A two dimensional histogram showing signal events for different values of $\mathrm{m_{rec}}$ and $\mathrm{p_t^*}(\mu^+\mu^-)$ after the multivariate selection was applied. 62

5.6 Signal and background distributions of the opening angle between μ^+ and μ^- in the CMS after preselections. 63

5.7 Signal and background distributions of θ_{rec} after preselections. 63

5.8 Two dimensional histograms showing background and signal events for different values of $\mathrm{m_{rec}}$ and θ_{rec} after preselection. 63

5.9 Two dimensional histograms showing background and signal events for different values of $\mathrm{m_{rec}}$ and θ_{rec} after the multivariate selection was applied. 64

5.10 Signal and background distributions of $\mathrm{E_{cluster}}$ for the positively charged muon after preselections. 65

5.11 Signal and background distributions of the ratio $\mathrm{E_{cluster}/p}$ from the positively charged muon after preselections. 65

5.12 The CMS energy of the lower energetic muon and the deposited ECL energy for tracks with $\mathrm{E_{cluster}/p} > 1$. For lower CMS energies a larger cluster energy is observed, consistent with the FSR hypothesis. 66

5.13 Signal and background distributions of the CMS energy from the higher energetic muon after all univariate selections. Lower values correspond to higher recoil masses. 66

5.14 Signal and background distributions of the recoil mass after all univariate selections, showing all backgrounds with remaining events. The filled histograms are stacked. 66

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

5.15 The most promising features and their correlations, selected from the 50 initial variables. They are different combinations of kinematic and topology related variables in different reference frames. 68

5.16 The feature importance ranking for one combination of variables, obtained with three different methods. Depending on what variables enter the training, the result of these rankings can differ quite significantly. 68

5.17 The distribution and pairwise correlations for the variables used in the multivariate selection. The shown signal events and SM backgrounds are selected from the respective Z' mass window. 70

5.18 A sketch of the ANN architecture. The first training with the BCE loss function was used to set the weights and biases of the net for the second training with the custom loss function based on the Punzi figure of merit. 71

5.19 The evolution of the BCE loss for 300 epochs in the first training. 73

5.20 The output of the ANN after the first training (BCE) for different values of the recoil mass. The red line indicates the cut which yields the largest Punzi figure of merit in the respective mass windows. The output distributions are shown before applying the sigmoid activation function on the output layer. 74

5.21 The evolution of the Punzi based loss for 2500 epochs in the second training. . . . 76

5.22 The output of the ANN after the second training (FOM) for different values of the recoil mass. The red line indicates the cut which yields the largest Punzi figure of merit in the respective mass windows. The output distributions are shown before applying the sigmoid activation function on the output layer. 77

5.23 The signal efficiency and the number of remaining background events for different working points. The background events are counted in the respective mass windows and a trigger efficiency of 89 % is applied on both signal and background. The classifier output before applying the sigmoid activation function is used. A cut at 0 will be chosen for the final selection. 79

5.24 The true positive and false positive rates of the classifier, shown for different Z' masses. Only events within the mass windows are used. The crosses mark the final selection. 79

5.25 The Punzi sensitivity region ('minimum detectable cross section') for different Z' masses before and after applying the ANN. 80

5.26 The Punzi FOM compared for three different selections: The optimised selection after the first training (BCE), the optimised selection after the second training (FOM) and the cut at a fixed value of the ANN output after the second training. . 80

5.27 A comparison of the ANN performance for the masses used in training and those left out. No difference can be observed between the samples. The shaded area indicates the expected $\pm 1\sigma$ statistical error. 81

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

5.28 A 3D scatter plot showing the input space of the ANN with $p_t^*(\mu^+\mu^-)$ set to 2.2 GeV/c. The separation boundary defined by the final selection (green sheet) separates the planes corresponding to different recoil masses in a way that optimises the selection for all Z' masses. 82

5.29 Background distributions of the recoil mass after applying all selections. 83

5.30 Fits of the signal distribution (after all selections) in order to compute the FWHM. The signal is modelled with the sum of a Gaussian and a Crystal Ball function. . . 89

5.31 The FWHM for each generated mass point and the thereof resulting width of the bins in the squared recoil mass distribution. 90

5.32 The bins defined for the recoil mass to be used for the fit. The bin width is calculated based on the FWHM of the signal. The vertical lines show the bin edges and the crosses indicate the mass points where a signal was generated. For visual clarity, the individual mass points are shifted with respect to the adjacent points. 90

5.33 The distribution of the estimated value $\widehat{\mu}$ around the true value μ' obtained with 50 000 toys. The hypothesised Z' mass is 2 GeV, and the chosen value for μ' is indicated by the vertical dotted line. 92

5.34 The test statistic $\tilde{t}_\mu$ depending on the estimated value $\widehat{\mu}$ for different combinations of the assumed value μ and the true value μ'. The asymptotic approximation based on Wald's formula is overlayed over the empirical values of $\tilde{t}_\mu$ using toy samples. σ is obtained from the HESSE method of MINUIT. For $\mu \neq \mu'$ also the approximation where σ is calculated from the Asimov data is shown. 93

5.35 The signal like distribution for three different values of m and μ, comparing toy samples and the asymptotic approximations, showing a good agreement between the distributions. 94

5.36 The background like distributions for three different values of m and μ, comparing toy samples and the asymptotic approximations. Here the asymptotic formulae fail to describe the distribution for increasing values of μ. 95

5.37 The p-value as a function of $\tilde{t}_\mu^{\text{obs}}$ for the signal like distribution for some combinations of m and μ. The asymptotic and the empirical distributions agree reasonably well. 96

5.38 The p-value as a function of $\tilde{t}_\mu^{\text{obs}}$ for the background like distribution for some combinations of m and μ. The asymptotic and empirical distributions divert for larger $\tilde{t}_\mu^{\text{obs}}$ except for $\mu = 0$. 97

5.39 The value of the p-value obtained with the asymptotic formula calculated at the 90% quantile of the empirical distribution for different values of m and μ. For the signal like distribution (a) the agreement is reasonably well. For the background like distribution (b) significant deviations are observed for $\mu > 0$. The mapping of the POI index to μ is given in Table A.6 and Table A.7. 97

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

5.40 The simulated dataset together with the nominal MC expectation for a total integrated luminosity of 50 fb^{-1}. Additionally to the background contributions a 3100 MeV Z' was injected with a cross section of 1.5 fb. In (b) also the signal MC which was used to sample the dataset is overlayed. 98

5.41 A pull plot showing on the y-axis the number of standard deviations the estimated nuisance parameters deviate from the value of the auxiliary measurement, which in this case also corresponds to the true value. 99

5.42 The simulated dataset with artificially introduced discrepancies together with the nominal MC expectation for a total integrated luminosity of 50 fb^{-1}. Additionally to the background contributions a 3100 MeV Z' was injected with a cross section of 1.5 fb. 100

5.43 The estimated values of $\boldsymbol{\nu}$ after performing the fit for the simulated dataset with discrepancies. The true value of $\boldsymbol{\nu}$ is correctly estimated within the errors. 100

5.44 The local p-values for different mass hypothesis for the simulated pseudo dataset corresponding to a total integrated luminosity of 50 fb^{-1}. p-values are obtained from toy samples as well as from the asymptotic formulae. 101

5.45 The interval construction for $m = 3100$ MeV. The shaded area corresponds to the observed two-sided confidence interval. The expected upper limit is given by the intersection of the blue line with a p-value of 0.1. 102

5.46 The confidence intervals for all mass points calculated with the simulated pseudo-data (shaded blue region). The colour bar indicates the number of signal events that would be expected for a given cross section μ. 102

5.47 The p-values when testing against the signal plus background hypothesis for all considered values of m and μ. The overlayed contours denote a p-value of 0.1, defining the borders of the 90% CL intervals. 103

5.48 The empirical distribution of $\tilde{t}_0(\widehat{m})$ overlayed with the distribution of $t_0(\widehat{m})$ divided by two. 105

5.49 The values of the test statistic $\tilde{t}_0$ overlayed with t_0 for different mass hypothesis, calculated for the same simulated sample. 105

5.50 The mean number of upcrossings and the value of $\mathcal{N}$ as a function of c_0. Only around $c_0 = 8$ is $\mathcal{N}$ independend of c_0. 105

5.51 A comparison of the p-values obtained empirically and using the asymptotic approximations as a function of $\tilde{t}_\mu^{\mathrm{obs}}$. 106

5.52 The relation between local and global significance. The empirical relation is due to limited sample size only usable up to $\mathrm{Z}_{\mathrm{local}} \sim 3$, while the asymptotic bound is valid for high significances. 107

5.53 The simulated dataset together with the nominal MC expectation for a total integrated luminosity of 500 fb^{-1}. Additionally to the background contributions a 3100 MeV Z' was injected with a cross section of 1.5 fb. 108

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

5.54 The local p-values for different mass hypothesis for the simulated pseudo dataset corresponding to a total integrated luminosity of 500 fb^{-1}. p-values are obtained from toy samples as well as from the asymptotic formulae. 108

6.1 The expected limits at 90 % CL on the cross section for a luminosity of 50 fb^{-1}. The blue diamonds mark the results of the statistical procedure applied on simulated pseudo-data, whereas a 3.1 GeV Z' was injected with a cross section of 1.5 fb. . . . 109

6.2 The expected limits at 90 % CL on g' for a luminosity of 50 fb^{-1}. The blue diamonds mark the results of the statistical procedure applied on simulated pseudo-data, whereas a 3.1 GeV Z' was injected with a cross section of 1.5 fb. 110

6.3 The expected limits at 90 % CL on the cross section, compared for a luminosity of 50 fb^{-1} and 500 fb^{-1}. 111

6.4 The expected limits at 90 % CL on g', compared for a luminosity of 50 fb^{-1} and 500 fb^{-1}. 111

B.1 A summary of the $\pi \to \mu$ mis-id rates for the dataset described in Chapter 4. The mis-id rates are shown in bins of momentum and polar angle, based on the binning scheme defined on page 50. 136

B.2 A summary of the $\pi \to e$ mis-id rates for the dataset described in Chapter 4. The mis-id rates are shown in bins of momentum and polar angle, based on the binning scheme defined on page 50. 137

B.3 A summary of the pion identification efficiencies for the dataset described in Chapter 4. The efficiencies are shown in bins of momentum and polar angle, based on the binning scheme defined on page 50. 138

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

References

[1] I. Adachi et al. (Belle II Collaboration), 'Search for an invisibly decaying Z' boson at Belle II in $e^+e^- \to \mu^+\mu^-(e^\pm\mu^\mp)$ plus missing energy final states', Phys. Rev. Lett. **124**, 141801 (2020).

[2] Cush, *Standard model of elementary particles*, public domain license; accessed Jan-2021, (2019) https://commons.wikimedia.org/w/index.php?curid=4286964.

[3] TriTertButoxy, *Summary of interactions between certain particles described by the standard model*, public domain license; accessed Jan-2021, (2013) https://commons.wikimedia.org/wiki/File:Elementary_particle_interactions.svg.

[4] M. Thomson, *Modern particle physics* (Cambridge University Press, 2013).

[5] K. G. Begeman, 'HI rotation curves of spiral galaxies. I. NGC 3198.', Astronomy and Astrophysics **223**, 47–60 (1989).

[6] T. S. van Albada, R. Sancisi, M. Petrou, R. J. Tayler, J. D. Barrow, P. J. E. Peebles and D. W. Sciama, 'Dark matter in spiral galaxies', Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences **320**, 447–464 (1986).

[7] G. Bertone, D. Hooper and J. Silk, 'Particle dark matter: evidence, candidates and constraints', Physics Reports **405**, 279–390 (2005).

[8] M. Tanabashi et al. (Particle Data Group), 'Review of particle physics', Phys. Rev. D **98**, 030001 (2018).

[9] R. Foot, 'New physics from electric charge quantization?', Modern Physics Letters A **06**, 527–529 (1991).

[10] R. Foot, X.-G. He, H. Lew and R. R. Volkas, 'Model for a light Z' boson', Physical Review D **50**, 4571–4580 (1994).

[11] X.-G. He, G. C. Joshi, H. Lew and R. R. Volkas, 'Simplest Z' model', Physical Review D **44**, 2118–2132 (1991).

[12] A. Dev, 'Gauged $L_\mu - L_\tau$ model with an inverse seesaw mechanism for neutrino masses', (2018).

[13] W. Altmannshofer, S. Gori, S. Profumo and F. S. Queiroz, 'Explaining dark matter and B decay anomalies with an $L_\mu - L_\tau$ model', Journal of High Energy Physics **2016** (2016) 10.1007/jhep12(2016)106.

[14] B. Shuve and I. Yavin, 'Dark matter progenitor: light vector boson decay into sterile neutrinos', Physical Review D **89** (2014) 10.1103/physrevd.89.113004.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

[15]W. Altmannshofer, M. J. Baker, S. Gori, R. Harnik, M. Pospelov, E. Stamou and A. Thamm, 'Light resonances and the low-q^2 bin of R_{K^*}', Journal of High Energy Physics **2018** (2018) 10.1007/jhep03(2018)188.

[16]D. Curtin, R. Essig, S. Gori and J. Shelton, 'Illuminating dark photons with high-energy colliders', Journal of High Energy Physics **2015** (2015) 10.1007/jhep02(2015)157.

[17]J. P. Lees, V. Poireau, V. Tisserand, E. Grauges, A. Palano, G. Eigen, D. N. Brown, Y. G. Kolomensky, H. Koch, T. Schroeder et al. (BaBar Collaboration), 'Search for a muonic dark force at BaBar', Physical Review D **94** (2016) 10.1103/physrevd.94.011102.

[18]O. Behnke and L. Moneta, 'Parameter estimation', in *Data analysis in high energy physics* (John Wiley & Sons, Ltd) Chap. 2, pp. 27–73.

[19]F. James, 'Minuit function minimization and error analysis: reference manual version 94.1', (1994).

[20]S. Baker and R. D. Cousins, 'Clarification of the use of chi-square and likelihood functions in fits to histograms', Nuclear Instruments and Methods in Physics Research **221**, 437–442 (1984).

[21]S. S. Wilks, 'The large-sample distribution of the likelihood ratio for testing composite hypotheses', Ann. Math. Statist. **9**, 60–62 (1938).

[22]J. Neyman, 'Outline of a theory of statistical estimation based on the classical theory of probability', Philosophical Transactions of the Royal Society of London Series A **236**, 333–380 (1937).

[23]G. Schott, 'Hypothesis testing', in *Data analysis in high energy physics* (John Wiley & Sons, Ltd) Chap. 3, pp. 75–105.

[24]G. Cowan, K. Cranmer, E. Gross and O. Vitells, 'Asymptotic formulae for likelihood-based tests of new physics', The European Physical Journal C **71**, 1554 (2011).

[25]J. Neyman and E. S. Pearson, 'On the problem of the most efficient tests of statistical hypotheses', Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character **231**, 289–337 (1933).

[26]G. J. Feldman and R. D. Cousins, 'Unified approach to the classical statistical analysis of small signals', Phys. Rev. D **57**, 3873–3889 (1998).

[27]E. Gross and O. Vitells, 'Trial factors for the look elsewhere effect in high energy physics', The European Physical Journal C **70**, 525–530 (2010).

[28]R. B. Davies, 'Hypothesis testing when a nuisance parameter is present only under the alternatives', Biometrika **74**, 33–43 (1987).

[29]G. Punzi, 'Sensitivity of searches for new signals and its optimization', Proceedings of PHYSTAT2003 **eConf C030908**, edited by L. Lyons, R. Mount and R. Reitmeyer (2003).

[30]B. Efron, 'Bayesians, frequentists, and physicists', Proceedings of PHYSTAT2003 **eConf C030908**, edited by L. Lyons, R. Mount and R. Reitmeyer (2003).

[31]*SuperKEKB and Belle II*, Belle II web page, accessed Jan-2021, https://www.belle2.org/project/super_kekb_and_belle_ii.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

[32]K. Akai, K. Furukawa and H. Koiso, 'SuperKEKB collider', Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **907**, 188–199 (2018).

[33]SuperB Collaboration, 'SuperB: a high-luminosity asymmetric e+ e- super flavor factory. conceptual design report', arXiv (2007).

[34]B. Aubert et al. (BABAR Collaboration), 'Observation of CP violation in the B^0 meson system', Phys. Rev. Lett. **87**, 091801 (2001).

[35]K. Abe et al. (Belle Collaboration), 'Observation of large CP violation in the neutral B meson system', Phys. Rev. Lett. **87**, 091802 (2001).

[36]E. Kou, P. Urquijo, W. Altmannshofer, F. Beaujean, G. Bell, M. Beneke, I. I. Bigi, F. Bishara, M. Blanke, C. Bobeth et al. (Belle II Collaboration), 'The Belle II physics book', Progress of Theoretical and Experimental Physics **2019** (2019) 10.1093/ptep/ptz106.

[37]T. Abe et al. (Belle II Collaboration), 'Belle II technical design report', arXiv (2010).

[38]V. Bertacchi, T. Bilka, N. Braun, G. Casarosa, L. Corona, S. Cunliffe, F. Dattola, G. De Marino, M. De Nuccio, G. De Pietro et al., 'Track finding at Belle II', Computer Physics Communications **259**, 107610 (2021).

[39]N. Braun, 'Momentum estimation of slow pions and improvements on the track finding in the central drift chamber for the Belle II experiment', MA thesis (Karlsruhe Institute of Technology (KIT), Karlsruhe, 2015).

[40]A. J. Bevan, B. Golob, T. Mannel, S. Prell, B. D. Yabsley, H. Aihara, F. Anulli, N. Arnaud, T. Aushev, M. Beneke et al., 'The physics of the B factories', The European Physical Journal C **74** (2014) 10.1140/epjc/s10052-014-3026-9.

[41]*Total recorded integrated luminosity*, Belle II confluence page, accessed Jan-2021, https://confluence.desy.de/display/BI/Belle+II+Luminosity.

[42]T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth and N. Braun, 'The Belle II core software', Computing and Software for Big Science **3** (2018) 10.1007/s41781-018-0017-9.

[43]S. Agostinelli et al. (GEANT4), 'GEANT4–a simulation toolkit', Nucl. Instrum. Meth. A **506**, 250–303 (2003).

[44]J. F. Krohn, P. Urquijo, F. Abudinén, S. Cunliffe, T. Ferber, M. Gelb, J. Gemmler, P. Goldenzweig, T. Keck, I. Komarov, T. Kuhr, L. Ligioi, M. Lubej, F. Meier, F. Metzner, C. Pulvermacher, M. Ritter, U. Tamponi, F. Tenchini and A. Zupanc, 'Global decay chain vertex fitting at b-factories', (2019) 10.1016/j.nima.2020.164269.

[45]G. Chikovani, M. Focacci, W. Kienzle, C. Lechanoine, B. Levrat, B. Maglic, M. Martin, P. Schübelin, L. Dubal, M. Fischer, P. Grieder, H. Neal and C. Nef, 'Evidence for a two-peak structure in the A2 meson', Physics Letters B **25**, 44–47 (1967).

[46]J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli and M. Zaro, 'The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations', Journal of High Energy Physics **2014**, 79 (2014).

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

[47]M. Bertemes et al., 'L1 CDC trigger performance study targeting dark sector analyses with 2019 data', Belle II internal document (2020).

[48]T. Keck, 'FastBDT: a speed-optimized multivariate classification algorithm for the Belle II experiment', Computing and Software for Big Science **1**, 2 (2017).

[49]F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, 'Scikit-learn: machine learning in python', Journal of Machine Learning Research **12**, 2825–2830 (2011).

[50]scikit-learn developers, *Scikit-learn user guide: permutation feature importance*, (2021) https://scikit-learn.org/stable/modules/permutation_importance.html.

[51]A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai and S. Chintala, 'PyTorch: an imperative style, high-performance deep learning library', in *Advances in neural information processing systems 32*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox and R. Garnett (Curran Associates, Inc., 2019), pp. 8024–8035.

[52]F. Abudinén et al. (Belle II Collaboration), 'Measurement of the integrated luminosity of the phase 2 data of the Belle II experiment', Chinese Physics C **44**, 021001 (2020).

[53]L. Heinrich, M. Feickert and G. Stark, *Scikit-hep/pyhf*, version v0.5.3, (Oct. 2020) https://doi.org/10.5281/zenodo.4110938.

[54]K. Cranmer, G. Lewis, L. Moneta, A. Shibata and W. Verkerke (ROOT), 'HistFactory: A tool for creating statistical models for use with RooFit and RooStats', (2012).

[55]A. Wald, 'Tests of statistical hypotheses concerning several parameters when the number of observations is large', Transactions of the American Mathematical Society **54**, 426–482 (1943).

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Appendix A

Tables

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Table A.1: Recoil mass windows around each generated Z' mass corresponding to ± 2 times the standard deviation of the respective signal peak. For each window the signal efficiency and the total number of expected background events are shown for a integrated luminosity of 50 fb^{-1}.

mass [GeV/c^2]	window [$\mathrm{m_{rec}}$]	window [$\mathrm{m^2_{rec}}$]	signal efficiency	expected background events
0.1	−0.68 - 0.69	−0.46 - 0.48	0.0240 ± 0.0010	18.4 ± 0.9
0.2	−0.66 - 0.72	−0.43 - 0.51	0.0313 ± 0.0012	19.0 ± 0.9
0.3	−0.63 - 0.76	−0.40 - 0.58	0.0391 ± 0.0013	19.6 ± 0.9
0.4	−0.57 - 0.80	−0.32 - 0.64	0.0450 ± 0.0014	20.7 ± 0.9
0.5	−0.52 - 0.88	−0.27 - 0.77	0.0474 ± 0.0015	21.8 ± 1.0
0.6	−0.47 - 0.97	−0.22 - 0.94	0.0528 ± 0.0015	23.2 ± 1.0
0.7	−0.33 - 1.04	−0.11 - 1.09	0.0582 ± 0.0016	23.4 ± 1.0
0.8	0.25 - 1.10	0.06 - 1.22	0.0587 ± 0.0016	20.5 ± 0.9
0.9	0.47 - 1.18	0.22 - 1.40	0.0645 ± 0.0017	18.1 ± 0.9
1.0	0.65 - 1.25	0.43 - 1.57	0.0626 ± 0.0017	15.2 ± 0.8
1.1	0.79 - 1.34	0.62 - 1.80	0.0658 ± 0.0017	13.4 ± 0.8
1.2	0.90 - 1.44	0.80 - 2.08	0.0664 ± 0.0017	11.9 ± 0.7
1.3	1.01 - 1.54	1.01 - 2.37	0.0680 ± 0.0017	10.5 ± 0.7
1.4	1.16 - 1.60	1.35 - 2.57	0.0718 ± 0.0018	7.4 ± 0.6
1.5	1.28 - 1.69	1.64 - 2.86	0.0713 ± 0.0018	6.6 ± 0.5
1.6	1.39 - 1.78	1.94 - 3.18	0.0708 ± 0.0018	5.6 ± 0.5
1.7	1.51 - 1.87	2.28 - 3.50	0.0683 ± 0.0018	4.6 ± 0.4
1.8	1.62 - 1.96	2.62 - 3.86	0.0684 ± 0.0018	4.0 ± 0.4
1.9	1.73 - 2.05	3.00 - 4.22	0.0712 ± 0.0018	3.6 ± 0.4
2.0	1.84 - 2.15	3.38 - 4.62	0.0713 ± 0.0018	3.9 ± 0.4
2.1	1.95 - 2.24	3.80 - 5.02	0.0763 ± 0.0019	4.1 ± 0.6
2.2	2.06 - 2.33	4.25 - 5.43	0.0766 ± 0.0019	3.8 ± 0.6
2.3	2.16 - 2.43	4.68 - 5.90	0.0729 ± 0.0018	3.5 ± 0.4
2.4	2.27 - 2.52	5.16 - 6.36	0.0776 ± 0.0019	3.5 ± 0.4
2.5	2.38 - 2.62	5.65 - 6.85	0.0762 ± 0.0019	3.3 ± 0.4
2.6	2.49 - 2.71	6.18 - 7.34	0.0734 ± 0.0018	3.06 ± 0.35
2.7	2.58 - 2.81	6.67 - 7.91	0.0746 ± 0.0018	3.2 ± 0.4
2.8	2.70 - 2.90	7.27 - 8.41	0.0748 ± 0.0018	2.94 ± 0.35
2.9	2.79 - 3.00	7.80 - 9.02	0.0761 ± 0.0019	3.1 ± 0.4
3.0	2.91 - 3.09	8.46 - 9.54	0.0762 ± 0.0019	2.94 ± 0.35
3.1	3.01 - 3.19	9.07 - 10.15	0.0770 ± 0.0019	2.47 ± 0.32
3.2	3.11 - 3.29	9.68 - 10.80	0.0723 ± 0.0018	2.89 ± 0.34
3.3	3.21 - 3.39	10.30 - 11.48	0.0747 ± 0.0018	3.4 ± 0.4
3.4	3.32 - 3.48	11.00 - 12.12	0.0739 ± 0.0018	2.71 ± 0.32
3.5	3.42 - 3.58	11.69 - 12.81	0.0744 ± 0.0018	2.50 ± 0.31
3.6	3.52 - 3.68	12.38 - 13.54	0.0770 ± 0.0019	2.68 ± 0.33
3.7	3.62 - 3.78	13.10 - 14.28	0.0783 ± 0.0019	2.98 ± 0.35
3.8	3.73 - 3.87	13.90 - 14.98	0.0760 ± 0.0018	2.95 ± 0.35
3.9	3.83 - 3.97	14.68 - 15.74	0.0727 ± 0.0018	2.44 ± 0.32
4.0	3.93 - 4.07	15.48 - 16.52	0.0763 ± 0.0019	2.73 ± 0.34
4.1	4.03 - 4.17	16.27 - 17.35	0.0765 ± 0.0019	3.1 ± 0.4
4.2	4.14 - 4.26	17.12 - 18.16	0.0747 ± 0.0018	3.12 ± 0.35
4.3	4.24 - 4.36	18.00 - 18.98	0.0731 ± 0.0018	3.6 ± 0.4

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table A.2: Continuation of Table A.1.

mass [GeV/c^2]	window [m_{rec}]	window [m^2_{rec}]	signal efficiency	expected background events
4.4	4.34 - 4.46	18.81 - 19.91	0.0732 ± 0.0018	4.5 ± 0.4
4.5	4.44 - 4.56	19.75 - 20.75	0.0753 ± 0.0018	3.8 ± 0.4
4.6	4.55 - 4.65	20.68 - 21.64	0.0750 ± 0.0018	4.1 ± 0.4
4.7	4.65 - 4.75	21.58 - 22.60	0.0733 ± 0.0018	4.0 ± 0.4
4.8	4.75 - 4.85	22.52 - 23.56	0.0743 ± 0.0018	5.1 ± 0.5
4.9	4.85 - 4.95	23.50 - 24.52	0.0726 ± 0.0018	3.9 ± 0.4
5.0	4.95 - 5.05	24.51 - 25.49	0.0746 ± 0.0018	5.3 ± 0.5
5.1	5.05 - 5.15	25.54 - 26.48	0.0724 ± 0.0018	5.0 ± 0.8
5.2	5.16 - 5.24	26.60 - 27.48	0.0711 ± 0.0018	5.6 ± 0.5
5.3	5.25 - 5.35	27.58 - 28.60	0.0716 ± 0.0018	4.6 ± 0.4
5.4	5.36 - 5.44	28.69 - 29.63	0.0742 ± 0.0018	4.7 ± 0.4
5.5	5.46 - 5.54	29.78 - 30.72	0.0720 ± 0.0018	6.4 ± 0.5
5.6	5.56 - 5.64	30.92 - 31.80	0.0723 ± 0.0018	5.6 ± 0.5
5.7	5.66 - 5.74	32.04 - 32.94	0.0684 ± 0.0018	6.3 ± 0.7
5.8	5.76 - 5.84	33.19 - 34.09	0.0688 ± 0.0018	5.6 ± 0.5
5.9	5.86 - 5.94	34.33 - 35.29	0.0737 ± 0.0018	7.4 ± 0.6
6.0	5.97 - 6.03	35.61 - 36.39	0.0723 ± 0.0018	7.3 ± 0.6
6.1	6.07 - 6.13	36.82 - 37.60	0.0687 ± 0.0018	7.2 ± 0.6
6.2	6.17 - 6.23	38.02 - 38.86	0.0718 ± 0.0018	8.0 ± 0.6
6.3	6.27 - 6.33	39.32 - 40.06	0.0703 ± 0.0018	8.0 ± 0.6
6.4	6.37 - 6.43	40.59 - 41.33	0.0695 ± 0.0018	9.8 ± 0.6
6.5	6.47 - 6.53	41.90 - 42.60	0.0706 ± 0.0018	11.4 ± 0.7
6.6	6.58 - 6.62	43.23 - 43.89	0.0720 ± 0.0018	12.6 ± 0.7
6.7	6.67 - 6.73	44.54 - 45.24	0.0716 ± 0.0018	17.5 ± 0.9
6.8	6.78 - 6.82	45.95 - 46.53	0.0686 ± 0.0018	18.9 ± 0.9
6.9	6.88 - 6.92	47.28 - 47.94	0.0716 ± 0.0018	25.5 ± 1.0
7.0	6.98 - 7.02	48.68 - 49.32	0.0729 ± 0.0018	31.2 ± 1.1
7.1	7.08 - 7.12	50.12 - 50.70	0.0692 ± 0.0018	41.3 ± 1.3
7.2	7.18 - 7.22	51.55 - 52.13	0.0717 ± 0.0018	54.9 ± 1.6
7.3	7.28 - 7.32	53.01 - 53.57	0.0726 ± 0.0018	66.9 ± 1.7
7.4	7.38 - 7.42	54.45 - 55.07	0.0705 ± 0.0018	100.1 ± 2.1
7.5	7.48 - 7.52	56.01 - 56.49	0.0696 ± 0.0018	103.0 ± 2.1
7.6	7.59 - 7.61	57.55 - 57.97	0.0648 ± 0.0017	127.4 ± 2.3
7.7	7.68 - 7.72	59.02 - 59.56	0.0710 ± 0.0018	226.9 ± 3.2
7.8	7.79 - 7.81	60.62 - 61.06	0.0664 ± 0.0017	233.1 ± 3.2
7.9	7.89 - 7.91	62.19 - 62.63	0.0671 ± 0.0017	311 ± 4
8.0	7.99 - 8.01	63.81 - 64.19	0.0631 ± 0.0017	354 ± 4
8.1	8.09 - 8.11	65.42 - 65.80	0.0656 ± 0.0017	470 ± 5
8.2	8.19 - 8.21	67.05 - 67.43	0.0620 ± 0.0017	590 ± 5
8.3	8.29 - 8.31	68.74 - 69.04	0.0577 ± 0.0016	630 ± 5
8.4	8.39 - 8.41	70.39 - 70.73	0.0579 ± 0.0016	883 ± 6
8.5	8.49 - 8.51	72.09 - 72.41	0.0575 ± 0.0016	1062 ± 7
8.6	8.59 - 8.61	73.83 - 74.09	0.0514 ± 0.0015	1035 ± 7
8.7	8.69 - 8.71	75.56 - 75.82	0.0504 ± 0.0015	1143 ± 7

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table A.3: Recoil mass windows around each generated Z' mass corresponding to ± 2 times the standard deviation of the respective signal peak. For each window the signal efficiency and the total number of expected background events are shown for a integrated luminosity of 500 fb^{-1}.

mass [GeV/c^2]	window [$\mathrm{m_{rec}}$]	window [$\mathrm{m^2_{rec}}$]	signal efficiency	expected background events
0.1	−0.68 - 0.69	−0.46 - 0.48	0.0168 ± 0.0009	154 ± 8
0.2	−0.66 - 0.72	−0.43 - 0.51	0.0237 ± 0.0010	160 ± 8
0.3	−0.63 - 0.76	−0.40 - 0.58	0.0318 ± 0.0012	167 ± 9
0.4	−0.57 - 0.80	−0.32 - 0.64	0.0381 ± 0.0013	175 ± 9
0.5	−0.52 - 0.88	−0.27 - 0.77	0.0422 ± 0.0014	185 ± 9
0.6	−0.47 - 0.97	−0.22 - 0.94	0.0489 ± 0.0015	196 ± 9
0.7	−0.33 - 1.04	−0.11 - 1.09	0.0539 ± 0.0016	199 ± 9
0.8	0.25 - 1.10	0.06 - 1.22	0.0566 ± 0.0016	174 ± 9
0.9	0.47 - 1.18	0.22 - 1.40	0.0627 ± 0.0017	151 ± 8
1.0	0.65 - 1.25	0.43 - 1.57	0.0635 ± 0.0017	129 ± 8
1.1	0.79 - 1.34	0.62 - 1.80	0.0668 ± 0.0017	115 ± 7
1.2	0.90 - 1.44	0.80 - 2.08	0.0653 ± 0.0017	109 ± 7
1.3	1.01 - 1.54	1.01 - 2.37	0.0682 ± 0.0018	99 ± 7
1.4	1.16 - 1.60	1.35 - 2.57	0.0711 ± 0.0018	74 ± 6
1.5	1.28 - 1.69	1.64 - 2.86	0.0705 ± 0.0018	66 ± 5
1.6	1.39 - 1.78	1.94 - 3.18	0.0689 ± 0.0018	57 ± 5
1.7	1.51 - 1.87	2.28 - 3.50	0.0674 ± 0.0017	48 ± 5
1.8	1.62 - 1.96	2.62 - 3.86	0.0679 ± 0.0017	40 ± 4
1.9	1.73 - 2.05	3.00 - 4.22	0.0700 ± 0.0018	36 ± 4
2.0	1.84 - 2.15	3.38 - 4.62	0.0707 ± 0.0018	37 ± 4
2.1	1.95 - 2.24	3.80 - 5.02	0.0747 ± 0.0018	34 ± 4
2.2	2.06 - 2.33	4.25 - 5.43	0.0743 ± 0.0018	29.1 ± 3.5
2.3	2.16 - 2.43	4.68 - 5.90	0.0710 ± 0.0018	27.0 ± 3.4
2.4	2.27 - 2.52	5.16 - 6.36	0.0747 ± 0.0018	27.8 ± 3.5
2.5	2.38 - 2.62	5.65 - 6.85	0.0731 ± 0.0018	24.3 ± 3.2
2.6	2.49 - 2.71	6.18 - 7.34	0.0714 ± 0.0018	21.3 ± 3.0
2.7	2.58 - 2.81	6.67 - 7.91	0.0722 ± 0.0018	22.4 ± 3.1
2.8	2.70 - 2.90	7.27 - 8.41	0.0723 ± 0.0018	23.4 ± 3.1
2.9	2.79 - 3.00	7.80 - 9.02	0.0738 ± 0.0018	24.6 ± 3.2
3.0	2.91 - 3.09	8.46 - 9.54	0.0728 ± 0.0018	22.1 ± 3.0
3.1	3.01 - 3.19	9.07 - 10.15	0.0737 ± 0.0018	21.3 ± 3.0
3.2	3.11 - 3.29	9.68 - 10.80	0.0699 ± 0.0018	23.7 ± 3.1
3.3	3.21 - 3.39	10.30 - 11.48	0.0716 ± 0.0018	23.5 ± 3.1
3.4	3.32 - 3.48	11.00 - 12.12	0.0706 ± 0.0018	21.7 ± 2.9
3.5	3.42 - 3.58	11.69 - 12.81	0.0716 ± 0.0018	20.6 ± 2.9
3.6	3.52 - 3.68	12.38 - 13.54	0.0740 ± 0.0018	22.3 ± 3.0
3.7	3.62 - 3.78	13.10 - 14.28	0.0764 ± 0.0019	23.9 ± 3.2
3.8	3.73 - 3.87	13.90 - 14.98	0.0742 ± 0.0018	23.0 ± 3.1
3.9	3.83 - 3.97	14.68 - 15.74	0.0703 ± 0.0018	18.9 ± 2.8
4.0	3.93 - 4.07	15.48 - 16.52	0.0743 ± 0.0018	20.4 ± 3.0
4.1	4.03 - 4.17	16.27 - 17.35	0.0747 ± 0.0018	22.5 ± 3.1
4.2	4.14 - 4.26	17.12 - 18.16	0.0730 ± 0.0018	24.2 ± 3.2
4.3	4.24 - 4.36	18.00 - 18.98	0.0712 ± 0.0018	28.1 ± 3.5

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table A.4: Continuation of Table A.3.

mass [GeV/c^2]	window [m_{rec}]	window [m^2_{rec}]	signal efficiency	expected background events
4.4	4.34 - 4.46	18.81 - 19.91	0.0712 ± 0.0018	34 ± 4
4.5	4.44 - 4.56	19.75 - 20.75	0.0743 ± 0.0018	32 ± 4
4.6	4.55 - 4.65	20.68 - 21.64	0.0735 ± 0.0018	36 ± 4
4.7	4.65 - 4.75	21.58 - 22.60	0.0725 ± 0.0018	35 ± 4
4.8	4.75 - 4.85	22.52 - 23.56	0.0734 ± 0.0018	45 ± 4
4.9	4.85 - 4.95	23.50 - 24.52	0.0715 ± 0.0018	34 ± 4
5.0	4.95 - 5.05	24.51 - 25.49	0.0733 ± 0.0018	47 ± 5
5.1	5.05 - 5.15	25.54 - 26.48	0.0714 ± 0.0018	43 ± 8
5.2	5.16 - 5.24	26.60 - 27.48	0.0693 ± 0.0018	49 ± 5
5.3	5.25 - 5.35	27.58 - 28.60	0.0692 ± 0.0018	44 ± 4
5.4	5.36 - 5.44	28.69 - 29.63	0.0717 ± 0.0018	41 ± 4
5.5	5.46 - 5.54	29.78 - 30.72	0.0687 ± 0.0018	58 ± 5
5.6	5.56 - 5.64	30.92 - 31.80	0.0682 ± 0.0018	49 ± 5
5.7	5.66 - 5.74	32.04 - 32.94	0.0652 ± 0.0017	51 ± 5
5.8	5.76 - 5.84	33.19 - 34.09	0.0648 ± 0.0017	47 ± 4
5.9	5.86 - 5.94	34.33 - 35.29	0.0691 ± 0.0018	66 ± 5
6.0	5.97 - 6.03	35.61 - 36.39	0.0667 ± 0.0017	64 ± 5
6.1	6.07 - 6.13	36.82 - 37.60	0.0628 ± 0.0017	53 ± 5
6.2	6.17 - 6.23	38.02 - 38.86	0.0663 ± 0.0017	60 ± 5
6.3	6.27 - 6.33	39.32 - 40.06	0.0625 ± 0.0017	59 ± 5
6.4	6.37 - 6.43	40.59 - 41.33	0.0623 ± 0.0017	65 ± 5
6.5	6.47 - 6.53	41.90 - 42.60	0.0628 ± 0.0017	79 ± 6
6.6	6.58 - 6.62	43.23 - 43.89	0.0635 ± 0.0017	83 ± 6
6.7	6.67 - 6.73	44.54 - 45.24	0.0625 ± 0.0017	104 ± 7
6.8	6.78 - 6.82	45.95 - 46.53	0.0591 ± 0.0016	115 ± 7
6.9	6.88 - 6.92	47.28 - 47.94	0.0624 ± 0.0017	145 ± 8
7.0	6.98 - 7.02	48.68 - 49.32	0.0618 ± 0.0017	182 ± 9
7.1	7.08 - 7.12	50.12 - 50.70	0.0589 ± 0.0016	242 ± 10
7.2	7.18 - 7.22	51.55 - 52.13	0.0609 ± 0.0017	305 ± 12
7.3	7.28 - 7.32	53.01 - 53.57	0.0627 ± 0.0017	399 ± 14
7.4	7.38 - 7.42	54.45 - 55.07	0.0609 ± 0.0017	590 ± 16
7.5	7.48 - 7.52	56.01 - 56.49	0.0618 ± 0.0017	645 ± 17
7.6	7.59 - 7.61	57.55 - 57.97	0.0565 ± 0.0016	834 ± 19
7.7	7.68 - 7.72	59.02 - 59.56	0.0629 ± 0.0017	1588 ± 27
7.8	7.79 - 7.81	60.62 - 61.06	0.0598 ± 0.0016	1774 ± 29
7.9	7.89 - 7.91	62.19 - 62.63	0.0627 ± 0.0017	2510 ± 33
8.0	7.99 - 8.01	63.81 - 64.19	0.0600 ± 0.0016	$(3.05 \pm 0.04) \times 10^3$
8.1	8.09 - 8.11	65.42 - 65.80	0.0638 ± 0.0017	$(4.27 \pm 0.04) \times 10^3$
8.2	8.19 - 8.21	67.05 - 67.43	0.0621 ± 0.0017	$(5.76 \pm 0.05) \times 10^3$
8.3	8.29 - 8.31	68.74 - 69.04	0.0590 ± 0.0016	$(6.48 \pm 0.05) \times 10^3$
8.4	8.39 - 8.41	70.39 - 70.73	0.0613 ± 0.0017	$(9.53 \pm 0.06) \times 10^3$
8.5	8.49 - 8.51	72.09 - 72.41	0.0621 ± 0.0017	$(1.193 \pm 0.007) \times 10^4$
8.6	8.59 - 8.61	73.83 - 74.09	0.0556 ± 0.0016	$(1.210 \pm 0.007) \times 10^4$
8.7	8.69 - 8.71	75.56 - 75.82	0.0580 ± 0.0016	$(1.384 \pm 0.008) \times 10^4$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table A.5: The bin edges defined over the squared recoil mass to perform the binned ML fit. The bin widths were calculated in such a way that they correspond to the FWHM of the signal distribution, which resulted in a total of 196 bins. (Section 5.5.3)

-0.5	0.08	0.67	1.26	1.85	2.45	3.06	3.66	4.26
4.87	5.47	6.08	6.68	7.29	7.89	8.48	9.08	9.67
10.26	10.85	11.43	12.01	12.59	13.16	13.72	14.29	14.85
15.4	15.95	16.5	17.04	17.57	18.1	18.63	19.16	19.67
20.19	20.7	21.21	21.71	22.21	22.7	23.19	23.67	24.16
24.63	25.11	25.58	26.04	26.51	26.96	27.42	27.87	28.32
28.76	29.2	29.64	30.07	30.5	30.93	31.36	31.78	32.19
32.61	33.02	33.43	33.83	34.24	34.64	35.03	35.43	35.82
36.2	36.59	36.97	37.35	37.73	38.1	38.47	38.84	39.2
39.57	39.93	40.29	40.64	40.99	41.34	41.69	42.04	42.38
42.72	43.06	43.39	43.73	44.06	44.39	44.71	45.04	45.36
45.68	45.99	46.31	46.62	46.93	47.24	47.55	47.85	48.15
48.45	48.75	49.04	49.34	49.63	49.92	50.21	50.49	50.78
51.06	51.34	51.62	51.89	52.17	52.44	52.71	52.98	53.25
53.51	53.77	54.04	54.3	54.56	54.81	55.07	55.32	55.57
55.82	56.07	56.32	56.56	56.81	57.05	57.29	57.53	57.77
58.01	58.24	58.48	58.71	58.94	59.17	59.4	59.63	59.85
60.08	60.3	60.53	60.75	60.97	61.18	61.4	61.62	61.83
62.05	62.26	62.47	62.68	62.89	63.1	63.31	63.52	63.72
63.93	64.13	64.33	64.54	64.74	64.94	65.14	65.33	65.53
65.73	65.92	66.12	66.31	66.5	66.7	66.89	67.08	67.27
67.46	67.64	67.83	68.02	68.2	68.39	68.57	68.76	68.94

Table A.6: The mapping of the POI index to the assumed signal strength μ for different values of m. The respective μ was used for calculating the p-values of the signal like distribution in Figure 5.39a.

m \ idx	0	1	5	10	15	20	25	30
0.1	0.0	0.1	3.5	7.8	12.1	16.4	20.7	25.0
3.1	0.0	0.1	0.7	1.4	2.1	2.8	3.5	4.2
6.1	0.0	0.1	0.9	1.9	3.0	4.0	5.0	6.0

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table A.7: The mapping of the POI index to the assumed signal strength μ for different values of m. The respective μ was used for calculating the p-values of the background like distribution in Figure 5.39b.

m \ idx	0	1	5	10	15	20	25	30
0.1	0.0	0.1	3.5	7.8	12.1	16.4	20.7	25.0
0.5	0.0	0.1	1.9	4.2	6.5	8.8	11.1	13.4
1.0	0.0	0.1	1.1	2.4	3.7	5.0	6.3	7.6
1.5	0.0	0.1	0.8	1.8	2.7	3.6	4.6	5.5
2.0	0.0	0.1	0.7	1.4	2.1	2.8	3.6	4.3
2.5	0.0	0.1	0.6	1.2	1.9	2.5	3.1	3.8
3.0	0.0	0.1	0.6	1.3	1.9	2.6	3.2	3.9
3.5	0.0	0.1	0.6	1.3	2.0	2.7	3.3	4.0
4.0	0.0	0.1	0.7	1.4	2.1	2.8	3.5	4.2
4.5	0.0	0.1	0.7	1.4	2.2	2.9	3.7	4.4
5.0	0.0	0.1	0.7	1.5	2.3	3.1	3.9	4.6
5.5	0.0	0.1	0.8	1.6	2.4	3.3	4.1	5.0
6.0	0.0	0.1	0.9	1.8	2.8	3.7	4.7	5.6
6.5	0.0	0.1	1.0	2.0	3.1	4.2	5.3	6.3
7.0	0.0	0.1	1.4	3.0	4.7	6.3	8.0	9.6
7.5	0.0	0.1	2.4	5.3	8.2	11.1	13.9	16.8
8.0	0.0	0.1	3.1	6.9	10.7	14.5	18.2	22.0

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Appendix B

PID summary plots

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

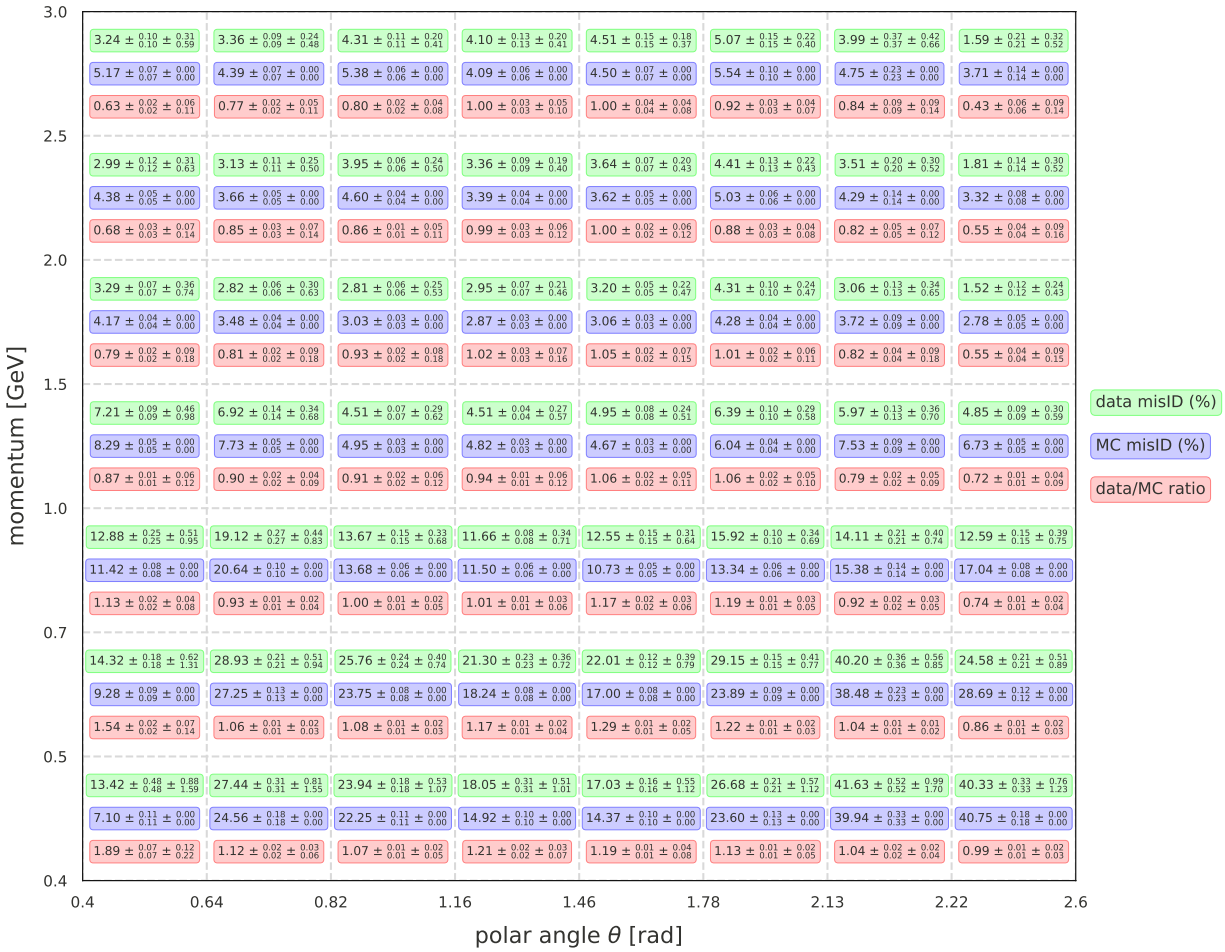


Figure B.1: A summary of the $\pi \to \mu$ mis-id rates for the dataset described in Chapter 4. The mis-id rates are shown in bins of momentum and polar angle, based on the binning scheme defined on page 50.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

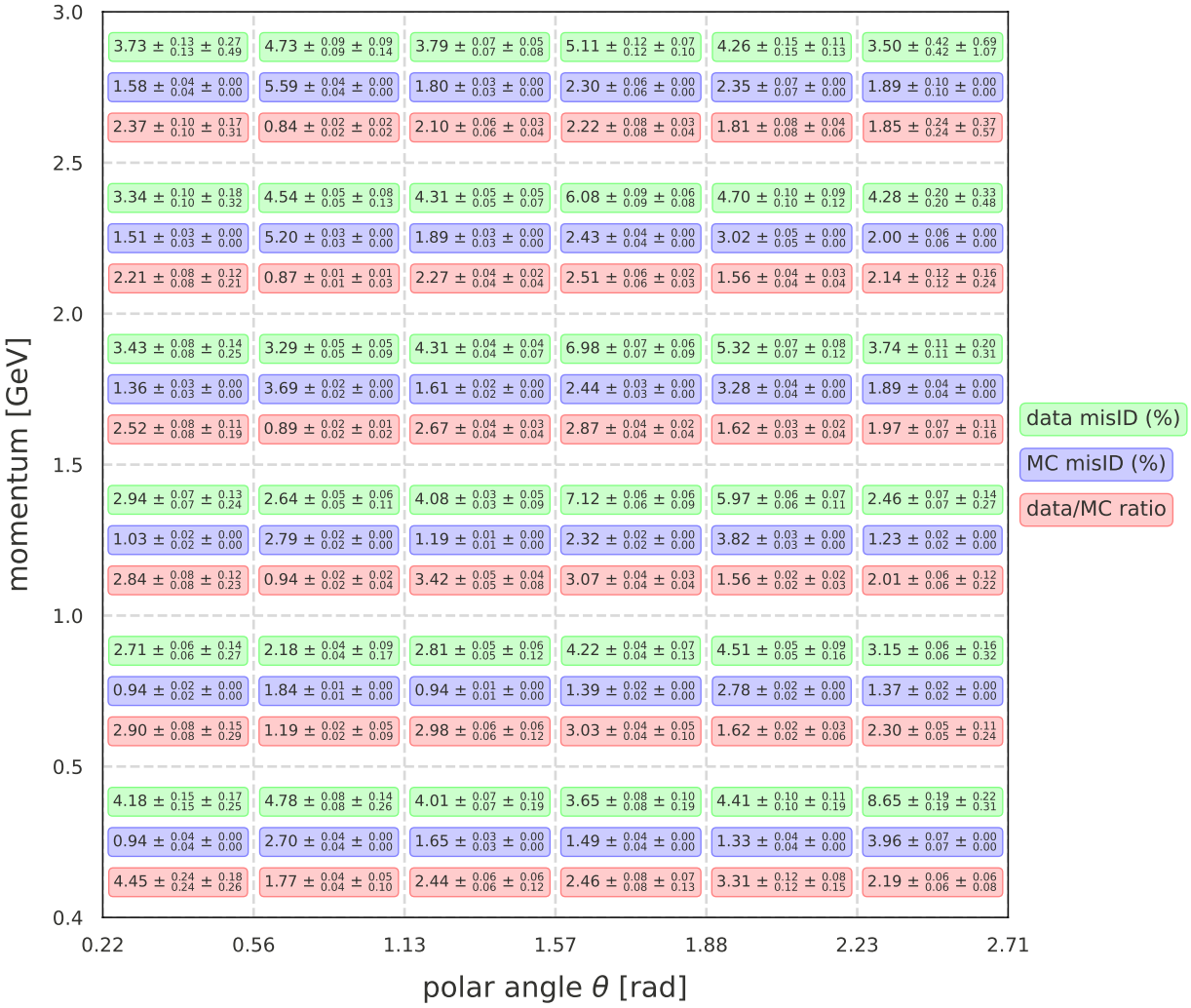


Figure B.2: A summary of the $\pi \to e$ mis-id rates for the dataset described in Chapter 4. The mis-id rates are shown in bins of momentum and polar angle, based on the binning scheme defined on page 50.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

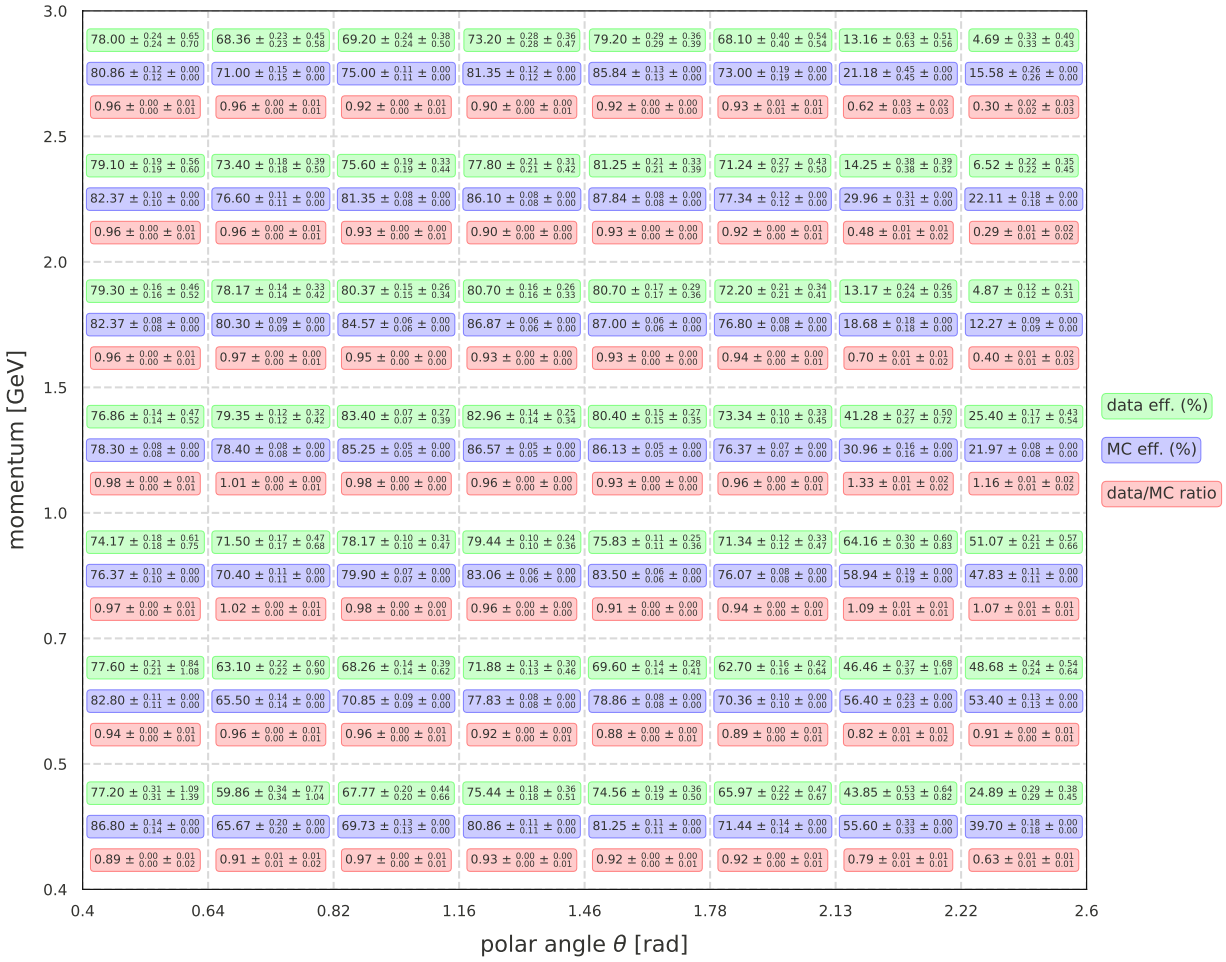


Figure B.3: A summary of the pion identification efficiencies for the dataset described in Chapter 4. The efficiencies are shown in bins of momentum and polar angle, based on the binning scheme defined on page 50.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN