TECHNISCHE UNIVERSITÄT WIEN

DISSERTATION

Test of light-lepton universality in tau decays with data from the Belle II experiment

zur Erlangung des akademischen Grades

Doktor der Technischen Wissenschaften

eingereicht von

Dipl.-Ing. Paul Feichtinger
Matrikelnummer 01425593

an der Fakultät für Physik der Technischen Universität Wien

in Zusammenarbeit mit dem Institut für Hochenergiephysik
der Österreichischen Akademie der Wissenschaften

Betreuung: **Privatdoz. Dipl.-Ing. Dr.techn. Christoph Schwanda**
E141 - Atominstitut, TU Wien

Mitbetreuung: **Univ.Lektor Dr. Gianluca Inguglia**

Wien, 5. Mai 2024

Paul Feichtinger

Christoph Schwanda

Kurzfassung

Diese Arbeit prüft die Universalität leichter Leptonen mit Hilfe des Belle II Experiments. Die verwendeten Daten stammen aus Elektron-Positron-Kollisionen des SuperKEKB-Beschleunigers in Tsukuba, Japan. Sie wurden bei einer Schwerpunktenergie von 10.58 GeV erfasst und entsprechen einer integrierten Luminosität von 362 fb^{-1}. Leptonen-Universalität ist ein grundlegendes Prinzip des Standardmodells der Teilchenphysik, welches besagt, dass die Kopplungsstärke der an der schwachen Wechselwirkung beteiligten Eichbosonen für alle Leptonen identisch ist. Dies ist nicht selbstverständlich, da zum Beispiel die Yukawa-Wechselwirkung mit dem Higgs-Feld unterschiedlich stark für die einzelnen Leptonen ist. Präzisionsmessungen im Flavour-Sektor könnten Unterschiede mit theoretischen Berechnungen aufzeigen, was wiederum Hinweise auf die Antwort fundamentaler Fragen geben kann, wie etwa, warum drei Generationen von Quarks und Leptonen existieren. Mit der Entdeckung von Neutrino-Oszillationen ist es bereits erwiesen, dass einige Eigenschaften des Standardmodells, die als wahr angenommen wurden, in der Natur nicht realisiert sind und mehr Erkenntnisse darüber notwendig sind. Diese Dissertation zielt darauf ab, die Universalität von Elektronen und Muonen in geladenen Strömen der schwachen Wechselwirkung zu überprüfen, indem das Verhältnis der Kopplungskonstanten g_μ/g_e gemessen wird. Dieser Quotient kann mit den Verzweigungsverhältnissen $\tau^\pm \to \mu^\pm\nu\bar{\nu}$ und $\tau^\pm \to e^\pm\nu\bar{\nu}$ berechnet werden. Diese Arbeit verwendet τ-Leptonen, die im Prozess $e^+e^- \to \tau^+\tau^-$ produziert wurden, wobei beide nur in ein geladenes Teilchen zerfallen. Jegliche signifikante Abweichung des Quotienten g_μ/g_e von 1 würde nahelegen, dass Leptonen-Universalität verletzt ist, zum Beispiel durch Wechselwirkungen mit bisher unbekannten Eichbosonen. Dies würde die bereits bestehenden Widersprüche im Flavour-Sektor erweitern, wie zum Beispiel die Hinweise auf Verletzung der Leptonen-Universalität in semileptonischen Zerfällen von B-Mesonen, die Unstimmigkeiten bezüglich des Cabibbowinkels oder des anomalen magnetischen Dipolmoments des Muons. Der in dieser Arbeit gemessene Wert von $g_\mu/g_e = 0.9974 \pm 0.0019$ ist in Übereinstimmung mit dem Standardmodell. Die Präzision ist limitiert durch systematische Unsicherheiten einhergehend mit der Identifikation von Leptonen. Diese Messung ist der genaueste Test der Leptonen-Universalität in Zerfällen von τ-Leptonen von einem einzelnen Experiment.

Abstract

This thesis investigates light-lepton universality with electron-positron collision data from the Belle II experiment at the SuperKEKB accelerator in Tsukuba, Japan. The data were recorded at a centre-of-mass energy of 10.58 GeV and are equivalent to an integrated luminosity of 362 fb^{-1}. Lepton universality is a property of the Standard Model of particle physics that implies that all leptons have equal couplings to the electroweak gauge bosons. Such an identity is not obvious; for example, the Yukawa couplings with the Higgs field differ between the leptons. Precision measurements in the flavour sector can potentially uncover discrepancies with theoretical predictions, which could help answer profound questions, such as why there are three generations of quarks and leptons. With the observation of large neutrino mixing, it has already become evident that some properties of the Standard Model that were thought to be true are violated in nature, and further insights are needed. This thesis aims to test the universality of electrons and muons in charged current interactions by measuring the gauge coupling ratio g_μ/g_e. This ratio can be obtained from the branching fractions of the decays $\tau^\pm \to \mu^\pm \nu\bar{\nu}$ and $\tau^\pm \to e^\pm \nu\bar{\nu}$. The approach for this work will be to measure these branching fractions using τ leptons produced in the process $e^+e^- \to \tau^+\tau^-$, with both decaying to only one charged particle. Any deviation of g_μ/g_e from unity would indicate that lepton universality is violated, for example, through interactions with new gauge bosons. This would complement existing anomalies in the flavour sector, such as indications of lepton universality violation in semileptonic B decays, the Cabibbo angle anomaly, or the inconsistencies related to the anomalous magnetic dipole moment of the muon. The result of this measurement is $g_\mu/g_e = 0.9974 \pm 0.0019$, which is consistent with the Standard Model expectation. Its precision is limited by systematic uncertainties related to lepton identification. At the moment, this measurement is the most precise test of lepton universality using tau decays obtained from a single experiment.

The best that most of us can hope to achieve in physics is simply to misunderstand at a deeper level.

WOLFGANG PAULI

Acknowledgements

I wish to express my gratitude to my supervisors, Gianluca and Christoph, for their consistent support, guidance, and encouragement throughout the course of my PhD. I especially thank Gianluca for his valuable counsel and for always envisioning a broader perspective. I appreciate the autonomy I was granted in my tasks, the endorsement to explore novel approaches, and the open discussions of new ideas. This welcoming environment helped me to grow both personally and professionally.

Special thanks also to all my colleagues and dear friends at HEPHY, who have made my time here a truly enriching experience. In particular, Huw, Philipp, Nadia, Géraldine, Daniel, Michel, and Rajesh for all the enlightening discussions, both those that were work-related and, more importantly, those that were not. It was a privilege to share an office with you, and your company during lunch breaks and after-work gatherings has been a constant source of joy.

My gratitude also extends to our τ LFU $|V_{us}|$ analysis team, which includes Gianluca, Ami, Petar, Géraldine, Zuzka, Michel, and Alberto. This joint effort between Vienna, Prague, and Hamburg has given rise to numerous productive concepts, and I appreciate the valuable insights, contributions, and suggestions. It was always a pleasure working with you. I owe particular thanks to Ami for thoroughly questioning many aspects of this work, which has contributed significantly to enhancing the quality of my research.

I am thankful to all Belle II colleagues who gave me advice, either directly through their insightful reviews and comments or indirectly by sharing their research and findings. This has greatly enriched my understanding and contributed to broadening my perspective. I particularly thank the tau working group for the numerous discussions that have been a cornerstone of my research. A special note of appreciation goes to the lepton ID group and specifically to Kenta-san and Philipp. Their prompt responses and dedicated assistance have been instrumental in preventing significant delays in this analysis.

I am also immensely grateful to the Belle II collaboration for fostering a conducive work environment characterised by cooperation, collegiality, professionalism, and infectious enthusiasm. I acknowledge the indispensable role of HEPHY in providing all the necessary infrastructure that has made this work possible. My sincere thanks also to the European Union for their financial support in the form of the ERC Starting Grant No. 947006.

Above all, I convey my deepest thanks to my family for their love, patience, and endless support: Danke, Mama, Papa und Susi, für euren steten Beistand und Vertrauen in mich. Eure Unterstützung während meines Studiums war essentiell, um meine Ziele zu erreichen. Ich bin auch zutiefst dankbar für die Zeit, die ich mit euch verbringen kann und die mir immer einen angenehmen Ausgleich zu meiner Arbeit bietet.

Finally, I thank Natasha for the precious moments we share together, which are of paramount importance to me. Your companionship has been invaluable, and I am grateful for your endurance in accepting my flaws and supporting me in times of despair. Thank you for consistently being a beacon of encouragement throughout this journey.

Contents

CHAPTER 1

Introduction

1.1 Overview

The discovery of the tau lepton in electron position collisions in 1975 [1–3] sparked the question of whether this newly found lepton was part of a new generation of leptons or whether it belonged to the electron or muon family [4]. It was soon concluded that the observations were incompatible with the tau lepton carrying a lepton number of the other generations, which meant that a new third generation of particles was discovered. One question which remains until today is whether the different lepton generations are entirely equal in terms of their fundamental interactions and whether the only differences we observe between them arise from their different masses. In the Standard Model (SM) of particle physics, this equivalence is a fundamental property known as lepton universality, or lepton flavour universality (LFU). Any deviation from this universality could indicate new physics beyond the SM. It has been tested experimentally through various processes, including charged current and flavour-changing neutral current transitions. While some anomalies in LFU measurements in the B sector [5–7] sparked some interest in recent years, otherwise no significant deviations from the SM expectation have been observed at the current level of precision, and high precision tests are needed to reach clarification.

Measurements with the leptonic decays of tau leptons allow for such high precision tests of LFU [8]. Testing the e-μ universality involves computing the ratio of the leptonic τ branching fractions,

$$R_\mu = \frac{\mathcal{B}(\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau)}{\mathcal{B}(\tau^- \to e^- \bar{\nu}_e \nu_\tau)}. \tag{1.1}$$

The ratio R_μ translates to the ratio of the couplings g_e and g_μ of the electron and muon to the charged gauge bosons $W^\pm$ of the electroweak interaction through the relation

$$\left|\frac{g_\mu}{g_e}\right|_\tau = \sqrt{R_\mu \frac{f(m_e^2/m_\tau^2)}{f(m_\mu^2/m_\tau^2)}}. \tag{1.2}$$

In this expression, m_e, m_μ, and m_τ represent the masses of the respective leptons and the function

$$f(x) = 1 - 8x + 8x^3 - x^4 - 12x^2 \ln x \tag{1.3}$$

stems from the phase space integral [9]. Therefore, as a consequence of the mass difference of the final state leptons, the ratio of branching fractions in the SM,

$$R_\mu^{\mathrm{SM}} = \frac{f(m_\mu^2/m_\tau^2)}{f(m_e^2/m_\tau^2)} \approx 0.9726, \tag{1.4}$$

deviates from unity, with values for the masses taken from the Particle Data Group (PDG) [10]. This prediction is theoretically very precise and is not limited by additional measurements of other quantities. Thus, the main challenge to test e-μ universality in tau decays is the experimental measurement of the leptonic branching fractions. Fortunately, one can measure $\mathcal{B}(\tau^- \to e^- \bar{\nu}_e \nu_\tau)$ and $\mathcal{B}(\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau)$ using the same experimental conditions, which results in a large cancellation of many systematic effects in their ratio. Direct measurements of R_μ have been performed by the CLEO [11] and *BABAR* [12] collaborations.

The CLEO collaboration measured R_μ with $e^+e^- \to \tau^+\tau^-$ (taupair) events in which both tau leptons decay to only one charged particle [11], which is referred to as the *1x1 topology*. The measurement used data corresponding to an integrated luminosity ($\mathcal{L}$) of $3.555\,\mathrm{fb}^{-1}$, taken at a centre-of-mass (CM) energy near 10.6 GeV. The result obtained by CLEO is

$$R_\mu = 0.9777 \pm 0.0063\,(\mathit{stat.}) \pm 0.0087\,(\mathit{sys.}), \qquad \text{(CLEO)} \tag{1.5}$$

where *stat.* indicates the statistical precision and *sys.* the uncertainty associated with the experimental systematics of the measurement. The CLEO collaboration did not only test e-μ universality but also e-τ and μ-τ universality and measured several absolute branching fractions of tau decays, as well as the mass of the tau lepton. The experimental measurement of the leptonic tau branching fractions and the tau mass, together with its lifetime, are essential inputs for the e-τ and μ-τ LFU tests, which currently limit their precision (see also Section 2.4).

A more precise measurement of R_μ was achieved by the *BABAR* collaboration using the *3x1 topology*, which refers to taupair events where one tau decays to three charged particles and the other to one charged particle [12]. *BABAR* used a data sample of $\mathcal{L} = 467\,\mathrm{fb}^{-1}$ at an e^+e^- CM energy near 10.58 GeV and obtained

$$R_\mu = 0.9796 \pm 0.0016\,(\mathit{stat.}) \pm 0.0036\,(\mathit{sys.}). \qquad \text{(\textit{BABAR})} \tag{1.6}$$

The total precision achieved by *BABAR* is 0.4 %, which propagates to 0.2 % precision on $|g_\mu/g_e|_\tau$. The systematic uncertainty associated with lepton identification at 0.32 % limits the precision of the measurement. This will also be the leading systematic uncertainty of this measurement with the Belle II experiment, which is similar to *BABAR* in many ways.

A global determination of R_μ, obtained from a fit from the Heavy Flavor Averaging Group (HFLAV) [13] yields

$$R_\mu = 0.9762 \pm 0.0028, \qquad \text{(HFLAV)} \tag{1.7}$$

which translates to the e-μ coupling ratio of

$$\left|\frac{g_\mu}{g_e}\right|_\tau = 1.0019 \pm 0.0014. \qquad \text{(HFLAV)} \tag{1.8}$$

The global fit uses 171 measurements to determine 125 quantities and is subject to 88 constraints. The result for $|g_\mu/g_e|_\tau$ is in accordance with the SM prediction of unity.

The only test of e-μ universality in charged currents that is more precise is from the PIENU collaboration using pion decays [14], which measures

$$\left|\frac{g_e}{g_\mu}\right|_\pi = 0.9996 \pm 0.0012. \qquad \text{(PIENU)} \tag{1.9}$$

More recently, measurements using $W^\pm$ decays have also become very precise. The latest result by the ATLAS collaboration [15] of $R_W^{\mu/e} = \mathcal{B}(W^\pm \to \mu^\pm \nu_\mu)/\mathcal{B}(W^\pm \to e^\pm \nu_e)$ is

$$R_W^{\mu/e} = 0.9995 \pm 0.0022\,(stat.) \pm 0.0036\,(sys.) \pm 0.0014\,(ext.), \tag{1.10}$$

where the last uncertainty indicates the uncertainty of an external measurement, which was used as a normalisation mode to reduce systematic effects. The total uncertainty of this measurement is 0.0045, corresponding to a precision on the ratio of the coupling constants of 0.225 %.

1.2 This work

This thesis will focus on measuring R_μ using the *1x1 topology* of taupair events, where one τ decays to an electron or a muon, while the other τ decays hadronically and to at least one π^0. This decay was chosen after estimating the sensitivity for different decay channels (see Section 4.3). Data for the measurement were obtained from the Belle II experiment, with e^+e^- collisions delivered by the SuperKEKB accelerator. The data were collected from 2019 to 2022 at a CM energy of 10.58 GeV, achieving an integrated luminosity of 362 fb^{-1}. This resulted in approximately 333×10^6 taupair events being produced.

The thesis starts with a general introduction to the underlying theoretical concepts and the motivation for the measurement, presented in Chapter 2. Then, in Chapter 3, the experimental setup is introduced, in particular the SuperKEKB accelerator and the Belle II experiment. Chapter 4 describes the general analysis procedure for this measurement, while in Chapter 5 several validation studies are discussed. Finally, the results are presented and interpreted in Chapter 6, and an outlook for future prospects is given.

Contributions by others It is important to keep in mind that this work uses Belle II data and was done as a member of the Belle II collaboration. In particular, it relies on the SuperKEKB collaboration for running the accelerator and the Belle II collaboration for the operation of the Belle II detector, as well as all the different aspects of data collection. In addition, many other tasks, such as performing Monte Carlo simulations, conducting performance studies, or maintaining the essential software and computing infrastructure, involve the hard work of many individuals who have dedicated their time and expertise to these duties. Although I have been involved in some of these areas, this analysis would not be possible without the collaborative effort of many people, which sometimes cannot be adequately cited. If some of the presented work is based on contributions made by others, this will be explicitly highlighted in the text.

Blind analysis This analysis is performed blind, meaning that the result is not obtained until the analysis procedure is carefully reviewed and fixed in place. Initially, the analysis

is set up only based on simulations, while later, data control samples are used to check the validity of the simulated samples or derive systematic uncertainties. Only after verifying all procedures is the result of the measurement *unblinded*, which is also done in several stages. This minimises the observer-expectancy effect, which involves subconsciously influencing the outcome based on observations and subjective expectations.

Natural unit convention In this work, the natural unit convention is used. Fundamental physical constants, such as the speed of light c and the reduced Planck constant $\hbar$, are set to unity. This convention simplifies calculations and allows to focus more on the underlying concepts. It implies that time and distance have the same units, which is the inverse of mass and energy, typically expressed in electronvolt.

Charge conjugates In some cases, the charge of the particles will be omitted from the respective symbols, or only one charge will be indicated. Unless explicitly stated otherwise, the charge conjugates are implied throughout this work.

CHAPTER 2

Theoretical overview

2.1 Introduction

In this chapter, an overview of the theoretical foundations important for this work is given. First, in Section 2.2, the Standard Model of particle physics is introduced, and Section 2.3 gives a general introduction to lepton universality. In Section 2.4, the tau lepton is discussed, in particular focusing on its leptonic decay rates and how they can be used to constrain lepton universality. Section 2.5 establishes some essential statistical methods that are commonly used in particle physics analyses. Finally, Section 2.6 gives a brief introduction to the basic principles of some computational methods used in this work, namely simulated annealing and artificial neural networks.

2.2 The Standard Model of particle physics

This section will give a brief overview of the Standard Model of particle physics. The SM is a very successful theory that gives some insight into how nature works at its most fundamental level. Underlying ideas, such as Lorentz and gauge symmetry principles, will be introduced, and particular emphasis will be given to concepts related to lepton universality. This is followed by a brief discussion on open questions in the SM, highlighting its shortcomings and some interesting features. This is not intended to be a comprehensive summary of the mathematical formulations which can be found elsewhere [16–20].

2.2.1 Lorentz invariance

Lorentz invariance is one of the most fundamental concepts that a theory of nature should fulfil. It states that the laws of nature do not change for different frames of reference, even at relativistic speeds. Any theory should yield the same results, independent of the frame in which it is calculated. The Schrödinger equation, for example, is not Lorentz invariant since it is first order in its time derivative and second order in the spatial derivatives. The corresponding relativistic wave equations of quantum mechanics are the Klein-Gordon and Dirac equations.

The Klein-Gordon equation reads

$$\left(\partial^\mu \partial_\mu + m^2\right) \psi = 0, \tag{2.1}$$

where $\psi(x,t)$ is the wavefunction and m is the mass of the particle it describes. The Dirac equation can be written as

$$\left(i\gamma^\mu \partial_\mu - m\right) \psi = 0, \tag{2.2}$$

where γ are the Dirac matrices which fulfil the anticommutation relations

$$\{\gamma^\mu, \gamma^\nu\} \equiv \gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = 2\eta^{\mu\nu}, \tag{2.3}$$

with $\eta^{\mu\nu}$ being the components of the Minkowski metric. The Dirac matrices generate a Clifford algebra, and every irreducible matrix representation of this algebra consists of 4×4 matrices. The elements of the vector space they act on are called spinors. A defining property of a spinor is that it changes sign after 2π rotations in space and only returns to its original state after a 4π rotation. The solutions of the Dirac equation already come with the concept of spin and antiparticles. It is often useful to use the adjoint spinor

$$\bar{\psi} \equiv \psi^\dagger \gamma^0. \tag{2.4}$$

With this definition, a four-vector (or four-current) can be defined as

$$j^\mu = \bar{\psi} \gamma^\mu \psi, \tag{2.5}$$

which satisfies the continuity equation $\partial_\mu j^\mu = 0$.

In the SM, these equations do not describe the dynamics of single-particle states, as the Schrödinger equation does, but rather quantum fields. This is necessary because the number of particles is not conserved anymore when relativistic particles are involved, since particle production and annihilation become meaningful. Particles are merely quantised excitations of the quantum fields, and interactions with other fields occur at quantised energies. The particles that make up our universe can be classified into fermions with half-odd integer spin and bosons with integer spin. Scalar fields with spin 0 are described by the Klein-Gordon equation. Spin-half fermions are the particles that constitute matter and are described by the Dirac equation. Spin-one particles are called vector bosons. They are the force carriers of fundamental interactions and are included in the Standard Model by requiring gauge invariance, which will be discussed in the next section.

To create Lorentz invariant quantities with fermion fields, they must appear in even numbers. This is because they carry an angular momentum of one-half, and Lorentz invariant quantities should have a total angular momentum of zero. Similarly to j^μ, which transforms as a four-vector, several other bilinear combinations of fermion fields with Lorentz transformation properties can be defined. All possible combinations can be built from linear combinations of the bilinear covariant currents listed in Table 2.1. The matrix γ^5 is a combination of the four Dirac matrices $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$. The bilinear fermion currents can be used to form Lorentz invariant quantities, together with scalar (spin-zero boson) or vector (spin-one boson) fields and derivatives, which also have Lorentz transformation properties [18]. For example, a spin-one boson field can only be combined with vector and axial vector currents.

Table 2.1: The most basic covariant currents formed from bilinear combinations of the fermion fields. They can be categorised by their Lorentz transformation properties.

Type	Form
Scalar	$\bar{\psi}\phi$
Pseudoscalar	$\bar{\psi}\gamma^5\phi$
Vector	$\bar{\psi}\gamma^\mu\phi$
Axial vector	$\bar{\psi}\gamma^\mu\gamma^5\phi$
Tensor	$\bar{\psi}\left(\gamma^\mu\gamma^\nu - \gamma^\nu\gamma^\mu\right)\phi$

Chirality An important property of spinors is called chirality. In the relativistic limit, it is equal to the helicity of particles, which is the normalised projection of its spin on its flight direction. Although helicity is not a Lorentz invariant concept, chirality is [21]. It is defined as the eigenstates of the γ^5 matrix, which are called right-handed and left-handed and denoted with subscripts R and L. In general, spinors have right-handed and left-handed components. With the projection operators

$$P_R = \frac{1}{2}\left(1+\gamma^5\right) \quad \text{and} \quad P_L = \frac{1}{2}\left(1-\gamma^5\right), \tag{2.6}$$

one of the components can be projected out. Thus, the combination of vector minus axial vector (V−A)

$$\bar{\psi}\gamma^\mu(1-\gamma^5)\phi = \bar{\psi}\gamma^\mu 2P_L\phi$$

is only affecting the left-handed component of the spinors. This is the form of the current involved in weak interactions, resulting in maximal parity violation, first observed by Wu et al. in beta decays [22].

2.2.2 Gauge invariance

Gauge symmetries are another important concept in the theoretical development of the Standard Model. It mandates that the Lagrangian remains invariant under local transformations from specific Lie groups. To guarantee this invariance, the derivatives appearing in the Lagrangian must be replaced by covariant derivatives

$$D_\mu \equiv \partial_\mu - igA_\mu^{\mathrm{C}}T^{\mathrm{C}},$$

where T are the generators of the Lie groups, each associated with a vector field A_μ, and g is a constant. The vector fields are called gauge fields, which generate the bosons that are the exchange particles for the fundamental forces. Thus, requiring gauge invariance results in all interactions between particles whose strength is given by g. In the SM, the fundamental interactions are the electromagnetic force transmitted through the photon, the strong force transmitted through gluons, and the weak force transmitted through the Z^0 and $W^\pm$ bosons.

The gluons arise from the requirement for local SU(3) gauge invariance of the so-called quark fields, which come in six distinct flavours. Every quark flavour consists of three fields with different colour charges, constituting a triplet that the SU(3) group acts on. Consequently, there are eight gluon fields, corresponding to the generators of the SU(3) group. Quarks only exist within composite states, such as mesons or baryons, that do not have a colour charge.

The electroweak sector of the SM arises from the requirement of invariance under $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ gauge transformations, first proposed by Glashow [23]. The SU(2) transformation acts on doublets of fields, which come in three generations for the quarks and three generations for the so-called leptons, which do not have a colour charge and thus do not participate in the strong interaction. For quarks, the three doublets are

$$\Omega_1 = \begin{bmatrix} u_L \\ d_L \end{bmatrix}, \quad \Omega_2 = \begin{bmatrix} c_L \\ s_L \end{bmatrix}, \quad \Omega_3 = \begin{bmatrix} t_L \\ b_L \end{bmatrix},$$

where the quark symbols use the first letter of the quark flavours (up, down, charm, strange, top, bottom). For leptons, the three generations consist of a charged lepton (the electron e, muon μ, and tau τ) and an associated neutral particle, the neutrino (ν) of the same flavour:

$$\Lambda_1 = \begin{bmatrix} (\nu_e)_L \\ e_L^- \end{bmatrix}, \quad \Lambda_2 = \begin{bmatrix} (\nu_\mu)_L \\ \mu_L^- \end{bmatrix}, \quad \Lambda_3 = \begin{bmatrix} (\nu_\tau)_L \\ \tau_L^- \end{bmatrix}.$$

Without mass, the three generations are entirely equivalent, and no fundamental principle explains why there are exactly three generations. The subscript L in the doublets indicates the chirality of the fermions. Only left-handed fermions participate in the weak interaction, which arises from the $\mathrm{SU}(2)_L$ transformations that mix the two components in each doublet. The charge related to the $\mathrm{SU}(2)_L$ gauge symmetry is the weak isospin T_3, with a value $+\frac{1}{2}$ for the upper entry in the doublets and $-\frac{1}{2}$ for the lower entry. Right-handed fermions form singlets under SU(2) and have a weak isospin 0. Requiring the $\mathrm{SU}(2)_L$ symmetry generates three new vector fields: $W_\mu^1, W_\mu^2, W_\mu^3$.

The $\mathrm{U}(1)_Y$ gauge transformation changes the phase of the fields according to a representation of the U(1) group. This representation is associated with a quantum number called the weak hypercharge Y. The vector field associated with the $\mathrm{U}(1)_Y$ symmetry is called B_μ.

The physical fields observed in nature are mixtures of the four fields associated with $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$. The fields associated with the charged W bosons have the form of ladder operators (they transform the field in the upper part of the doublet to the lower part and vice versa):

$$W_\mu^\pm = \left(W_\mu^1 \pm i W_\mu^2\right)/\sqrt{2}.$$

The W_μ^3 and the B_μ field instead mix together with the weak mixing angle θ_W to form the field of the Z^0 boson and the photon field A_μ:

$$\begin{bmatrix} Z_\mu^0 \\ A_\mu \end{bmatrix} \equiv \begin{bmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{bmatrix} \begin{bmatrix} W_\mu^3 \\ B_\mu \end{bmatrix}.$$

This mixing of fields can be explained by the Higgs mechanism [24–27]. It works by introducing another doublet of fields to the Standard Model, charged under weak isospin and hypercharge, only this time consisting of two complex scalar fields. The form of the potential of these fields can be chosen in such a way that, even though the theory as a whole is symmetric under $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$, the ground state of the system breaks the symmetry. In this process, called spontaneous symmetry breaking, one massive scalar field and three so-called Goldstone boson fields appear. The massive scalar field is the Higgs field, while the fields of the Goldstone bosons are unphysical and can be removed using the freedom of local gauge invariance, called

the unitary gauge. During this transformation, the gauge bosons responsible for the weak interaction become massive, whereas the photon remains massless even after the spontaneous symmetry breaking. From the four degrees of freedom of the two complex scalar fields initially added to the theory, three end up in longitudinal polarisation of the massive gauge bosons, while the fourth becomes the scalar Higgs field with a non-zero vacuum expectation value of 246 GeV everywhere in spacetime. The particle associated with this field is the Higgs boson H.

The photon can be seen as the gauge boson of a new symmetry $\mathrm{U}(1)_{\mathrm{EM}}$, which describes electromagnetism and acts on fields with a new quantum number Q. This is the electric charge, which can be derived from weak isospin and hypercharge as $Q = T_3 + Y/2$. The electric charge is conserved by all interactions, which also means that only overall electrically neutral terms are allowed in the Lagrangian. However, the weak isospin and hypercharge are not conserved by the interactions with the Higgs field. In other words, the Higgs mechanism breaks the electroweak $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ symmetry into $\mathrm{U}(1)_{\mathrm{EM}}$. The quantum numbers for charge, weak isospin, and weak hypercharge for all particles in the SM are shown in Table 2.2. The $W^{\pm}$ bosons only interact with particles charged under weak isospin (the left-handed particles in the isospin doublets), while the Z^0 boson interacts with all electrically charged particles *and* particles with non-zero weak isospin.

Table 2.2: The electric charge Q, weak isospin T_3 and weak hypercharge Y of all the particles in the SM. The right-handed neutrinos do not have any charge and it is unclear whether they exist.

	Particle symbols	Q	T_3	Y
left-handed leptons	$\nu_{\mathrm{e}}, \nu_\mu, \nu_\tau$	0	$+1/2$	-1
	$\mathrm{e}^-, \mu^-, \tau^-$	-1	$-1/2$	-1
left-handed quarks	$\mathrm{u}, \mathrm{c}, \mathrm{t}$	$+2/3$	$+1/2$	$+1/3$
	$\mathrm{d}, \mathrm{s}, \mathrm{b}$	$-1/3$	$-1/2$	$+1/3$
right-handed leptons	ν_R	0	0	0
	$\mathrm{e}_R^-, \mu_R^-, \tau_R^-$	-1	0	-2
right-handed quarks	$\mathrm{u}_R, \mathrm{c}_R, \mathrm{t}_R$	$+2/3$	0	$+4/3$
	$\mathrm{d}_R, \mathrm{s}_R, \mathrm{b}_R$	$-1/3$	0	$-2/3$
Weak bosons	$W^{\pm}$	± 1	± 1	0
	Z^0	0	0	0
Photon	γ	0	0	0
Gluon	g	0	0	0
Higgs boson	H	0	$-1/2$	$+1$

It is also interactions with the Higgs field that allows fermions to have mass in the SM. The strength of this interaction, quantified by the so-called Yukawa couplings, determines the masses. In fact, without the Higgs mechanism, none of the particles in the SM could have mass without breaking gauge invariance. As a result, the Higgs boson interacts with all massive particles.

In addition to Yukawa interactions, interactions with the Higgs field also allow for the mixing of the generations of quarks and leptons, which is described by the CKM and PMNS matrix. The mixing and Yukawa coupling parameters, as with all other couplings, are only determined

Table 2.3: Measured and estimated values for the fine structure constants of electromagnetic and strong forces, along with the weak mixing angle across various energy scales.

	$q^2 \approx 0$	m_τ	m_Z	m_W
α [29, 30]	$\frac{1}{137}$	$\frac{1}{134}$	$\frac{1}{128}$	
α_S [28]	$O(1)$	0.35	0.12	
$\sin^2\theta_W$ [10, p. 179]			0.23	0.22

by experiments. Measuring the CKM parameters and studying the resulting phenomena is an important task in flavour physics.

2.2.3 Coupling constants

The strength of particle interactions is determined by the coupling constants of the electromagnetic force (e), the strong force (g_S), and the weak force (g_W, g_Z). These constants are fundamental to our understanding of these interactions and play a crucial role in many areas of modern physics. Experimental results have demonstrated that the weak coupling is equal for all lepton flavours, a principle known as lepton flavour universality, which will be elaborated on further in the next section.

The couplings are commonly written as the fine-structure constants

$$\alpha_{EM} = \frac{e^2}{4\pi}, \quad \alpha_S = \frac{g_S^2}{4\pi}, \quad \text{and} \quad \alpha_W = \frac{g_W^2}{4\pi}.$$

Virtual processes can affect the value of the coupling, which we effectively observe in experiments (which differs from the underlying 'bare' coupling). This results in a dependency on the energy scale of the interactions, referred to as the 'running of the coupling'. Table 2.3 gives the values of the effective couplings for different energy scales [10, 28–30]. The couplings of the weak force are related to the electromagnetic coupling through the weak mixing angle by

$$g_W = \frac{e}{\sin\theta_W} = g_Z \cos\theta_W, \tag{2.7}$$

which results in $\alpha_W \approx \frac{1}{30}$.

For couplings that are significantly smaller than one, perturbation theory can be used to perform practical calculations. This works by expanding the interactions in powers of g with so-called Feynman diagrams. The diagrams can be categorised according to the order in which the coupling appears, corresponding to the number of vertices in the diagrams. Usually, considering only the lowest-order diagrams already results in reasonable estimates of the calculated quantities. The fundamental building blocks of Feynman diagrams are shown in Figure 2.1. This also highlights all the possible interactions between fermions and bosons in the SM. For large couplings, as is the case for the strong interaction and low energies, non-perturbative methods, such as calculations on a discrete lattice, must be used.

Because weak and strong gauge bosons are self-interacting, the corresponding couplings decrease with higher energy. On the other hand, in abelian gauge theories such as quantum

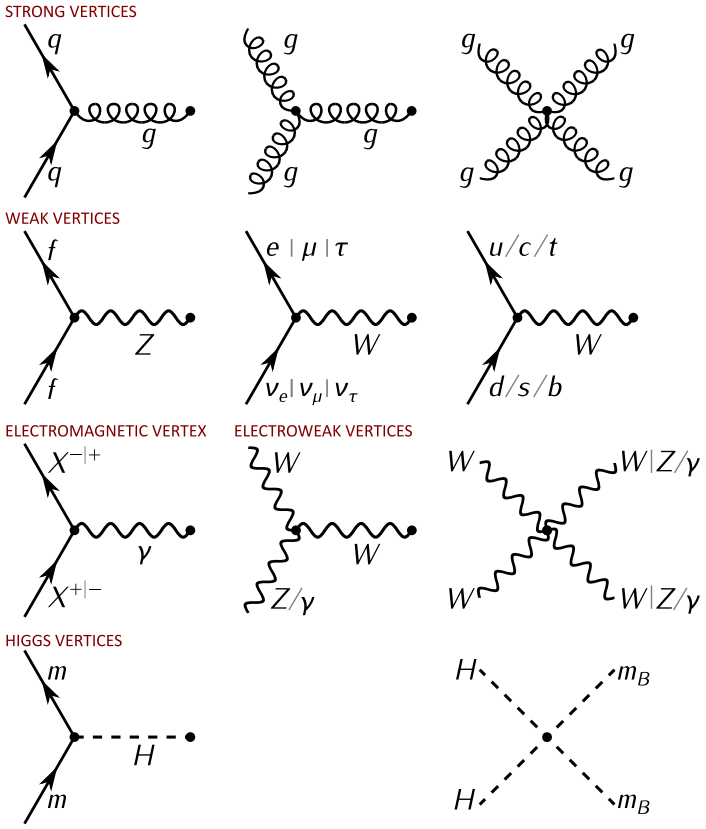


Figure 2.1: All possible vertices in the SM, indicating the allowed interactions. Here, q denotes quarks, g denotes gluons, f denotes fermions, X denotes charged particles, m denotes massive particles, and m_B denotes massive bosons. When particles are listed with /, they can be used interchangeably, whereas when separated with |, they can only be used with the corresponding partner. Taken from Wikimedia Commons: Standard Model - All Feynman diagram vertices from Ponor, licensed under CC BY-SA 4.0 [31], based on work by Lindon [32].

electrodynamics (QED), self-interactions are absent, leading to an increase in the effective coupling at higher energies. This suggests a convergence of all couplings at some higher energy scale, as indicated in Figure 2.2. In the SM, this unification does not emerge, but in several extensions, such as supersymmetry, the couplings would converge in a single point at high energies.

2.2.4 Open questions

The SM of particle physics is a highly successful theory, and many of its properties can be derived from a few fundamental concepts. However, there are still many parameters and properties that cannot be derived from first principles. Without neutrino masses, it has 19 free parameters (26 with neutrino masses), which must be obtained experimentally. Moreover, it contains 37 quantum fields, which exhibit seemingly arbitrary features, such as the three generations of fermions or the form of the potential for the Higgs field.

Each particle type in the SM and their masses are shown in Figure 2.3. Fermions can be categorised into quarks and leptons. Additionally, each particle is paired with an antiparticle

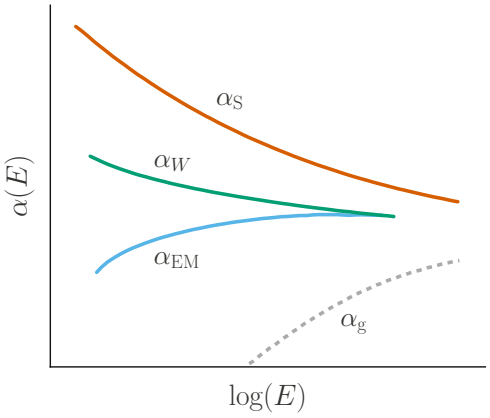


Figure 2.2: An illustration depicting the energy dependence of coupling constants for strong, weak, and electromagnetic interactions. Although not the case in the SM, in some theoretical models the interactions unify at a certain energy. A hypothetical increase in gravitational interaction is shown, suggesting that it could also be included in a *theory of everything*.

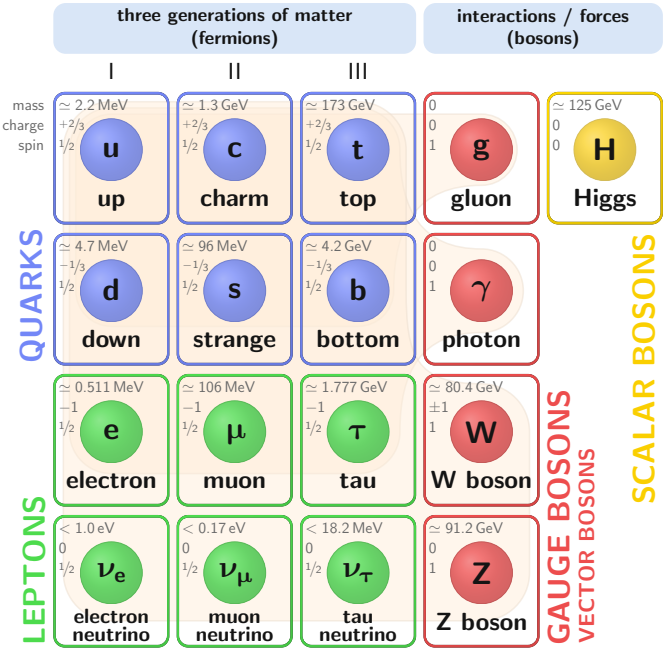


Figure 2.3: The Standard Model of particle physics: SM particles from Izaak Neutelings, licensed under CC BY-SA 4.0 [33]

that has identical mass but opposite charge. One interesting feature that the parameters exhibit, which is not explained by the SM, is that the masses of fermions increase with each generation. This can be clearly seen from Figure 2.4. This suggests that there may be an underlying reason for this pattern that explains why the generations seem to behave the same in every aspect not related to their mass.

There are also multiple other reasons why the SM cannot be the complete picture. The SM does not provide a promising dark matter candidate and falls short in describing gravity.

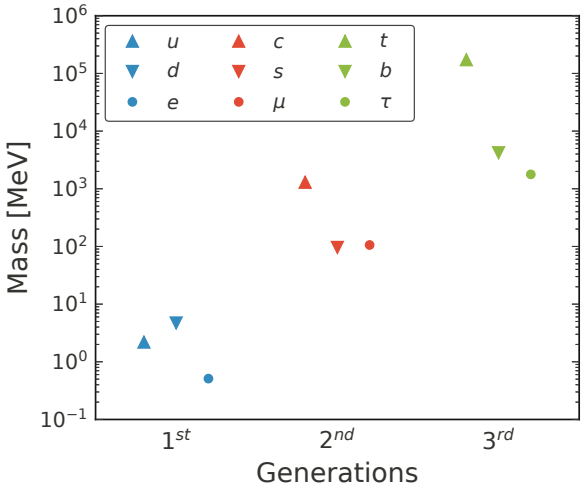


Figure 2.4: The masses of the charged fermions in the three generations of matter.

Moreover, despite the inclusion of CP violation in the SM, it fails to adequately account for the observed matter-antimatter asymmetry in the universe [34]. Furthermore, SM neutrinos have no mass, while observing neutrino oscillations implies a finite mass for at least two neutrinos. The mechanism of how these neutrinos acquire mass and why their mass differs significantly from that of other leptons remains unsolved. Several approaches have been proposed to expand the SM to address these questions, but none have succeeded so far.

2.3 Lepton universality

The concept of lepton universality is a basic principle in particle physics that plays a crucial role in our understanding of fundamental interactions between particles. It states that the strength of the weak coupling is the same for all lepton flavours. In other words, the interactions of the different leptons are identical, and in the SM, there is no fundamental difference between the lepton flavours. The parameter that describes the strength of the interactions of leptons through the weak force is the weak coupling g_W, which is related to the Fermi constant

$$G_F = \frac{g_W^2}{4\sqrt{2}M_W^2}. \tag{2.8}$$

The value of this coupling can be measured experimentally using different approaches, all of which should yield the same result if LFU holds. In practice, this is done by measuring the ratio of couplings of the different leptons to the $W^\pm$ boson, of which there are three possibilities.

Lepton universality has been experimentally tested and confirmed with a high degree of precision in many experiments during the last decades. These tests have been conducted using ratios of various processes, utilising pions [14], kaons [35, 36], B mesons [37–40], on-shell W bosons [15, 41, 42], and τ leptons [11, 12]. A summary of the most accurate experimental measurements of the coupling ratios with different decays can be found in [8, p. 45].

Interestingly, recent measurements of LFU using B meson decays show some tensions with

respect to the theoretical predictions. In particular, the ratio

$$R\left(D^{(*)}\right) = \mathcal{B}\left(\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau\right)/\mathcal{B}\left(\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell\right) \tag{2.9}$$

with $\ell = e$ or μ, was measured by several experiments to be incompatible with the SM expectation [5, 6]. Other anomalies in the B sector were observed in $b \to s\ell\ell$ transitions, in particular when comparing $B^+ \to K^+\mu^+\mu^-$ and $B^+ \to K^+e^+e^-$ [7]. A more recent measurement by LHCb [40], which supersedes the previous measurements, agrees with the SM, indicating that the source of the discrepancies is of a different nature. Another long-standing discrepancy between theory prediction and experiment concerns the anomalous magnetic moment of the muon. The most recent data from the Fermilab Muon $g-2$ experiment [43] indicate a $5\,\sigma$ tension with a theory prediction that uses data-driven calculations of hadronic vacuum polarization (HVP) [44]. However, more recent predictions of the anomalous magnetic moment based on lattice QCD calculations [45, 46] of HVP are in agreement with the measurements, and the reason for the discrepancies still has to be understood.

These results have generated significant interest [47–52], and ongoing research is being done to verify their validity and understand the implications of any possible deviations from lepton universality. A broad class of SM extensions explains the mechanism that leads to the violation of the universality of leptons. These extensions often assume the existence of new interactions or particles that couple to the leptons differently. Examples are two-Higgs-doublet models [53, 54] or a singly charged-scalar singlet [55]. More precise measurements of LFU are needed to test and exclude these models. Tests of lepton universality via $W^\pm$ boson decays in LHC experiments are limited to detecting charged currents, whereas decays of the τ lepton also are sensitive to flavour-changing neutral currents [56, 57].

2.4 The tau lepton

Tau leptons are fundamental particles and belong to the third generation of leptons. Being leptons, they do not interact via the strong interaction and only decay via the weak interaction, which makes them useful to probe many properties of the SM [58] and to search for new physics beyond it. They are the heaviest of the three charged leptons, with a mass of $1776.86 \pm 0.12\,\text{MeV}$ [10], which allows them to decay into not only electrons and muons, but also hadrons. The tau has a relatively short lifetime ($\tau_\tau = 290\,\text{fs}$) compared to other leptons ($\tau_\mu = 2.2\,\mu\text{s}$), which makes them almost decay immediately, and usually only their decay products can be detected. In Figure 2.5, the most common tau decays and their relative fraction are shown on a pie chart, where h denotes a hadron, which is either a pion or a kaon. The decays to pions dominate over kaons since the latter is Cabibbo-suppressed, by a factor of about $\sin^2\theta_c/\cos^2\theta_c \approx 0.053$ with $\theta_c = 13.02°$ being the Cabbibo angle. The hadronic decays of tau leptons are of great interest since they also provide a unique way to study the strong interaction responsible for binding quarks together to form hadrons [4, 8]. The leptonic decays of the tau can instead be used to test the Lorentz structure of the charged current in the theory of weak interaction. The observable parameters are called Michel parameters, which are bilinear combinations of coupling constants that arise in the most general expression for the decay matrix element [4]. In the SM, the only non-zero coupling due to the V−A nature of the weak interaction is the left-handed vector coupling g^V_{LL}. The leptonic decays of the tau make up about 35.2 % of all decays. In Figure 2.6, Feynman diagrams are shown for a generic hadronic decay (2.6a) and a leptonic decay (2.6b), where ℓ denotes either an electron or a muon.

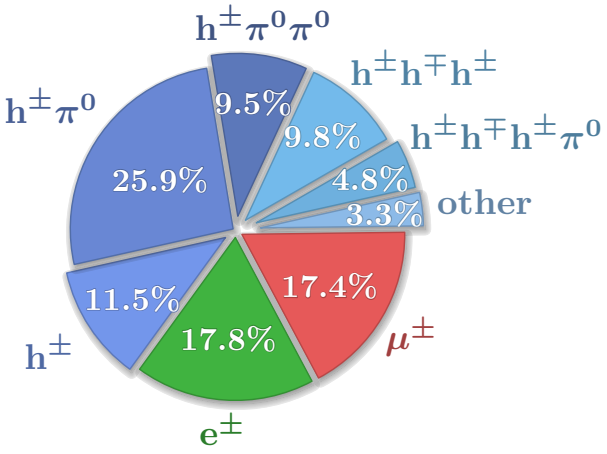
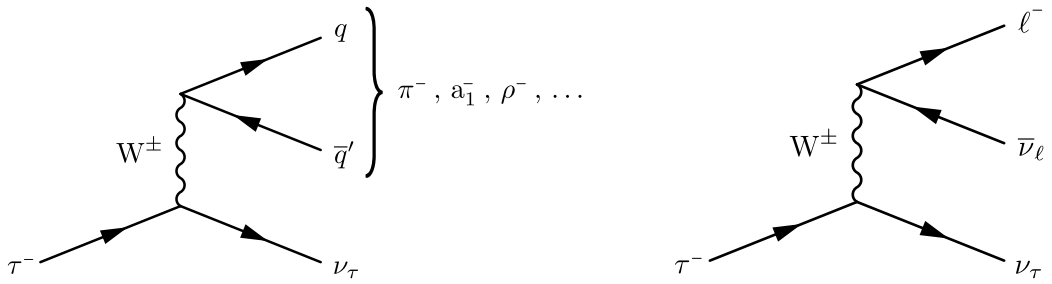


Figure 2.5: A pie chart of tau branching fractions: Piechart of SM decays from Izaak Neutelings, licensed under CC BY-SA 4.0 [59]



(a) (b)

Figure 2.6: Feynman diagrams for a hadronic decay (a) and a leptonic decay (b) of the tau lepton.

2.4.1 Leptonic decay rates

The decay rates of the tau lepton to an electron or muon can be computed from the Feynman diagram shown in Figure 2.6b. This process can be treated as an effective theory in which the interactions include four fermions. This is known as the four-fermion interaction or Fermi theory, and the corresponding process is shown in Figure 2.7. Neglecting the $W^{\pm}$ propagator is justified when the involved masses are much below the $W^{\pm}$ mass. In this case, the impact on the decay width is about 0.03 %, for which a correction term will be added later.

The Lagrangian for this decay, assuming a pure V−A interaction, is

$$\begin{aligned}\mathcal{L} &= \frac{G_F}{\sqrt{2}}\left(\bar{\nu}_\tau\gamma^\alpha\left(1-\gamma^5\right)\tau\right)\left(\bar{\ell}\gamma_\alpha\left(1-\gamma^5\right)\nu_\ell\right) \\ &= \frac{G_F}{\sqrt{2}}\left(\bar{\ell}\gamma^\alpha\left(1-\gamma^5\right)\tau\right)\left(\bar{\nu}_\tau\gamma_\alpha\left(1-\gamma^5\right)\nu_\ell\right),\end{aligned} \tag{2.10}$$

where G_F is the Fermi constant defined in Equation 2.8, γ^α and γ^5 are the Dirac gamma matrices introduces in Section 2.2, and ν_τ, τ, ℓ and ν_ℓ indicate the spinors of the corresponding

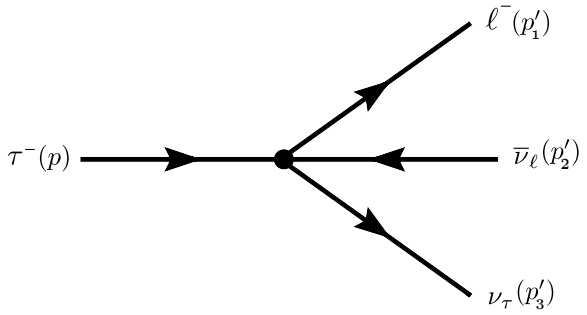


Figure 2.7: A diagram showing a leptonic decay of the tau within the Fermi theory.

particles with the adjoint spinors indicated by a bar as defined in Equation 2.4. To compute the decay rates, the phasespace integral [4.11 18, p. 95]

$$\Gamma = \frac{1}{2E}\left(\prod_a \int \frac{d^3p_a'}{(2\pi)^3 2E_a'}\right)(2\pi)^4\delta^4\left(p - \sum_a p_a'\right)|\mathcal{M}|^2 \tag{2.11}$$

has to be evaluated, with the momenta p_a and energies E_a corresponding to the particles as shown in Figure 2.7, and the matrix element $\mathcal{M}$. The spin-averaged matrix element is

$$\begin{aligned}\sum_{\text{spins}}|\mathcal{M}|^2 = \frac{G_F^2}{2}\,\mathrm{Tr}\left[\left(\not{p} + m_\tau\right)\gamma^\alpha\left(1-\gamma^5\right)\not{p}_3'\gamma^\beta\left(1-\gamma^5\right)\right]\times \\ \mathrm{Tr}\left[\not{p}_2'\gamma_\alpha\left(1-\gamma^5\right)\left(\not{p}_1' + m_\ell\right)\gamma_\beta\left(1-\gamma^5\right)\right],\end{aligned} \tag{2.12}$$

with, m_ℓ and m_τ being the masses of the corresponding leptons.

This results in the partial decay width

$$\frac{\mathcal{B}(\tau\to\ell\nu_\ell\nu_\tau)}{\tau_\tau} = \Gamma(\tau\to\ell\nu_\ell\nu_\tau) = \frac{G_F^2 m_\tau^5}{192\pi^3}f(m_\ell^2/m_\tau^2), \tag{2.13}$$

where τ_τ is the lifetimes of τ lepton,

$$f(x) = 1 - 8x + 8x^3 - x^4 - 12x^2\log x, \tag{2.14}$$

is due to the phase space integral, which was first derived by Tsai [9] assuming massless neutrinos. The effect of non-zero neutrino masses on this prediction is negligible at the current level of precision [60]. This decay width still has to be corrected to account for radiative effects, with a correction factor

$$R_\gamma^\tau = 1 + \frac{\alpha\,(m_\tau)}{2\pi}\left(\frac{25}{4} - \pi^2 + \mathcal{O}\left(\frac{m_\ell^2}{m_\tau^2}\right)\right) + \cdots, \tag{2.15}$$

which does not depend on the final state lepton up to order m_ℓ^2/m_τ^2 [3.2.2 61].

The correction due to the nonlocal structure of the W boson propagator [62–64] is

$$R_W^{\tau\ell} = 1 + \frac{3}{5}\frac{m_\tau^2}{M_W^2} + \frac{9}{5}\frac{m_\ell^2}{M_W^2} + \mathcal{O}\left(\frac{m_\ell^4}{M_W^2 m_\tau^2}\right), \tag{2.16}$$

which can be reabsorbed in the effective coupling together with higher-order electroweak corrections (Δr) [8, p. 43]:

$$G_{\ell\tau}^2 = \left[\frac{g^2}{4\sqrt{2}M_W^2}(1+\Delta r)\right]^2 \left[1+\frac{3}{5}\frac{m_\tau^2}{M_W^2}+\frac{9}{5}\frac{m_\ell^2}{M_W^2}+\mathcal{O}\left(\frac{m_\ell^4}{M_W^2 m_\tau^2}\right)\right]. \tag{2.17}$$

With this definition, $G_{e\mu}$ is the Fermi coupling defined in the V−A theory, which is measured to a very high degree of precision using the lifetime of the muon and the masses of the electron and the muon [8]:

$$G_{e\mu} = 1.166\,378\,7 \pm 0.000\,000\,6 \times 10^{-5}\,\mathrm{GeV}^{-2}. \tag{2.18}$$

The size of the radiative correction from Equation 2.15 and due to the nonlocal structure of the W boson propagator from Equation 2.16 is shown in Table 2.4 for the decay width of $\tau \to e\nu_e\nu_\tau$ and $\tau \to \mu\nu_\mu\nu_\tau$, when computed with Equation 2.13. The value quoted in the table for radiative corrections was computed by Marciano and Sirlin [65]. The radiative correction is sizeable but only depends on the lepton in the final state at higher orders. The correction associated with the W boson propagator [62] is much smaller and differs only slightly for $\tau \to e\nu_e\nu_\tau$ and $\tau \to \mu\nu_\mu\nu_\tau$.

Table 2.4: The size of the correction factors to the decay width of $\tau \to e\nu_e\nu_\tau$ and $\tau \to \mu\nu_\mu\nu_\tau$ for the QED radiative corrections and the correction due to neglecting the W boson propagator.

	$\tau \to e\nu_e\nu_\tau$	$\tau \to \mu\nu_\mu\nu_\tau$
QED radiative corrections [65]	−0.432 %	−0.432 %
$W^\pm$ boson propagator [62]	+0.0293 %	+0.0296 %

2.4.2 Lepton universality tests with tau decays

The leptonic decays of tau leptons allow for high precision tests of lepton universality, which states that the three charged leptons couple to the W bosons in the same way (see Section 2.3). Until now, LFU was implicitly assumed by having the couplings g_e, g_μ, and g_τ equal. When this assumption is not used, the theoretical prediction for the partial decay widths is

$$\frac{\mathcal{B}(\tau \to \ell\nu_\ell\nu_\tau)}{\tau_\tau} = \frac{g_\tau^2 g_\ell^2 m_\tau^5}{6144\pi^3 M_W^4} f(m_\ell^2/m_\tau^2) R_\gamma^\tau R_W^{\tau\ell}. \tag{2.19}$$

Using these partial decay widths for the leptonic tau decays and the analogous expressions for the decay $\mu \to e\nu_e\nu_\mu$, three ratios of the couplings can be obtained:

$$\left(\frac{g_\mu}{g_e}\right)^2 = \frac{\mathcal{B}(\tau \to \mu\nu_\mu\nu_\tau)}{\mathcal{B}(\tau \to e\nu_e\nu_\tau)} \frac{f(m_e^2/m_\tau^2)}{f(m_\mu^2/m_\tau^2)} \frac{R_W^{\mu e}}{R_W^{\tau\mu}}, \tag{2.20}$$

$$\left(\frac{g_\tau}{g_\mu}\right)^2 = \frac{\mathcal{B}(\tau \to e\nu_e\nu_\tau)}{\mathcal{B}(\mu \to e\nu_e\nu_\mu)} \frac{f(m_e^2/m_\mu^2)}{f(m_e^2/m_\tau^2)} \frac{\tau_\mu m_\mu^5}{\tau_\tau m_\tau^5} \frac{R_\gamma^\mu R_W^{\mu e}}{R_\gamma^\tau R_W^{\tau e}}, \tag{2.21}$$

and

$$\left(\frac{g_\tau}{g_e}\right)^2 = \frac{\mathcal{B}(\tau \to \mu\nu_\mu\nu_\tau)}{\mathcal{B}(\mu \to e\nu_e\nu_\mu)} \frac{f(m_e^2/m_\mu^2)}{f(m_\mu^2/m_\tau^2)} \frac{\tau_\mu m_\mu^5}{\tau_\tau m_\tau^5} \frac{R_\gamma^\mu R_W^{\mu e}}{R_\gamma^\tau R_W^{\tau\mu}}. \tag{2.22}$$

Table 2.5: The world averaged values for some basic parameters needed for tests of LFU, taken from the PDG (2022) [10].

m_τ	$1776.86 \pm 0.12\,\text{MeV}$
τ_τ	$290.3 \pm 0.5\,\text{fs}$
$\mathcal{B}(\tau \to e\nu_e\nu_\tau)$	$17.82 \pm 0.04\,\%$
$\mathcal{B}(\tau \to \mu\nu_\mu\nu_\tau)$	$17.39 \pm 0.04\,\%$
m_μ	$105.658\,375\,5 \pm 0.000\,002\,3\,\text{MeV}$
τ_μ	$2.196\,981\,1 \pm 0.000\,002\,2\,\mu\text{s}$
$\mathcal{B}(\mu \to e\nu_e\nu_\mu)$	$100\,\%$
m_e	$0.510\,998\,950\,00 \pm 0.000\,000\,000\,15\,\text{MeV}$
m_W	$80.377 \pm 0.012\,\text{GeV}$

For Equation 2.20, the last term that originates from the difference in corrections associated with the $W^\pm$ boson propagator can be neglected. Using the numbers given in Table 2.4, the ratio $\frac{R_W^{\mu e}}{R_W^{\tau\mu}}$ deviates from unity less than 3×10^{-6}. This results in the SM prediction for the ratio of the leptonic tau branching fractions of $0.972\,564 \pm 0.000\,004$, using the current best world averaged values for the particle masses from the PDG (2022), which are also shown in Table 2.5 [10]. This ratio g_μ/g_e, which tests e-μ universality, has the advantage that it can be obtained experimentally using only $\mathcal{B}(\tau \to \mu\nu_\mu\nu_\tau)$ and $\mathcal{B}(\tau \to e\nu_e\nu_\tau)$ without additional input. Moreover, both branching fractions can be measured using similar conditions in the same experiment, which results in a significant cancellation of experimental systematics. This ratio will be measured in this work with Belle II data.

Equations 2.21 and Equation 2.22 can instead be used to test μ-τ and e-τ universality. Essential inputs for these tests are the absolute branching fractions of the τ decaying to lighter leptons, the measurements of the τ mass, and its lifetime. They also rely on the same quantities for the muon and the mass of the $W^\pm$ boson, but they are known much more precisely. These quantities as listed in the PDG (2022) are shown in Table 2.5 [10]. The parameters associated with the tau lepton are expected to be improved by Belle II. For the tau mass, there is already a new result from Belle II, published in 2023 [66]. Using Equation 2.21 or Equation 2.22, a linear relation between the τ lifetime and the leptonic branching ratios of the tau lepton can be obtained [8, p. 44]. With the SM prediction of $g_\tau = g_\mu = g_e$ and inserting values from Table 2.5, this results in

$$\mathcal{B}(\tau \to e\nu_e\nu_\tau) = \frac{\mathcal{B}(\tau \to \mu\nu_\mu\nu_\tau)}{0.972564 \pm 0.000004} = \frac{\tau_\tau}{1632.76 \pm 0.55\,\text{fs}}. \tag{2.23}$$

This relation is also shown in Figure 2.8 as the blue band, with the red markers indicating the measured values of the τ lifetime and $\mathcal{B}(\tau \to e\nu_e\nu_\tau)$ as well as $\mathcal{B}(\tau \to \mu\nu_\mu\nu_\tau)$ from Table 2.5. The value of $\mathcal{B}(\tau \to \mu\nu_\mu\nu_\tau)$ was translated into $\mathcal{B}(\tau \to e\nu_e\nu_\tau)$ using the SM prediction from Equation 2.23. This figure combines all LFU tests that use the leptonic tau branching fractions in one image. The two markers agreeing with each other corresponds to the test of e-μ universality; however, one has to keep in mind that the shown uncertainties of $\mathcal{B}(\tau \to e\nu_e\nu_\tau)$ and $\mathcal{B}(\tau \to \mu\nu_\mu\nu_\tau)$ are correlated with each other. For the shown values from the fit quoted by the PDG, the correlation coefficient is 18 % [10]. The agreement of the individual markers with the blue line indicates e-τ (light red) and μ-τ (dark red) universality. Currently, all measured

values agree with each other under the SM assumptions, but more precise determinations of the individual inputs are needed for a more powerful test of LFU.

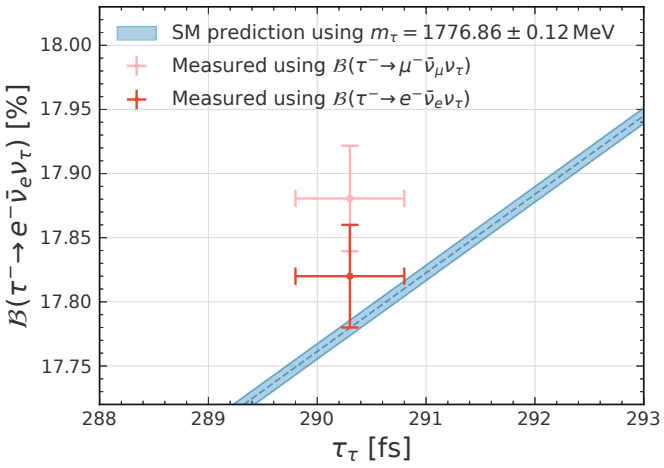


Figure 2.8: The relation between tau lifetime and leptonic branching fractions, as predicted by the SM, shown as a blue band. The width of the band originates mainly from the uncertainty of the measurement of the tau mass. The global combinations of the measured values from experiments are shown as red crosses [10], where the darker uses the measured value of $\mathcal{B}(\tau \to e\nu_e\nu_\tau)$ and the lighter uses the measured value of $\mathcal{B}(\tau \to \mu\nu_\mu\nu_\tau)$, which is then translated to $\mathcal{B}(\tau \to e\nu_e\nu_\tau)$.

2.5 Statistical methods

This section introduces some fundamental statistical concepts frequently employed in high energy physics (HEP) and also relevant to this analysis. It starts with a brief discussion of the maximum likelihood method, followed by a description of hypothesis testing, particularly the profile likelihood ratio method. Although many of the concepts introduced here are also important in Bayesian statistics and can be interpreted from that perspective, this work focuses more on the frequentist approach.

2.5.1 Maximum likelihood method

Most physics analyses aim to determine certain characteristics of the underlying distribution of a particular dataset. This underlying distribution can be described with a statistical model, denoted $P_{\text{model}}(\boldsymbol{x}|\boldsymbol{\theta})$. Here, $\boldsymbol{x}$ are the observations that can depend on the model parameters, denoted as $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)$. The model P_{model} gives the probability of observing $\boldsymbol{x}$, given any instance of the parameters, so it can be thought of as the data-generating process. The model will never perfectly describe the true underlying distribution, but it can describe it sufficiently closely [67].

When the model is evaluated with observed data $\boldsymbol{x}_0$, the likelihood function

$$L(\boldsymbol{\theta}) = P_{\text{model}}(\boldsymbol{x}_0|\boldsymbol{\theta}), \tag{2.24}$$

is obtained, which now only depends on $\boldsymbol{\theta}$. By maximising the likelihood function with respect to $\boldsymbol{\theta}$, the parameter values that best describe the observations can be estimated. In general,

such a function is called an estimator $\hat{\theta}(\boldsymbol{x})$ and its value at the observed data is the estimate $\hat{\theta}$. An estimator can be constructed in many ways, but some of its properties, such as consistency, bias, efficiency, and robustness, are important criteria to consider [10, p. 660]. This particular estimator is the maximum likelihood (ML) estimator. It is frequently used in HEP because of certain characteristics it exhibits when dealing with large data sets, also referred to as the asymptotic limit. That is, the ML estimator is

- *consistent*: With enough data, the estimate $\hat{\theta}$ converges to the true value [68],
- *normal*: The estimate $\hat{\theta}$ adheres to a normal distribution [68],
- and *efficient*: The estimator has the smallest possible variance, given by the Rao-Cramér-Fréchet bound [69, 70].

With these properties, the inverse of the covariance matrix for the parameters $\hat{\theta}$ corresponds to the curvature of the negative logarithm of the likelihood function, evaluated at its minimum [10]:

$$\left(\widehat{V}^{-1}\right)_{ij} = -\left.\frac{\partial^2 \ln L}{\partial\theta_i \partial\theta_j}\right|_{\hat{\boldsymbol{\theta}}}. \tag{2.25}$$

This feature is extensively used to assess the uncertainty of estimates, for instance, via the HESSE technique of the MINUIT minimisation software [71]. When these asymptotic properties become valid depends, in general, on the model [72]. Before using these properties, it is important to understand the probability distribution of the estimator, which can be obtained by calculating the estimator for each possible realisation of the observed data.

2.5.2 Hypothesis testing

Likelihood ratio tests

The significance of the difference between two models can be tested using the likelihood ratio as a test statistic:

$$\lambda = \frac{L(\boldsymbol{\theta}_1)}{L(\boldsymbol{\theta}_2)}, \tag{2.26}$$

where $\boldsymbol{\theta}_0 \in \Theta_0$ and $\boldsymbol{\theta}_1 \in \Theta$ are two realisations of the possible parameters. Θ is the total parameter space available and Θ_0 is a subset of the parameter space associated with the null hypothesis H_0. Using the likelihood ratio results in the most powerful statistical test, which was first shown by Neyman and Pearson [73].

In addition, Wilks' theorem [74] states that the quantity

$$-2\ln\lambda$$

follows a χ^2 distribution with degrees of freedom corresponding to the difference in the amount of free parameters between the numerator and the denominator. This property is only true when in the asymptotic regime and assuming that the null hypothesis is valid. Knowing this distribution makes the likelihood ratio a convenient choice for a test statistic.

The profile likelihood ratio

Often, the so-called profile likelihood ratio [75]

$$\lambda_p(\mu) = \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})}, \tag{2.27}$$

is used as a test statistic. The numerator is the maximal value of the likelihood when one of the parameters μ is fixed to a specific value. This parameter is called the parameter of interest, and the maximum likelihood estimate (MLE) for the remaining parameters is denoted as $\hat{\hat{\boldsymbol{\theta}}}$. The quantity in the numerator depends on the chosen value of μ and is also called the profile likelihood since, when plotted as a function of μ, it gives a profile of the likelihood function. The quantity in the denominator is instead the maximum of the likelihood when the parameter of interest is also at its MLE. In this work, the test statistic t_μ, defined as

$$t_\mu = -2 \ln \lambda_p(\mu), \tag{2.28}$$

will be used. The value of t_μ depends on the value of the parameter μ. Its properties and asymptotic distribution are described by Cowan et al. [75] and are based on the results of Wilks [74] and Wald [76]. When μ coincides with the actual parameter value and the model is correct, t_μ will asymptotically be distributed according to a χ^2 distribution with one degree of freedom. Using this knowledge about the asymptotic distribution, a confidence interval (CI) with a certain confidence level (CL) can be constructed for the parameter of interest by selecting the range of μ where the p-value of the observed value of t_μ is greater than $1 - \mathrm{CL}$. This is called the test-inversion method, which is similar to the interval construction by Neyman [77]. Depending on the test statistic, various intervals with the same coverage can be obtained [78], similar to the ordering rule in Neyman's method.

2.6 Computational methods

2.6.1 Simulated Annealing

Simulated annealing is a stochastic optimisation algorithm that can be used to find the global minimum of a function [79]. The algorithm is especially useful when the function is multidimensional and has many local minima, which can make it challenging to find the true global minimum using other methods. The algorithm resembles the physical process of annealing, which refers to heating and then cooling a material to reduce its defects and thus increase its stability.

The algorithm is based on the Metropolis-Hastings [80, 81] algorithm. It starts with an initial solution and then iteratively generates new candidate solutions by making small random changes to the current solution. These candidate solutions are evaluated using the objective function that one wants to optimise, which in this case is also called the energy E, since it resembles a potential energy function in materials. The new solution is accepted or rejected based on a probability that depends on the energy difference ΔE between the current solution and the candidate solution and a parameter called temperature T. This function is called the acceptance probability function and has the form

$$p_{\text{accept}} = \min\left(1, e^{\frac{\Delta E}{T}}\right). \tag{2.29}$$

It is based on the Boltzmann distribution from statistical mechanics, which relates the energy of a system to the probability of being in a particular state.

At high temperatures, the algorithm is more likely to accept candidate solutions that worsen the objective function, while at low temperatures, it becomes more selective and only accepts solutions that are better than the current solution. The temperature parameter is gradually reduced over time, following a cooling schedule that starts with a high temperature and slowly decreases to a low temperature. This allows the algorithm to explore the search space, since the possibility of accepting worse solutions prevents it from getting stuck in local minima. Then, the algorithm can gradually converge towards the global minimum as the temperature decreases.

2.6.2 Artificial neural networks

Artificial neural networks (ANNs) are functions inspired by the structure of biological neural networks in the human nervous system. They can solve complex problems by learning from data, like the learning process of the human brain [82].

An ANN consists of layers of interconnected neurons [83]. In the simplest architecture, these layers comprise an input layer, one or more hidden layers, and an output layer. The input layer receives input data, which is then passed through the hidden layers, and the output layer produces the final output of the network. Each neuron in the network receives input from other neurons, processes that input, and then transmits its output to other neurons in the network. One of the most basic types of ANNs is also called a multilayer perceptron (MLP), where every neuron in one layer connects to every neuron in the neighbouring layers. This is the type that will be used in this work, and in the following, it will simply be referred to as ANN.

The connections between neurons in an ANN are characterised by weights, which can be any real number and determine the strength of the connection between two neurons. Furthermore, each neuron is assigned a specific number known as the bias [84]. With these parameters, the output a_j^l from the j^{th} neuron in the l^{th} layer is determined as

$$a_j^l = \sigma\left(\sum_k w_{jk}^l a_k^{l-1} + b_j^l\right), \tag{2.30}$$

with b_j^l denoting the bias for neuron j in layer l, and w_{jk}^l representing the weight corresponding to the connection from neuron k in layer $l-1$ to neuron j in layer l. The bias term is added to the weighted sum of the inputs to the neuron, and the result is then passed to the activation function σ, another key component of ANNs. It determines what value is passed to the subsequent neurons. It often has the form of a continuous step function, defining a threshold for which the neuron produces an output, similar to neuronal firing in the brain [85]. One commonly used activation function is the sigmoid function (or Fermi function),

$$\sigma(z) = \frac{1}{1+e^{-z}}, \tag{2.31}$$

which maps the input to a value between zero and one, or the hyperbolic tangent,

$$\tanh(z) = \frac{\mathrm{e}^{z} - \mathrm{e}^{-z}}{\mathrm{e}^{z} + \mathrm{e}^{-z}}, \tag{2.32}$$

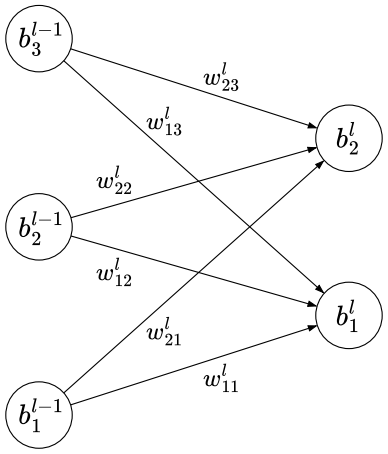


Figure 2.9: A schematic of a part of a neural network, with three neurons shown for layer $l-1$ and two neurons for layer l. The weights for the connections between neurons and the biases associated with each neuron are indicated.

which maps the input to a value between minus one and one. Figure 2.9 illustrates the parameters of the network and how they can be attributed to the nodes and edges of a graph. With Equation 2.30, it is possible to map any given input to a specific output of the network.

The parameters of the network are adjusted during the learning process to improve its performance. The training of ANNs may occur supervised or unsupervised [86]. In the supervised approach, the network learns from labelled data consisting of the input-output pairs $\vec{x}$ and y. The input is fed to the network and the obtained output $\vec{a}(\boldsymbol{w}, \boldsymbol{b}, \vec{x})$ is compared with the desired output. To quantify how well the network performs, a loss function, such as the quadratic loss for regression tasks, is used.

One commonly used loss function for classification tasks, where the outputs of the ANN correspond to different classes, is cross-entropy, which has a close relationship with the maximum likelihood method [86, Chapter 6]. It measures the difference between two probability distributions: the predicted probabilities yielded by the model and the actual distribution of the labels in the training data. When the cross-entropy is minimised, the neural network learns to maximise the likelihood of observing the training data. When the output of the neural network is normalised by a softmax function, it can be interpreted as probabilities [87]. The actual probability is 1 for the class to which the data point corresponds and 0 for all other classes. For the general case of multiclass classification, the loss is then defined as

$$L_{\mathrm{CE}} = -\frac{1}{N}\sum_{n=1}^{N} w_n \log \frac{\exp\left(a_{n,y_n}\right)}{\sum_{c=1}^{C}\exp\left(a_{n,c}\right)} \tag{2.33}$$

where $a_{n,\,c}$ is the value of the c^{th} output neuron for the n^{th} training sample, y_n is the index of the class to which the training sample belongs, C is the total number of classes, which the denominator is summed over, N is the number of training samples, and w_n is a weight associated with the n^{th} training sample. This weight can be used to change the importance of the individual training samples given during the training, which is especially important when training with data that consists of different classes, but are in the wrong proportions. When only two classes are provided, the cross-entropy reduces to the binary cross-entropy

(BCE) function. In unsupervised learning, the network is trained using unlabelled data and the weights are adjusted to identify patterns and structures in the data.

The parameters are adjusted using an algorithm called gradient descent. It works by iteratively changing the parameters in the negative direction of the loss function gradient. In each iteration, a step is taken in the direction of the steepest downward slope of the loss function, resulting in the name gradient descent. The size of the step is determined by an additional parameter called the learning rate (LR) η, which can also vary throughout the training process of the network. Thus, the rules for updating the parameters $\boldsymbol{w}$ and $\boldsymbol{b}$ in each iteration are

$$w_{jk}^l \rightarrow \tilde{w}_{jk}^l = w_{jk}^l - \eta \frac{\partial C}{\partial w_{jk}^l} \quad \text{and} \tag{2.34}$$

$$b_j^l \rightarrow \tilde{b}_j^l = b_j^l - \eta \frac{\partial C}{\partial b_j^l}. \tag{2.35}$$

To efficiently calculate the derivatives required for gradient descent, the backpropagation algorithm [88] is used. Using the chain rule, this algorithm computes derivatives starting at the last layer and moving backwards through the network. This approach avoids redundant calculations and allows all derivatives to be obtained in a single backward pass. Automatic differentiation is a generalisation of the backpropagation algorithm and is widely used in various applications [89].

For training ANNs with large amounts of data [90], stochastic gradient descent is used, which is a modification of gradient descent. It uses random subsets of data called mini-batches to calculate the gradient of the loss function. This will result in stochastic fluctuations of the gradient vector with respect to the one obtained using the entire sample, which in some cases can also help to escape local minima. When all mini-batches that comprise the training sample have been used, one training epoch is considered complete. This process is repeated until the loss stagnates or a certain number of training epochs are reached. The size of individual batches must be optimised to balance performance and execution speed.

Overtraining of the ANN can occur when the network picks up specific features of the training data and performs poorly on new data. This is more likely to occur when the amount of training data is too small for the complexity of the model, such as when the ANN has many free parameters. To check for overtraining, it is helpful to split the data into a train sample to train the network and a test sample, which is only used to evaluate its performance. The network's performance on the test sample can be used to infer the actual performance on unseen data. During training, this can be done by monitoring the loss value as a function of the epochs. Regularisation techniques and early stopping of the training process can also be used to prevent overtraining [91].

CHAPTER 3

Experimental setup

3.1 Introduction

In this chapter, the experimental setup is outlined. This includes some fundamental concepts presented in Section 3.2, an introduction to SuperKEKB in Section 3.3, and the most important aspects of the Belle II experiment in Section 3.3. Section 3.5 summarises the used dataset and simulated samples, and Section 3.6 discusses the identification of charged particles at Belle II, which is particularly important for this measurement.

3.2 Fundamental concepts

3.2.1 Luminosity

Luminosity is an essential concept in HEP that quantifies the rate of interactions per unit cross-section in particle collisions. Thus, multiplying the luminosity by the cross-section σ of a certain physics process yields the total number of expected events N of that process.

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

Here, $L(t)$ denotes the instantaneous luminosity per unit of time, resulting in the total luminosity when integrated over a time period.

The instantaneous luminosity is given by the expression

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

with f being the revolution frequency, n the total bunches per beam in the storage ring, and n_1 and n_2 denoting the particle count in each bunch. In the denominator, the terms σ_x^* and σ_y^* denote the beam sizes at the interaction point (IP) in the horizontal and vertical directions, while $\mathcal{F}$ includes additional factors [10, 92]. In practice, the computation of luminosity based on the beam parameters is quite complicated because of the limited knowledge and the time dependence of some quantities. Thus, the luminosity is usually obtained using a reference

process with a known cross-section. The number of produced events can be computed from the observed events by reconstructing this process with a known efficiency, for example, obtained from simulations. Using this information, the luminosity can be calculated with Equation 3.1. Typically, units of $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ are used to report the instantaneous luminosity, while the integrated luminosity is usually expressed in inverse barns with 1 barn being equal to $1 \times 10^{-24}\,\mathrm{cm}^2$.

3.2.2 Energy loss of charged particles

When charged particles travel through matter, they can lose energy by interacting with the particles, atoms, and molecules of the material (which can be a solid, liquid, or gas). This can lead to ionisation, atomic, or collective excitations. The involved process and the amount of lost energy depend heavily on the momentum and mass of the traversing particle, based on which different effects have to be accounted for.

Ionisation

The Bethe-Bloch equation [93, 94] quantifies the mean rate of energy loss $\left\langle -\frac{dE}{dx} \right\rangle$ due to ionisation for moderately relativistic particles. Here, moderately relativistic refers to particles with approximately $0.1 \leq \beta\gamma \leq 10000$, where β is the velocity of the particle normalised to the speed of light and γ is the Lorentz factor. The range in which the equation is valid and when other effects become important also depends on the atomic number Z of the material. Additional atomic effects need to be considered at low energies, while radiative effects start to play a role at high energies. The energy loss in the regime described by the Bethe-Bloch formula depends only on β and is qualitatively the same for all particles and target materials (with some scale dependence on Z) [10].

The Bethe-Bloch equation is

$$\left\langle -\frac{dE}{dx} \right\rangle = Kz^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \frac{2m_e c^2 \beta^2 \gamma^2 W_{\max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right] \tag{3.3}$$

with the meaning of the appearing symbols summarised in Table 3.1. At low velocities, the

Table 3.1: Definition of the symbols appearing in the Bethe-Bloch equation, together with the usually used units. The table was adapted from „Review of Particle Physics“ [10, p. 549].

Symbol	Definition
E	energy of the traversing particle [MeV]
x	mass per unit area [$\mathrm{g/cm}^2$]
$W_{\max}$	maximum possible energy transfer to e^- in a single collision [MeV]
z	charge number of the traversing particle
Z	atomic number of the target
A	mass number of the target [g/mol]
K	coefficient $= 0.307\,075\,\mathrm{MeV mol}^{-1}\mathrm{cm}^{-2}$
I	mean excitation potential of the target [eV]
$\delta(\beta\gamma)$	correction due to the density effect

energy loss decreases approximately as $1/\beta^2$ with increasing energy and reaches a minimum

at around $\beta\gamma = p/m \approx 3$, where m is the mass of the particle and p its momentum. When the velocity of the particle becomes more relativistic, the first term in the bracket becomes important, and the energy loss increases logarithmically. This logarithmic increase is partially cancelled by the correction term due to the density effect ($\delta(\beta\gamma)$), which is due to the polarisation of the medium due to the extended electric field of the particle [10]. The functional form of the Bethe-Bloch equation can be seen in Figure 3.1 for a positively charged muon traversing through copper. The dashed and dotted lines indicate the individual contributions, while the solid curve gives the total effect. The Bethe-Bloch equation only describes the range between $\beta\gamma \sim 0.1$ and $\beta\gamma \sim 1000$, indicated by the vertical lines. Very relativistic particles also experience energy loss due to the emission of Bremsstrahlung, which eventually becomes dominant and is referred to as the radiative regime. The energy after which the energy loss due to ionisation becomes equal to the radiation losses is called the critical energy, denoted $E_{\mu c}$ in the figure. For electrons, this critical energy is at the order of 10 MeV for most materials.

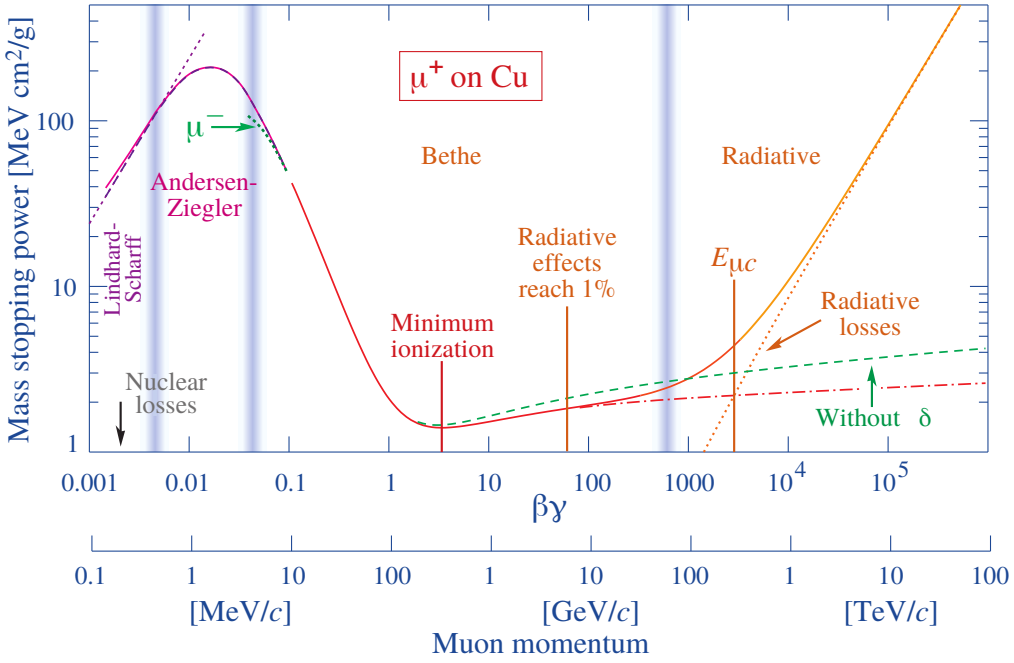


Figure 3.1: The mean energy loss $\left\langle -\frac{dE}{dx} \right\rangle$ of a positively charged muon in copper as a function of $\beta\gamma = p/m$, taken from „Review of Particle Physics“ by the PDG [10, p. 551]. The vertical lines indicate the boundaries between different approximations with the Bethe-Bloch equation only describing the range between $\beta\gamma \sim 0.1$ and $\beta\gamma \sim 1000$. This is also the range most critical for particles produced in collisions at the Belle II experiment. Although the Bethe-Bloch equation depends only on the velocity β of the particle, this is no longer the case for the radiative regime, where the primary cause of energy loss is Bremsstrahlung.

Electromagnetic showers

When an electron traverses a material above the critical energy, it can create a cascade of secondary electrons and photons. This cascade of particles is known as a particle shower. In this process, first, a photon is generated via Bremsstrahlung, which then creates an e^+e^- pair via pair production. The produced electron and positron can, in turn, create a photon, and

the process repeats. The length scale of this process is called one radiation length, defined as the mean distance over which the energy of the electron is reduced by a factor of $1/e$. It depends on the density and composition of the material, with materials with a higher atomic number and a higher density having a shorter radiation length. For elements with $Z > 4$ it is approximately

$$X_0 = \frac{716.406A}{Z^2\left(\ln\left(184/Z^{\frac{1}{3}}\right) - f\right) + Z\ln\left(1194/Z^{\frac{2}{3}}\right)}\ \mathrm{g\,cm^{-2}}. \tag{3.4}$$

with Z being the atomic number, A the atomic mass in units of $\mathrm{g\,mol^{-1}}$ and f a Coulomb correction factor. More details and radiation lengths for many materials can be found in [95].

Pair production is the predominant cause of energy loss for photons with high enough energies (above 10 MeV to 100 MeV, depending on the material). It happens through the interaction of the photon with the electric field of an atomic nucleus or the field of the electrons, whereas the latter has a much smaller cross-section. At lower photon energies, other processes take over, such as Compton scattering or the photoelectric effect [10].

Although electromagnetic showers are primarily relevant for electrons and photons, all charged particles undergo radiative processes at sufficiently high energies. However, the critical energy above which this occurs for pions and muons (when traversing iron) is only several hundred GeV [10].

Hadronic showers

Hadronic showers are created when highly energetic hadrons interact with matter through inelastic strong interactions. This interaction generates a cascade of secondary particles that continue to interact with the material. As the shower progresses, the number of particles and their energy decreases until the remaining particles are fully absorbed.

The development of hadronic showers depends on various factors, such as the energy and type of the initial hadron and the density of the material. Since many different physical processes are involved, it is more complex to describe and detect than electromagnetic showers. Additionally, some of the initial hadrons energy is transferred to nuclear recoil, excitation, or break-up, which often does not produce a detectable signal [21]. The amount of undetected energy can vary between showers and is a limiting factor in the resolution of hadronic calorimeters [10, p. 600].

Hadronic showers are always accompanied by electromagnetic showers caused by some of the secondary particles, such as neutral pions that decay into two photons. Compared to electromagnetic showers, hadronic showers take longer to develop and have a larger size [96].

3.3 The SuperKEKB accelerator

SuperKEKB [97] is a particle collider located in Tsukuba, Japan. In nominal operation, it accelerates electron and positron beams to energies of 7 GeV and 4 GeV, which then collide at the Belle II detector at a CM energy, often denoted as $\sqrt{s}$, of 10.58 GeV. The asymmetric beams lead to a boost with a Lorentz factor of $\beta\gamma = 0.28$ of the CM frame relative to the laboratory frame. This boost is important for studying time-dependent CP violation with B

mesons due to their larger decay length away from the collision point. The beams of KEKB, the predecessor of SuperKEKB, were more asymmetric, which resulted in a Lorentz boost of 0.425.

The SuperKEKB accelerator is shown in Figure 3.2, with the electron ring coloured red and the positron ring blue. The main storage rings, which have a circumference of 3.016 km, can accommodate 14 776 bunches of particles that collide every 6.5 ns at the interaction point. Each bunch is approximately 0.6 cm in length and contains roughly 3×10^{10} particles [10, p. 545].

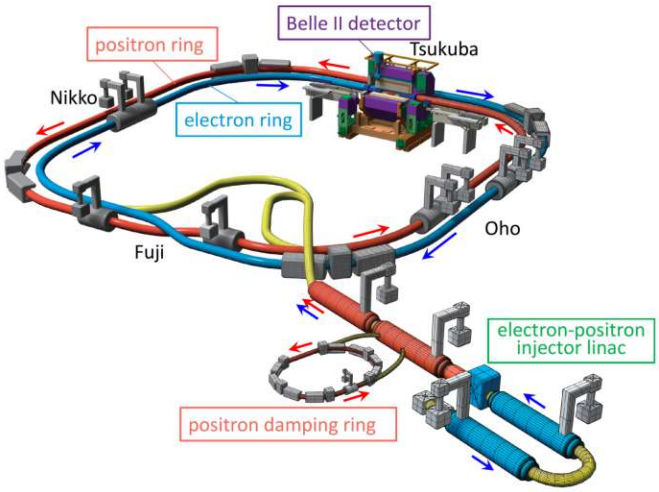


Figure 3.2: An overview of the SuperKEKB accelerator, taken from „SuperKEKB collider“ [97].

The electrons are generated using a photocathode high-current radio-frequency gun, which can achieve low emittance by irradiating a small cold cathode with short laser pulses. Emittance is the area of phase space that describes the transverse particle motion of particles and is populated by noninteracting and nonradiating particles. Low emittance of the particle beams is required to achieve high luminosities. The produced electron beams have a bunch charge of 5 nC and are accelerated to 7 GeV using a 600 m long linear accelerator and injected into the high energy storage ring (HER) [98].

Another electron beam with 10 nC is created using a thermionic gun previously used in KEKB. This beam is used to generate positrons through electromagnetic showers by firing it at a 14 mm thick tungsten target. The positrons are then circulated in a damping ring with an energy of 1.1 GeV to achieve a low emittance. After being accelerated to 4 GeV, they are injected into the low energy storage ring (LER) with a bunch charge of 4 nC [97, 98].

The design luminosity of SuperKEKB is a factor of 40 larger than what was achieved by its predecessor, KEKB. Reaching this goal requires reducing the beam size at the interaction region from 1 µm to approximately 50 nm. SuperKEKB uses the ‘crab waist’ scheme, also known as the nano-beam scheme, invented by Pantaleo Raimondi for the super B-factory [99]. The idea is to reduce the longitudinal size of the overlap region of the two beams at the IP to be much smaller than the bunch length. This allows the bunches to intersect only in a very focused region at the IP, where the vertical beam size is minimal, as it gets squeezed by superconducting magnets at the interaction region. The size of the overlap region depends on

the crossing angle and the horizontal beam size. The crossing angle at the interaction point of SuperKEKB is 4.76° [98].

Another critical factor is the increase of the beam currents by a factor of 2 to 3. Additional key parameters of SuperKEKB can be found in Table 3.2, and a comparison of selected parameters between what was achieved by KEKB and the design of SuperKEKB is provided in Table 3.3. The beta functions are a measure of the transverse beam size and its divergence and represent how well the beam is focused. The beam energy asymmetry was reduced to minimise losses

Table 3.2: Important parameters of SuperKEKB [92, 100], with updated numbers from [10, p. 545], reflecting conditions in March 2022.

Parameter	LER (e^+)	HER (e^-)	Unit
Energy	4.000	7.007	GeV
Half crossing angle	41.5		mrad
Horizontal emittance	4.0	4.6	nm
Vertical emittance	0.058	0.058	nm
Beta functions at IP (β_x^*/β_y^*)	80/1	60/1	mm
Beam currents	0.819	0.64	A
Luminosity	4.7×10^{34}		$\mathrm{cm}^{-2}\ \mathrm{s}^{-1}$

due to Touschek scattering in the lower energy beam, resulting in a smaller boost but a higher solid angle acceptance. This is particularly important for physics analyses that rely on the precise knowledge of missing energy [100]. At the time of writing this, SuperKEKB achieved a

Table 3.3: A comparison between some important parameters for KEKB (achieved) and SuperKEKB (design).

	KEKB	SuperKEKB	
E [GeV]	3.5 / 8.0	4.0 / 7.0	LER/HER
β_y^* [mm]	5.9 / 5.9	0.27 / 0.30	LER/HER
β_x^* [cm]	120 / 120	3.2 / 2.5	LER/HER
I [A]	1.6 / 1.2	3.6 / 2.6	LER/HER
φ [mrad]	11	41.5	
$\mathcal{L}$ [$\mathrm{cm}^{-2}\mathrm{s}^{-2}$]	2.1×10^{34}	8×10^{35}	

maximum instantaneous luminosity of $4.7 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-2}$, which is the highest ever recorded at a particle collider. However, SuperKEKB has not reached its designed goal, and higher instantaneous luminosities are expected in the future.

3.3.1 Beam-induced backgrounds

The Belle II detector has to deal with many beam-induced backgrounds that can potentially affect its performance. These backgrounds arise from a variety of processes. The most significant are Touschek scattering, beam-gas scattering, synchrotron radiation, radiative Bhabha scattering, and e^+e^- pair production via the two-photon process. The last two mentioned backgrounds scale directly with luminosity, since they result from the colliding beams. This is why they are sometimes referred to as luminosity backgrounds, whereas the

others are single-beam backgrounds. Single-beam backgrounds can be avoided or reduced to some degree with beam-steering, collimators, vacuum-scrubbing, shielding, or a well-designed interaction region (IR). Luminosity backgrounds, on the other hand, cannot be reduced.

Touschek scattering Touschek scattering occurs when particles within a bunch collide and subsequently strike the beam pipe. This results in a particle shower that may hit the Belle II detector. The rate of Touschek scattering depends on the beam current, the total number of bunches, and is inversely proportional to the beam size [92]. Because of the small beam sizes in SuperKEKB, Touschek scattering is a major background source. However, it is only relevant for the LER since the scattering rate is also proportional to $1/E^3$ with E being the beam energy, and the HER also has a lower beam current [98].

Beam-gas scattering Beam-gas scattering occurs when the beams interact with residual gas in the vacuum tubes and then hit the wall of the vacuum chamber, producing a shower of particles. The beams can interact with the gas molecules via Coulomb scattering and Bremsstrahlung. The beam-gas rate depends on the beam current, the number of gas molecules (determined by the vacuum pressure), and the magnetic field strength of the focusing magnets [98].

Synchrotron radiation Synchrotron radiation results from the radial acceleration of charged particles in the magnetic field of bending and focusing magnets. The synchrotron photons can have energies on the order of keV and potentially impact the inner detectors. This can be substantially mitigated by specially designing the IR to counteract such effects. The intensity of synchrotron radiation increases with the square of the beam energy, which means that it occurs predominantly in the HER [100].

Injection background The injection background is due to the continuous injection of charge into the storage rings, which modifies the beam bunch. During injection, a veto signal is sent that prevents Belle II from recording data, but studies have shown that the background conditions increase shortly after an injection.

Radiative Bhabha scattering Photons from radiative Bhabha scattering $e^+e^- \to e^+e^-\gamma$ are emitted primarily along the beam direction, where they can interact with the iron of the magnets. This can create a large number of neutrons, which are the primary background for the K_L - Muon detector. Additionally, the electrons and positrons participating in the scattering may undergo energy losses, which results in them being overbent by the magnets. Consequently, they may collide with the magnet walls, generating electromagnetic showers [98].

Two photon process Via the two-photon process ($e^+e^- \to e^+e^-\ell^+\ell^-$), a very low momentum lepton pair can be produced, primarily affecting the detectors very close to the collisions. In the first layer of the Pixel detector, which is located only 1.3 cm away from the IP, a track rate of 700 kHz/cm^2 up to 10 MHz/cm^2 is estimated due to this process alone. This would correspond to about 900 to 14 000 lepton pairs in each event [98]. The rate drops with the inverse distance to the IP, but the inner detectors need to be designed in a way to be able to deal with this constant background. Moreover, the produced leptons can spiral in the magnetic field due to their low momentum, resulting in many detector hits [100].

3.4 The Belle II experiment

The Belle II detector [98] is an upgrade of the Belle detector and introduces a new generation of B factories [101]. It is specially designed to handle up to 40 times higher event rates and 10 to 20 times higher background rates. Higher backgrounds result in increased occupancy, radiation damage, fake signals, and noise in several sub-detectors. The increased event rates require an improved trigger system, an efficient data acquisition (DAQ), and a high-performance distributed computing system. At the same time, Belle II still has to meet the requirements of a B-factory, such as an excellent vertex resolution (around 50 µm), a good reconstruction efficiency and momentum resolution of both charged and neutral particles across the full kinematic range of the experiment, effective particle identification, and a hermetic detector [100]. Although the primary objective of Belle II is to study B mesons, it is also a multipurpose detector. With a very large cross-section for D meson and τ lepton production, Belle II is also a charm and tau factory, enabling world-leading measurements in these areas. Furthermore, due to the ability to trigger low-multiplicity events and reconstruct missing momentum (due to precise knowledge of the initial and final states), Belle II is well suited for searches with missing energy, as it is the case in dark sector searches.

Figure 3.3 shows a slice of the Belle II detector, as seen from above. In the centre, the e^+ and e^- beams cross in the interaction region, around which the sub-detector systems are located. The following will give a brief description of each sub-detector. Most of the sub-detectors have a cylindrical design around the beam directions, and due to the higher energy of the e^- beam, the detector exhibits a forward-backward asymmetry. The laboratory coordinate system has its origin at the interaction point, with the z axis pointing in the direction of the electron beam (the forward direction), the x axis being horizontal, and the y axis pointing vertically upward [98]. A superconducting solenoid outside the electromagnetic calorimeter generates a homogeneous 1.5 T magnetic field parallel to the z-axis.

3.4.1 Vertex detectors

The Belle II vertex detectors are the Pixel detector (PXD) and the Silicon Vertex detector (SVD). The PXD was a new introduction in Belle II to achieve better vertex resolution, which is necessary due to the smaller boost of the collision products compared to Belle. These detectors are placed around the beryllium beam pipe, which has a diameter of approximately 10 mm. The PXD is composed of two layers situated at a radius of 14 and 22 mm. It is made up of nearly 8 million pixels and employs depleted p-channel field effect transistor (DEPFET) sensors, with the advantage that they can be manufactured extremely thin (around 50 µm) to minimise multiple scattering.

The SVD comprises four layers, each located at a radius of 38, 80, 115, and 140 mm [102]. Every layer is outfitted with double-sided silicon strip sensors, reducing the number of readout channels compared to those needed by the PXD. A cross section of the vertex detectors is shown in Figure 3.4, with the PXD modules close to the collision point and the four SVD layers positioned around it. There are a total of 20 diamond sensors placed surrounding the beam pipe, 4 of which are located at each end of the PXD, as well as 6 sensors at the forward and backward region of the SVD support structure. They are used to monitor the beam background. High backgrounds could damage the SVD, which is why the beam is automatically aborted if the diamond sensors measure background levels above some threshold [103].

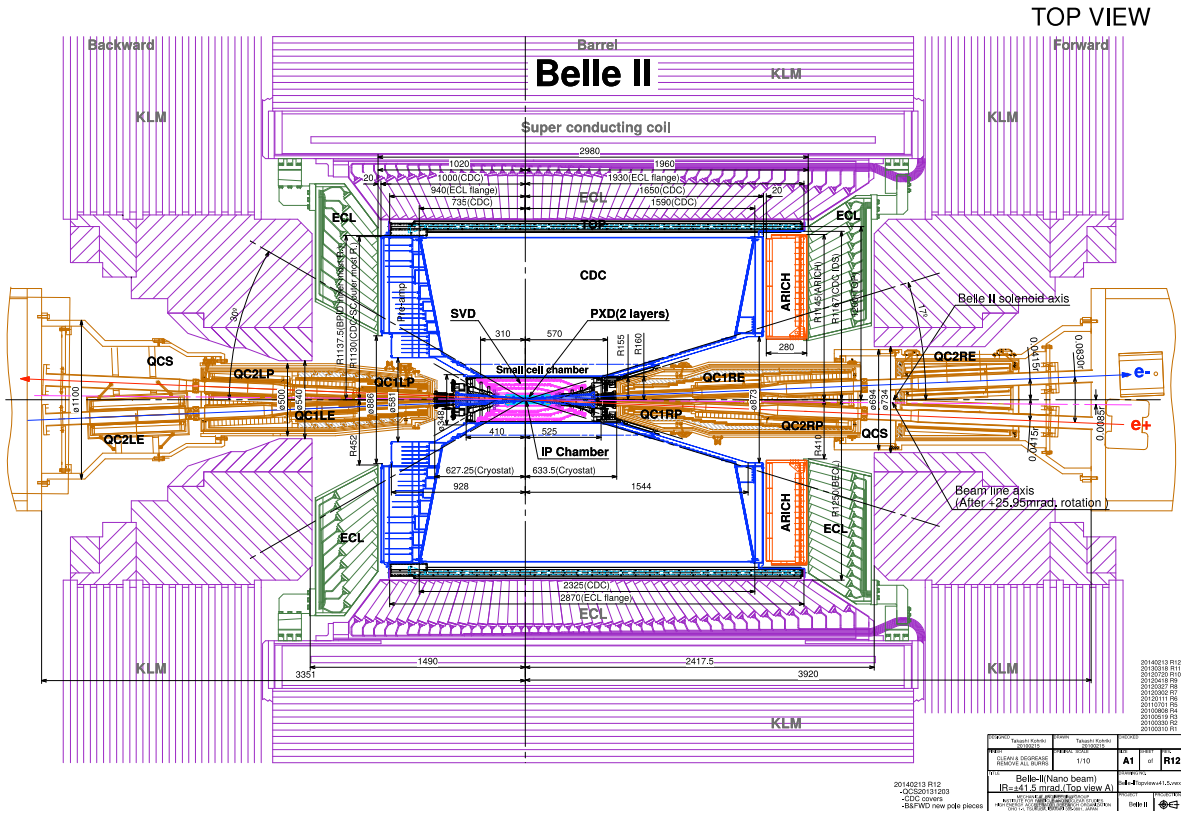


Figure 3.3: An overhead view of a slice of the Belle II detector, illustrating the interaction region and the different sub-detector systems. The drawing is taken from „The Belle II Physics Book“ [100, p. 37].

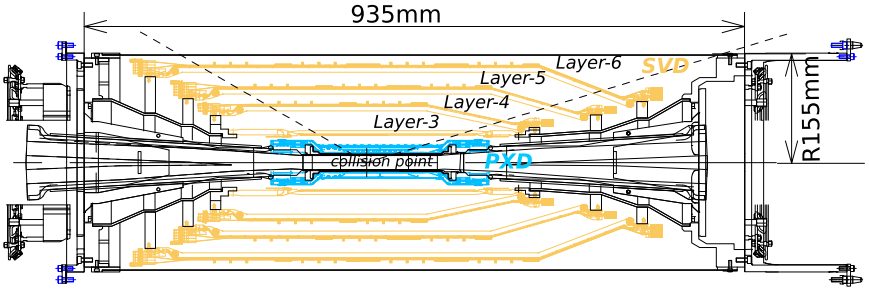


Figure 3.4: A technical drawing of the cross section of the vertex detectors, showing the PXD in blue and the SVD in orange. The PXD consists of two layers and the SVD comprises four layers. The image is adapted from [103].

The vertex detectors are used primarily to accurately measure the decay vertices of particles. Initially, particle tracks are reconstructed using only information from the Central Drift Chamber (CDC). Subsequently, these tracks are extrapolated to the PXD with the help of data from the SVD. In addition, the SVD is also capable of reconstructing particles that decay beyond the PXD, like $K^0_S \to \pi^+\pi^-$, as well as low momentum particles (as low as $p_T \sim 10\,\text{MeV}$) that fail to reach the CDC.

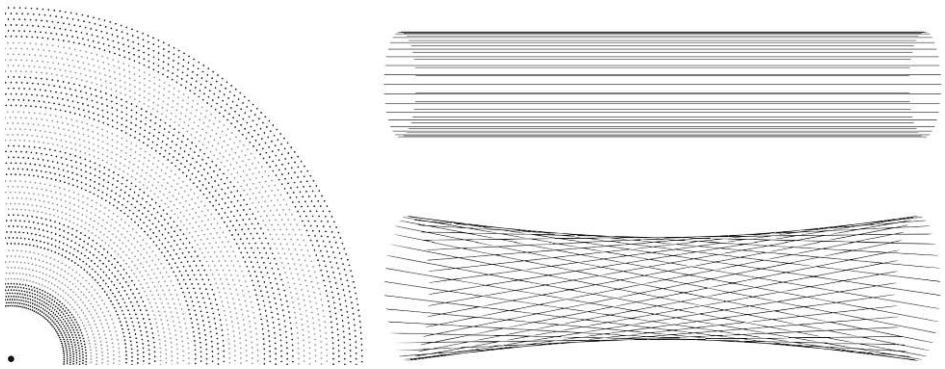


Figure 3.5: The CDC viewed in the r-φ plane (left) with alternating axial and stereo layers, which are depicted on the right. The axial layers are aligned with the magnetic field direction (z direction), while the stereo layers are twisted. The image is taken from [105].

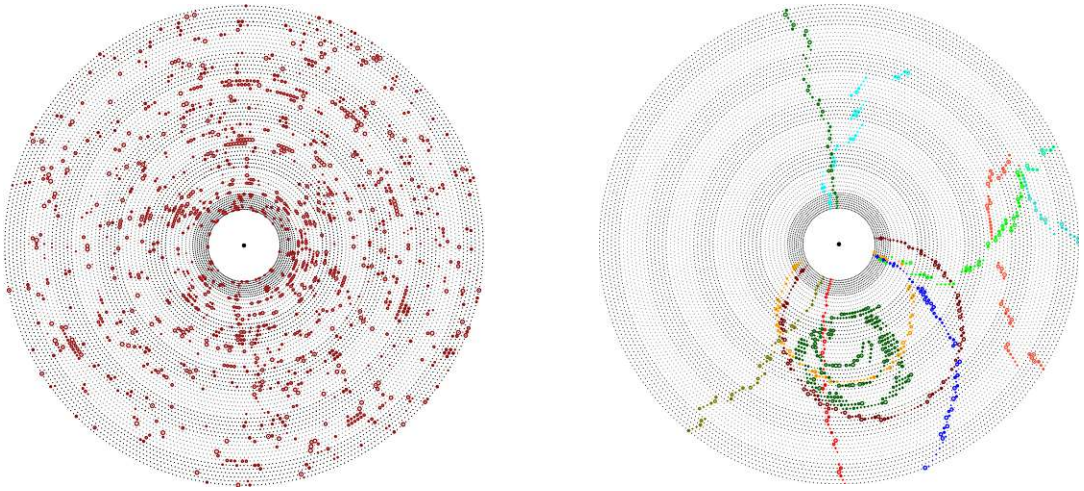
3.4.2 Central Drift Chamber

The CDC in Belle II is essential for reconstructing charged particles and their momenta by detecting their traces. This is called tracking, and such reconstructed particles are called tracks. The momentum of the reconstructed particles can be obtained by their curvature in the magnetic field, with the radius of the circle in the transverse plane being $r = p_t/B$, where p_t is the transverse component of the momentum vector and B is the strength of the homogeneous magnetic field parallel to the z-axis [104]. The CDC is composed of 56 layers arranged from $r = 160\,\text{mm}$ up to $r = 1130\,\text{mm}$, each containing a specific number of wires with 30 µm diameter. These layers are grouped into super-layers, consisting of either axial or stereo layers. Axial layers align with the direction of the magnetic field, whereas stereo layers are twisted to enable the determination of the z position, which allows for a three-dimensional track reconstruction. The structure of the layers can be seen on the left in Figure 3.5, where the CDC is shown in the r-φ plane. The upper right of the figure shows an axial layer, and the lower right shows a stereo layer. The gas in the CDC is a mixture consisting of equal parts of helium (He) and ethane (C_2H_6), with an average drift velocity of $3.3\,\text{cm/µs}$. Figure 3.6 displays the wire arrangement in the r-φ plane, with simulated hits on the wires shown as circles. The diameter of each circle reflects the intensity of the detected current. Figure 3.6a shows a simulation of the anticipated beam background with the design luminosity. In contrast, Figure 3.6b shows a simulated physics event in which the CDC signals are classified according to the track they originate from. This illustrates the difficulties of accurately and reliably identifying and reconstructing tracks among significant background noise.

3.4.3 Cherenkov detectors

The two Cherenkov detectors, the Time-of-propagation counter (TOP) and the Aerogel Ring Imaging Cherenkov counter (ARICH), enhance the accuracy of particle identification between kaons and pions in a momentum range from 0.4 GeV to 4 GeV. They also help to differentiate pions, muons, and electrons with momenta under 1 GeV [100]. The operational principle of both detectors is based on the emission of Cherenkov radiation by a charged particle travelling at a velocity β, which exceeds the speed of light in a medium, given by the equation

$$\cos\theta_c = \frac{1}{n(\lambda)\beta}, \tag{3.5}$$



(a) A simulation of the hits expected from backgrounds at the design luminosity of SuperKEKB. The image is taken from [105].

(b) A simulated physics event where the CDC signals are classified according to their origin. The image is taken from [106, p. 57].

Figure 3.6: CDC wires viewed in the r-φ plane, and the resulting signal from simulated events indicated as circles.

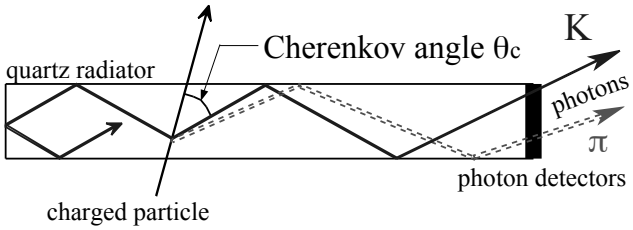


Figure 3.7: A sketch of a quartz bar as it is used in the TOP detector. When a charged particle transverses the detector, Cherenkov photons get emitted at a specific angle depending on its velocity. Reflection occurs within the bar until the photons arrive at the photodetector. The image is taken from [98, p. 220].

where n is the refractive index of the medium and θ_c is the angle at which the Cherenkov radiation is emitted. The Cherenkov angles can be used to distinguish particles with the same momentum but different masses.

The TOP detector covers the entire barrel region at $r = 120\,\text{cm}$ and consists of 16 quartz bars that are 2.6 m long, 45 cm wide and 2 cm thick [98]. As particles traverse the bar, they emit Cherenkov photons, which undergo total internal reflection within the material until they arrive at the side of the bar. Here, a position-sensitive microchannel plate photomultiplier tube (PMT) captures them, featuring a time resolution of approximately 100 ps [107, 108]. The angle θ_c can be reconstructed by measuring the time of propagation and the impact position of the Cherenkov photons. In addition, the production time of the particles has to be known within about 50 ps, which was already utilised by the Belle experiment [109]. Figure 3.7 shows a sketch of a quartz bar, indicating how it is used for particle identification.

The ARICH consists of an aerogel plane and a detector plane with tiles that form 4 and 7 rings, respectively, organised in a disk at the forward direction (positive z) of Belle II [110]. Particles

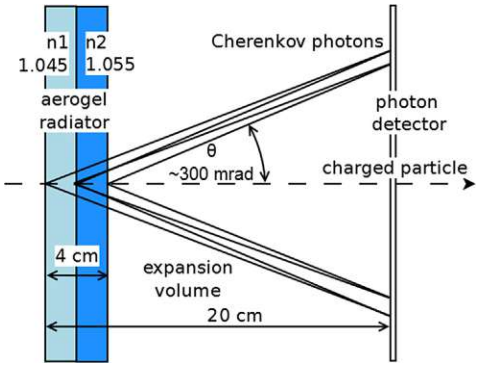


Figure 3.8: An illustration depicting two aerogel layers with varying refractive indices, which increases the number of Cherenkov photons at the detector plane without significantly degrading the resolution. The image is taken from [110].

Table 3.4: The polar angle coverage of the ECL detector and its division into segments, with the resulting number of CsI(Tl) crystals [112].

	θ coverage	θ segments	φ segments	# of crystals
Forward endcap	$12.01° - 31.36°$	13	$48 - 144$	1152
Barrel	$32.2° - 128.7°$	46	144	6624
Backward endcap	$131.5° - 155.03°$	10	$64 - 144$	960

that traverse the aerogel tiles generate Cherenkov photons, which subsequently travel through a 20 cm expansion volume before reaching detectors that are sensitive to individual photons [98]. The radius of the photon ring at the detector plane allows for direct reconstruction of the Cherenkov angle. Increasing the thickness of the aerogel layer enhances the photon detection rate but also increases the uncertainty of the photon emission point, affecting the precision with which the Cherenkov angle can be determined. By stacking two aerogel layers with varying refractive indices the photon emission can be enhanced without significantly degrading the resolution [111]. This method is illustrated in Figure 3.8.

3.4.4 Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECL) is composed of a barrel and endcaps containing 8736 thallium-doped caesium iodide (CsI(Tl)) crystals, which cover about 90 % of the solid angle in the CM system. The crystals are positioned at distances between $r = 125\,\text{cm}$ and $r = 162\,\text{cm}$ within the barrel and at $z = -102\,\text{cm}$ and $z = 196\,\text{cm}$ for the endcaps. Table 3.4 summarises the polar angle coverage of the ECL for the different detector regions and the number of crystal segments in θ and φ. Each crystal has a truncated pyramid shape and is approximately 6 cm × 6 cm wide, and 30 cm in length, . They are oriented towards the IP but slightly tilted in φ, so the gaps do not align with the IP. CsI(Tl) was chosen because it is a bright scintillator with a short radiation length and good mechanical properties [98]. The energy depositions in the crystals excite metastable states that emit scintillation light. The thallium doping shifts the light from UV to visible and increases the characteristic time to about 1 µs. Two photodiodes at the back of each crystal detect the scintillation light. While the crystals have been reused from Belle, the readout electronics have been upgraded to allow

waveform sampling. This enables the measurement of the signal waveform and separation of signals based on their scintillation response to particles with different ionisation properties, called pulse shape discrimination (PSD). The primary function of the ECL is to identify photons and electrons. This is possible since they release most of their energy into the crystals. Precise measurement of the location and energy of photons, which usually ranges from 20 MeV to 4 GeV, is a requirement for many analyses at Belle II. With an increase in event rate and background levels the performance of the ECL is expected to deteriorate, primarily due to the long decay time of scintillations in CsI(Tl) crystals. The resulting pileup and overlapping of pulses can only be partially mitigated with PSD.

3.4.5 K_L - Muon detector

The K_L - Muon detector (KLM) features a structure composed of iron plates and active detector materials that surround the barrel and the endcaps. It spans approximately from 200 cm to 250 cm within the barrel. The thickness of the iron plates is 4.6 cm, corresponding to 3.9 interaction lengths, enabling hadronic showers of K_L^0 mesons. Additionally, these plates act as a magnetic field flux return, as they are positioned outside of the solenoid [100]. The detectors are Resistive Plate Chambers (RPCs) [104] or scintillator strips. RPCs were already used in Belle and are a type of proportional counter based on the principle of the ionisation chamber. In Belle II, RPCs are restricted to the outer layers of the barrel because of their significant dead time, which is insufficient for managing the high neutron counts from electromagnetic showers caused by the higher backgrounds. Consequently, scintillator strips equipped with silicon photomultipliers were mounted in the two innermost layers of the barrel and in the endcaps, where these backgrounds are expected to be the largest [98]. The KLM is used primarily to detect and identify K_L^0 and muons. A K_L^0 typically undergo hadronic showers within the ECL or KLM. In contrast, the penetration of muons through the KLM layers varies based on their momentum and the amount of electromagnetic energy they deposit. Muons with less than 0.7 GeV generally fail to reach the KLM, whereas those exceeding 1.5 GeV often pass through all layers and exit the Belle II detector.

3.4.6 Reconstruction software

The Belle II analysis software framework (basf2) [113, 114] is designed to handle both online and offline data processing efficiently. It uses modules to perform small tasks in sequential order, each communicating information to and from a common object-store. Steering files written in Python define the sequence and configuration of the modules, while the modules are mainly written in C++. One sequence is called a path, of which more than one can be specified simultaneously. The framework is used to process the low-level data from the detectors to define intermediate objects such as track hits from the vertex detectors and the CDC or calorimeter clusters from the ECL. This information is then used to construct higher-level objects, such as charged tracks or decay vertices, often involving elaborate algorithms that must cope with many background signals. By storing only the higher-level objects (the lower-level data is still kept but stored elsewhere), event size can be reduced by a factor of approximately 40. The reduced information can then be used to determine quantities associated with the reconstructed particles, such as their four-momentum or particle identification information. Even higher-level quantities about the event as a whole, or the outputs of specially trained multivariate classifiers, can be conveniently computed within basf2 [see 100, Section 5]. The end products are saved

to so-called mini Data Summary Tables (mDST) files, where a set of features is stored for each physics event. These features, also called variables, are mainly associated with the reconstructed energy and momenta of the observed particles, but higher-level quantities such as event variables, particle identification variables, or outputs from multivariate classifiers can also be obtained. The exact variables that are computed and stored are selected by the analysts and depend on the needs of the respective physics analysis. Since Belle II has such a large data output, having all the information available as an analyst is impossible. Therefore, one must carefully consider how information is compressed into higher-level objects at each step, since, as with all compression algorithms, this results in some loss of information.

Another relatively straightforward way to reduce the volume of data each analyst needs to process and store is by using so-called skims [115, p. 51]. Skims select a subset of the available data, already excluding events that are irrelevant for specific physics analyses. For example, the skims should select $e^+e^- \to \tau^+\tau^-$ events for tau physics and mostly remove other processes. For the skim selection, the focus is more on high signal efficiency rather than retaining a very pure sample, meaning that only parts of the data with almost no signal events should be cut out. Multiple people can then use the skims as a starting point for a more refined analysis without processing the entire dataset. Skims can also be defined and processed within basf2.

The development of basf2 is ongoing and it is steadily improved with enhanced algorithms and new features. For this thesis, the used basf2 version is `light-2305-korat` [116].

3.4.7 Trigger System

The Belle II trigger system ensures that most desired events are recorded while not triggering on beam backgrounds or processes with a high cross-section, such as Bhabha events. All this is done under the constraints of limited storage capabilities and bandwidth of the DAQ system. The trigger is essential for low-multiplicity processes such as $e^+e^- \to \tau^+\tau^-$ or dark sector searches that can have as little as a single high energy photon in the final state. The trigger system comprises two main stages: the hardware-based trigger system, or level 1 (L1), and the software-based high level trigger (HLT).

The L1 trigger system is required to make a retention decision for an event in under 5 µs and still must also be capable of distinguishing between two consecutive events spaced as closely as 200 ns. The maximum output rate of the L1 trigger system is 30 kHz. At the design luminosity of SuperKEKB, the frequency of events for a process with a cross-section of 1 nb will be around 800 Hz [100]. For comparison, radiative Bhabha scattering, which is one of the most prominent physics backgrounds for many low-multiplicity processes, has a cross-section of 74 nb in the acceptance region of ECL. Therefore, the L1 trigger system must be incredibly efficient and still provide good discrimination of background sources.

Several sub-detectors provide information to the L1, most notably the CDC and the ECL. The detector signals are processed using Field Programmable Gate Arrays (FPGAs), which allows for fast low-level particle and event reconstruction. The CDC is used to reconstruct tracks from charged particles, which are either 2D or 3D tracks. A 2D track is a reconstructed track that uses only information from its projection to the r-φ plane, allowing a faster reconstruction. Novel 3D tracking algorithms using neural networks have been developed for the L1 trigger [117, 118], to also provide information about the z coordinate of tracks. This helps to distinguish background tracks not starting from the IP. The associated tracks on the trigger level are

called Neuro 3D. These 3D tracks can also match with ECL or KLM clusters, resulting in a better classification of the events. The ECL provides information on the position of the cluster and the energy deposited to the L1. This is especially important for vetoing Bhabha events, for which a special logic based on back-to-back ECL clusters was developed (this is a separate trigger called 3D Bhabha). Unfortunately, this Bhabha veto also removes some of the interesting low-multiplicity processes, so there is always a trade-off which can be controlled by the tightness of the veto.

Following the L1 trigger, events undergo a assessment through the software-driven HLT. In this phase, full event reconstruction is accessible using all sub-detectors, with the exception of the PXD, due to its extremely high data rates. The HLT is also responsible for identifying regions of interest for the PXD readout based on reconstructed tracks. Data storage limitations require the HLT to lower the event rates to approximately 10 kHz. To process the rate of 30 kHz acquired by the L1, 6000 CPU cores have to be used [100].

For this work, only a few selected L1 trigger lines will be used. These triggers have specific requirements and names, which are listed in Table 3.5 for triggers using CDC information and in Table 3.6 for ECL-based trigger lines. In addition to the listed criteria, all trigger lines require the absence of an injection veto from SuperKEKB to avoid data recording during periods of increased noise following beam injections. For CDC triggers, a full track is defined as one that passes through all the CDC layers, whereas a reduced (or 'short') track may end or exit the CDC before the last layer. This is especially important for particles with momentum in the forward or backward direction. In general, low-multiplicity events with tracks in the barrel region will have a higher trigger efficiency since the fraction of undesired background events with tracks in the endcap regions is significantly larger.

Some trigger lines may have been inactive or prescaled during some data-taking periods. A prescale of n means that only every n^{th} event triggered by the associated trigger line is kept, often used for triggers with a high rate when the overall rate is at its limit. More about the used trigger lines and their efficiency for this analysis will be provided in Section 4.4.3.

Table 3.5: Trigger names and activation conditions for different CDC-based trigger lines.

Trigger	Criteria	Comments
`ffo`	number of full 2D tracks ≥ 2 r-φ opening angle $> 90\,^\circ$ Not an ECL 3D Bhabha	prescale 100 since experiment 17
`fyo`	number of full 2D tracks ≥ 1 number of Neuro 3D tracks ≥ 1 r-φ opening angle $> 90\,^\circ$ Not an ECL 3D Bhabha	only active since experiment 12
`fso`	number of full 2D tracks ≥ 1 number of reduced 2D tracks ≥ 1 r-φ opening angle $> 90\,^\circ$ Not an ECL 3D Bhabha	only active since experiment 12 prescale 100 since experiment 17
`syo`	number of reduced 2D tracks ≥ 1 number of Neuro 3D tracks ≥ 1 r-φ opening angle $> 90\,^\circ$ Not an ECL 3D Bhabha	only active since experiment 16
`stt`	number of Neuro 3D tracks ≥ 1 $\lvert z\rvert < 15\,\mathrm{cm}$ $p > 0.7\,\mathrm{GeV}$ Not an ECL 3D Bhabha	single-track trigger for minimum bias only active since experiment 17

Table 3.6: Trigger names and activation conditions for different ECL-based trigger lines.

Trigger	Criteria	Comments
`hie`	Total ECL energy $> 1\,\mathrm{GeV}$ Not an ECL 3D Bhabha	
`lml0`	$N_{\mathrm{cluster}} \geq 3$, one with $E_{\mathrm{lab}} \geq 300\,\mathrm{MeV}$ $1 \leq \theta_{\mathrm{id}} \leq 17 \equiv 12.4^\circ < \theta_{\mathrm{lab}} < 154.7^\circ$	prescale 10 since experiment 20
`lml6`	Only one cluster with $E_{\mathrm{CM}} \geq 1\,\mathrm{GeV}$, $4 \leq \theta_{\mathrm{id}} \leq 15$ No other cluster with $E_{\mathrm{lab}} \geq 300\,\mathrm{MeV}$ anywhere	
`lml8`	Two clusters with $170^\circ < \Delta\phi_{\mathrm{CM}} < 190^\circ$ and $E_{\mathrm{lab}} > 250\,\mathrm{MeV}$ No other cluster with $E_{\mathrm{CM}} \geq 2\,\mathrm{GeV}$ anywhere	
`lml9`	Two clusters with $170^\circ < \Delta\phi_{\mathrm{CM}} < 190^\circ$ One cluster with $E_{\mathrm{lab}} > 250\,\mathrm{MeV}$, the other with $E_{\mathrm{lab}} < 250\,\mathrm{MeV}$ No other cluster with $E_{\mathrm{CM}} \geq 2\,\mathrm{GeV}$ anywhere	
`lml10`	Two clusters with $160^\circ < \Delta\phi_{\mathrm{CM}} < 200^\circ$ and $160^\circ < \sum \theta_{\mathrm{CM}} < 200^\circ$ No other cluster with $E_{\mathrm{CM}} \geq 2\,\mathrm{GeV}$ anywhere	
`lml12`	$N_{\mathrm{cluster}} \geq 3$, one with $E_{\mathrm{lab}} \geq 500\,\mathrm{MeV}$ $2 \leq \theta_{\mathrm{id}} \leq 16 \equiv 18.5^\circ < \theta_{\mathrm{lab}} < 139.3^\circ$	Tighter selection of `lml0`

3.5 Data and simulation

This section introduces the current Belle II dataset and the simulated samples used for the analysis.

3.5.1 Data

Collisions with the entire Belle II detector (except for a part of the PXD) began on March 25$^{\text{th}}$ 2019 and ended in fall 2022 for long shutdown one (LS1). The dataset collected in this period is also called pre-LS1 data. Data collection is organised into sequentially numbered experiments. Table 3.7 shows the relevant experiments (7 to 26) and the year it was taken, together with the CM energy used in the respective periods and the corresponding integrated luminosities. Figure 3.9 shows the weekly integrated recorded luminosity usable for physics

Table 3.7: The data-taking periods of Belle II and the corresponding experiment number up to 2022, together with the CM energy at which the data was taken and the corresponding integrated luminosity, where the first uncertainty is statistical while the second is systematic.

Period	Exp.	CM energy	Integrated luminosity [fb^{-1}]
early 2019	7	$\Upsilon(4S)$	$0.506 \pm 0.002 \pm 0.004$
	8	$\Upsilon(4S)$	$1.663 \pm 0.003 \pm 0.011$
		off-resonance	$0.808 \pm 0.001 \pm 0.006$
		$\Upsilon(4S)$ scan	$0.038 \pm 0.001 \pm 0.001$
late 2019	10	$\Upsilon(4S)$	$3.655 \pm 0.002 \pm 0.024$
early 2020	12	$\Upsilon(4S)$	$54.573 \pm 0.005 \pm 0.350$
		off-resonance	$8.691 \pm 0.002 \pm 0.056$
late 2020	14	$\Upsilon(4S)$	$16.500 \pm 0.006 \pm 0.106$
early 2021	16	$\Upsilon(4S)$	$10.294 \pm 0.004 \pm 0.066$
	17	$\Upsilon(4S)$	$10.715 \pm 0.004 \pm 0.069$
	18	$\Upsilon(4S)$	$89.900 \pm 0.011 \pm 0.576$
		off-resonance	$8.482 \pm 0.004 \pm 0.055$
late 2021	20	$\Upsilon(4S)$	$3.788 \pm 0.003 \pm 0.025$
	21	$\Upsilon(5S)$ scan	$19.750 \pm 0.004 \pm 0.147$
	22	$\Upsilon(4S)$	$32.060 \pm 0.007 \pm 0.206$
		$\Upsilon(5S)$ scan	$0.041 \pm 0.001 \pm 0.001$
early 2022	24	$\Upsilon(4S)$	$85.642 \pm 0.011 \pm 0.549$
	25	off-resonance	$24.574 \pm 0.006 \pm 0.158$
	26	$\Upsilon(4S)$	$54.795 \pm 0.009 \pm 0.351$

analyses and the overall recorded luminosity.

Data collected with the Belle II detector between 2019 and 2022 at a CM energy corresponding to the $\Upsilon(4S)$ resonance amounts to a total integrated luminosity of

$$\mathcal{L}_{\text{int}} = 361.654 \pm 0.021 \pm 2.170 \text{fb}^{-1}, \tag{3.6}$$

where the first uncertainty is statistical and the second is systematic. This luminosity measurement, which was obtained using Bhabha events, is provided by Belle II colleagues for the whole collaboration [120].

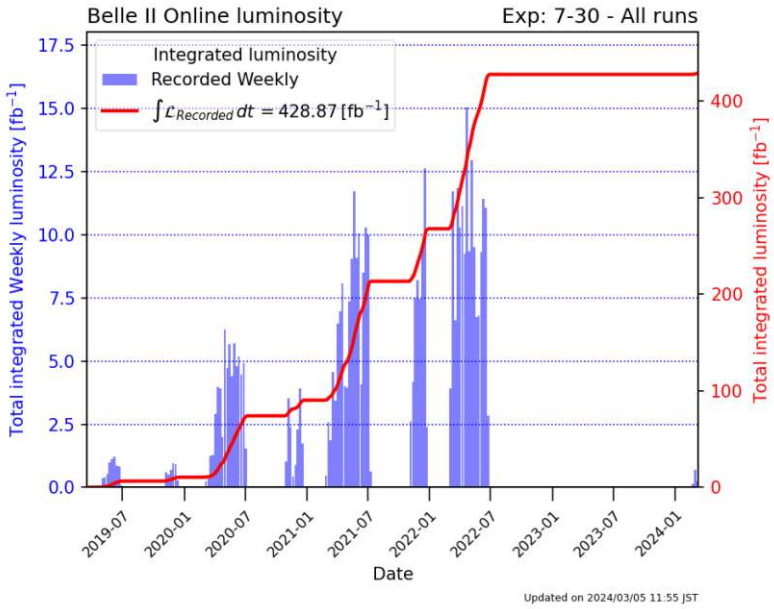


Figure 3.9: The integrated luminosity recorded by the Belle II detector between 2019 and 2024. The histogram shows the weekly luminosity, while the red line shows the total integrated luminosity. After July 2022, Belle II and SuperKEKB entered LS1, only continuing data-taking in February 2024. The image is taken from [119].

3.5.2 Cross-sections

The cross-sections for the dominant physics processes in electron-positron collisions at Belle II energies are given in Table 3.8. In some cases, restrictions to the kinematics of the final state particles are applied, such as for their momentum to lie inside the detector acceptance. These selections are often imposed at the generator level when simulating samples to limit their size [100]. The quoted cross-sections are obtained with Monte Carlo simulations using different generators (mentioned in the next section), with the quoted uncertainties only due to the limited size of the produced samples. The cross-section for $q\bar{q}$ finals states or $e^+e^- \to \Upsilon(4S)$ [10, p. 908] is not very well known and is sometimes obtained using data [115, 121].

3.5.3 Monte Carlo simulation

Monte Carlo (MC) simulations are critical in order to understand the kinematics and response of the detector for different physics events. It typically involves three steps: event generation, detector interaction simulation, and detector response digitisation.

During the event generation step, the production vertices of particles, their four-momenta, and decays are produced for particular physics processes, either following the SM or a new physics model. Several SM physics processes are used in this work to study both signal and background, using different generators that specialise in specific processes.

The $e^+e^- \to \tau^+\tau^-$ events are generated by the KKMC generator [122, 123]. Following this, the decay of tau leptons is modelled using the TAUOLA software package [124–127], and radiative corrections are applied via PHOTOS [128]. In TAUOLA, 92 decay modes are active and initialised with branching fractions following the particle listing from the PDG [10]. They are

Table 3.8: The primary physics processes occurring in e^+e^- collisions at a CM energy of 10.58 GeV, taken from „The Belle II Physics Book“ [100, p. 48].

	Process	Restrictions	Cross-section [nb]
$e^+e^- \to$	$\Upsilon(4S)$		1.110 ± 0.008
	$u\bar{u}(\gamma)$		1.61
	$d\bar{d}(\gamma)$		0.40
	$s\bar{s}(\gamma)$		0.38
	$c\bar{c}(\gamma)$		1.30
$e^+e^- \to$	$\tau^+\tau^-(\gamma)$		0.919
	$e^+e^-e^+e^-$	$W_{\ell\ell} > 0.5\,\text{GeV}$	39.7 ± 0.1
	$e^+e^-\mu^+\mu^-$	$W_{\ell\ell} > 0.5\,\text{GeV}$	18.9 ± 0.1
$e^+e^- \to$	$e^+e^-(\gamma)$	$10° < \theta_e^* < 170°$, $E_e^* > 0.15\,\text{GeV}$	300 ± 3
		$p_e > 0.5\,\text{GeV}$ and e in ECL	74.4
$e^+e^- \to$	$\gamma\gamma(\gamma)$	$10° < \theta_\gamma^* < 170°$, $E_\gamma^* > 0.15\,\text{GeV}$	4.99 ± 0.05
		$E_\gamma > 0.5\,\text{GeV}$ in ECL	3.30
$e^+e^- \to$	$\mu^+\mu^-(\gamma)$	None	1.148
		$p_\mu > 0.5\,\text{GeV}$ in CDC	0.831
$e^+e^- \to$	$\mu^+\mu^-\gamma(\gamma)$	$p_\mu > 0.5\,\text{GeV}$ in CDC, $\geq 1\gamma\,(E_\gamma > 0.5\,\text{GeV})$ in ECL	0.242
$e^+e^- \to$	$\nu\bar{\nu}(\gamma)$		0.25×10^{-3}

determined from a global fit to all available branching fraction measurements while considering the unitarity constraint. Depending on the decay channels, these τ decays compose both the signal and the background for this analysis.

KKMC is also used for the production of $\mu^+\mu^-(\gamma)$ and $q\bar{q}$, while the fragmentation of the $q\bar{q}$ pairs is handled by the PYTHIA computer program [129]. The event generator BabaYaga.NLO [130–134] is employed for $e^+e^- \to e^+e^-(\gamma)$ events, also known as Bhabha scattering, and EvtGen is utilised for the decay of B and D mesons. For the creation of $e^+e^- \to \ell^+\ell^-\ell^+\ell^-$ and $e^+e^- \to e^+e^-h'^+h'^-$, AAFH [135–137] and TREPS [138] are employed. During the generation of these events, referred to as two-photon processes, no radiative effects are included due to limitations of the generators. A comprehensive list of all generated samples and details about their size and cross-sections used in generation are provided in Table 4.1 in Section 4.2.

The generated particles undergo a complete simulation of their interactions with the detector via the GEANT4 [139] software. This process includes the creation of authentic trajectories, secondary particles, and energy depositions within the detector, also taking into account electronic effects and stochastic noise. The precise geometry and material composition of the Belle II detector are employed for these simulations.

As a last step, the consequent detector signals are used to reconstruct the events using basf2 [113]. The detector signals are acquired from the GEANT4 hits by individual modules for each detector. The L1 trigger system is also simulated, which is referred to as trigger

simulation (TSIM). Contrary to data, simulated events are not rejected, neither from the L1 nor from the HLT trigger, and only the trigger decision on whether it would keep the event is stored. This is also referred to as trigger 'flags'.

The resulting simulated samples will be referred to as MC samples. They are used to approximate the distributions of the underlying model, since an analytical description of the whole process is not feasible. However, the simulation will never be perfect and the discrepancies between reality and simulation must be carefully evaluated and, if needed, corrected.

At Belle II there are two types of MC samples: run-dependent and run-independent. Run-dependent MC uses randomly triggered data events to include beam backgrounds and evolving detector conditions that match those present during the different data-taking periods. On the other hand, run-independent MC uses simulated beam backgrounds and fixed detector conditions.

3.6 Charged particle identification

Particle identification (PID), particularly lepton identification (lepton ID), is a crucial component in this measurement. It is essential to maintain high identification efficiency while keeping the misidentification (mis-id) rate low. Although the particle identification system is also simulated in the MC samples, imperfections in the simulations must be corrected. To address this issue, data-driven corrections from calibration channels are used. However, there is an associated uncertainty with these corrections, which is expected to be the leading systematic in this measurement. This uncertainty will be independent between muons and electrons and thus not cancel in the ratio of branching fractions.

In this section, the charged particle identification of Belle II and the calibration procedure are introduced. A summary of the correction factors and associated uncertainties will be given in Sections 3.6.5 and 3.6.6. Other, more analysis-specific aspects of particle identification will only be discussed in Chapter 4. The overall performance of particle identification for the analysis will be summarised in Section 4.4.6, particularly in Table 4.14. Implementing the systematic uncertainties in the analysis will be described in Section 4.6.2.

3.6.1 Sub-detector likelihoods

The nominal particle identification at Belle II uses likelihoods to discriminate long-lived charged particles that are commonly produced: electrons, muons, pions, kaons, protons, and deuterons. Every sub-detector contributes data useful for particle identification, regardless of whether the particle leaves a detectable trace. The likelihood

$$\mathcal{L}_{\text{det}}(i) \coloneqq \mathcal{L}_{\text{det}}(\boldsymbol{x}|i) \quad \text{with} \quad i \in \{e, \mu, \pi, K, p, d\} \tag{3.7}$$

is defined for each sub-detector, det $\in$ {CDC, TOP, ARICH, ECL, KLM}[1], utilising a probability density function (PDF) for the observations $\boldsymbol{x}$, derived from simulations, data calibration channels, or analytical methods. The global likelihood for the hypothesis i combines the particle identification information from all sub-detectors and is calculated as the product $\prod_{\text{det}} \mathcal{L}_{\text{det}}(i)$, assuming the independence of each of the individual likelihoods.

[1]While in principle possible, currently no information from the SVD is used.

Using this product, the likelihood ratios

$$L(i) := \frac{\prod_{\text{det}} \mathcal{L}_{\text{det}}(i)}{\sum_j \prod_{\text{det}} \mathcal{L}_{\text{det}}(j)} \tag{3.8}$$

can be established, where the summation includes all possible particle hypotheses, and i represents the hypothesis under examination. According to Bayes' theorem [140],

$$P(i|\boldsymbol{x}) = \frac{P(\boldsymbol{x}|i)\pi(i)}{\sum_j P(\boldsymbol{x}|j)\pi(j)}, \tag{3.9}$$

the ratio $L(i)$ can be interpreted as the probability P of observing a particle of type i, when the prior probabilities $\pi(i)$ are equal. This would correspond to the case where all particle types appear with the same probability in the detector, which is usually not the case. In practice, this assumption is still used since otherwise the likelihood ratio $L(i)$ would become analysis dependent, and correction factors would have to be obtained for each analysis individually.

Some examples of inputs from sub-detectors that are used to define the likelihoods relevant to this work are:

- The CDC allows for the measurement of the characteristic energy loss dE/dx at specific momenta. This measurement is crucial for tracks with low momentum (under 1 GeV), which do not extend to external detectors intended for identifying particles.
- The relation between the deposited energy in the ECL (E_{ECL}) and the momentum of the associated track. In particular, electrons can be easily identified because their ratio E_{ECL}/p is close to one, which is not true for all other charged particles. This is because nearly all of the energy of the electron is absorbed in the ECL.
- The distinctions in longitudinal penetration depth and transverse scattering in the KLM due to differences between hadronic and electromagnetic showers. By comparing the expected patterns in the KLM with those observed from tracks extrapolated from the CDC under a muon hypothesis, it is possible to compute the probability that a track is indeed a muon. This is especially important for separating muons and pions at high momentum.

A more detailed description of all the sub-detector inputs can be found elsewhere [100].

3.6.2 Particle identification variables

Equation 3.8 establishes the *global* particle identification variables at Belle II, restricted to values between 0 and 1. In the following measurement, this is used to identify muons and will be referred to as P_μ. Electron identification is performed using a boosted decision tree (BDT) classifier, specifically trained to distinguish electrons from other charged particles [141]. This method has proven to be more effective than the purely likelihood-based approach, especially in differentiating electrons from pions, achieving lower mis-id rates at equivalent efficiency levels. The BDT utilises the likelihoods from each sub-detector and additional observables from the ECL, including variables that characterise electromagnetic showers. In the following measurement, the BDT output serves as the criterion for the identification of electrons and is denoted as P_e.

3.6.3 Calibration channels

Imperfections in the simulation are corrected on the basis of the observed differences between data and MC, measured in several calibration channels. Efficiency measurements utilise decays such as $J/\psi \to \ell^+\ell^-$ or the processes $e^+e^- \to \ell^+\ell^-\gamma$ and $e^+e^- \to e^+e^-\ell^+\ell^-$, where ℓ represents either an electron or a muon. For assessing mis-id rates, decays such as $K^0_S \to \pi^+\pi^-$ and $e^+e^- \to \tau^+\tau^-$ events, with the decay $\tau^\pm \to \pi^\pm\pi^\mp\pi^\pm\nu_\tau$, are used. These correction factors are obtained globally by the Belle II lepton ID performance group and then provided to all analysts.

The efficiencies and mis-id rates can be obtained by fitting a resonance peak over a flat background in an invariant mass spectrum, such as for the J/ψ or K^0_S, which could be produced from different processes. However, this technique is not possible for calibration channels using low-multiplicity events because there is no discriminating mass peak. Instead, a method called tag and probe is used. For this, a pure sample of one particle species is obtained (the tag selection), and the efficiency is calculated using a PID requirement on this particle candidate (the probe selection). For this method to work, the probe selection must be applied to a particle (the probe track) with no particle identification bias due to the tag selection. If the probe selection is independent of the tag selection, any influence from the tag selection should be cancelled in the efficiency calculation. Often, the tag selection involves another particle identification requirement on a different track (the tag track) since the same particle types are frequently produced in pairs. The following gives two examples of calibration channels where this technique is used. This should not go into great detail and only selected results will be shown. I was involved in both performance studies, particularly the taupair study was set up and maintained by me [142].

Tau calibration channel

The idea of the tau calibration channel is to select taupair events in which one of the taus decays into three pions (and possibly additional π^0s) and the other with a decay that only includes one charged particle. These events can be easily identified because of the distinct topology. Because the decay products of the tau are boosted with respect to the CM frame, these events have three tracks with small opening angles and one track in the opposite direction. The relevant tau decays to three charged particles and their branching fractions are listed in Table 3.9. The decay is dominated by pions with a small fraction of charged kaons. Since the decay $\tau^- \to \pi^- K^+ \pi^- \nu_\tau$ is heavily suppressed, one can further suppress a lot of kaon contamination by tagging the two same charged tracks as pions and probing the differently charged track. With this strategy and an optimised selection, a sample consisting of about 98 % pions, with the largest contamination from kaons (0.9 %) was obtained. The total number of pions in this sample corresponds to around 4 300 000 per 100 fb^{-1}.

The momentum and polar angle distributions of the probe track are shown in Figure 3.10. Since the PID performance is highly dependent on the momenta and polar angle of the particles, the correction factors are binned in these two dimensions and separated according to charge. As apparent from the distributions, this study has only a sufficient number of events from 200 MeV up to $\sim$ 4 GeV, and more toward the barrel region of the detector.

Table 3.9: Tau decays into three charged particles and their branching fractions [10]. Charge-conjugated modes are implied.

Tau decay	Branching fraction
$\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$	$9.31 \pm 0.05\,\%$
$\tau^- \to \pi^-\pi^+\pi^-\pi^0\nu_\tau$	$4.62 \pm 0.05\,\%$
$\tau^- \to K^-\pi^+\pi^-\nu_\tau$	$0.345 \pm 0.007\,\%$
$\tau^- \to K^-\pi^+\pi^-\pi^0\nu_\tau$	$0.131 \pm 0.012\,\%$
$\tau^- \to K^-K^+\pi^-\nu_\tau$	$0.1435 \pm 0.0027\,\%$
$\tau^- \to K^-K^+\pi^-\pi^0\nu_\tau$	$0.0061 \pm 0.0018\,\%$

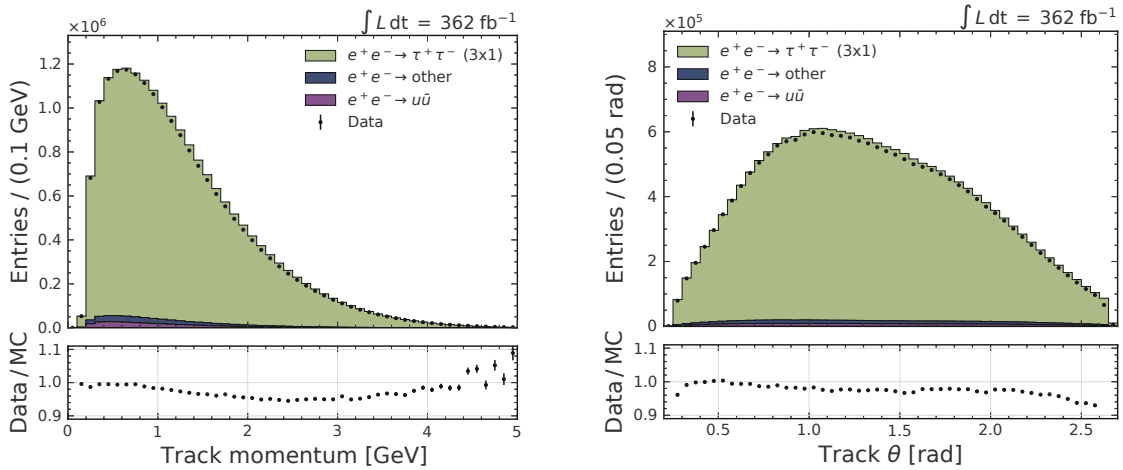


Figure 3.10: The momentum (left) and polar angle (right) of the pion candidate selected from the differently charged track when reconstructing $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$. The stacked histograms show taupair events that match the desired topology, $e^+e^- \to u\bar{u}$ events, and all other events.

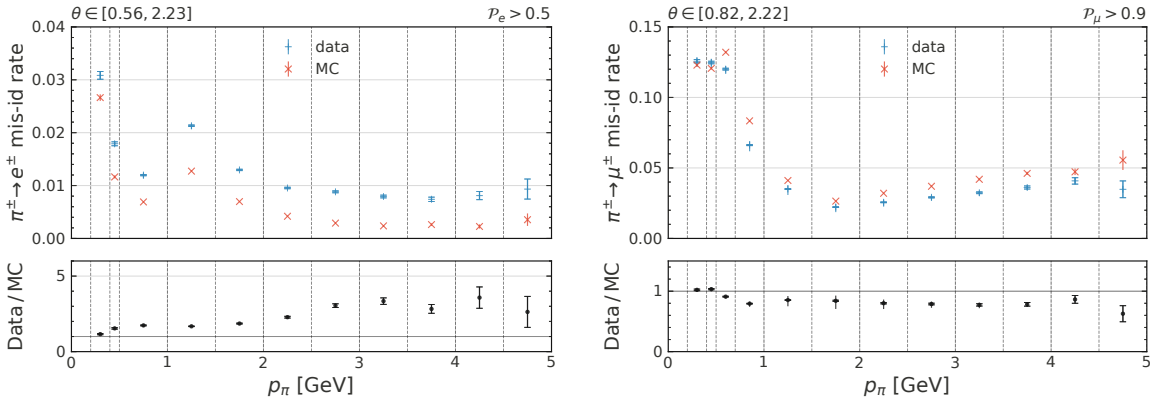
For every correction bin, the mis-id rates are computed in the data using

$$\epsilon_{\text{data}} = \frac{N_{\text{pid}} - \sum_j n^j_{\text{pid}} \cdot r_j}{N - \sum_j n^j} \quad \text{with} \quad j \in \{e, \mu, K, p, d\}, \tag{3.10}$$

where N represents the total number of events in data (after analysis and tag selections), and N_{pid} indicates the number of data events passing the cut on the PID variable associated with the probe track. This number of events has to be corrected for their contamination with different particle types. These corrections are derived from MC, where n^j represents the count of background events for each particle type j, and n^j_{pid} indicates the events passing the probe selection. Since the cut on the PID variable will also introduce some data/MC discrepancies for n^j_{pid}, it has to be corrected with the factor r_i This correction factor is the data/MC ratio of mis-id rates for the particle of type j to be misidentified as a pion, which can be obtained from other performance studies. If this correction factor is not available, a value of one with conservative uncertainties can be used instead. Although background contamination is small, the uncertainty on the correction factor r_i is one of the dominant systematic uncertainties for ϵ_{data}, together with the systematic arising from the trigger. The mis-id rates in MC can be calculated by restricting the sample to true pions. The statistical uncertainties, which become

dominant at high momentum, are propagated for both the mis-id rate in data and in MC.

An example of such obtained mis-id rates and their ratio is shown in Figure 3.11, for $\pi^\pm \to e^\pm$ on the left and $\pi^\pm \to \mu^\pm$ on the right. The mis-id rates are displayed across different momentum bins and integrated over the whole barrel region of the ECL ($\theta \in [0.56, 2.23]$) and the KLM ($\theta \in [0.82, 2.22]$), respectively. The horizontal lines on the error bar indicate the statistical uncertainty from data, while the total error bar is the quadratic sum of the statistical and systematic components. At the selected working points, the $\pi^\pm \to e^\pm$ mis-id rate rate for high momentum is below 1 %, while for $\pi^\pm \to \mu^\pm$ it ranges from 3 % to 5 %. Especially for the $\pi^\pm \to \mu^\pm$ case, a significant increase in mis-id rates can be observed for momenta below 1 GeV. This is due to tracks not reaching the KLM and the pion muon separation predominantly relying on the TOP and ARICH detectors, as well as the ECL. For the $\pi^\pm \to e^\pm$ case instead, the separation power relies mainly on ECL information and only at low momentum the CDC becomes important, where also a rise in mis-id rates can be observed. The correction factors for $\pi^\pm \to e^\pm$ are quite large (a factor of 3 for high momentum), which is due to the imperfect simulation of the ECL and CDC. However, it is also important to consider the size of the mis-id rates, which in this case is relatively small. Thus, the large correction factors and associated uncertainties will have a reduced impact when selecting electrons.



(a) $\pi^\pm \to e^\pm$ mis-id rate

(b) $\pi^\pm \to \mu^\pm$ mis-id rate

Figure 3.11: The mis-id rates and their ratio in different momentum bins, as obtained from the taupair calibration channel. On the left, a PID selection of $P_e > 0.5$ is applied, while on the right, $P_\mu > 0.9$ is required. The mis-id rates are shown integrated over the barrel region of the ECL and the KLM, respectively. The error bar shows the total uncertainty, with horizontal lines indicating the statistical uncertainty from the data.

Bhabha calibration channel

The Bhabha calibration channel uses the abundant amount of Bhabha events produced at Belle II, which can be selected to obtain a very clean sample of electrons. The main challenge of this study is to account for the L1 and HLT triggers, which are designed to veto Bhabha events. With only $e^+e^- \to e^+e^-$ events, the momentum of the electrons is very restricted and, because of the boost of the CM frame, correlated with their polar angle. Therefore, for lower momentum electrons, this study relies mainly on radiative Bhabha events ($e^+e^- \to e^+e^-\gamma$),

where the γ takes away momentum from the electrons, which allows them to cover a larger phase space.

The calculation is very similar to the taupair study, with the advantage that the purity of the Bhabha sample is around 99.99 %. With no neutrinos in the final state, one can also require closed kinematics by constraining the invariant mass or total CM energy of the $e^+e^-(\gamma)$ system to the CM energy of 10.58 GeV. Due to the high purity, there is almost no systematics due to background contamination, and the main systematic uncertainty originates from observed differences between correction factors when requiring different L1 or HLT selections. This does not necessarily imply that the trigger lines cause these differences, but it could also be due to different effects arising from the selection conditions.

Nevertheless, the usage of triggers for this study imposes some complications, such as accounting for the prescale factors of the triggers or the definitions themselves, which can change during the data-taking. The varying prescales can impact the computed efficiencies since, when they are not accounted for, different weights are given to the various periods, resulting in biased corrections since the PID performance is not constant over time. This is especially important since the prescales for Bhabha-related triggers can be very large and vary between 1 and 1000. Accounting for prescales becomes even more involved since some of the already prescaled HLT lines require a specific L1 line, which can also be prescaled. Since trigger conditions constantly evolve, this study must be re-examined frequently to ensure that everything is accounted for. Several checks were performed on the dataset used in this analysis, indicating that all prescales are taken into account.

The different conditions for which the electron efficiency correction factors are calculated are summarised in Table 3.10. The trigger conditions for the L1 and HLT lines mentioned in the table are described in the following, where $\Sigma\theta_{\mathrm{CM}}$ is the sum of the polar angles of 2 ECL clusters in the CM frame, $\Delta\phi_{\mathrm{CM}}$ is the difference of the azimuthal angles of 2 ECL clusters in the CM frame and $E(\mathrm{CL}X)$ is energy of the X^{th} ECL cluster in the CM frame, with $X \in \{1, 2\}$.

L1 trigger: 3D Bhabha

This trigger requires two back-to-back tracks based on polar and azimuthal angles and thresholds on the energy of the associated clusters in the ECL. Starting from experiment 10, this is equivalent to the 3D Bhabha veto.

- $165° < \Sigma\theta_{\mathrm{CM}} < 190°$
- $160° < \Delta\phi_{\mathrm{CM}} < 200°$
- $E(\mathrm{CL1}) > 3\,\mathrm{GeV}$ *and* $E(\mathrm{CL2}) > 3\,\mathrm{GeV}$
- $E(\mathrm{CL1}) > 4.5\,\mathrm{GeV}$ *or* $E(\mathrm{CL2}) > 4.5\,\mathrm{GeV}$

L1 trigger: Bhapur

The requirement for kinematics is looser than that of the 3D Bhabha bit, resulting in lower purity and higher efficiency.

- $160° < \Sigma\theta_{\mathrm{CM}} < 200°$
- $140° < \Delta\phi_{\mathrm{CM}} < 220°$

Table 3.10: The different configurations to select $e^+e^- \to e^+e^-(\gamma)$ events. They are used to derive a systematic uncertainty for the electron efficiency corrections obtained with the Bhabha calibration channel.

Name	Requirements
eegamma+hie	L1: hie HLT: n2GeVPhotonBarrel ≥ 1 *or* radBhabha $(E/p)_{\text{tag}} > 0.8$ - radBhabha condition (when active) b2b veto (not passing conditions of L1 3D Bhabha)
eegamma+hie (no bkg correction)	L1: hie HLT: n2GeVPhotonBarrel ≥ 1 *or* radBhabha $(E/p)_{\text{tag}} > 0.8$ - radBhabha condition (when active) b2b veto (not passing conditions of L1 3D Bhabha) ignore backgrounds in efficiency computation
eegamma+lml1	L1: lml1 HLT: n2GeVPhotonBarrel ≥ 1 *or* radBhabha $E_{\text{tag}}^{\text{ECL}} > 2\,\text{GeV}, 0.61 < \theta_{\text{tag}} < 2.15$ - lml1 condition $(E/p)_{\text{tag}} > 0.8$ - radBhabha condition (when active) b2b veto (not passing conditions of L1 3D Bhabha)
eegamma+lml1 (allow b2b)	L1: lml1 HLT: n2GeVPhotonBarrel ≥ 1 *or* radBhabha $E_{\text{tag}}^{\text{ECL}} > 2\,\text{GeV}, 0.61 < \theta_{\text{tag}} < 2.15$ - lml1 condition $(E/p)_{\text{tag}} > 0.8$ - radBhabha condition (when active)
selectee+bhapur	L1: bhapur HLT: selectee
selectee+bhapur+lml1	L1: bhapur *and* lml1 HLT: selectee $E_{\text{tag}}^{\text{ECL}} > 2\,\text{GeV}, 0.61 < \theta_{\text{tag}} < 2.15$ - lml1 condition
ee_flat+lml1	L1: lml1 *and* un-prescaled bhapur (via HLT) HLT: exactly one of the eeFlat lines above 30° $E_{\text{tag}}^{\text{ECL}} > 2\,\text{GeV}, 0.61 < \theta_{\text{tag}} < 2.15$ - lml1 condition

- $E(\text{CL1}) > 2.5\,\text{GeV}$ *and* $E(\text{CL2}) > 2.5\,\text{GeV}$
- $E(\text{CL1}) > 4\,\text{GeV}$ *or* $E(\text{CL2}) > 4\,\text{GeV}$

L1 trigger: lml1

- one ECL cluster with $E(\text{CL1}) > 2\,\text{GeV}$ and $4 \leq \theta_{\text{id}} \leq 14$ (ECL barrel)

HLT line: radBhabha

This trigger selects radiative Bhabha events without explicitly reconstructing the photon but using requirements for the missing momentum.

- two tracks with $p^{\text{CM}} > 1\,\text{GeV}$
- $p^{\text{CM}}_{\text{missing}} > 1\,\text{GeV}$ *and* $10° < \theta^{\text{CM}}_{\text{missing}} < 170°$
- $\Delta\phi_{\text{CM}} < 170°$
- E_{ECL}/p (CM) > 0.8 (one of the tracks)

HLT line: n2GeVPhotonBarrel > N

- more than N ECL clusters with energy $> 2\,\text{GeV}$ and $32\,° < \theta < 130\,°$

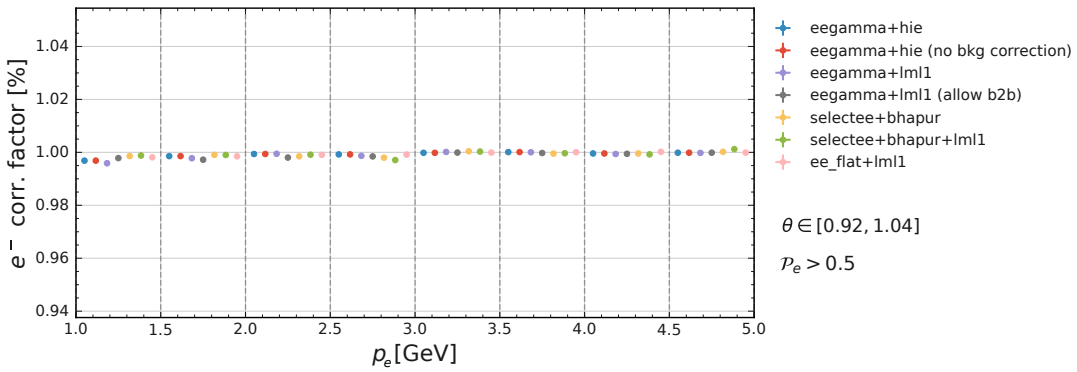
HLT lines: selectee and eeFlat

The main difference between selectee and eeFlat is that eeFlat is divided into slices of polar angle, with prescales depending on the expected rate of Bhabha tracks in these regions, while selectee is for the entire polar angle region and only has one prescale factor. In addition, they have different requirements on the L1, which also come with different prescales. All of this results in significantly fewer events kept by the selectee trigger in the barrel region. The selection requirements, however, are identical:

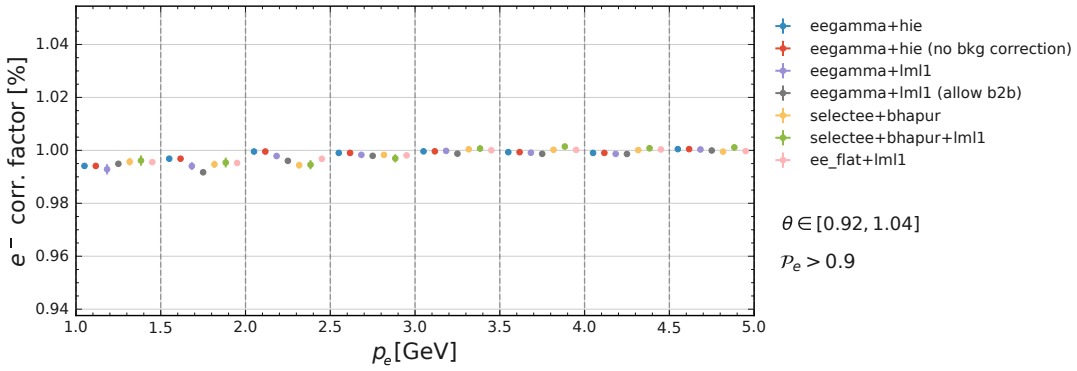
- $M(\text{track}^+, \text{track}^-) > 5.29\,\text{GeV}$
- $E(\text{CL1}) > 1.5\,\text{GeV}$ *or* $E(\text{CL2}) > 1.5\,\text{GeV}$

The conditions in Table 3.10 can roughly be categorised into requirements for back-to-back tracks (when using 'selectee' or 'eeFlat') and the orthogonal selection that vetoes back-to-back tracks since it is intended to select radiative Bhahba events ('eegamma'). It is important to note that events with back-to-back tracks where the 4-momenta of the electrons do not sum up to the 4-momentum of the e^+e^- collisions will likely be radiative Bhabha events. This will be predominantly the case when computing electron efficiency corrections with electron momentum below 5 GeV, which are most relevant for many analyses.

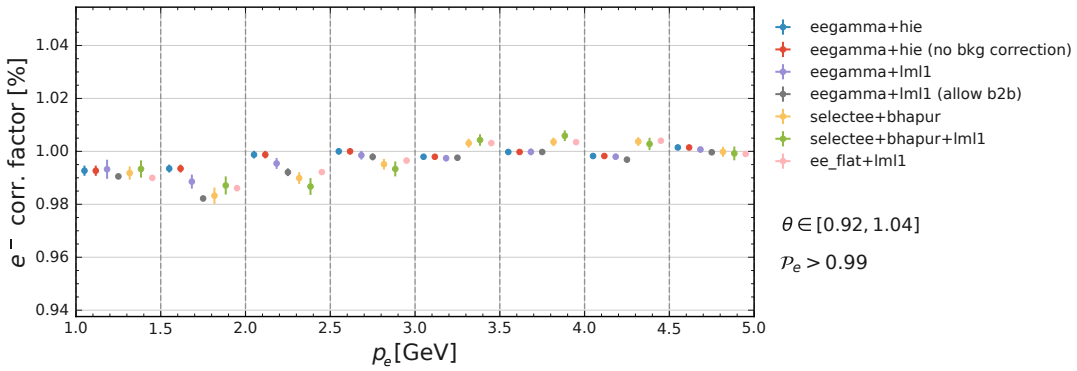
The correction factors obtained for each of the different requirements are shown in Figure 3.12. The correction factors are shown between 1 GeV and 5 GeV, in bins of 0.5 GeV (the requirements are shown next to each other in each bin only for presentation purposes) and only for a small polar angle range $\theta \in [0.92, 1.04]$. This example was only chosen to illustrate the behaviour. In other polar angle regions, the observations are very similar. For loose cuts on the electron identification variable, such as $P_e > 0.5$ (3.12a), the differently selected samples are in relatively good agreement with each other, and the correction factor is very close to one. Still, there are some slight differences at the level of $\sim 0.2\,\%$. When tightening the selection requirements to



(a) The correction factors using $P_e > 0.5$.



(b) The correction factors using $P_e > 0.9$.



(c) The correction factors using $P_e > 0.99$.

Figure 3.12: The electron efficiency correction factors obtained with different selection requirements imposed by the choice of the L1 and HLT lines. The correction factors are shown in bins of 0.5 GeV, ranging from 1 GeV to 5 GeV and for electrons with $\theta \in [0.92, 1.04]$. Three selections on the electron identification variable P_e are compared, with increasing tightness from top to bottom.

Table 3.11: The polar angle binning used for lepton ID corrections. The correspondence between the bins and the detector regions is illustrated in Figure 3.13.

θ [rad]	θ [deg]	$\cos\theta$	Correspondence
(0.22, 0.40)	(12.61°, 22.92°)	(0.98, 0.92)	1
(0.40, 0.56)	(22.92°, 32.09°)	(0.92, 0.85)	2
(0.56, 0.64)	(32.09°, 36.67°)	(0.85, 0.80)	3
(0.64, 0.73)	(36.67°, 41.83°)	(0.80, 0.75)	4
(0.73, 0.82)	(41.83°, 46.98°)	(0.75, 0.68)	5
(0.82, 0.92)	(46.98°, 52.71°)	(0.68, 0.61)	6
(0.92, 1.04)	(52.71°, 59.59°)	(0.61, 0.51)	7
(1.04, 1.16)	(59.59°, 66.46°)	(0.51, 0.40)	8
(1.16, 1.31)	(66.46°, 75.06°)	(0.40, 0.26)	9
(1.31, 1.46)	(75.06°, 83.65°)	(0.26, 0.11)	10
(1.46, 1.54)	(83.65°, 88.24°)	(0.11, 0.03)	11
(1.54, 1.60)	(88.24°, 91.67°)	(0.03, −0.03)	12
(1.60, 1.78)	(91.67°, 101.99°)	(−0.03, −0.21)	13
(1.78, 1.94)	(101.99°, 111.15°)	(−0.21, −0.36)	14
(1.94, 2.04)	(111.15°, 116.88°)	(−0.36, −0.45)	15
(2.04, 2.13)	(116.88°, 122.04°)	(−0.45, −0.53)	16
(2.13, 2.22)	(122.04°, 127.20°)	(−0.53, −0.60)	17
(2.22, 2.40)	(127.20°, 137.51°)	(−0.60, −0.74)	18
(2.40, 2.60)	(137.51°, 148.97°)	(−0.74, −0.86)	19
(2.60, 2.71)	(148.97°, 155.27°)	(−0.86, −0.91)	20

$P_e > 0.9$ (3.12b) or $P_e > 0.99$ (3.12c), the discrepancies become more significant and are at the percent level. In particular, there seems to be a difference between the back-to-back events and the complementary selection. One hypothesis is that the back-to-back requirement selects events where the radiative photon is collinear with one of the tracks, or there could be two (or more) radiated photos whose momentum sum coincides with the direction of one of the electrons. When these photons are close to the ECL cluster of the electron, this could alter the PID performance [see 142, Chapter 4]. More studies are needed to fully understand these discrepancies. The nominal correction factor is derived from eegamma+hie, and a systematic uncertainty based on these observed differences is assigned.

3.6.4 Binning of the correction factors

As mentioned, the lepton ID correction factors are binned in momentum, polar angle, and charge. The polar angle binning is presented in Table 3.11. The selection of bin edges is influenced by key areas (like gaps or instrumentation) and the segmentation of the sub-detectors. A vertical cross section of the Belle II detector, which highlights the areas belonging to the various bins, is depicted in Figure 3.13.

The momentum binning differs slightly between electron and muon-related corrections at low momentum. While for electron-related corrections the bin edges are $\{0.2, 0.4, 0.5, 1.0\}$GeV, for

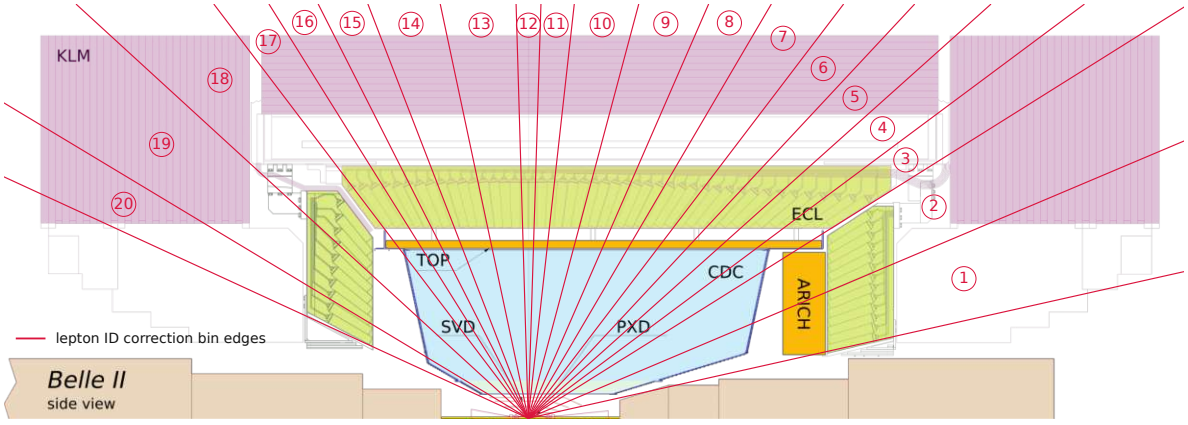


Figure 3.13: A cross section of the Belle II detector viewed from the side, showing the lepton ID segments in polar angle. The highlighted edges of the correction bins match those listed in Table 3.11. The illustration of the Belle II detector is adapted from [98, p. 14].

muon-related corrections they are $\{0.2, 0.4, 0.5, 0.7, 1.0\}$GeV. The additional bin is substantial, given that muons with momentum less than 0.7 GeV typically fail to reach the KLM. This significantly impacts the identification efficiencies and correction factors, shown in the next section. Above 1 GeV, the binning is regular with a bin width of 0.5 GeV. While efficiency corrections are available up to 7 GeV, the $h^{\pm} \to \ell^{\pm}$ mis-id rate corrections are only available up to 5 GeV, and also only for the polar angle bins in the forward direction. This is due to the limited number of high momentum tracks in the corresponding calibration channels, such as the taupair channel presented in Section 3.6.3.

3.6.5 Correction factors

The data/MC efficiency corrections for electrons and muons are shown in Figure 3.14, with the binning defined in Section 3.6.4. The regions where different calibration channels were used to compute the corrections are highlighted in the figure. The kinematic reach is limited for some calibration channels, especially for momentum above 3 GeV, where only the $\mu\mu\gamma$ and $ee\gamma$ channels provide corrections. The correction factor is close to 1.0 for the electron efficiency, with the exception of the endcap regions. For muon efficiency, the correction factor is close to 1.0 in the barrel region above 1 GeV. The drop in the data/MC ratio below 1 GeV can be attributed to the fact that muon identification and its simulation become more challenging at low momentum because particles no longer traverse all KLM layers. The discrimination also relies on other sub-detectors likelihoods, which are not so well modelled in MC.

In Figure 3.15, the mis-id rate data/MC correction factors are shown for pions that are misidentified as electrons or muons.

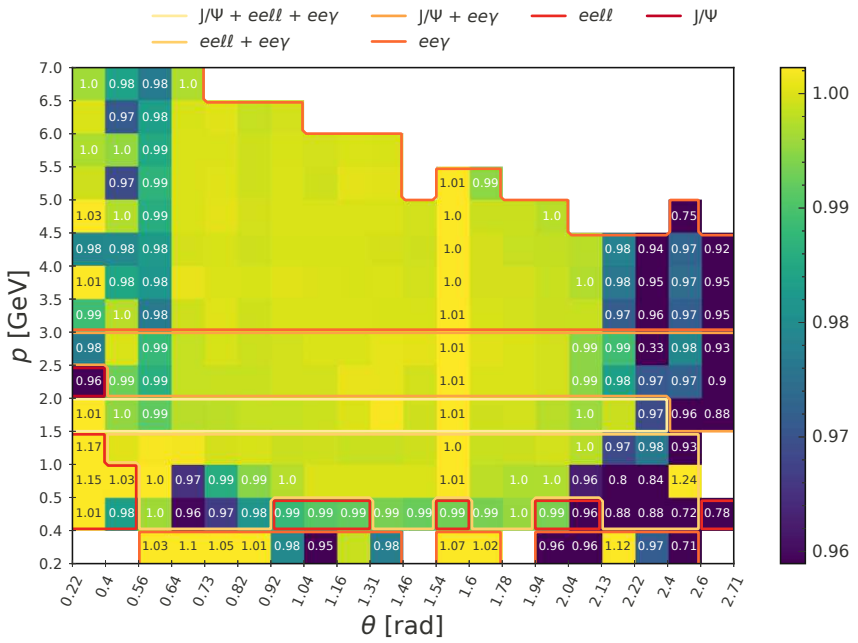
3.6.6 Uncertainties of the corrections

The correction factors come with an associated uncertainty, divided into statistical and systematic components. The systematic component contains the differences between the correction factors obtained from the individual calibration channels when they are incompatible. This can be attributed to effects of the event topology, such as track isolation. The correlation

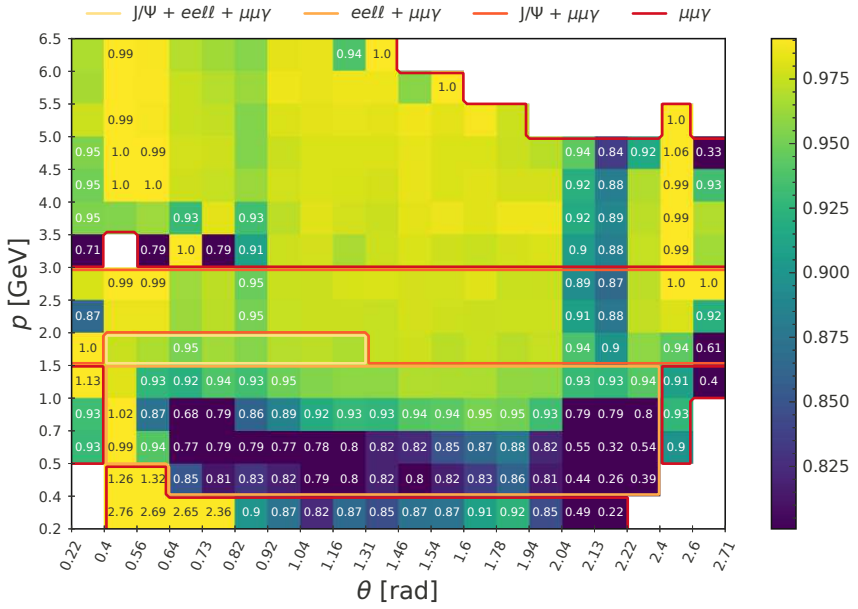
of systematic uncertainties between the correction bins is currently not known, and some conservative assumptions will have to be made, described in more detail in Section 4.6.2. The size of the total uncertainty (the two components added in quadrature) of electron efficiency and muon efficiency corrections is shown in Figure 3.16. In general, the uncertainties can be asymmetric, with different values for the up and down components. Only the larger components are shown, which will be used later to obtain symmetric uncertainties. The uncertainties are expressed as a percentage of the correction factor.

The uncertainty is relatively small in most bins, at the order of 0.1 %. For electron efficiency corrections, there is a significant increase in uncertainties in the endcap regions ($\theta \in [0.22, 0.56]$ and $\theta \in [2.23, 2.71]$). For muon efficiency corrections, the uncertainty increases in regions with limited KLM coverage and below 1 GeV. This coincides with regions with a correction factor away from 1 or regions where inefficiencies are observed in MC. The larger uncertainties just above 3 GeV are due to increased statistical uncertainties from the $e^+e^- \to \mu\mu\gamma$ calibration channel, which requires that at least one of the muon candidates has momentum below 3 GeV. This means that just above that threshold, both muons need to lose some of their energy to pass the selection. For higher momentum muons (with p_μ^{CM} around half of the CM energy), the number of available events increases again, and the uncertainty is similar to right below 3 GeV.

In Figure 3.15, the uncertainties associated with the mis-id rate correction factors are shown for pions that are misidentified as electrons or muons.

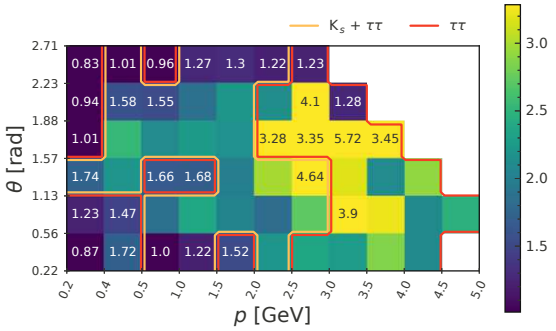


(a) The e^- efficiency data/MC correction factors when using $P_e > 0.5$.

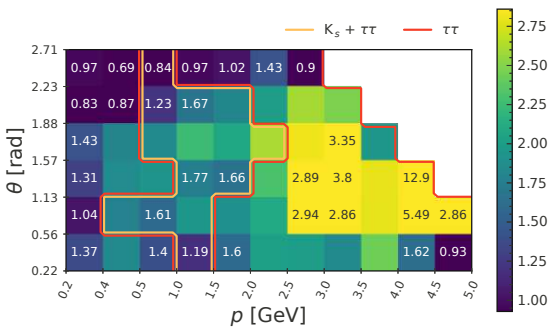


(b) The μ^- efficiency data/MC correction factors when using $P_\mu > 0.9$.

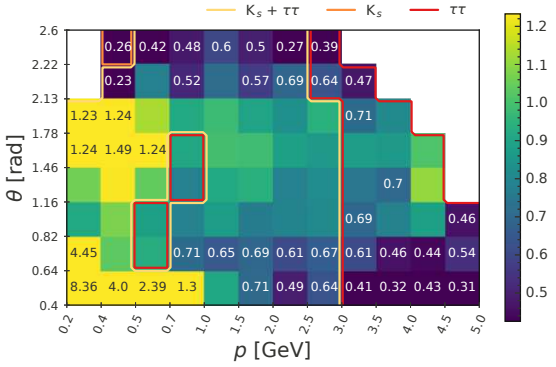
Figure 3.14: data/MC correction factors for identification of electrons (top) and muons (bottom). The highlighted regions indicate where different calibration channels were used to obtain the correction. In the barrel region above 1 GeV the correction factors are very close to 1.



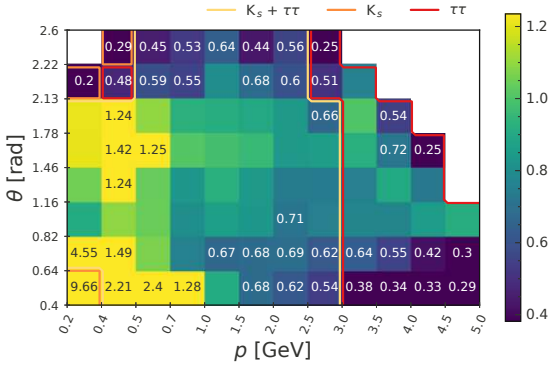
(a) The $\pi^- \to e^-$ mis-id rate data/MC correction factors when using $P_e > 0.5$.



(b) The $\pi^+ \to e^+$ mis-id rate data/MC correction factors when using $P_e > 0.5$.

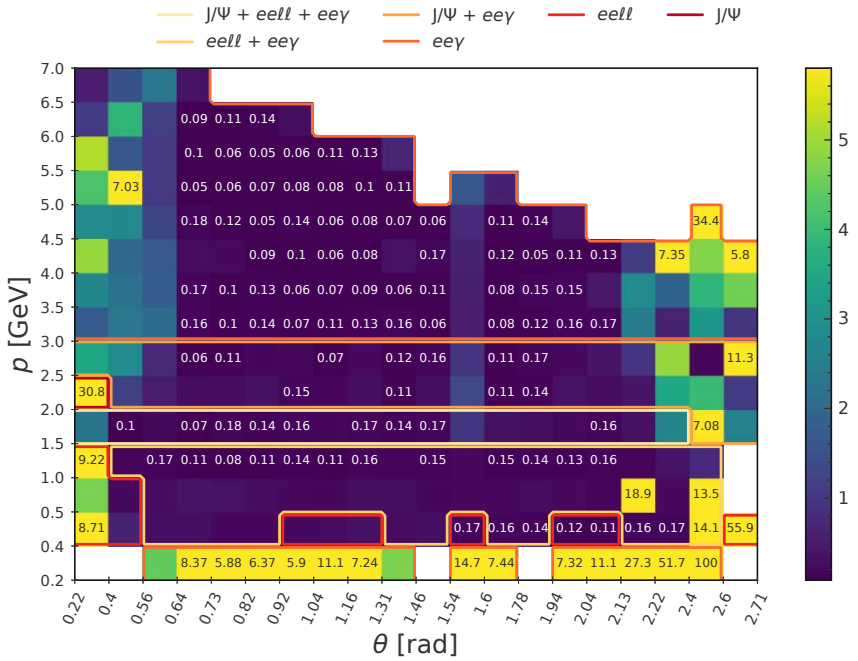


(c) The $\pi^- \to \mu^-$ mis-id rate data/MC correction factors when using $P_\mu > 0.9$.

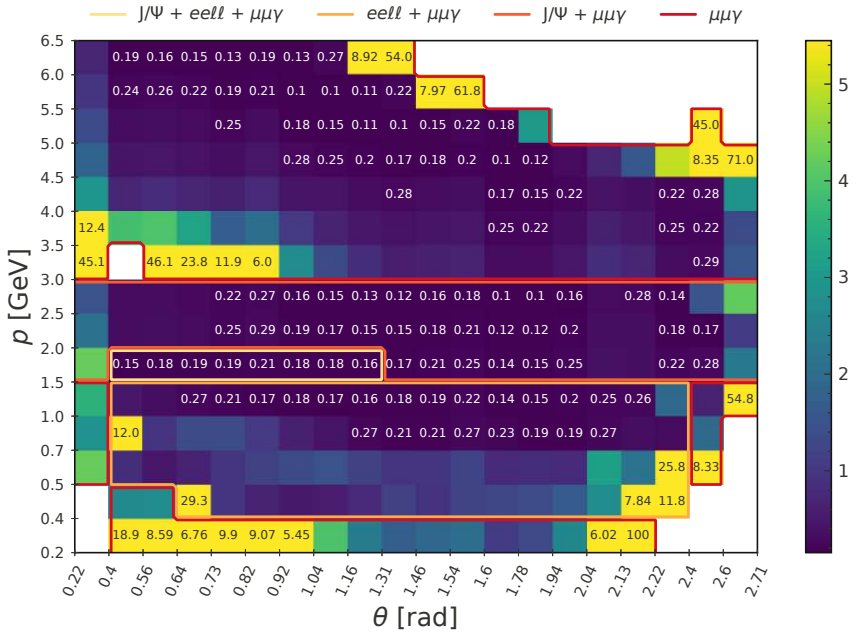


(d) The $\pi^+ \to \mu^+$ mis-id rate data/MC correction factors when using $P_\mu > 0.9$.

Figure 3.15: data/MC correction factors for the mis-id rate of identifying pions as electrons (top) or muons (bottom). Correction factors are shown for negatively charged (left) and positively charged (right) pions. The highlighted regions indicate where different calibration channels were used to obtain the correction.

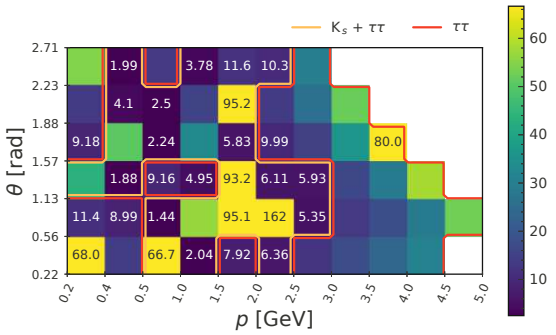


(a) The e^- efficiency correction uncertainties when using $P_e > 0.5$.

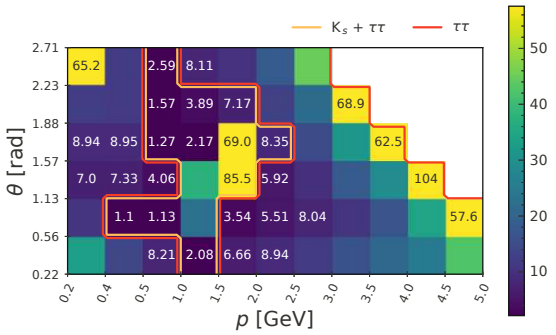


(b) The μ^- efficiency correction uncertainties when using $P_\mu > 0.9$.

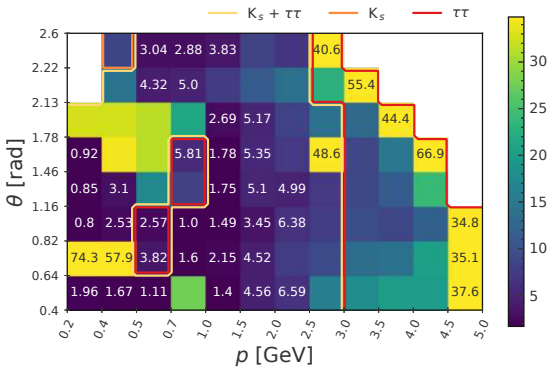
Figure 3.16: The relative uncertainties associated with electron (top) and muon (bottom) efficiency corrections in percent. The statistical and systematic uncertainties are added in quadrature, and only the larger of the up or down components is shown. The highlighted regions indicate where different calibration channels were used to obtain the correction.



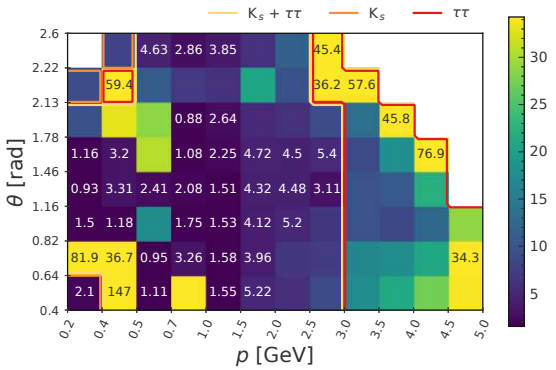
(a) The $\pi^- \to e^-$ mis-id rate correction uncertainties when using $P_e > 0.5$.



(b) The $\pi^+ \to e^+$ mis-id rate correction uncertainties when using $P_e > 0.5$.



(c) The $\pi^- \to \mu^-$ mis-id rate correction uncertainties when using $P_\mu > 0.9$.



(d) The $\pi^+ \to \mu^+$ mis-id rate correction uncertainties when using $P_\mu > 0.9$.

Figure 3.17: The relative uncertainties associated with the mis-id rate of identifying pions as electrons (top) or muons (bottom) in percent. The uncertainties are shown for negatively charged (left) and positively charged (right) pions. The statistical and systematic uncertainties are added in quadrature, and only the larger of the up or down components is shown. The highlighted regions indicate where different calibration channels were used to obtain the correction.

CHAPTER 4

Analysis procedure

4.1 Introduction

In this chapter, the analysis procedure is presented. First, Section 4.2 describes the datasets used for this analysis. In Section 4.3, the different decay channels that could be used to measure R_μ with the 1x1 topology and their advantages and drawbacks are discussed. Then, Section 4.4 elaborates on event selection and background suppression. Section 4.5 introduces the statistical model that will be used to extract R_μ from the data, and in Section 4.6, the systematic uncertainties and how they are included in the fit model are discussed. Finally, in Section 4.7, the impact of these systematic effects on the measurement is evaluated.

4.2 Datasets

Data This analysis uses all available data taken at the $\Upsilon(4S)$ resonance before fall 2022 when Belle II entered LS1. More details on the data-taking periods and the integrated luminosities were provided in Section 3.5.1. The data were reprocessed with the 13^{th} processing campaign of Belle II, with prompt calibration after experiment 20. It corresponds to a total integrated luminosity of about $362\,\text{fb}^{-1}$ (see also 3.6). Only $\Upsilon(4S)$ data, also referred to as on-resonance data, are used for this work, since the main limitations on the precision of the results come from systematic effects, and the addition of off-resonance data would only introduce complications due to the change in CM energy. Most importantly, the correction factors to account for differences in particle identification performance between data and simulation are only available for the on-resonance dataset.

Simulation The MC samples used for this work originate from the 15^{th} MC production campaign of Belle II, using run-independent settings. More details on the MC simulation at Belle II were provided in Section 3.5.3. An overview of the simulated processes and the size of the generated samples is given in Table 4.1. After the initial event selection described in Section 4.4, which was optimised with all the listed samples, it was found that most of the processes can be completely suppressed. In the end, the only simulated samples that contribute to the analysis are $q\bar{q}$, $\tau\tau$, ee, $\mu\mu$, $eeee$, $ee\mu\mu$ and $ee\tau\tau$.

Table 4.1: The simulated processes along with their respective cross-sections, number of generated events and luminosity equivalent. The charge of the particles is omitted. Only the processes up to $ee \to ee\tau\tau$ contribute to the background after an initial analysis selection.

Process	Cross-section [nb]	Number of events	MC luminosity [fb^{-1}]
$ee \to \tau\tau$	0.919	919×10^6	1000
$ee \to d\bar{d}$	0.401	401×10^6	1000
$ee \to c\bar{c}$	1.329	1329×10^6	1000
$ee \to s\bar{s}$	0.383	383×10^6	1000
$ee \to u\bar{u}$	1.605	1605×10^6	1000
$ee \to ee(\gamma)$	295.8	$29\,580 \times 10^6$	100
$ee \to eeee$	39.55	7910×10^6	200
$ee \to ee\mu\mu$	18.83	3766×10^6	200
$ee \to \mu\mu(\gamma)$	1.148	1148×10^6	1000
$ee \to ee\tau\tau$	0.01836	37×10^6	2000
$ee \to B^+B^-$	0.54	108×10^6	200
$ee \to B^0\overline{B}^0$	0.51	102×10^6	200
$ee \to \gamma\gamma$	5.1	2550×10^6	500
$ee \to ee\pi\pi$	1.895	1895×10^6	1000
$ee \to \pi\pi\gamma$	1.667×10^{-1}	333×10^6	2000
$ee \to eeKK$	7.98×10^{-2}	160×10^6	2000
$ee \to \pi\pi\pi^0\gamma$	2.378×10^{-2}	47.6×10^6	2000
$ee \to KK\gamma$	1.63×10^{-2}	32.6×10^6	2000
$ee \to eepp$	1.17×10^{-2}	23×10^6	2000
$ee \to K^0\bar{K}^0\gamma$	8.864×10^{-3}	17.7×10^6	2000
$ee \to \mu\mu\mu\mu$	3.512×10^{-4}	0.7×10^6	2000
$ee \to \mu\mu\tau\tau$	1.441×10^{-4}	0.29×10^6	2000
$ee \to \tau\tau\tau\tau$	0.2114×10^{-6}	0.002×10^6	10000

Currently, $e^+e^- \to e^+e^-Nh$ processes with $h \in \{\pi, K\}$ and $N > 2$ are not simulated, even though they can contribute to the background of this measurement. This background will be removed using data-driven selections, described in Section 4.4.4.

4.3 Sensitivity of different decay channels

This section compares the different possible τ decay channels with the 1x1 topology concerning their sensitivity to measure R_μ. Although an inclusive measurement could be attempted that requires a leptonic decay for one of the τ leptons and a generic decay for the other, this has multiple drawbacks. For one, it can result in higher backgrounds as it is more difficult to optimise the selection compared to explicitly requiring a specific decay channel. Moreover, it results in less control over the systematic effects, which have to be considered for each of the contributing channels, as they might differ between them. While it would result in a larger dataset, there is no merit in these complications for a systematically limited measurement. It is better to split the dataset into specific channels which can be optimised and treated individually and always be combined later on. Here and in the following, the different requirements for tau

decays in the 1x1 topology will be simply referred to as channels.

The studies presented here were carried out early in the analysis before a clear measurement strategy was developed. The main goal was to decide which channels are the most promising for performing the measurement or whether all channels should be used simultaneously. It is important to note that the lepton ID corrections and uncertainties initially used for the comparisons underwent significant changes in the following years until they reached the precision they had at the time of writing. To have a consistent picture, all studies were updated with the most recent lepton ID corrections. This reduced the estimated systematic uncertainties associated with lepton ID compared to the original study, but the conclusions remain the same. The selection initially used to acquire the samples used for this study was left unchanged as it does not significantly impact the validity of the findings. However, as it was understood at a later stage, for an accurate measurement of R_μ, more care must be taken about what features are used for the selection of events. What is shown here should be considered an optimal selection based only on MC simulations without considering the accurate modelling of the chosen features or the systematic uncertainties associated with imposing a selection on them. In total, seven different taupair channels were selected:

1. $(\tau \to \mu\nu_\mu\nu_\tau) \times (\tau \to \mu\nu_\mu\nu_\tau)$
2. $(\tau \to e\nu_e\nu_\tau) \times (\tau \to \mu\nu_\mu\nu_\tau)$
3. $(\tau \to e\nu_e\nu_\tau) \times (\tau \to e\nu_e\nu_\tau)$
4. $(\tau \to \mu\nu_\mu\nu_\tau) \times (\tau \to h\nu_\tau)$, $h \in \{\pi, K\}$
5. $(\tau \to e\nu_e\nu_\tau) \times (\tau \to h\nu_\tau)$, $h \in \{\pi, K\}$
6. $(\tau \to \mu\nu_\mu\nu_\tau) \times (\tau \to hn\pi^0\nu_\tau)$, $h \in \{\pi, K\}$ and $n > 0$
7. $(\tau \to e\nu_e\nu_\tau) \times (\tau \to hn\pi^0\nu_\tau)$, $h \in \{\pi, K\}$ and $n > 0$

With these seven channels, in the following abbreviated only by using the charged particles and by $hn\pi^0$ for the decay $\tau \to hn\pi^0\nu_\tau$, four different ratios of products of branching fractions can be formed to compute R_μ:

1. $\mathcal{B}(\mu) \times \mathcal{B}(\mu)$ / $\mathcal{B}(e) \times \mathcal{B}(\mu)$
2. $\mathcal{B}(e) \times \mathcal{B}(\mu)$ / $\mathcal{B}(e) \times \mathcal{B}(e)$
3. $\mathcal{B}(\mu) \times \mathcal{B}(h)$ / $\mathcal{B}(e) \times \mathcal{B}(h)$
4. $\mathcal{B}(\mu) \times \mathcal{B}(hn\pi^0)$ / $\mathcal{B}(e) \times \mathcal{B}(hn\pi^0)$

In each of these ratios, which are not independent of each other, one of the branching fractions cancels. The corresponding tau decay will be termed the tag decay. The sensitivity of each of these ratios to measure R_μ will depend on the size of the samples, their purity, and their sensitivity to systematic effects, particularly lepton ID correction uncertainties. Many common systematic effects, such as luminosity uncertainty or systematics associated with the tag decay, will cancel when computing R_μ.

Table 4.2: Preselection criteria for different taupair channels used to study their sensitivity. The selection for $\ell \times hn\pi^0$ is the same as for $\ell \times h$. The meaning of the variables and the units used are detailed separately in Table 4.3.

$\mu \times \mu$	$e \times \mu$	$e \times e$	$\mu \times h$	$e \times h$		Variable		$\mu \times \mu$	$e \times \mu$	$e \times e$	$\mu \times h$	$e \times h$
1.33	0.49	0.31	0.22	0.21	$\leq$	$p^{\mathrm{CM}}_{t,\mathrm{max}}$	$<$	5.06	5.05	5.17	5.04	5.16
1.65	0.75	7.04	-13.29	-12.89	$\leq$	m^2_{missing}	$<$	59.27	97.39	71.50	99.49	83.08
0.41	0.01	1.28	0.11	0.03	$\leq$	$p^{\mathrm{CM}}_{t,\mathrm{missing}}$	$<$	4.46	4.35	3.88	4.79	4.76
0.51	0.23	0.46	0.20	0.38	$\leq$	$\theta^{\mathrm{CM}}_{\mathrm{missing}}$	$<$	2.64	2.90	2.61	2.88	2.77
2.51	1.08	1.98	2.65	2.47	$\leq$	$E^{\mathrm{CM}}_{\mathrm{visible}}$	$<$	8.75	9.39	10.00	9.10	8.62
1.66	1.37	1.50	1.09	1.24	$\leq$	$\alpha^*_{(\mathrm{sig,tag})}$	$<$	3.13	3.14	2.91	3.14	3.00
0.10	0.00	0.02	0.02	0.00	$\leq$	$\theta^{\mathrm{CM}}_{\mathrm{min,missing}}$	$<$	1.32	1.53	1.51	1.47	1.52

4.3.1 Background suppression

Preselected samples for each of the seven channels were obtained using the selection requirements listed in Table 4.2. The meaning of the symbols used, the reference frame in which they are computed, and the units used are all detailed separately in Table 4.3. Here and in the following, they will be referred to as features or variables, which are used interchangeably. The smallest possible polar angle of any of two missing particles, $\theta^{\mathrm{CM}}_{\mathrm{min,missing}}$, is useful for suppressing the final states of $e^+e^-\ell^+\ell^-$ or $\ell^+\ell^-\gamma\gamma$ with ℓ being a muon or an electron [11]. It is obtained using

$$\theta^{\mathrm{CM}}_{\mathrm{min,missing}} = \sin^{-1}\left(\frac{p^{\mathrm{CM}}_{t,\mathrm{missing}}}{\sqrt{s} - E^{\mathrm{CM}}_{\mathrm{visible}}}\right). \tag{4.1}$$

Table 4.3: The description of the variables used for preselecting different taupair channels, as well as the reference frame in which they are computed, and the units in which the selection requirements of Table 4.2 are provided.

Variable	Description	Frame	Unit
$p^{\mathrm{CM}}_{t,\mathrm{max}}$	higher transverse momentum of both tracks	CM	GeV/c
m^2_{missing}	missing mass squared	-	GeV$^2/c^4$
$p^{\mathrm{CM}}_{t,\mathrm{missing}}$	missing transverse momentum	CM	GeV/c
$\theta^{\mathrm{CM}}_{\mathrm{missing}}$	polar angel of missing momentum	CM	GeV/c
$E^{\mathrm{CM}}_{\mathrm{visible}}$	total visible energy	CM	GeV
$\alpha^*_{(\mathrm{sig,tag})}$	opening angle between the tracks	CM	rad
$\theta^{\mathrm{CM}}_{\mathrm{min,missing}}$	minimum polar angle of any of two missing particles	CM	rad

Using these preselections as a starting point, for each channel, an ANN was trained to suppress the backgrounds. For the training of the neural networks, 28 features associated with event kinematics, the kinematics of the tracks in the individual hemispheres, and the reconstruction of neutral particles were used. In addition, the ratio of their deposited energy in the ECL to their reconstructed momentum (E_{ECL}/p) was used as an input for the charged particles.

This is generally not recommended, as it is a discriminating feature for particle identification, whose mismodelling the measurement of R_μ is particularly sensitive to. Using it in an ANN makes the estimation of the associated systematics very difficult. Since the goal of this study is only to provide an estimate of the sensitivity, not an actual measurement, these systematics can be ignored. The achieved selection is used as an 'optimal' selection, which is also why a relatively large number of features were used. Later, the systematics associated with lepton ID are calculated only based on the selections on the variables P_e and P_μ, which were defined in Section 3.6.2. The receiver operating characteristic (ROC) curves of the ANNs are shown in Figure 4.1 in terms of their signal efficiency and purity. The starting points of the curves indicate the efficiency and purity of the preselection. The signal efficiency is computed as

$$\varepsilon_{\text{sig}} = \frac{N_{\text{sig}}^{\text{sel}}}{N_{\text{sig}}^{\text{gen}}} = \frac{N_{\text{sig}}^{\text{sel}}}{N_{\tau\tau}^{\text{gen}} \cdot \mathcal{B}_i^{\text{gen}} \cdot \mathcal{B}_j^{\text{gen}} \cdot (2 - \delta_j^i)}, \tag{4.2}$$

where $N_{\text{sig}}^{\text{sel}}$ is the number of remaining signal events in MC after applying all selections, but already corrected for the differences between data and simulation introduced by the lepton ID selection. $N_{\text{sig}}^{\text{gen}}$ indicates the number of generated signal events, which can be calculated from the product of the total number of generated taupair events, $N_{\tau\tau}^{\text{gen}}$, with the branching fractions of the tau decays selected in the corresponding channel: $\mathcal{B}_i^{\text{gen}}$ and $\mathcal{B}_j^{\text{gen}}$. The factor $(2 - \delta_j^i)$ is equal to 1 when the signal is defined as both tau leptons decaying to the same final state, while it is 0 when they are different final states. This accounts for the fact that each final state can be associated with the positively charged or negatively charged tau. For the branching fractions, it is important to note that they should correspond to what was used in the generation of the MC samples. The purity is defined as the fraction of signal events present in the sample

$$p_{\text{sig}} = \frac{N_{\text{sig}}^{\text{sel}}}{N_{\text{sig}}^{\text{sel}} + N_{\text{bkg}}^{\text{sel}}}, \tag{4.3}$$

where $N_{\text{bkg}}^{\text{sel}}$ denotes the total number of background events. When estimating these numbers from MC, it is important to take into account the different luminosity equivalents of the samples by scaling the events accordingly.

To obtain the final samples, a cut is placed on the output of each ANN, which yields an efficiency similar to that achieved by CLEO in their combined measurement of branching fractions using the same channels [11]. The purity, signal efficiency, and the corresponding number of expected events for $100\,\text{fb}^{-1}$ are shown in Table 4.4. A more detailed breakdown of the composition of the samples in terms of different tau decays and non-taupair backgrounds, also compared to the composition of the samples obtained by CLEO, is shown in Table 4.5 and Table 4.6. In the tables, the green dots (●) indicate which tau decays are defined as signal for the respective samples, while the orange dots (●) indicate signal tau decays from other samples that are background for the particular sample.

4.3.2 Calculation of R_μ

Using a simple counting approach, R_μ could be obtained from the relation between the observed number of signal events, $N_{\text{sig}}^{\text{obs}}$, and the number of signal events produced in the data, $N_{\text{sig}}^{\text{prod}}$, multiplied with the efficiencies associated with the selection of these events, estimated from the simulation:

$$N_{\text{sig}}^{\text{obs}} = \varepsilon_{\text{sig}} \cdot \varepsilon_{\text{sig}}^{\text{trig}} \cdot N_{\text{sig}}^{\text{prod}}. \tag{4.4}$$

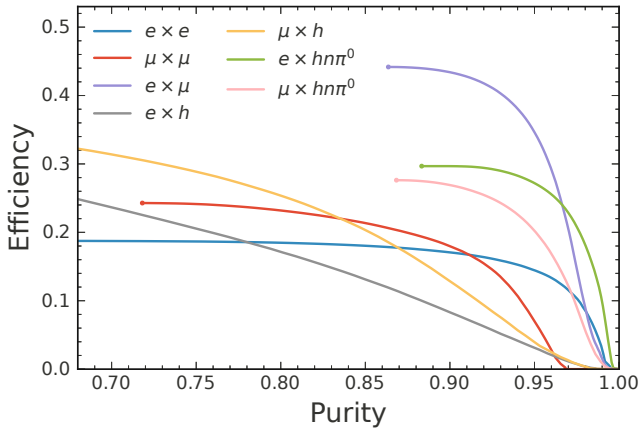


Figure 4.1: ROC curves of the ANNs trained to suppress background for each of the investigated taupair channels, showing the trade-off between signal efficiency and purity. The starting points of the curves indicate the efficiency and purity of the preselection.

Table 4.4: A summary of the selection results obtained for the different channels. The purity of the samples, the signal efficiency, and the corresponding number of expected events for 100 fb^{-1} are shown. The selection on the output of the neural networks was chosen so that the efficiency of the signal is comparable to what was achieved by the CLEO analysis [11].

Mode	Purity	Efficiency	Events per 100 fb^{-1}
$\mu \times \mu$	95.4 %	5.3 %	148 220
$e \times \mu$	97.8 %	10.6 %	603 890
$e \times e$	97.2 %	11.0 %	320 371
$\mu \times h$	94.3 %	5.2 %	190 522
$e \times h$	92.9 %	5.3 %	198 989
$\mu \times hn\pi^0$	98.0 %	5.0 %	580 796
$e \times hn\pi^0$	99.3 %	5.1 %	612 057

Here, the two efficiencies are the signal efficiency ε_{sig}, which incorporates effects from acceptance and event selection, and the trigger efficiency $\varepsilon_{\text{sig}}^{\text{trig}}$, which could, for example, be estimated from data. In general, this product of efficiencies should contain all efficiencies or data-based correction factors that can alter the estimated signal efficiency. The exact form of this product depends on what is already contained in the calculation of ε_{sig}. In the following, the signal efficiency is obtained as defined in Equation 4.2. The number of signal events produced in data can instead be computed as

$$N_{\text{sig}}^{\text{prod}} = \sigma_{\tau\tau} \cdot \mathcal{L} \cdot \mathcal{B}_i \cdot \mathcal{B}_j \cdot (2 - \delta_j^i), \tag{4.5}$$

where $\sigma_{\tau\tau}$ is the cross-section of the process $e^+e^- \to \tau^+\tau^-$, $\mathcal{L}$ is the total integrated luminosity as measured in data and $\mathcal{B}_i$ and $\mathcal{B}_j$ are the true branching fractions of the tau decays in the selected channel.

Inserting these quantities into Equation 4.4 and we obtain

$$N^{\text{obs}} - N_{\text{bkg}}^{\text{sel}} \cdot \varepsilon_{\text{bkg}}^{\text{trig}} \equiv \frac{N_{\text{sig}}^{\text{sel}} \cdot \varepsilon_{\text{sig}}^{\text{trig}}}{N_{\tau\tau}^{\text{gen}} \cdot \mathcal{B}_i^{\text{gen}} \cdot \mathcal{B}_j^{\text{gen}} \cdot (2 - \delta_j^i)} \cdot \sigma_{\tau\tau} \cdot \mathcal{L} \cdot \mathcal{B}_i \cdot \mathcal{B}_j \cdot (2 - \delta_j^i). \tag{4.6}$$

Table 4.5: A detailed breakdown of the composition of the samples $\mu \times \mu$, $e \times e$, $e \times \mu$ and $\mu \times h$ in terms of different tau decays and non-taupair backgrounds. The tau decay for the signal and tag side is shown if applicable. Additionally, the corresponding number of expected events for 100 fb^{-1}, the fraction of these events in the total sample, and a comparison to the sample composition obtained by CLEO [11] are shown.

		Signal side	Tag side	Events/100 fb^{-1}	Fraction [%]	CLEO [%]
$\mu \times \mu$	●	$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to \mu\nu_\mu\nu_\tau$	148 220.0	95.4	95.9
	●	$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to \pi\nu_\tau$	5215.0	3.4	3.37
	●	$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to \pi\pi^0\nu_\tau$	655.0	0.4	0.18
		$ee\mu\mu$		414.5	0.3	-
		$ee\tau\tau$		225.2	0.1	0.11
		$\mu\mu(\gamma)$		189.5	0.1	0.22
		cosmic		-	-	0.15
$e \times e$	●	$\tau \to e\nu_e\nu_\tau$	$\tau \to e\nu_e\nu_\tau$	320 371.0	97.2	98.38
	●	$\tau \to e\nu_e\nu_\tau$	$\tau \to h\nu_\tau$	-	-	0.78
		$ee(\gamma)$		5683.0	1.7	0.32
		$eeee$		1433.5	0.4	0.23
		$ee\tau\tau$		485.7	0.1	0.24
$e \times \mu$	●	$\tau \to e\nu_e\nu_\tau$	$\tau \to \mu\nu_\mu\nu_\tau$	603 890.0	97.8	97.62
	●	$\tau \to e\nu_e\nu_\tau$	$\tau \to \pi\nu_\tau$	8170.0	1.3	1.96
	●	$\tau \to e\nu_e\nu_\tau$	$\tau \to \pi\pi^0\nu_\tau$	1316.0	0.2	0.11
		$ee\mu\mu$		1450.5	0.2	0.13
		$ee\tau\tau$		989.0	0.2	0.12
$\mu \times h$	●	$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to \pi\nu_\tau$	178 938.0	88.6	79.99
	●	$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to K\nu_\tau$	11 584.0	5.7	
	●	$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to \pi\pi^0\nu_\tau$	4339.0	2.1	3.71
		$\tau \to \pi\pi^0\nu_\tau$	$\tau \to \pi\nu_\tau$	262.0	0.1	
		$\tau \to \pi\nu_\tau$	$\tau \to \pi\nu_\tau$	2558.0	1.3	1.83
	●	$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to \mu\nu_\mu\nu_\tau$	1979.0	1.0	9.14
	●	$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to e\nu_e\nu_\tau$	518.0	0.3	2.62
		$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to hK^0_L\nu_\tau$	-	-	1.02
		$ee\tau\tau$		385.4	0.2	0.12
		cosmic		-	-	0.13
		$ee\mu\mu$		-	-	0.9
		$\mu\mu(\gamma)$		-	-	0.34

Table 4.6: A detailed breakdown of the composition of the samples $e \times h$, $\mu \times hn\pi^0$ and $e \times hn\pi^0$ in terms of different tau decays and non-taupair backgrounds. The tau decay for the signal and tag side is shown if applicable. Additionally, the corresponding number of expected events for 100 fb^{-1}, the fraction of these events in the total sample, and a comparison to the sample composition obtained by CLEO [11] are shown.

		Signal side	Tag side	Events/100 fb^{-1}	Fraction [%]	CLEO [%]
$e \times h$	●	$\tau \to e\nu_e\nu_\tau$	$\tau \to \pi\nu_\tau$	186 991.0	87.3	82.57
	●	$\tau \to e\nu_e\nu_\tau$	$\tau \to K\nu_\tau$	11 998.0	5.6	
	●	$\tau \to e\nu_e\nu_\tau$	$\tau \to \pi\pi^0\nu_\tau$	9273.0	4.3	3.61
	●	$\tau \to e\nu_e\nu_\tau$	$\tau \to \mu\nu_\mu\nu_\tau$	1760.0	0.8	9.02
	●	$\tau \to e\nu_e\nu_\tau$	$\tau \to e\nu_e\nu_\tau$	1128.0	0.5	2.58
		$\tau \to e\nu_e\nu_\tau$	$\tau \to \pi K_L^0\nu_\tau$	334.0	0.2	1.03
		$\tau \to h\nu_\tau$	$\tau \to h\nu_\tau$	-	-	0.36
		$ee(\gamma)$		874.0	0.4	-
		$ee\tau\tau$		416.0	0.2	0.32
		$eeee$		339.5	0.2	0.33
$\mu \times hn\pi^0$	●	$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to \pi\pi^0\nu_\tau$	543 966.0	91.8	96.91
	●	$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to \pi\pi^0\pi^0\nu_\tau$	33 605.0	5.7	
	●	$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to K\pi^0\nu_\tau$	3225.0	0.5	
		$\tau \to \pi\nu_\tau$	$\tau \to \pi\pi^0\nu_\tau$	4849.0	0.8	1.76
	●	$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to h\nu_\tau$	-	-	
		$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to \pi\pi^0\gamma\nu_\tau$	1871.0	0.3	-
		$\tau \to \pi\pi^0\nu_\tau$	$\tau \to \pi\pi^0\nu_\tau$	1205.0	0.2	0.08
		$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to h\pi^0 K_L^0\nu_\tau$	-	-	0.75
		$ee\tau\tau$		1156.1	0.2	-
$e \times hn\pi^0$	●	$\tau \to e\nu_e\nu_\tau$	$\tau \to \pi\pi^0\nu_\tau$	596 663.0	96.8	98.32
	●	$\tau \to e\nu_e\nu_\tau$	$\tau \to \pi\pi^0\pi^0\nu_\tau$	13 499.0	2.2	
	●	$\tau \to e\nu_e\nu_\tau$	$\tau \to K\pi^0\nu_\tau$	1895.0	0.3	
		$\tau \to e\nu_e\nu_\tau$	$\tau \to \pi\pi^0\gamma\nu_\tau$	1904.0	0.3	< 0.34
	●	$\tau \to e\nu_e\nu_\tau$	$\tau \to h\nu_\tau$	-	-	0.32
		$\tau \to e\nu_e\nu_\tau$	$\tau \to h\pi^0 K_L^0\nu_\tau$	-	-	0.78
		$\tau \to e\nu_e\nu_\tau$	$\tau \to hK_L^0\nu_\tau$	-	-	0.1
		$ee\tau\tau$		924.9	0.2	0.1

Here, the observed number of signal events was estimated as the difference between the total number of observed events, N^{obs}, and the product of the background events estimated from MC, $N^{\text{sel}}_{\text{bkg}}$, with the relevant efficiencies, $\varepsilon^{\text{trig}}_{\text{bkg}}$. Rearranging Equation 4.6 to get an expression for the product of the branching fractions, we obtain

$$\mathcal{B}_i \cdot \mathcal{B}_j = \frac{N^{\text{obs}} - N^{\text{sel}}_{\text{bkg}} \cdot \varepsilon^{\text{trig}}_{\text{bkg}}}{N^{\text{sel}}_{\text{sig}} \cdot \varepsilon^{\text{trig}}_{\text{sig}}} \cdot \frac{N^{\text{gen}}_{\tau\tau}}{\sigma_{\tau\tau} \cdot \mathcal{L}} \cdot \mathcal{B}^{\text{gen}}_i \cdot \mathcal{B}^{\text{gen}}_j. \tag{4.7}$$

This can now be used to compute R_μ as the ratio of two products of branching fractions,

$$R_\mu = \frac{\mathcal{B}_{\tau\to\mu\nu\nu}}{\mathcal{B}_{\tau\to e\nu\nu}} = \frac{\frac{N^{\text{obs}} - N^{\text{sel}}_{\text{bkg}} \cdot \varepsilon^{\text{trig}}_{\text{bkg}}}{N^{\text{sel}}_{\text{sig}} \cdot \varepsilon^{\text{trig}}_{\text{sig}}}(\tau \to \mu\nu_\mu\nu_\tau, \tau \to X)}{\frac{N^{\text{obs}} - N^{\text{sel}}_{\text{bkg}} \cdot \varepsilon^{\text{trig}}_{\text{bkg}}}{N^{\text{sel}}_{\text{sig}} \cdot \varepsilon^{\text{trig}}_{\text{sig}}}(\tau \to e\nu_e\nu_\tau, \tau \to X)} \times \frac{\mathcal{B}^{\text{gen}}_{\tau\to\mu\nu\nu}}{\mathcal{B}^{\text{gen}}_{\tau\to e\nu\nu}}, \tag{4.8}$$

where one of the tau decays in the numerator should be $\tau \to \mu\nu_\mu\nu_\tau$, one of the tau decays in the denominator should be $\tau \to e\nu_e\nu_\tau$, and the respective other tau decay should be to the same final state, here denoted as $\tau \to X$.

4.3.3 Uncertainty estimation

In the following, Equation 4.8 is used only to estimate the precision of R_μ obtained with different channels. For this, the number of observed events is set equal to the total number of events expected from MC, scaled to the target luminosity. The expected statistical precision is then obtained from the standard deviation of a Poisson distribution with a mean equal to this number of events, propagated to R_μ.

Similarly, to see the impact of the uncertainties associated with the lepton ID corrections, the corresponding uncertainties on $N^{\text{sel}}_{\text{sig}}$ and $N^{\text{sel}}_{\text{bkg}}$ are propagated to R_μ. Different values can be obtained depending on the assumptions on the correlations of the systematics uncertainties between the individual correction bins, denoted $\rho^{\text{bins}}_{\text{sys}}$. This correlation is currently unknown. An optimistic assumption of a correlation of $\rho^{\text{bins}}_{\text{sys}} = 0.5$ and the most conservative assumption of $\rho^{\text{bins}}_{\text{sys}} = 1$ will be tested for the sensitivity estimation. This matter of bin-by-bin correlations of the lepton ID correction uncertainties will be further discussed in Section 4.6.2. To estimate the impact on R_μ, it is essential to take into account the correlations between the numerator and the denominator in Equation 4.8, when the decay $\tau \to X$ also includes a final state lepton. The correlations between all seven product branching fractions, indicated by the name of the corresponding channel, are shown in Figure 4.2 for both cases of $\rho^{\text{bins}}_{\text{sys}} = 0.5$ and $\rho^{\text{bins}}_{\text{sys}} = 1.0$. These correlations were obtained with toy samples, analogous to the procedure described in Section 4.6.2. The most relevant correlations in Figure 4.2 are between $\mu \times \mu$ and $e \times \mu$, as well as $e \times \mu$ and $e \times e$ since the channels with hadrons are completely independent when the accompanying leptons are of a different type. The correlation between $\mu \times \mu$ and $e \times \mu$ is less than between $e \times \mu$ and $e \times e$, since the lepton ID uncertainties associated to electron efficiency dominates over the muon efficiency uncertainties. An interesting feature of taking the ratios between channels with two leptons is that while assuming a larger bin-by-bin correlation will result in larger uncertainties for each branching fraction product, this also results in a larger correlation between the numerator and the denominator, and thus more cancellation for R_μ. However, in the present cases the assumption of $\rho^{\text{bins}}_{\text{sys}} = 1.0$, still results in the largest uncertainty on R_μ, but the difference between $\rho^{\text{bins}}_{\text{sys}} = 0.5$ and $\rho^{\text{bins}}_{\text{sys}} = 1.0$ is not as significant as

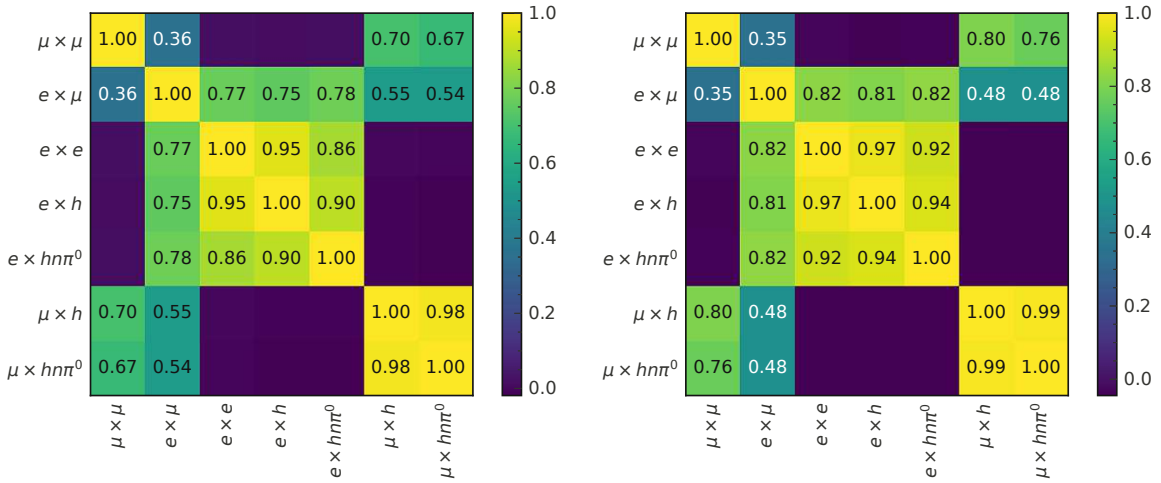


Figure 4.2: Correlations of lepton ID uncertainties between the seven different products of branching fractions, using the optimistic assumption of the bin-by-bin correlations of $\rho_{\text{sys}}^{\text{bins}} = 0.5$ (left) and the most conservative assumption of $\rho_{\text{sys}}^{\text{bins}} = 1.0$ (right). The correlation is important to consider when computing R_μ as the ratio of two of the branching fraction products.

for the channels with hadrons. This can be seen in Figure 4.3, where the estimated lepton ID systematic is shown as a hatched bar for each ratio. The right end of the bar shows the uncertainty of R_μ obtained with $\rho_{\text{sys}}^{\text{bins}} = 1.0$, while the left end shows the uncertainty of R_μ obtained with $\rho_{\text{sys}}^{\text{bins}} = 0.5$. In addition, the figure shows the estimated statistical precision using the selections obtained with the cut on the ANN output, with the event yields summarised in Table 4.4. For this, three different integrated luminosities are compared: $362\,\text{fb}^{-1}$, which corresponds to the luminosity used for the present measurement, $1\,\text{fb}^{-1}$ and $5\,\text{fb}^{-1}$. The latter are reasonable estimates for the datasets of Belle II in the near future. The figure also indicates the uncertainty due to lepton ID systematics and the statistical precision of the *BABAR* measurement [12] as well as the uncertainty due to lepton ID systematics of the CLEO measurement [11]. It is important to note that these uncertainties for the different channels should only give an indication of their differences. The actual values of the uncertainties are only very rough estimates and are based on many assumptions. The selections also do not take into account other systematics, such as background normalisation or the mismodelling of some of the features used. The selection cannot only have a significant impact on the estimated statistical precision, but also on the systematics associated with lepton ID since the lepton ID uncertainties depend on the kinematics of the leptons. Another critical factor is the number of pions that fake leptons in the individual channels, since in general the mis-id rate uncertainties are larger than those associated with lepton efficiencies. Nevertheless, the present study indicates that computing R_μ from the ratio $\mu \times hn\pi^0/e \times hn\pi^0$ results in the best precision, both for lepton ID systematics as well as the statistical uncertainty. This can be explained by the fact that with the $\tau \to hn\pi^0\nu_\tau$ decay, a very pure signal sample can be selected and that the corresponding tau decays have a substantial branching fraction (35.4 %, see Figure 2.5 in Section 2.4). In addition, the presence of only a single lepton in the event simplifies the analysis in terms of taking into account lepton ID effects.

With respect to the actual measurement with the $\ell \times hn\pi^0$ channels, described in the rest of this work, the uncertainties shown in Figure 4.3 are different for two main reasons. The

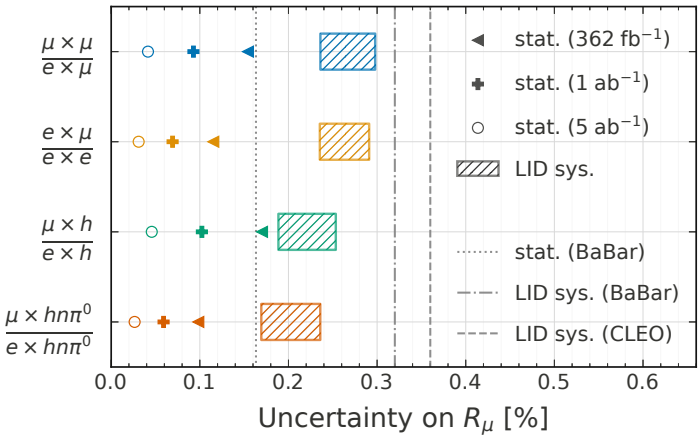


Figure 4.3: The estimated statistical precision and systematic uncertainty associated with lepton ID corrections while calculating R_μ with different ratios of branching fraction products using Equation 4.8. The hatched bars depict the range of systematic uncertainties on R_μ due to the uncertainties associated with lepton ID corrections, obtained with a bin-by-bin correlation of their systematic uncertainties of $\rho_{\text{sys}}^{\text{bins}} = 0.5$ for the lower end and $\rho_{\text{sys}}^{\text{bins}} = 1.0$ for the higher end of the bars. The markers indicate the estimated statistical uncertainty using datasets with different integrated luminosities. The horizontal lines indicate the statistical uncertainty of the *BABAR* measurement (represented by a dotted line) and the uncertainties due to lepton ID systematics from both the *BABAR* measurement (represented by a dash-dotted line) and the CLEO measurement (represented by a dashed line).

statistical uncertainty of the final results will be lower than that estimated here for 362 fb^{-1}, since the signal efficiency will be twice as high in the actual selection used for the measurement. This is because the cut on the output of the ANNs used here was based on the efficiency achieved by CLEO, which is unreasonably tight for this analysis. The second reason is that the selection presented here vetoes the π^0s on the signal side, since information about the reconstructed neutral particles is passed to the ANN. This results in significantly less fake pion background and reduced lepton ID systematics than for the actual measurement, where such a veto is not used. This will be elaborated on more later on.

4.3.4 Trigger efficiencies

The trigger is another aspect to consider when selecting the channels. Here, the efficiency of triggering the channels and its possible systematic effects should be considered. The uncertainty associated with the trigger was with 0.1 % the second largest systematic error in the *BABAR* measurement [12]. The efficiencies of several trigger lines, based on the inputs of ECL, CDC, or KLM, are compared in Table 4.7. A description of the trigger lines can be found in Section 3.4.7.

The channels $\mu \times h n \pi^0$ and $e \times h n \pi^0$ have a very high efficiency for the ECL-based trigger lines. The $\mu \times \mu$ and $\mu \times h$ channels have very low efficiency for ECL-based triggers due to the small amount of energy deposited in the calorimeter and CDC-based trigger lines should be included. However, using CDC-based triggers can come with some additional problems since their modelling in the simulation is much worse compared to the ECL triggers. The $e \times \mu$ and

Table 4.7: A comparison of the trigger efficiencies for different channels and trigger lines commonly used in the experiment.

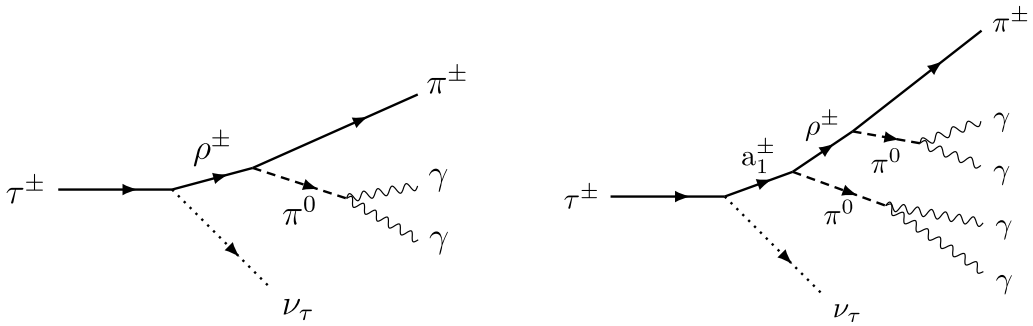
Trigger lines	$\mu \times \mu$	$e \times \mu$	$e \times e$	$\mu \times h$	$e \times h$	$\mu \times hn\pi^0$	$e \times hn\pi^0$
low multi (ECL)	31.1 %	86.7 %	20.5 %	62.4 %	30.5 %	95.4 %	80.4 %
high E (ECL)	0.1 %	88.5 %	99.6 %	48.6 %	95.8 %	93.7 %	99.7 %
high E + low multi	31.1 %	91.6 %	100.0 %	66.3 %	96.8 %	97.9 %	99.9 %
two-track (CDC)	84.2 %	83.4 %	83.1 %	82.4 %	81.5 %	74.5 %	75.3 %
CDC + KLM	95.2 %	82.7 %	34.1 %	92.8 %	79.8 %	87.0 %	57.9 %

$e \times e$ channels have a higher efficiency with the ECL-based trigger lines due to the presence of electrons, but $e \times \mu$ still has a lower efficiency than $\mu \times hn\pi^0$. One advantage of having the trigger efficiency driven by the decay $\mu \times hn\pi^0$ is that more systematic effects associated with it will cancel. So, also from the perspective of the trigger, computing R_μ from the ratio $\mu \times hn\pi^0 / e \times hn\pi^0$ is the best option.

4.4 Event selection

4.4.1 Topology selection

In this section, the basic selection requirements to obtain the desired event topology are specified. This analysis uses the 1x1 topology, and as presented in the previous section, the chosen channels involve the decay $\tau \to hn\pi^0\nu_\tau$ with $n > 0$ for one of the tau leptons. This decay will be referred to as the tag side decay since it is only used to tag the event, with the decay of interest being the decay of the other tau lepton (either $\tau \to e\nu_e\nu_\tau$ or $\tau \to \mu\nu_\mu\nu_\tau$). The dominant tag side decays with this channel selection are depicted in Figure 4.4, whereas the leptonic tau decays were already shown in Figure 2.6b in Section 2.4.



(a) $\tau^\pm \to \rho^\pm\nu_\tau \to \pi^\pm\pi^0\nu_\tau$ (b) $\tau^\pm \to a_1^\pm\nu_\tau \to \pi^\pm\pi^0\pi^0\nu_\tau$

Figure 4.4: The dominant tau decays on the tag side used for this analysis. The decay on the left into a charged pion and one π^0 has a branching fraction of 25.5 %, wile the decay on the right with two π^0s has a branching fraction of 9.3 % [10].

The event selection is based on tracks where the closest approach to the interaction point in the r-ϕ plane (dr) is below 1 cm, and the difference in z-direction (dz) ranges from −3 cm to 3 cm. Neutral particles reconstructed for this analysis are π^0s and photons.

Photon candidates are selected from ECL clusters with $17° < \theta < 150°$, which corresponds to the acceptance of CDC. This is done to make sure that the clusters do not originate from charged particles. These photons are used to reconstruct $\pi^0 \to \gamma\gamma$ candidates, for which multiple selection requirements were tested. The one resulting in a π^0 reconstruction efficiencies of 30 % was found to give the best results. This selection identifies neutral pion candidates from photons whose invariant mass lies in the range of 120 MeV to 145 MeV, equivalent to a range of ± 2 times the mass resolution. Additional criteria are applied to photon candidates, such as an energy requirement. It changes according to the detector region to minimise the combinatorial background from low energy photons originating from beam backgrounds. This is especially crucial in the endcap regions, where these backgrounds are more significant. For the forward region with $17° < \theta < 31.4°$ a threshold of 80 MeV is required, while for the barrel region with $32.2° < \theta < 128.7°$ a threshold of 30 MeV, and for the backward region with $130.7° < \theta < 150°$ a threshold of 60 MeV is used. In addition, the angle between the momentum directions of the photons that form a π^0 candidate is required to be below 1.4 rad and the difference of their azimuthal angles should be less than 1.5 rad.

All remaining photon candidates which do not form a π^0 candidate need to have an energy larger than 150 MeV before being used in this analysis. While it is still required that they are reconstructed from clusters within the CDC acceptance, there are additional timing requirements for the photon concerning the time of the collision. After correct calibration and considering the flight time, the photons originating from the IP should have a timing difference consistent with 0. Due to limited resolution, photons are kept within a window of ± 200 ns around the time of the collision. Moreover, the ratio between the timing difference and the uncertainty of the cluster timing (defined from the timing window that contains 99 % of all actual photons) is required to be less than 2.

Such identified photon candidates are used to compute variables associated with the event shape and event kinematics. It is worth noting that, except for the definition of those variables, the photon candidates are not used elsewhere. In particular, there is no selection on the photon multiplicity to avoid being sensitive to additional systematics associated with that selection. This is because the decay $\tau \to e\nu_e\nu_\tau$ is more likely to be accompanied by photons than the decay $\tau \to \mu\nu_\mu\nu_\tau$ due to final-state radiation (FSR) and Bremsstrahlung. This would make R_μ sensitive to systematics associated with the simulation or reconstruction of such photons. In addition, the misreconstruction of the decay $\tau \to \pi\pi^0\nu_\tau$ instead of the signal decays is one of the main backgrounds. Since this involves a misidentification of a pion as a lepton, which is more likely when selecting muons, this background is more prominent in the muon sample. This could again introduce a bias if only one of the photons originating from the π^0 is correctly reconstructed.

Tau leptons produced in $e^+e^- \to \tau^+\tau^-$ are back-to-back in the e^+e^- CM frame and have a considerable boost, which results in the τ decay products being isolated in two distinct hemispheres. The hemispheres are defined using the plane orthogonal to the thrust vector $\hat{t}$, which is obtained by maximising the thrust value

$$V_{\text{thrust}} \overset{\text{max}}{=} \sum_i \frac{|\vec{p}_i^{\,\text{CM}} \cdot \hat{t}|}{\sum_j |\vec{p}_j^{\,\text{CM}}|}, \tag{4.9}$$

where $\vec{p}^{\,\text{CM}}$ represents the three-momentum in the CM frame, and the summation includes all detected particles [143, 144]. The thrust vector approximates the τ flight direction, which cannot be uniquely obtained experimentally due to the presence of neutrinos.

Table 4.8: The fraction of events where a fake π^0 was reconstructed from a photon inside a cone of 0.1 rad around the momentum direction of the tracks originating from a J/ψ.

decay	$\varepsilon^{fake}_{J/\psi,data}$	$\varepsilon^{fake}_{J/\psi,MC}$	$\varepsilon^{fake}_{J/\psi,data}/\varepsilon^{fake}_{J/\psi,MC}$
$J/\psi \to ee$	$20.5 \pm 0.1\%$	$17.5 \pm 0.1\%$	1.170 ± 0.006
$J/\psi \to \mu\mu$	$11.1 \pm 0.1\%$	$9.86 \pm 0.03\%$	1.126 ± 0.007

To select the desired topology, one charged track must be in one hemisphere and another oppositely charged track must be in the other. The hemisphere known as the signal side contains a lepton candidate which is assumed to originate from $\tau^- \to e^-\bar{\nu}_e\nu_\tau$ or $\tau^- \to \mu^-\bar{\nu}_\mu\nu_\tau$ decays. Any number of π^0 and photon candidates is allowed in the signal hemisphere. Initially, a π^0 veto was used, which helps to reduce the background from $\tau \to \pi\pi^0(\pi^0)\nu$, but after estimating the impact of correcting for differences between data and MC associated with fake π^0s (see Section 4.4.2), this veto was removed. In the opposite hemisphere, the tag side, only one charged particle is allowed with the addition of at least one π^0 candidate. The reconstructed τ on the tag side includes only the two most energetic π^0 candidates, even if there are more than two π^0 candidates in the event. This affects only some specific tag side variables, like the invariant mass of the reconstructed tau. Both charged tracks are assigned the pion mass hypothesis, and the track on the tag side additionally has the requirement of $E_{\text{ECL}} < 0.8$ to reject Bhabha events. Instead, no initial lepton ID requirement is imposed for the signal track. In a later stage of the analysis, the signal track will be required to pass a requirement on P_e or P_μ to only select leptonic tau decays. The reason for using the pion mass hypothesis for all tracks is to avoid dependence on PID for the computation of event-based quantities such as the visible energy, since it would not be trivial to correct for the mismodelling due to PID.

4.4.2 π^0 veto and fake π^0 correction

To obtain a rough estimate of the size of the correction associated with fake π^0s, a J/ψ sample is used, where π^0 candidates are reconstructed using the same requirements as for the analysis. The main difference between the e-channel and the μ-channel for the fake π^0s is the increased presence of FSR photons with the electron, which results in a higher rate of fake π^0s. In the signal MC, the fraction of events removed by the π^0 veto is 3 % and 1.5 % for the electron and the muon signal, respectively.

For the following estimation, it will be assumed that in 1.5 % of events for both channels, the fake π^0s are of the same origin, and thus R_μ will not be sensitive to it. We then want to estimate the size of a correction to the remaining 1.5 % in the electron channel. In the J/ψ sample, thus, only π^0 candidates reconstructed from at least one photon identified as a FSR candidate are considered. FSR candidates are defined as all photons inside a cone of 0.1 rad around the direction of the momentum of the tracks. In addition, both tracks must be within the polar angle region of the analysis $0.82 < \theta < 2.13$. Then, the efficiency of reconstructing at least one π^0 in data ($\varepsilon^{fake}_{data}$) and MC ($\varepsilon^{fake}_{MC}$) is calculated. The results are listed in Table 4.8, together with their statistical uncertainties.

These efficiencies are still made up of different types of fake π^0s: First, some use actual FSR photons, which are of interest for this estimation ($\varepsilon^{fake,FSR}_{J/\psi}$), but there will also be a

Table 4.9: The fraction of fake π^0s reconstructed from FSR photons in data, MC, and their ratio. The uncertainties are statistical only.

$\varepsilon_{data}^{fake,FSR}$	$\varepsilon_{MC}^{fake,FSR}$	$\varepsilon_{data}^{fake,FSR}/\varepsilon_{MC}^{fake,FSR}$
4.81 ± 0.06	3.91 ± 0.03	1.232 ± 0.017

contribution of fake π^0s not from FSR (e.g. only from beam background), where one photon happens to be within the cone used to define FSR ($\varepsilon_{J/\psi}^{fake,bkg}$). In addition to that, there will be some real π^0s where one of the photons is inside the cone ($\varepsilon_{J/\psi}^{real}$). Thus, the fake efficiencies computed above can be written as

$$\varepsilon_{J/\psi}^{fake} = \varepsilon_{J/\psi}^{real} + \varepsilon_{J/\psi}^{fake,bkg} + \varepsilon_{J/\psi}^{fake,FSR}. \tag{4.10}$$

Since for the decay $J/\psi \to \mu\mu$ FSR photons are suppressed, it can be used to estimate $\varepsilon^{real} + \varepsilon^{fake,bkg}$. Thus, we can directly compute

$$\varepsilon_{J/\psi}^{fake,FSR} = \varepsilon_{J/\psi\to ee}^{fake} - \varepsilon_{J/\psi\to\mu\mu}^{fake}. \tag{4.11}$$

It still needs to be taken into account that the calculated $\varepsilon_{J/\psi}^{fake,FSR}$ is for two tracks from the J/ψ, but we are interested in the efficiency for a single track ($\varepsilon^{fake,FSR}$). Assuming that the cones are not overlapping, $\varepsilon_{J/\psi}^{fake,FSR}$ is almost twice as large as $\varepsilon^{fake,FSR}$, but the fact that there are events where both tracks have a FSR photon and a π^0 candidate reconstructed from it still has to be taken into account, which is $\varepsilon^{fake,FSR} \times \varepsilon^{fake,FSR}$. Finally, the relation

$$\varepsilon_{J/\psi}^{fake,FSR} = 1 - (1 - \varepsilon^{fake,FSR})^2, \tag{4.12}$$

and consequently

$$\varepsilon^{fake,FSR} = 1 - \sqrt{1 - \varepsilon_{J/\psi}^{fake,FSR}} \tag{4.13}$$

is obtained.

Using these considerations, $\varepsilon^{fake,FSR}$ can be computed in data and MC, which together with their ratio is shown in Table 4.9. The estimated correction would be around 23 %. This would result in a change of signal MC in the electron channel of around 0.35 %, which would directly propagate to a change of R_μ of the same order. A detailed study is needed to estimate this correction factor and associated uncertainty carefully. This was not possible at the timescale of this analysis. Thus, it was chosen to drop the π^0 veto at the cost of increased backgrounds from $\tau \to \pi\pi^0(\pi^0)\nu$ on the signal side.

4.4.3 Trigger

For this analysis, several ECL trigger bits are considered to trigger the desired events. They use only information from the ECL sub-detector, with conditions defined in Table 3.6 in Section 3.4.7. All the listed ECL trigger bits in Table 3.6, except for `lml0` and `lml12`, were active during the entire data-taking period of interest for this measurement. The `lml0` trigger has a prescale factor of 10 for a significant fraction of the dataset, starting from experiment 20, which is why it is not used. Instead, `lml12` was inactive for the first data-taking periods

before experiment 10. The corresponding runs only amount to an integrated luminosity of $2.2\,\mathrm{fb}^{-1}$, which is why it is still used for this analysis.

At least one of the mentioned trigger bits must be active in data and TSIM. Any differences between the trigger efficiency in data and TSIM will be corrected, as described in Section 4.6.4. This difference amounts to about 1.2 % higher efficiency in data. In data, the trigger efficiency is computed with the help of an orthogonal reference trigger as

$$\varepsilon_{\mathrm{trg}} = \frac{N_{\mathrm{trg}\wedge\mathrm{ref}}}{N_{\mathrm{ref}}}. \tag{4.14}$$

Here, $N_{\mathrm{trg}\wedge\mathrm{ref}}$ is the number of events triggered by both the selected trigger bits and the reference bits, while N_{ref} is the number of events only triggered by the reference bits. In this way, the trigger efficiency can be measured in the subset of events triggered by the reference trigger, which should give a good estimate of the actual trigger efficiency. The current analysis uses the CDC-based triggers `ffo` and `fyo` to define the reference sample. Their trigger conditions were introduced in Table 3.5 in Section 3.4.7.

In Figure 4.5, the efficiency of the aforementioned ECL trigger bits in data is shown, split according to experiments. The efficiency is computed after applying all selections, on the left for the e-channel and on the right for the μ-channel. The efficiency for the relevant bits is very stable for all run periods. The horizontal dashed lines show the trigger efficiencies in TSIM, which are in good agreement for all shown trigger lines. Figure 4.6 shows the same trigger

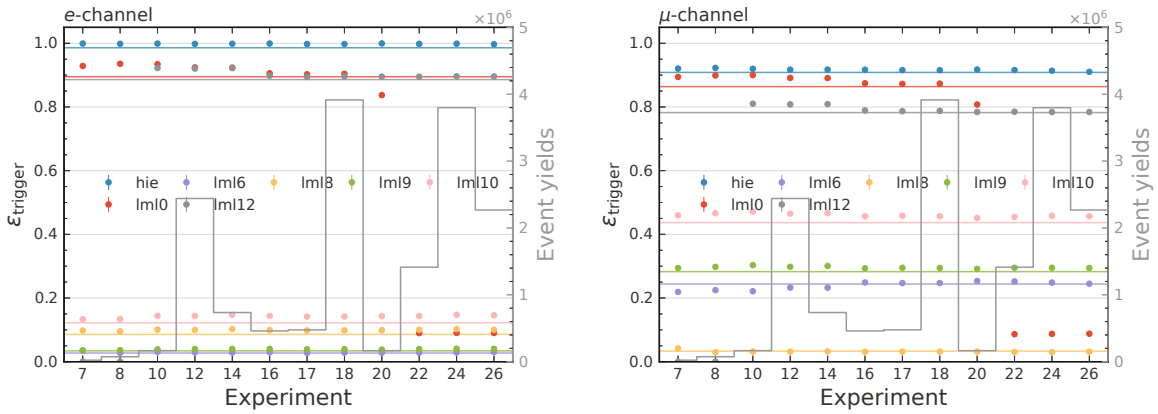


Figure 4.5: The trigger efficiency for some ECL trigger bits in data, split according to experiment number, both for the e-channel (left) and the μ channel (right). The prescale of `lml0` during experiment 20 can be observed, as well as the active period of `lml12`. The overlayed histogram shows the event yields in the respective bins, and the horizontal lines show the trigger efficiencies in TSIM.

efficiencies as a function of signal momentum. Here, a characteristic momentum dependence for some of the `lml` bits is observed, which comes from the activation requirements. The by far most important trigger bit is `hie`. It relies mainly on the π^0s from the tag side, which is also why the efficiency is relatively flat over the whole momentum region.

It is also desirable to examine the CDC trigger bits used to define the reference sample. The experiment dependence of the efficiency is shown in Figure 4.7 for the combined e and μ sample. None of the trigger bits were active during the entire data-taking. The `ffo` bit was

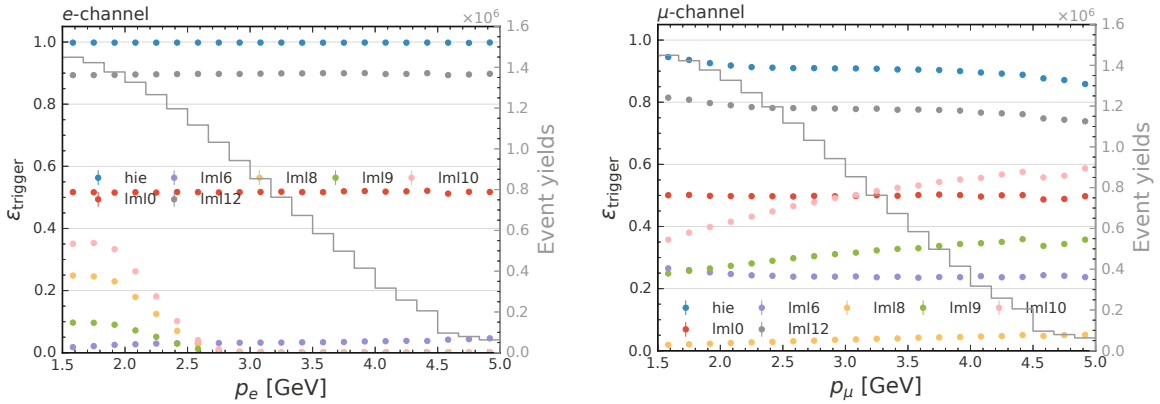


Figure 4.6: The trigger efficiency for some ECL trigger bits in data, as a function of signal momentum, for the e-channel (left) and the μ-channel (right). The overlayed histogram shows the event yields in the respective bins.

replaced by `fyo`, and the `fso` bit was replaced by `syo`. This was only a transition to include ANN-based trigger information, and they should trigger very similar event types. The `stt` bit was only activated during experiment 16 and showed a significantly higher efficiency, especially in TSIM. The `stt` trigger is not used to define the reference sample since this could bias the calculated trigger efficiency, giving more weight to later run periods.

The CDC bits are generally much worse modelled in TSIM than the ECL bits, as seen in Figure 4.7. In some experiments, the efficiency in data deviates up to 20 % from the MC efficiency, and a strong experiment dependence is observed. In addition, the reference bits will not trigger certain event types, which could cause an intrinsic bias of the reference method due to correlations between the ECL and CDC triggers. A systematic uncertainty that accounts for these effects will be derived in Section 4.6.4.

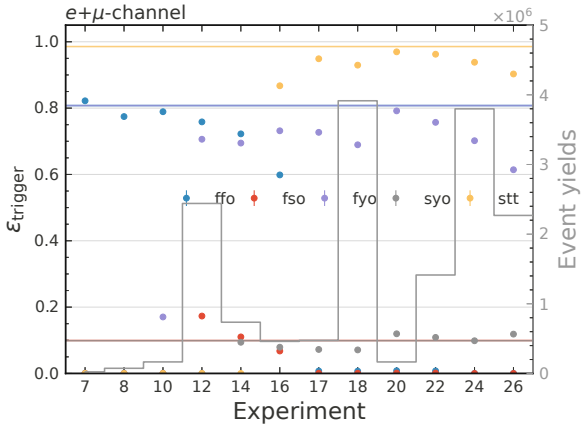


Figure 4.7: The experiment dependence of the CDC trigger bits used to define the reference sample. The trigger efficiency is calculated for the combined e and μ channel. The overlayed histogram shows the event yields in the respective bins, and the horizontal lines show the trigger efficiencies in TSIM.

4.4.4 Data driven cuts

A first set of cuts is determined to remove a data excess, attributed to a hadronic two-photon process ($e^+e^- \to e^+e^-n'h'$ with $n' > 2$), which are not simulated in MC. As expected for this process, the excess concentrates at low visible CM energy of the event, shown in Figure 4.8. There is also a data excess observed for high values of $E^{\text{CM}}_{\text{visible}}$, originating from a wrong normalisation of the Bhabha background.

In this figure and in all the following which also show data, the MC histograms are stacked on top of each other. When no data are shown, the histograms are generally not stacked. The legend in these figures shows the individual contributions from the underlying physics processes:

- $e^+e^- \to \tau^+\tau^-$ (shown as $\tau\tau$)
- $e^+e^- \to e^+e^-\mu^+\mu^-$ (shown as $ee\mu\mu$)
- $e^+e^- \to e^+e^-e^+e^-$ (shown as $eeee$)
- $e^+e^- \to e^+e^-(\gamma)$ (shown as $ee\gamma$).

All MC samples not belonging to any of the shown categories are contained in *others*. In some cases, the $\tau\tau$ sample is further split according to the decay products of the tau on the signal and tag side (labelled $\tau\tau \to$ signal decay $\times$ tag decay). For brevity, neutrinos are neglected, h denotes a pion or a kaon, and $hn\pi^0$ indicates a hadron with at least one π^0 (the tag side definition). Table 4.10 defines the symbols for the different variables used for event selection.

Table 4.10: The symbols for different variables, together with a brief description.

Variable	Description
p_ℓ, p_e, p_μ	magnitude
θ_ℓ, θ_e, θ_μ	polar angle
ϕ_ℓ, ϕ_e, ϕ_μ	azimuthal angle
	} of the momentum vector of the track on the signal side, where the subset indicates the selected particle type
$E^{\text{CM}}_{\text{visible}}$	total visible energy in the CM frame
θ_{missing}	polar angle of the missing momentum direction
$\theta^{\text{CM}}_{\text{missing}}$	polar angle of the missing momentum direction in the CM frame
$p^{\text{CM}}_{t,\text{missing}}$	transverse component of the missing momentum direction in the CM frame
m^2_{missing}	square of the missing mass of the event
V_{thrust}	the thrust value
$\cos(\theta_{\text{thrust}})$	cosine of the polar angle of the thrust vector $\hat{t}$
$p^{\text{CM}}_{\text{tag}}$	magnitude of the momentum of the tag side track in the CM frame
$m_{\tau(\text{tag})}$	invariant mass of the reconstructed tag side tau
$p^{\text{CM}}_{\tau(\text{tag})}$	momentum magnitude of the reconstructed tag side tau in the CM frame
$\cos(\theta^{\text{CM}}_{\tau(\text{tag})})$	cosine of the polar angle of the tag side tau momentum in the CM frame
$p^{\text{CM}}_{t,\text{max}}$	the larger of the transverse momenta of the tracks in the CM frame
$p^{\text{thrust frame}}_{t,\text{min}}$	The smaller of the transverse momenta of the tracks in the rotated CM frame such that the z direction coincides with the thrust axis

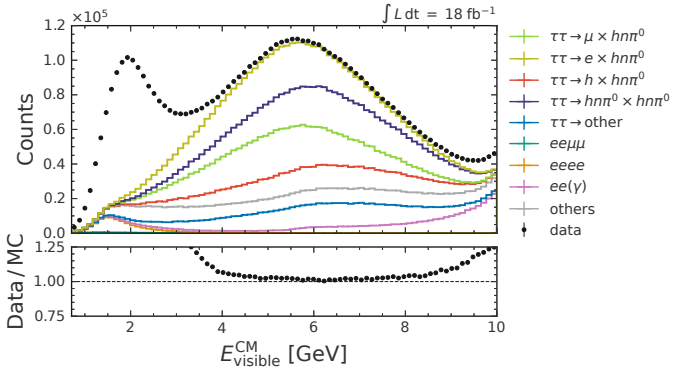


Figure 4.8: The visible energy in the CM frame before any preselection. The MC histograms are stacked.

The excessive events also have low thrust values, as shown in Figure 4.9a, which is characteristic of the hadronic two-photon process. Another indication that the origin is a two-photon process is that the missing momentum vector of the events in question points towards the beam pipe, as seen in Figure 4.9b. It also manifests itself in large values for the squared missing mass and low missing transverse CM momentum; see Figures 4.9c and 4.9d. Since the signature of the excess is very distinct in multiple variables, it is possible to completely remove it with a data-driven selection.

There are many possibilities for this, but a 2-dimensional cut in the plane of missing mass squared and transverse missing momentum, together with a lower cut on the thrust value, was found to entirely remove the excess while having the least impact on the signal region. Nevertheless, these cuts still remove some events from the tails of the signal distributions, which results in a reduction of 11 % of the reconstructed signal events. This is a change in the signal efficiency of about 4.8 %, going from 43 % to 38.2 %. However, since no background suppression has been applied yet, those events were likely to be removed because the other two-photon processes peak in the same region.

The 2-dimensional cut is shown in Figure 4.10 as a black dashed line. The histogram on the left shows the data/MC ratio, where only bins up to 1.35 are displayed on the colour bar, and the histogram on the right shows the data-MC difference. The cuts are as follows:

- $m^2_{\text{missing}} < (20 + 40 \cdot p^{\text{CM}}_{t,\text{missing}})$
- $V_{\text{thrust}} > 0.85$

The data and MC distributions of $E^{\text{CM}}_{\text{visible}}$ and θ_{missing} after the data-driven cuts are shown in Figure 4.11. The previously observed excess has almost completely disappeared. The analysis selections will eventually remove the remaining data excess at large values of visible energy due to the wrong normalisation of the Bhabha background.

4.4.5 Preselection cuts

After applying data-driven cuts, the simulated samples are used to identify discriminating features to suppress the background. First, a cut-based preselection is applied to remove

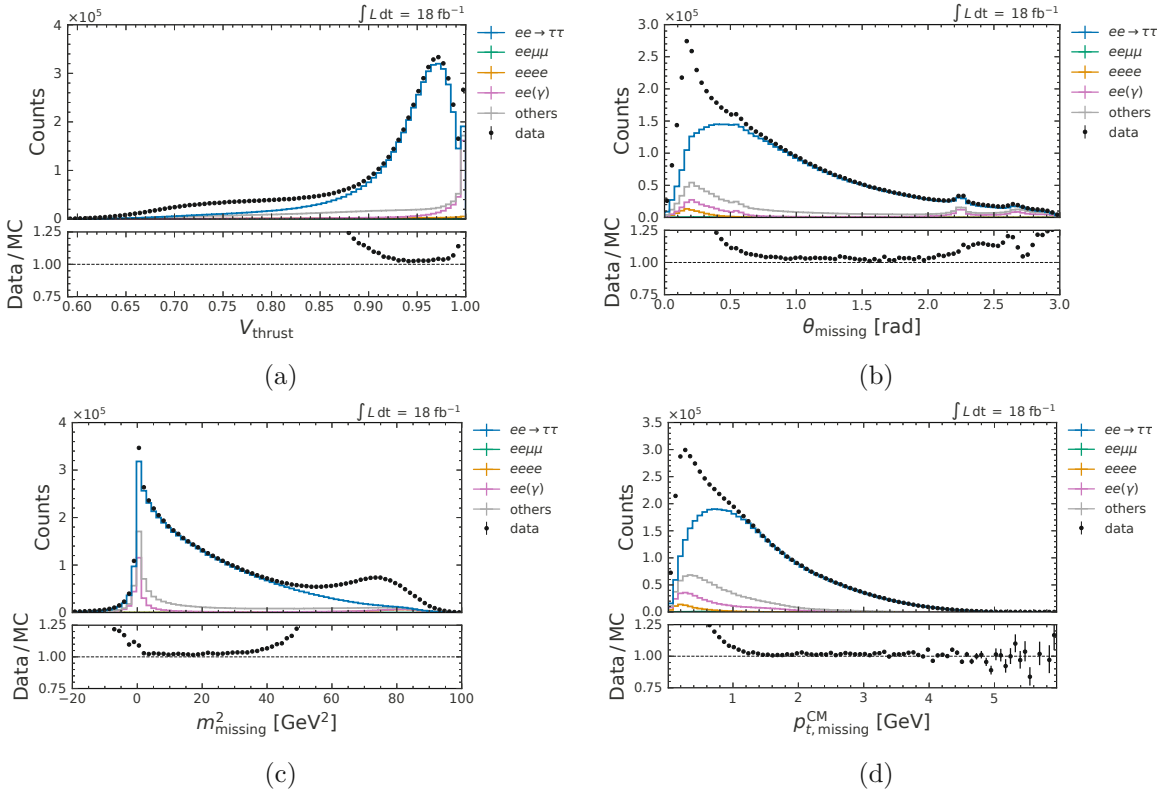


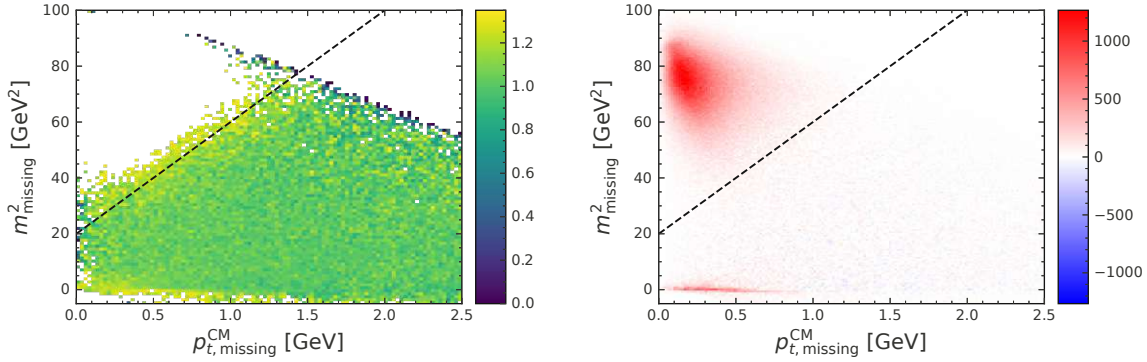
Figure 4.9: The thrust value (a), the polar angle of the missing momentum vector (b), the missing mass squared of the event (c) and the transverse component of the missing momentum in the CM (d) before any preselection. These features prominently show the data events not present in the simulation, attributed to a hadronic two-photon process.

a significant fraction of the dominant background events. Later, an ANN is used for a more refined selection. The preselection was found to have a positive impact on the ANN's performance. This is because the classifier can focus more on the less trivial backgrounds and the duration of training is significantly reduced. In addition, this preselection serves as a private skim selection to reduce the data sample size, save disk space, and avoid memory problems.

The selection was optimised simultaneously using the following six features:

1. $E^{\text{CM}}_{\text{visible}}$
2. $\theta^{\text{CM}}_{\text{missing}}$
3. V_{thrust}
4. $p^{\text{CM}}_{\tau(\text{tag})}$
5. $\cos(\theta^{\text{CM}}_{\tau(\text{tag})})$
6. $m_{\tau(\text{tag})}$

The signal and background distributions for these variables are shown in Figure 4.12. For the optimisation of this preselection, only 1 % of the events in all the available MC samples are used, which was found to be sufficient.



(a) data/MC ratio (b) data-MC difference

Figure 4.10: The data excess shown in the 2-dimensional plane of missing mass squared and missing transverse momentum in the CM. The figure on the left shows the data/MC ratio, where only bins with a ratio below 1.35 are displayed, and the figure on the right shows the difference between data and MC. The black dashed line indicates the cut applied in the analysis.

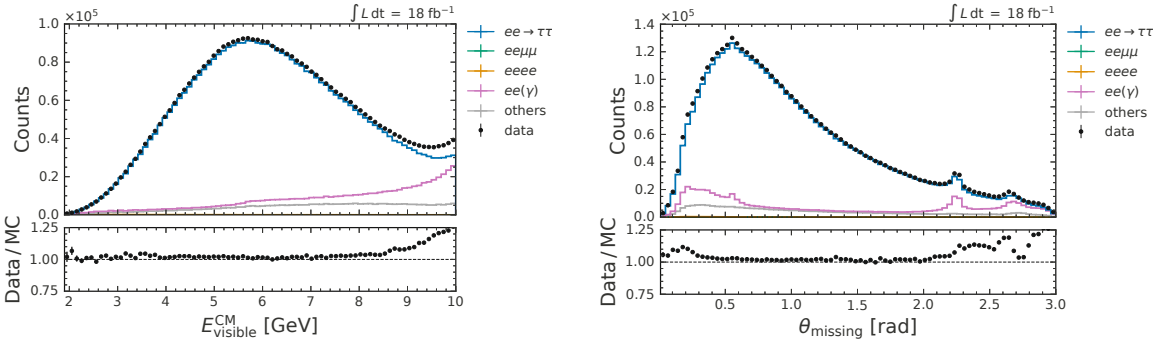


Figure 4.11: The visible energy in the CM frame (left) and the polar angle of the missing momentum (right) after the data-driven selection.

The optimal set of all possible rectangular cuts is found with simulated annealing, for which the basic principles were introduced in Section 2.6.1. More details on the implementation of the algorithm and its workings can be found in the publicly available repository [145]. The quantity that is minimised is

$$-p_{\text{sig}} \cdot H(\varepsilon_{\text{sig}} - \varepsilon_{\text{ref}}) \tag{4.15}$$

for different fixed values of ε_{ref}. Here, p_{sig} is the purity of the signal defined in Equation 4.3, H is the Heaviside step function, and ε_{sig} is the signal efficiency, as defined in Equation 4.2.

It is important to note that for this simultaneous optimisation, the signal sample is defined as the combined electron and muon channels. This is to ensure that both channels have a common selection and to avoid accidental unblinding of the ratio. The common selection is desired because it results in the cancellation of most of the systematics associated with the selection because of the similar kinematics of the electron and muon channels.

For the optimisation, a particle identification requirement is imposed. Since electron and muon channels are kept combined, events are selected that either pass $P_\mu > 0.9$ or $P_e > 0.5$. The

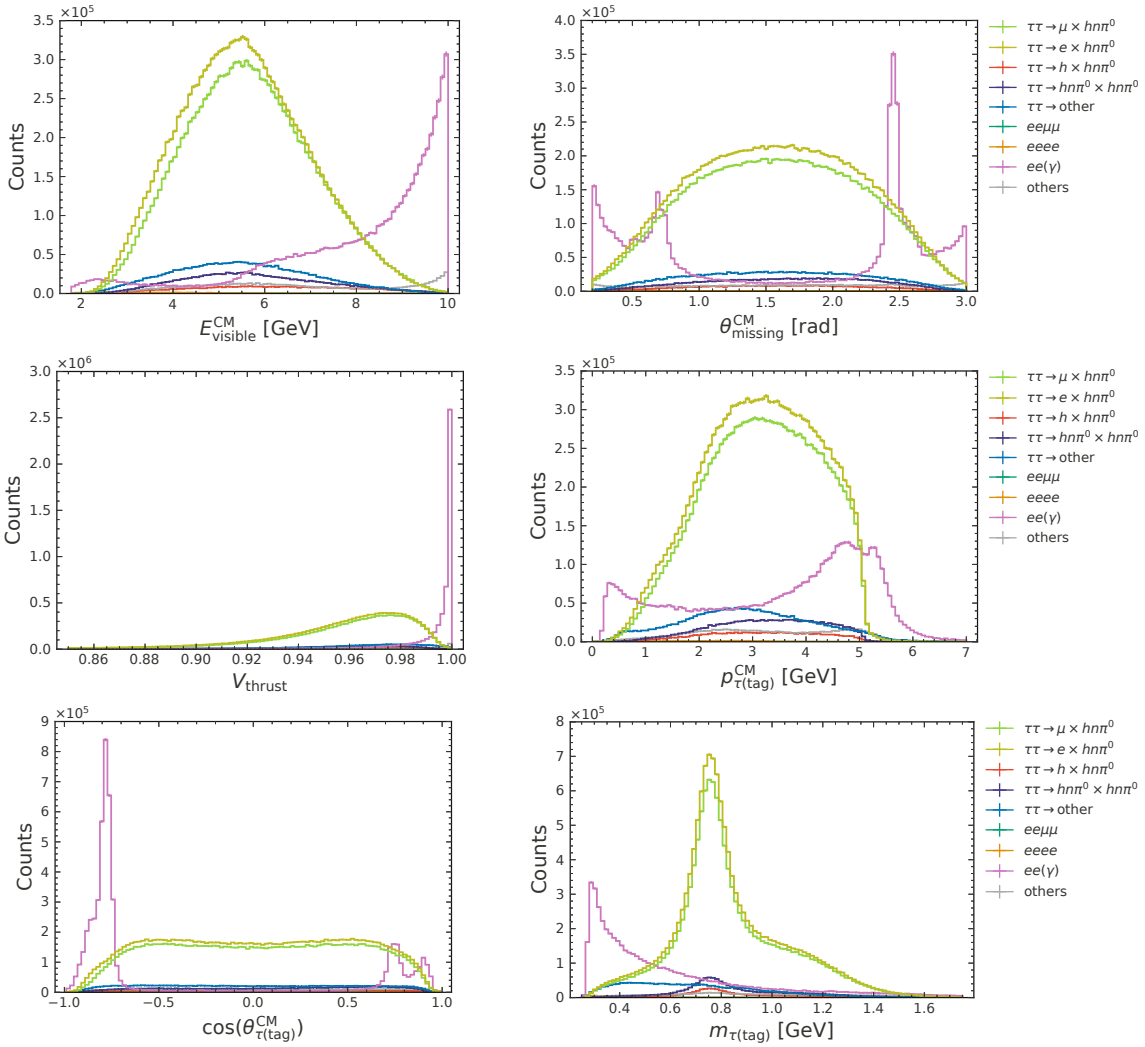


Figure 4.12: The signal and background distributions for the variables used to optimise the preselection. The data-driven cuts have already been applied. The histograms are not stacked.

variable P_μ is the global likelihood-based muon identification variable and P_e is the BDT-based electron identification variable, as described in Section 3.6.2.

The tau decays defined as the signal and tag side, together with the decay mode and branching fraction assigned to it in the TAUOLA generator, are listed in Table 4.11. These branching fractions are used to compute the signal efficiency. The values of the branching fractions originate from a global fit by HFLAV [13] that uses 170 measurements of τ branching fractions or branching fraction ratios and is also subject to the unitarity constraint. The values are taken from the 2020 version of the PDG [146].

From the optimisation procedure, a set of selections is obtained for each value of ε_{ref}, for which 250 values are chosen uniformly distributed between 1 and 0. These selections are shown on an ROC curve with purity on the x-axis and efficiency on the y-axis in Figure 4.13. From all

Table 4.11: The decay mode and branching fraction assigned to signal and tag side tau decays in TAUOLA. The branching fractions are obtained from a global fit to 170 measurements of 129 different τ branching fractions, which are listed in the 2020 version of the PDG [146].

Decay mode	Decay	BF	Classification
1	$\tau \to e\nu_e\nu_\tau$	0.178175	signal
2	$\tau \to \mu\nu_\mu\nu_\tau$	0.173937	signal
163	$\tau \to \pi\pi^0\nu_\tau$	0.254941	tag
165	$\tau \to K\pi^0\nu_\tau$	0.004328	tag
4	$\tau \to \pi 3\pi^0\nu_\tau$	0.010429	tag
106	$\tau \to K\pi^0\pi^0\nu_\tau$	0.000647	tag
110	$\tau \to \pi\pi^0\gamma\nu_\tau$	0.001789	tag
111	$\tau \to \pi\pi^0\pi^0\nu_\tau$	0.092595	tag

these selections, the one that maximises the following figure of merit (FOM) is chosen:

$$\mathrm{FOM} = \frac{N_{\mathrm{sig}}}{\sqrt{N_{\mathrm{sig}} + N_{\mathrm{bkg}}}}, \tag{4.16}$$

with N_{sig} being the number of remaining signal events and N_{bkg} the number of background events. This FOM minimises the standard deviation on the estimated signal strength when performing a Poisson counting experiment, which only takes into account the statistical uncertainty [10, p. 674]. Since the measurement will be systematically limited, optimising for the statistical uncertainty is not optimal. However, the such obtained selection is very loose and can be used as a starting point for a more refined selection, which will be shown in the following. The final selection will also take into account the systematic uncertainties and would in any case supersede this preselection. In Figure 4.13, the chosen preselection is marked with a black cross. It corresponds to a purity of 84.4 % and a combined signal efficiency of 29.9 % (with a figure or merit of 4643).

Alternatively, the FOM (4.16) could have been optimised directly with the simulated annealing algorithm, but to compare the cut-based approach to a multivariate classifier, having the ROC curve available can be helpful. It was also verified that the selection obtained with this procedure gives the same results as when optimising the FOM directly.

The fraction of signal events and the most dominant background processes after this selection are listed in Table 4.12. The most dominant background is $e^+e^- \to \tau^+\tau^-$, where the signal and tag side do not match the definition in Table 4.11. This includes several tau decays, but the largest fraction comes from decays in which a pion on the signal side is misidentified as a muon. A more detailed breakdown of the background will be given later in the note (Table 4.16 in Section 4.4.9).

Table 4.12: The most prominent processes after applying the preselection. The combined signal events sum up to a purity of 84.4 %.

Process	Fraction
$ee \to (\tau \to e\nu_e\nu_\tau)(\tau \to hn\pi^0\nu_\tau)$	45.3%
$ee \to (\tau \to \mu\nu_\mu\nu_\tau)(\tau \to hn\pi^0\nu_\tau)$	39.1%
$ee \to \tau\tau$ (other)	5.5%
$ee \to (\tau \to hn\pi^0\nu_\tau)(\tau \to hn\pi^0\nu_\tau)$	3.9%
$ee \to ee(\gamma)$	2.4%
$ee \to (\tau \to h\nu_\tau)(\tau \to hn\pi^0\nu_\tau)$	1.6%
$ee \to c\bar{c}$	0.7%
$ee \to u\bar{u}$	0.6%
$ee \to ee\tau\tau$	0.3%

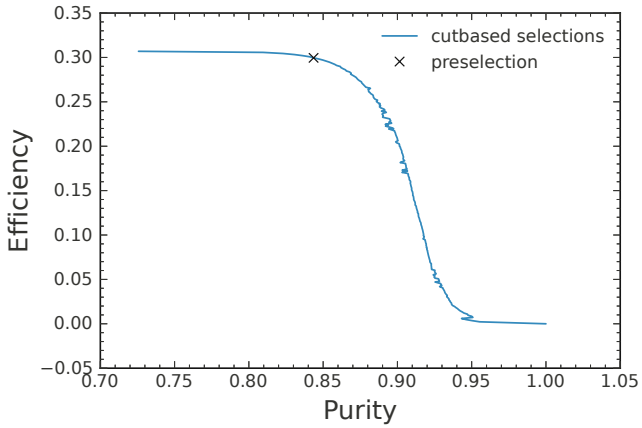


Figure 4.13: A ROC curve obtained by optimising cuts for maximum purity with a fixed efficiency. The set of cuts chosen for the preselection is marked with a black cross.

The exact cuts of the preselection are:

$$
\begin{aligned}
2.1 &\leq E^{\mathrm{CM}}_{\mathrm{visible}} &&< 9.2 \\
0.23 &\leq \theta^{\mathrm{CM}}_{\mathrm{missing}} &&< 2.97 \\
0.85 &\leq V_{\mathrm{thrust}} &&< 0.996 \\
0.62 &\leq p^{\mathrm{CM}}_{\tau(\mathrm{tag})} & & \\
-0.91 &\leq \cos(\theta^{\mathrm{CM}}_{\tau(\mathrm{tag})}) &&< 0.91 \\
0.32 &\leq m_{\tau(\mathrm{tag})} & &
\end{aligned}
$$

In Figure 4.14, the 1D histograms of these variables are shown with the corresponding cuts indicated by red arrows. All the cuts on the other variables are applied to show how much background is removed by the individual variables ('n-1 plots').

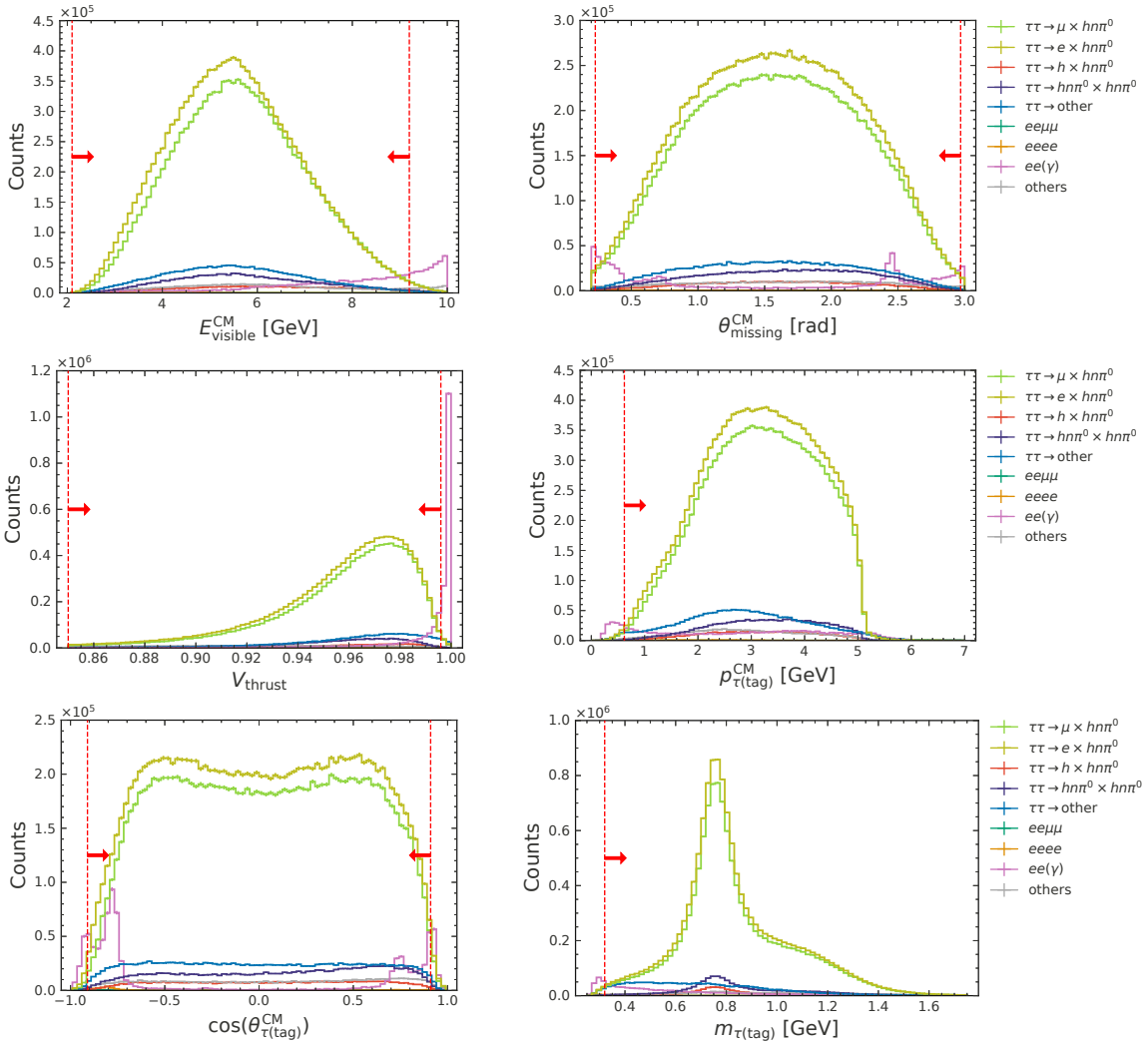


Figure 4.14: The variables used for the preselection, with the rectangular cuts indicated as red arrows. For each variable, all the cuts on the other variables are applied ('n-1 plots').

4.4.6 Lepton efficiencies

Using the cuts on the lepton ID variables $P_\mu > 0.9$ and $P_e > 0.5$, the associated efficiencies and mis-id rates in MC can be computed using truth information. In Figure 4.15, the efficiencies are shown as a function of signal momentum (left) and polar angle (right). In Figure 4.16, the electron identification efficiencies (left) and the muon identification efficiencies (right) are shown as two-dimensional distributions over momentum and polar angle. The horizontal and vertical dashed lines correspond to the binning defined above. A significant efficiency drop is observed for momentum less than 1 GeV and in some distinct regions of polar angle, coming from gaps and inefficient regions in the detectors. The integrated efficiencies and mis-id rates are shown in Table 4.13 for MC and corrected MC, using the data/MC correction factors from the Belle II lepton ID performance group.

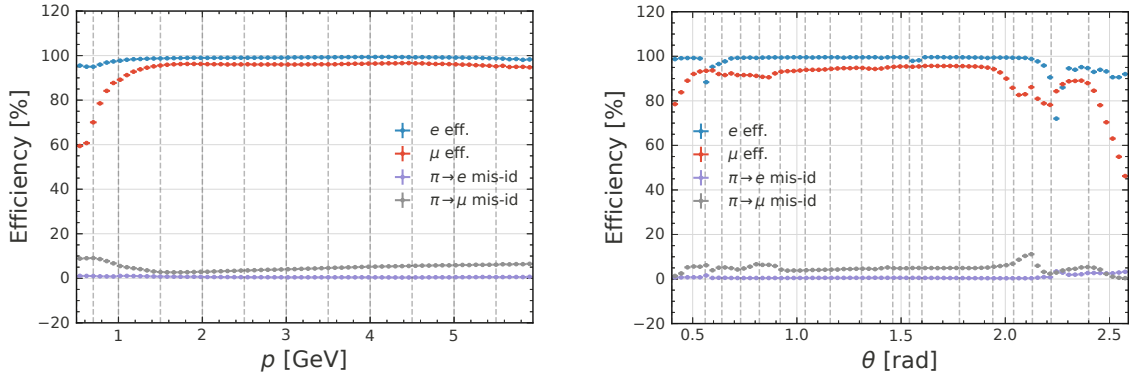


Figure 4.15: 1D distributions of lepton efficiencies and mis-id rates binned in momentum (left) and polar angle (right). The efficiencies are computed in MC using truth information. The vertical dashed lines indicate the binning used for the lepton ID corrections.

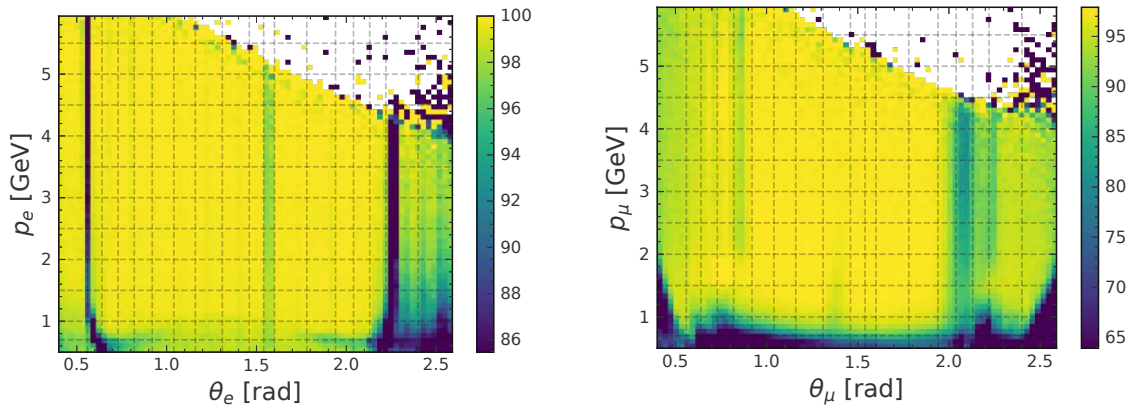


Figure 4.16: 2D distributions of electron (left) and muon (right) efficiencies in bins of momentum and polar angle, computed in MC using truth information. The vertical and horizontal dashed lines indicate the binning used for the lepton ID corrections.

Restricted signal kinematics

Based on the information in this section, the analysis will be limited to some regions of signal kinematics where lepton ID is best understood. In general, lepton ID in the endcap regions has larger associated uncertainties and suffers from steep decreases in efficiency (see Figure 4.15). Thus, the signal momentum is restricted to the KLM barrel,

$$0.82 < \theta_{\text{lepton}} < 2.13,$$

so as not to pollute the sample with these events.

Similarly, the lepton efficiency drops for momenta below 1.5 GeV and especially for muon identification, large corrections must be applied at low momentum. In principle, since a template fit binned in signal momentum is performed, the low momentum region would not affect the precision of the measurement, even if there are large uncertainties associated with it. However, this region would also not contribute to the measurement, so it is better to restrict the signal kinematics instead. Furthermore, the correction factor itself changes significantly as

type	MC	corrected MC
e efficiency	98.79 %	99.06 %
μ efficiency	94.17 %	90.30 %
$\pi \to e$ mis-id	0.63 %	1.21 %
$\pi \to \mu$ mis-id	4.62 %	3.87 %

Table 4.13: The integrated lepton identification efficiencies and mis-id rates in truth matched MC. The corrected MC corresponds to efficiencies computed with the data/MC correction factors applied, so it should be equivalent to the efficiencies expected in data.

type	MC	corrected MC
e efficiency	99.71 %	99.73 %
μ efficiency	96.81 %	93.91 %
$\pi \to e$ mis-id	0.35 %	0.91 %
$\pi \to \mu$ mis-id	3.81 %	3.11 %
$K \to e$ mis-id	0.17 %	0.55 %
$K \to \mu$ mis-id	2.37 %	3.18 %

Table 4.14: The integrated lepton identification efficiencies and mis-id rates in truth matched MC after restricting the kinematics of the lepton candidate. The corrected MC corresponds to efficiencies computed with the data/MC correction factors applied.

a function of the momentum below 1.5 GeV. If this change of the correction factor is also large within the bin and the kinematics of the leptons used in this analysis do not correspond to the kinematics of the lepton in the calibration sample that was used to derive the correction factors, this could lead to a bias in the measurement. Some effects that indicate this difference in muon efficiencies between data and MC as a function of momentum can be seen from the control channel, presented in Section 5.2.5. The data/MC comparison for the muon candidate shown there also includes the bin from 1.0 GeV to 1.5 GeV. Therefore, in order to be insensitive to unexpected systematic effects, the requirement

$$1.5\,\text{GeV} < p_\ell < 5\,\text{GeV}$$

is imposed on the signal lepton.

Even though restricting the lepton kinematics means cutting away a large chunk of signal events, the measurement is still systematically limited and in terms of statistical precision, this restriction has little impact. In addition, events that are not covered by the $\pi \to \ell$ mis-id rate corrections are removed.

In Table 4.14, the integrated lepton identification efficiencies and mis-id rates are shown after applying all selections and restricting the kinematics of the lepton candidate. Since this kinematic range also corresponds to the region with the highest efficiency, the integrated values increased with respect to Table 4.13. Although these efficiencies and mis-id rates are computed after the multivariate selection presented in the next section, they are unaffected because it hardly influences the kinematics of the signal track.

4.4.7 Multivariate selection

For a more refined background suppression, an ANN is trained with all the events remaining after the preselection and the restriction of the signal kinematics. The architecture is a simple, fully connected ANN with two hidden layers containing 20 neurons respectively, each with a tanh activation function. The network is trained using the stochastic gradient descent algorithm with cross-entropy as a loss function. Other hyperparameters used to train the network are:

- batch size: 2048
- epochs: 200
- learning rate: 0.1
- momentum: 0.15.

Seven input features related to the kinematics of the event are passed to the network and mapped to four output neurons. One output neuron represents the combined signal sample of the electron plus muon channel, while the other three represent different background events that are still dominant after preselection. The backgrounds are separated into taupair events that do not pass the signal requirement, Bhabha events, and everything else. Before the input features are passed to the ANN, they are standardised using a scaling method robust to outliers. This is a common preprocessing step applied when using machine learning, which helps to improve its performance. In particular, ANNs are very sensitive to the scale of the input data. The features are transformed by subtracting their median value and dividing by their interquartile range, which is the range between the 25 % and 75 % percentiles. Both quantities are obtained from the training data.

The seven features used in training are:

1. $m_{\tau(\text{tag})}$
2. $\cos(\theta^{\text{CM}}_{\tau(\text{tag})})$
3. V_{thrust}
4. $\cos(\theta_{\text{thrust}})$
5. $E^{\text{CM}}_{\text{visible}}$
6. $p^{\text{CM}}_{\text{tag}}$
7. $p^{\text{CM}}_{t,\text{missing}}$

One important criterion for selecting these features was that there is no strong dependence on the signal side kinematics. Even though the kinematics of the signal leptons are very similar, using associated variables could introduce a bias that does not cancel when estimating R_μ. Any mismodelling of the input features and thus of the ANN output that only originates from the tag side will almost completely cancel for the measurement of R_μ. Any residual effect induced by the slightly different signal kinematics will be estimated in Section 4.6.13 with an embedded sample. In addition, systematic effects associated with mismodelling that depend on the background composition will not completely cancel, but they will be treated separately in Section 4.6.9.

In Figure 4.17, the feature importance of the selected variables is shown, calculated with two different methods based on a forest of trees [147]. On the left, the ranking is based on the mean decrease in impurity, while on the right it is based on permutation importance, for which the values of a feature are randomly shuffled and the resulting loss in accuracy is used as a metric [147, 148]. Both scores were computed using scikit-learn [149], and more information about the advantages of the respective methods can be found at [150].

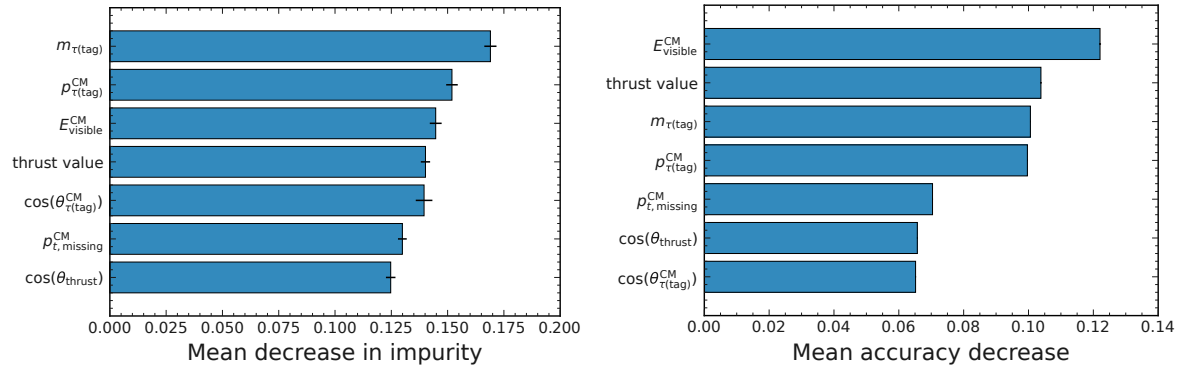


Figure 4.17: The feature importance of the variables passed to the ANN, estimated with a forest of trees. Different methods are used to compute a score that reflects the importance of the different features, namely the mean decrease in impurity (left) and the mean accuracy decrease with feature permutation (right).

Since an ANN is used, correlated input features are not a concern, but it is still desirable to keep the amount of redundant information passed to the network to a minimum. This is mainly to keep the model simple and reduce unnecessary dependency on features that could introduce some mismodelling. The features have been chosen to avoid large correlations and it has been found that the addition of more input features does not result in a performance gain of the ANN. The correlations between the ANN input variables are shown in Figure 4.18.

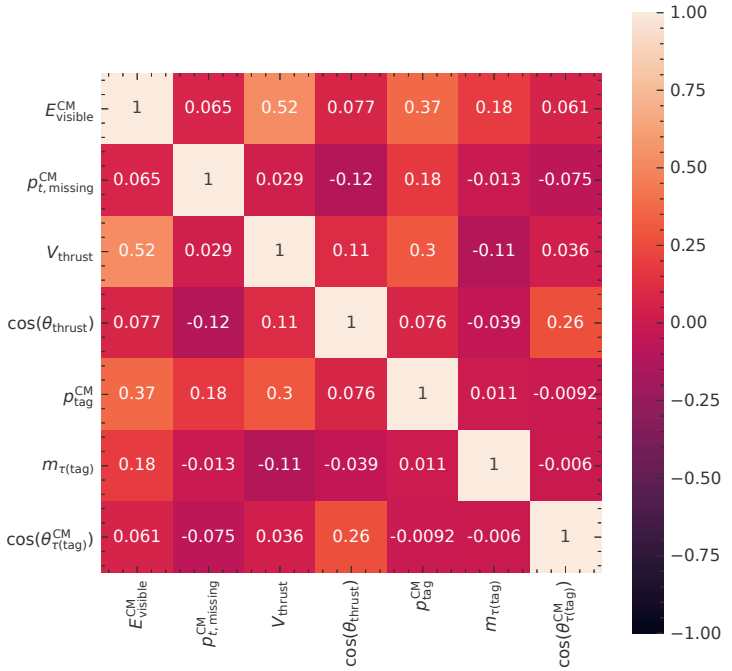


Figure 4.18: The correlation of the variables used as input for the ANN.

In Figures 4.19 and 4.20 the distributions of the input variables for the ANN are shown, both in data and in MC. To show the distributions, no PID cut was imposed on the signal track. There are some differences in the shape of the distributions between data and MC for the

ANN inputs. As long as this mismodelling is common for the e-channel and the μ-channel, this is not concerning as it will cancel when measuring R_μ. This is expected to be the case, since it can be attributed to the tag side variables. Some of the observed discrepancies also originate from mismodelled backgrounds, which will be almost completely removed with the selection of the ANN.

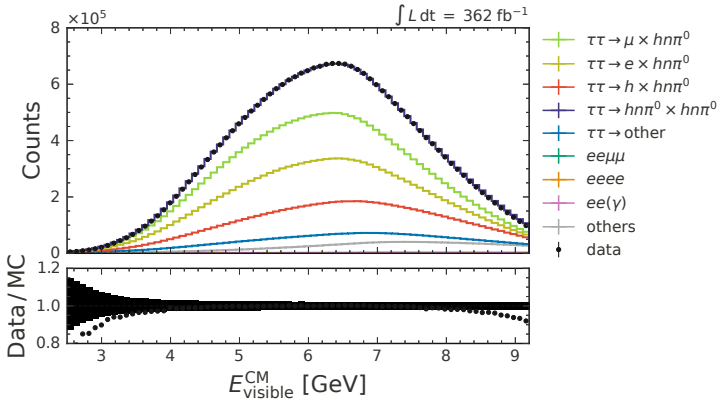


Figure 4.19: Data and MC distribution for the visible energy in the CM frame, used as a ANN input, without any PID selection applied on the signal track. The hatched area in the subplot indicates the uncertainty coming from π^0 and photon efficiency corrections.

The evolution of the loss during the training of the network is shown in Figure 4.21, for both the training data (12% of the available MC) and a test dataset (8% of the available MC). The loss computed in the test data closely follows the training loss, indicating that there is no overtraining. Since a large dataset is used for training and the network does not have a large number of parameters, no overtraining issues are expected.

Figure 4.22 shows the ROC curve of the ANN with signal efficiency on the y-axis and purity on the x-axis. Both efficiency and purity are computed for the combined signal sample. The two lines correspond to different cut values on the ANN output for the combined signal, once computed without applying corrections for lepton identification and once with corrections. The network was trained with the correction weights applied. For this analysis, a cut at 0.9 is applied to the output associated with the combined signal, which is indicated by the black cross in the figure. The choice of this value will be justified in Section 5.3.3, where it is shown that the overall uncertainty on R_μ reaches a minimum around that threshold value. This will be confirmed again in Section 5.5.6 using actual data.

Figure 4.23 shows the distribution of the four output neurons of the ANN. The neuron responsible for classifying the signal (4.23a) is shown, with a vertical line indicating the applied cut. The other neurons are associated with the other taupair background (4.23b), the Bhabha background (4.23c) and all other backgrounds (4.23d).

The agreement between data and simulation for high outputs of the ANN can be seen in Figure 4.24, where the ANN output for values greater than 0.8 is shown.

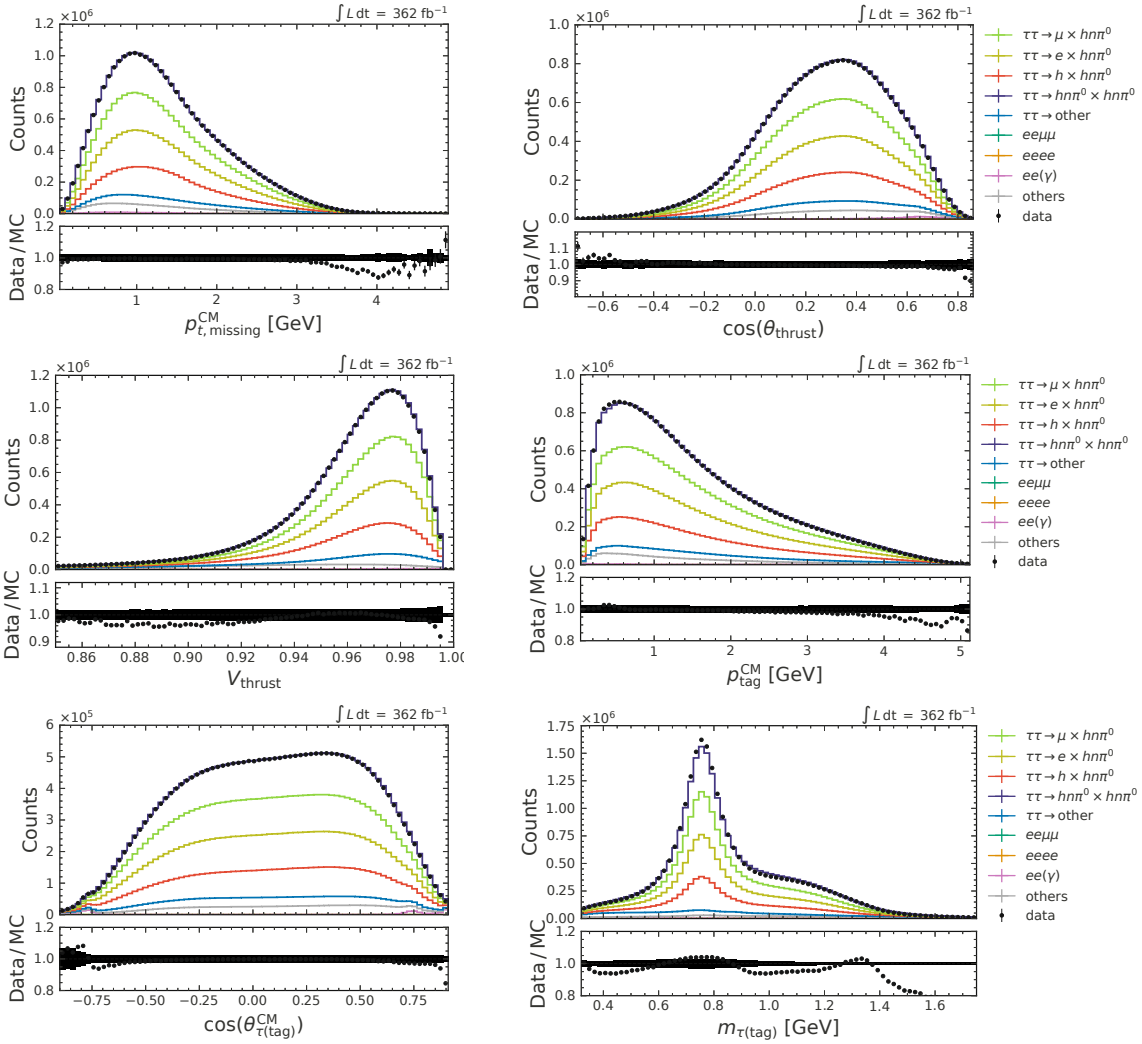


Figure 4.20: Data and MC distribution for the ANN inputs, without any PID selection applied on the signal track. The hatched area in the subplot indicates the uncertainty coming from π^0 and photon efficiency corrections.

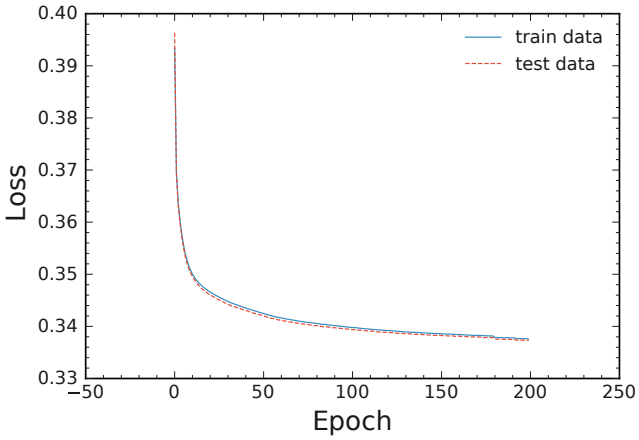


Figure 4.21: The evolution of the cross-entropy loss over 200 training epochs, shown for both the training and the test datasets.

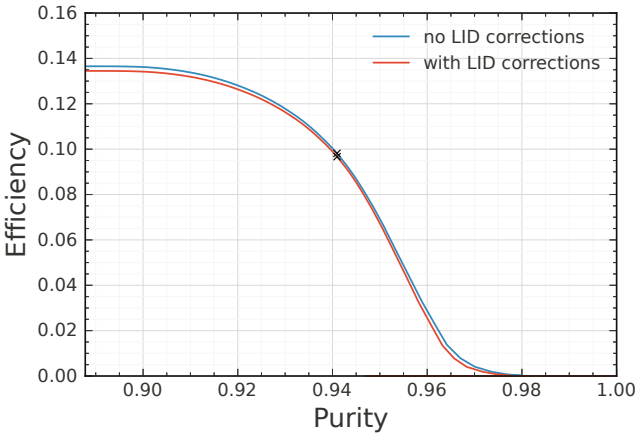
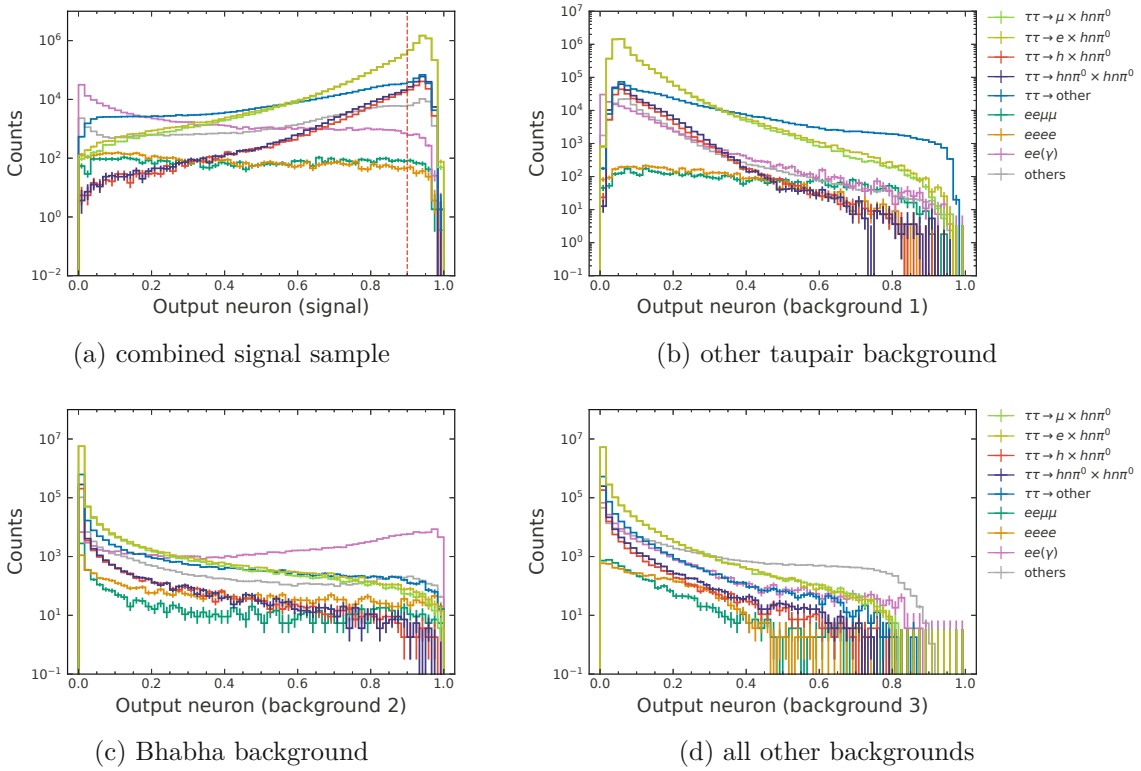


Figure 4.22: The ROC curve of the ANN with and without lepton ID corrections applied. The black crosses indicate the chosen cut value of 0.9. Efficiency and purity are computed for the combined signal sample.



(a) combined signal sample

(b) other taupair background

(c) Bhabha background

(d) all other backgrounds

Figure 4.23: The distributions of the four output neurons of the ANN, trained to distinguish between the combined signal and different types of background. The output neuron responsible for classifying the combined signal is shown in 4.23a. The vertical dashed line indicates a cut value of 0.9, which is used for the analysis.

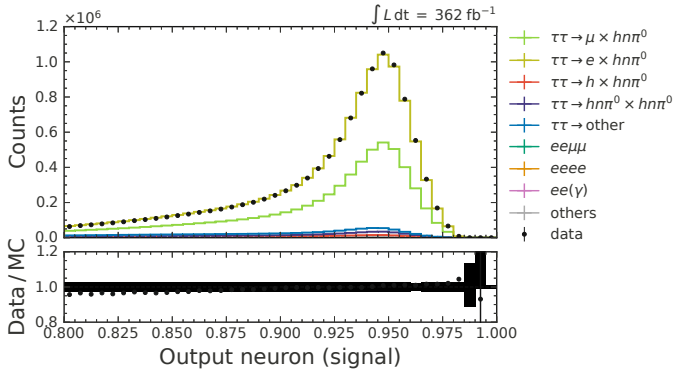


Figure 4.24: A comparison between data and simulation for the ANN output responsible for classifying the combined signal between 0.8 and 1.0. A threshold of 0.9 is used in the analysis to select signal events.

4.4.8 Sample before lepton selection

In this section, some data and MC distributions are shown after having the ANN cut applied but without any lepton ID requirement for the signal track. The uncertainty band in the data/MC subplot indicates the size of the π^0 and photon efficiency correction uncertainties. Figure 4.25 shows the distribution of the visible energy in the CM frame and Figure 4.26 shows the momentum and polar angle distributions of the signal track. In general, a good data/MC agreement is observed.

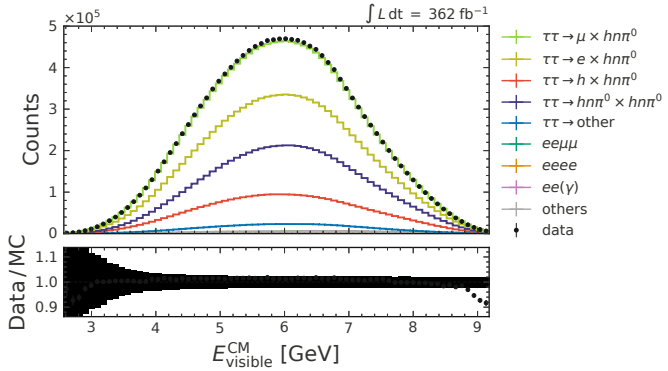


Figure 4.25: Data and MC distributions of the visible energy in the CM frame after applying the ANN cut but without any lepton ID selection on the signal track.

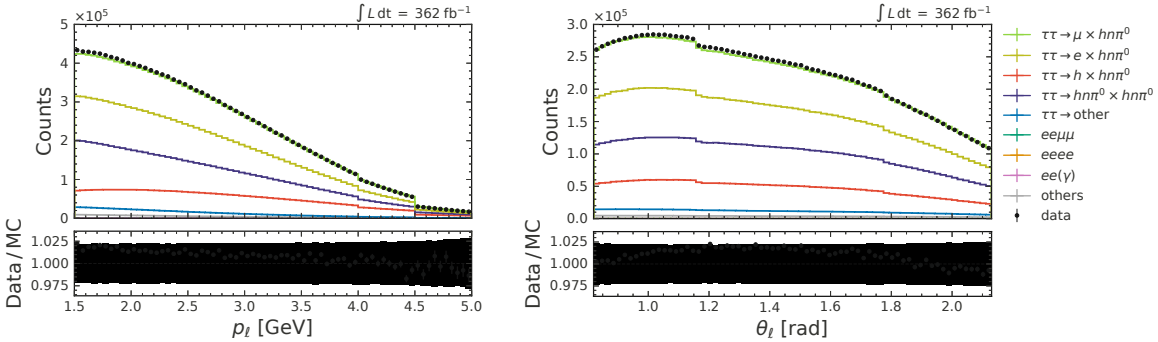


Figure 4.26: Data and MC distributions of the signal momentum in the laboratory frame (left) and the polar angle of the signal track (right) after applying the ANN cut but without any lepton ID selection on the signal track.

4.4.9 Sample after lepton selection

This section shows the distributions after applying the lepton ID requirements while keeping the channels combined. With this combined sample, all variables that cannot discriminate significantly between the e-channel and the μ-channel do not provide information about R_μ when compared to data, so there is little risk of being influenced by the observer-expectancy effect. Since some kinematic variables still have slightly different shapes for the two channels, originally only 20 % of the available dataset were used for these comparisons. For presentational purposes and completeness, in this thesis the full dataset is shown, obtained only after the final results were already unblinded. In Figure 4.27, the distributions of the input variables for

the ANN are shown, both in data and MC, after having applied the cut on the ANN output. The hatched area in the ratio indicates the uncertainty originating from lepton ID, π^0 and photon efficiency corrections. Figure 4.28 shows the distribution of the visible energy in the

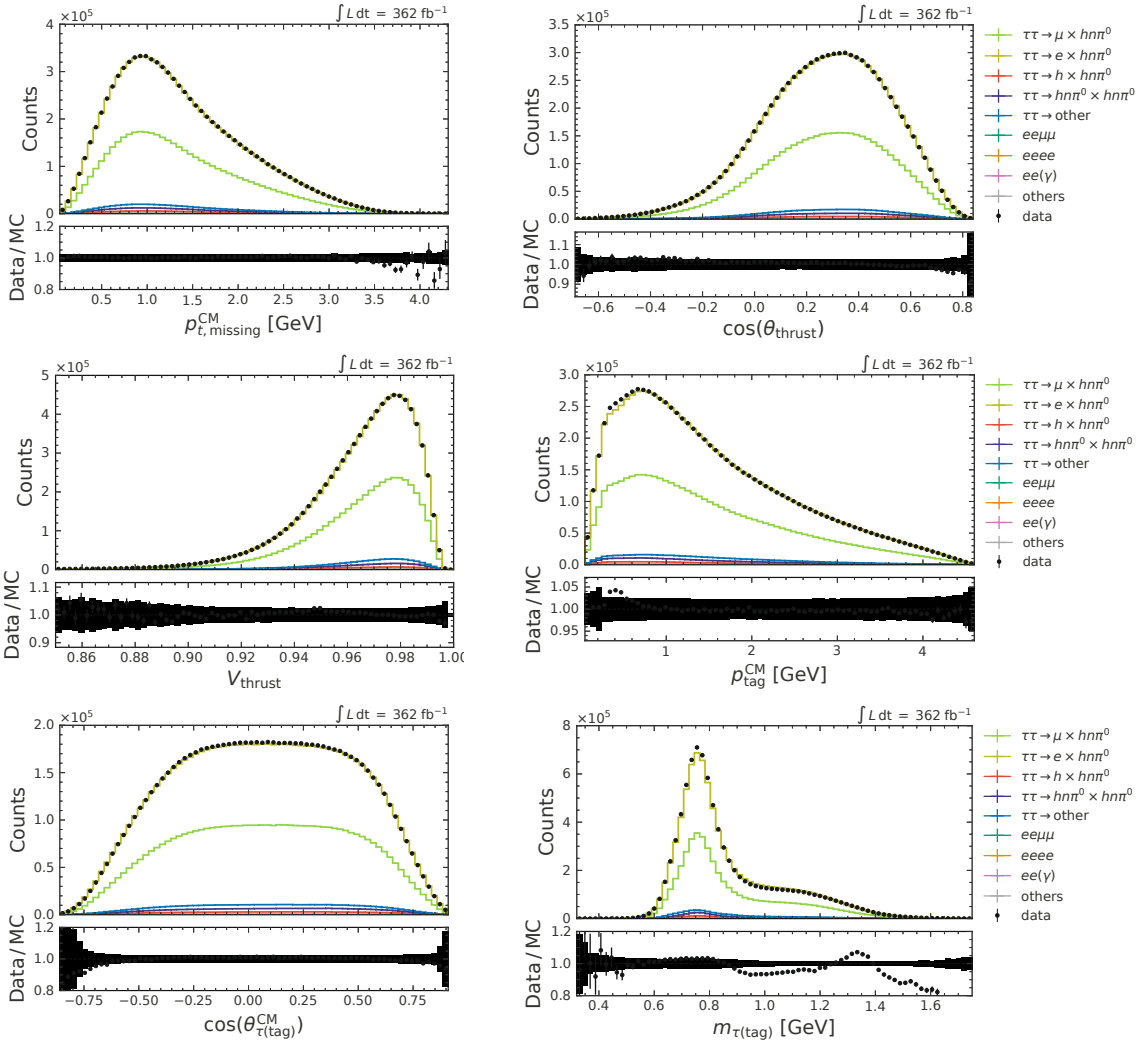


Figure 4.27: Data and MC distributions for the ANN inputs, after the cut on the ANN output, and with the combined lepton ID cut applied on the signal track. The hatched area in the subplot indicates the uncertainty originating from lepton ID, π^0 and photon efficiency corrections.

CM frame, the polar angle of the signal track and the momentum of the signal track. The left figures show only simulated distributions to better compare the shapes of the MC samples, while the right figures compare data and simulation.

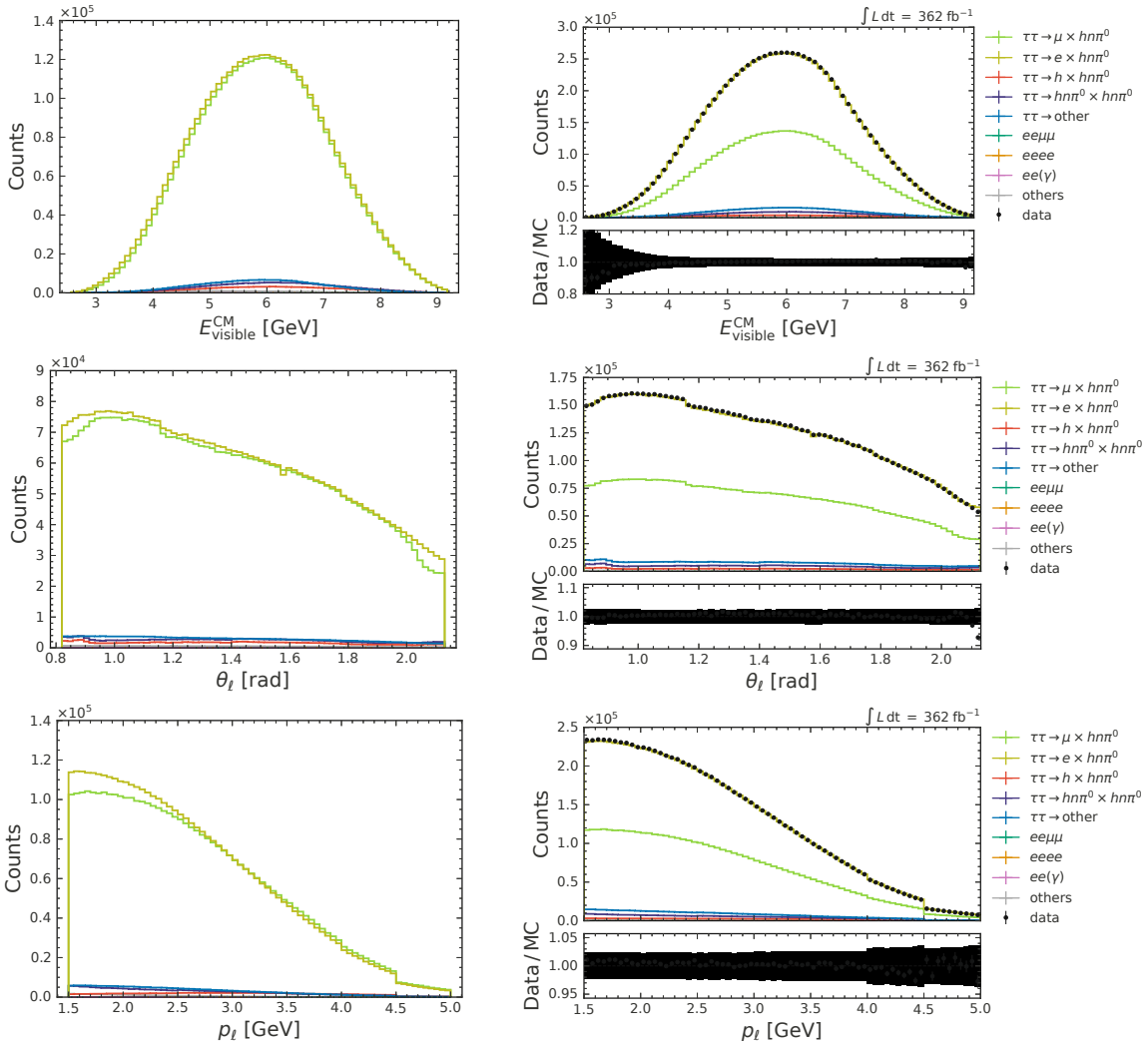


Figure 4.28: Distributions of the visible energy in the CM frame, the polar angle of the signal track and the signal momentum in the laboratory frame, after applying the ANN cut and with the combined lepton ID selection applied on the signal track. The left figures show only the MC distributions (not stacked), while the right figures show data and simulation, where the simulated samples are stacked on each other.

4.4.10 Selection summary

After applying all selection requirements, the signal reconstruction efficiency is found to be 9.48 % for $\tau^- \to e^- \bar{\nu}_e \nu_\tau$ and 9.44 % for $\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$. The e sample has a purity of 96 %, while the μ sample has a purity of 92 %. A summary of the signal efficiency, the number of surviving signal and background events scaled to the target luminosity, and the composition of the sample after all selections can be found in Table 4.15.

Table 4.15: A summary of the samples after all selection requirements for the e-channel, the μ-channel, and the combined $e + \mu$-channel. It includes the number of surviving signal and background events scaled to the target luminosity of 362 fb^{-1}, the resulting FOM, the signal efficiency, and the sample composition in terms of taupair and non-taupair events for each channel.

	e-channel	μ-channel	e+μ-channel
N_{sig}	4 096 340	3 978 925	8 075 306
N_{bkg}	170 197	340 719	510 875
FOM	1983.2	1914.4	2755.9
signal efficiency	9.48 %	9.44 %	9.46 %
$(\tau_{\text{sig}} \to e\nu_e\nu_\tau)(\tau_{\text{tag}} \to hn\pi^0\nu_\tau)$	96.01 %	< 0.01 %	47.71 %
$(\tau_{\text{sig}} \to \mu\nu_\mu\nu_\tau)(\tau_{\text{tag}} \to hn\pi^0\nu_\tau)$	< 0.01 %	92.11 %	46.34 %
$(\tau_{\text{sig}} \to hn\pi^0\nu_\tau)(\tau_{\text{tag}} \to hn\pi^0\nu_\tau)$	0.93 %	2.99 %	1.97 %
$(\tau_{\text{sig}} \to h\nu_\tau)(\tau_{\text{tag}} \to hn\pi^0\nu_\tau)$	0.36 %	2.16 %	1.27 %
other taupair events	2.29 %	2.30 %	2.29 %
$ee\tau\tau$	0.17 %	0.18 %	0.18 %
$c\bar{c}$	0.10 %	0.10 %	0.10 %
$u\bar{u}$	0.06 %	0.10 %	0.08 %

The exact composition of the sample after all selections and the combined lepton ID cut is broken down in Table 4.16. For taupair events, the true tau decay on the signal and tag sides is specified, and the identifier assigned to the respective tau decays in TAUOLA is listed (MCModes). The shown composition is obtained after applying all corrections to the simulated samples, and the shown number of events is scaled to the target luminosity of 362 fb^{-1}. The most significant background within the taupair sample comes from pions misidentified as muons, with $\tau^- \to h^- \nu_\tau$ or $\tau^- \to h^- n\pi^0 \nu_\tau$ decays on the signal side. These backgrounds with a charged pion as the decay product of the signal tau account for about 3.3 % of the total sample (1.3 % of the e sample and 5.2 % of the μ sample). Most of the remaining 2.7 % of the sample that is classified as background comes from events in which the decay of the tau on the tag side does not match the tag definition (see Table 4.11), such as $\tau \to \pi K^0_S \nu_\tau$. Most of these events still have the correct leptonic tau decay on the signal side. This corresponds to about 2.3 % of the sample in both the e-channel and the μ-channel. The largest non-taupair contribution is 0.2 % and comes from the two-photon process $e^+e^- \to e^+e^-\tau^+\tau^-$, which has the same contribution in both samples.

In addition, the composition of the sample in terms of the true particle type on the signal side in MC after applying all the selections and for different tightnesses of the lepton ID requirements is shown in Table 4.17.

Table 4.16: Signal and background for the $\tau(\to \ell\bar{\nu}\nu) \times \tau(\to h\pi^0(\pi^0)\bar{\nu}\nu)$ channel, after all selections and including the combined lepton ID cut. The number of events is scaled to the target luminosity of 362 fb^{-1}.

Signal track	Tag track	Events	Fraction [%]	MCModes
$\tau \to e\nu_e\nu_\tau$	$\tau \to \pi\pi^0\nu_\tau$	2 796 564.78	32.09	(1, 163)
$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to \pi\pi^0\nu_\tau$	2 767 863.19	31.76	(2, 163)
$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to \pi\pi^0\pi^0\nu_\tau$	1 161 607.69	13.33	(2, 111)
$\tau \to e\nu_e\nu_\tau$	$\tau \to \pi\pi^0\pi^0\nu_\tau$	1 130 932.56	12.98	(1, 111)
$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to \pi 3\pi^0\nu_\tau$	106 283.96	1.22	(2, 4)
$\tau \to e\nu_e\nu_\tau$	$\tau \to \pi 3\pi^0\nu_\tau$	102 140.49	1.17	(1, 4)
$\tau \to \pi\pi^0\nu_\tau$	$\tau \to \pi\pi^0\nu_\tau$	84 322.16	0.97	(163, 163)
$\tau \to \pi\nu_\tau$	$\tau \to \pi\pi^0\nu_\tau$	76 371.20	0.88	(303, 163)
$\tau \to \pi\pi^0\nu_\tau$	$\tau \to \pi\pi^0\pi^0\nu_\tau$	36 471.36	0.42	(163, 111)
$\tau \to e\nu_e\nu_\tau$	$\tau \to K\pi^0\nu_\tau$	33 218.64	0.38	(1, 165)
$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to K\pi^0\nu_\tau$	32 021.93	0.37	(2, 165)
$\tau \to \pi\nu_\tau$	$\tau \to \pi\pi^0\pi^0\nu_\tau$	29 565.58	0.34	(303, 111)
$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to \pi\pi^0\gamma\nu_\tau$	25 495.52	0.29	(2, 110)
$\tau \to e\nu_e\nu_\tau$	$\tau \to \pi\pi^0\gamma\nu_\tau$	24 825.38	0.28	(1, 110)
$\tau \to \pi\pi^0\pi^0\nu_\tau$	$\tau \to \pi\pi^0\nu_\tau$	23 510.04	0.27	(111, 163)
$\tau \to e\nu_e\nu_\tau$	$\tau \to \pi\nu_\tau$	14 096.19	0.16	(1, 303)
$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to \pi\nu_\tau$	12 224.63	0.14	(2, 303)
$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to \pi K^0_S\nu_\tau$	11 880.33	0.14	(2, 226)
$\tau \to e\nu_e\nu_\tau$	$\tau \to \pi K^0_S\nu_\tau$	11 427.18	0.13	(1, 226)
$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to \pi\pi^0 K^0_L\nu_\tau$	11 124.84	0.13	(2, 127)
$\tau \to e\nu_e\nu_\tau$	$\tau \to \pi\pi^0 K^0_L\nu_\tau$	10 838.05	0.12	(1, 127)
$\tau \to \pi\pi^0\pi^0\nu_\tau$	$\tau \to \pi\pi^0\pi^0\nu_\tau$	10 707.13	0.12	(111, 111)
$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to \pi 4\pi^0\nu_\tau$	8910.07	0.10	(2, 67)
$\tau \to e\nu_e\nu_\tau$	$\tau \to \pi 4\pi^0\nu_\tau$	8622.92	0.10	(1, 67)
$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to K\pi^0\pi^0\nu_\tau$	8264.88	0.09	(2, 106)
$\tau \to e\nu_e\nu_\tau$	$\tau \to K\pi^0\pi^0\nu_\tau$	7980.26	0.09	(1, 106)
$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to \pi\pi^0 K^0_S\nu_\tau$	7817.87	0.09	(2, 126)
$\tau \to e\nu_e\nu_\tau$	$\tau \to \pi\pi^0 K^0_S\nu_\tau$	7476.11	0.09	(1, 126)
$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to K\pi^0 K^0_L\nu_\tau$	5392.26	0.06	(2, 125)
$\tau \to e\nu_e\nu_\tau$	$\tau \to \mu\nu_\mu\nu_\tau$	5302.57	0.06	(1, 2)
$\tau \to e\nu_e\nu_\tau$	$\tau \to K\pi^0 K^0_L\nu_\tau$	5269.30	0.06	(1, 125)
$\tau \to \mu\nu_\mu\nu_\tau$	$\tau \to K 3\pi^0\nu_\tau$	4938.02	0.06	(2, 9)
$\tau \to e\nu_e\nu_\tau$	$\tau \to K 3\pi^0\nu_\tau$	4811.81	0.06	(1, 9)
$ee\tau\tau$		15 621.83	0.18	
$c\bar{c}$		8856.54	0.10	
$u\bar{u}$		6529.66	0.07	
$ee(\gamma)$		2046.96	0.02	
$s\bar{s}$		1033.25	0.01	
$d\bar{d}$		902.69	0.01	

Table 4.17: The sample composition in terms of the type of charged particle in the signal hemisphere. All analysis selections and corrections to the simulation have been applied. The composition is shown for varying levels of tightness of the lepton ID requirements. The nominal requirements used in the analysis are $P_e > 0.5$ for the e-channel channel and $P_\mu > 0.9$ for the μ-channel. Also shown is the fraction of events where the particle type in MC could not be retrieved due to a failed truth matching.

	$e^\pm$ [%]	$\mu^\pm$ [%]	$\pi^\pm$ [%]	$K^\pm$ [%]	Failed matching [%]
$P_e > 0.5$	98.57	< 0.01	1.36	0.05	0.01
$P_e > 0.9$	99.40	< 0.01	0.57	0.01	< 0.01
$P_e > 0.95$	99.56	< 0.01	0.43	< 0.01	< 0.01
$P_e > 0.99$	99.78	< 0.01	0.21	< 0.01	< 0.01
$P_\mu > 0.5$	< 0.01	93.09	6.19	0.28	0.43
$P_\mu > 0.9$	< 0.01	94.74	4.56	0.26	0.44
$P_\mu > 0.95$	< 0.01	95.04	4.26	0.25	0.44
$P_\mu > 0.99$	< 0.01	95.47	3.85	0.24	0.44

4.5 Statistical model

4.5.1 Fit strategy

The binned maximum likelihood method will be used to extract R_μ and its associated uncertainty from the data. It is a particular case of the ML method (see Section 2.5.1) where the observed events are organised into a total number of N bins. In this case, the data vector indicates the count of entries per bin. When the overall count of events is also random, the observed events adhere to a Poisson distribution[1] (Pois) with a mean equivalent to the intrinsic expected number of events in that bin. Given that the data are independent, the likelihood function can be obtained by multiplying the individual Poisson distributions [151].

The binned ML method has several advantages, although there is always a loss of information when using bins. Numerical optimisation is generally cheaper using the binned ML method, as the CPU usage tends to rise linearly with the event count in the unbinned case. Additionally, the binned approach allows the direct use of simulation templates for the distributions when an analytical expression is not present. For this analysis, the templates are defined from MC simulations, and in the following the fit will also be referred to as template fit.

The `pyhf` library [152], a Python adaptation of HistFactory [153], is used for the binned maximum likelihood fit. This library allows for the construction of complex statistical models that incorporate systematic effects through nuisance parameters. It uses the JSON data format to declare the full statistical model in a single document. This document can then be easily preserved, shared, and utilised to replicate the outcome of the analysis with various statistical techniques. Furthermore, the availability of published likelihoods will simplify the global combination of different measurements.

[1] If the total number of events is predefined by the design of the experiment, then the distribution of the events in each bin would be multinomial.

Table 4.18: Symbol notation for the fit function defined by a `pyhf` model, adapted from [154].

Symbol	Description
$\boldsymbol{n}$	observed event counts in data
$\boldsymbol{a}$	auxiliary data
$\boldsymbol{\eta}$	free parameters
$\boldsymbol{\chi}$	constrained parameters
$\nu(\boldsymbol{\eta}, \boldsymbol{\chi})$	calculated event rates
ν^0	nominal event rates
$\boldsymbol{\kappa}(\boldsymbol{\eta}, \boldsymbol{\chi})$	multiplicative rate modifier
$\boldsymbol{\Delta}(\boldsymbol{\eta}, \boldsymbol{\chi})$	additive rate modifier
$c_\chi\left(a_\chi \mid \chi\right)$	constraint term for constrained parameter χ
a_χ	auxiliary data associated with constrained parameter χ
Pois	the Poisson distribution

4.5.2 Likelihood specification

The general form of the fit function used by the `pyhf` library is [154]

$$f(\boldsymbol{n}, \boldsymbol{a} \mid \boldsymbol{\eta}, \boldsymbol{\chi}) = \prod_{c \in \text{channels},\, b \in \text{bins}_c} \text{Pois}\left(n_{cb} \mid \nu_{cb}(\boldsymbol{\eta}, \boldsymbol{\chi})\right) \prod_{\chi \in \boldsymbol{\chi}} c_\chi\left(a_\chi \mid \chi\right), \tag{4.17}$$

with

$$\nu_{cb}(\boldsymbol{\eta}, \boldsymbol{\chi}) = \sum_{s \in \text{samples}} \nu_{scb}(\boldsymbol{\eta}, \boldsymbol{\chi}) = \sum_{s \in \text{samples}} \left(\prod_{\kappa \in \boldsymbol{\kappa}} \kappa_{scb}(\boldsymbol{\eta}, \boldsymbol{\chi})\right) \left(\nu^0_{scb} + \sum_{\Delta \in \boldsymbol{\Delta}} \Delta_{scb}(\boldsymbol{\eta}, \boldsymbol{\chi})\right). \tag{4.18}$$

The meaning of the symbols is given in Table 4.18. Here, a bold symbol denotes a parameter set, which can be further specified by indices s, c and b. The index s denotes a *sample*, c indicates a *channel*, and b defines the individual bins. The samples correspond to different templates that are summed over to obtain the total calculated event rates in each bin. The event rates from each sample can be modified by multiplicative or additive rate modifiers, which are functions that are parameterised by nuisance parameters. In this measurement, all parameters $\boldsymbol{\eta}$ and $\boldsymbol{\chi}$ are nuisance parameters, except for R_μ, one of the free parameters and our parameter of interest. The fit function comprises a product of Poisson distributions, one for each bin in each channel. The channels are a collection of bins, defined over statistically independent datasets that are fit simultaneously. For the present measurement, two channels are needed to fit R_μ directly: one will be referred to as the electron channel (e-channel) and the other as the muon channel (μ-channel). The cut on the lepton ID variable for the lepton candidate in the signal hemisphere defines these channels in data and MC: $P_e > 0.5$ for the e-channel and $P_\mu > 0.9$ for the μ-channel.

4.5.3 Binning

The templates are defined over the momentum of the signal lepton candidate in the laboratory frame. The main reason for this is that lepton ID corrections and associated uncertainties are binned in momentum, and the impact of momentum bins where lepton ID is not well understood and the uncertainties are large can be reduced. One can think of the fit as

doing a combined measurement of R_μ in all momentum bins simultaneously. The additional information on the kinematic dependence would be lost in an integrated measurement over the full momentum range. This method can only be more precise than the integrated measurement, with the only disadvantage being that the systematic effects must be evaluated depending on the momentum, and their correlation across the momentum bins has to be accurately modelled.

As mentioned in Section 3.6.4, the lepton ID corrections from 1.5 GeV to 5.0 GeV are binned in steps of 0.5 GeV. To define the templates, these bins are further split into three bins of equal size, resulting in a total of 21 bins. This allows us to incorporate additional information from the shape of the templates in the fit. In general, having more bins will capture more information and result in a better sensitivity. On the other hand, having more bins increases the number of nuisance parameters and, consequently, the optimisation time. Also, having too many bins would make the fit more sensitive to a potential mismodelling of the templates and systematic effects. The split into three bins was chosen since it seemed to be a good trade-off between the disadvantages and benefits. It showed a significant improvement in sensitivity to R_μ with respect to only one or two bins while keeping the bin width (167 MeV) well above the momentum resolution, which depends on the momenta of the particles but is below 30 MeV [155]. Further increasing the bin count showed some improvement, but the increase in precision was less significant and started to plateau. Further discussion on the impact of the binning on the results can be found in Section 5.5.

4.5.4 Template definition

To measure R_μ, one template needs to define the signal component, and one or more templates describe the background component for each channel. Since most of the background (about 95 %) originates from $e^+e^- \to \tau^+\tau^-$ events, splitting the background templates according to the physics process makes little sense. Due to the importance of lepton ID uncertainty, it is beneficial to separate the templates according to the particle type of the signal lepton. A part of the background events will have the signal lepton correctly identified (as either electron or muon), while the other part will have misidentified particles ('fakes'). For the background with correctly identified leptons, the lepton ID uncertainty (from efficiency corrections) will be fully correlated with the signal template. In contrast, independent mis-id rate corrections and uncertainties, estimated from control samples with the corresponding particle type, will be used for the template with fakes. This distinction gives us three templates for each channel, for which the expected yields are shown in Figure 4.29. While the signal template and the background template with the correct lepton type are similar for both channels, the background template with misidentified particles is larger for the μ-channel. This is because almost all of these fake particles are pions, which are more likely to be misidentified as muons than electrons.

This results in the following event rates for each channel, which are used in the fit function:

$$\begin{aligned} \nu_b^e(\boldsymbol{\eta},\boldsymbol{\chi}) &= \kappa_e \cdot \nu_b^{e,\,\mathrm{sig}}(\boldsymbol{\eta},\boldsymbol{\chi}) + \nu_b^{e,\,\mathrm{bkg(true)}}(\boldsymbol{\eta},\boldsymbol{\chi}) + \nu_b^{e,\,\mathrm{bkg(fake)}}(\boldsymbol{\eta},\boldsymbol{\chi}) \quad \text{and} \\ \nu_b^\mu(\boldsymbol{\eta},\boldsymbol{\chi}) &= R_\mu \cdot \kappa_{e/\mu}^{\mathrm{gen}} \cdot \kappa_e \cdot \nu_b^{\mu,\,\mathrm{sig}}(\boldsymbol{\eta},\boldsymbol{\chi}) + \nu_b^{\mu,\,\mathrm{bkg(true)}}(\boldsymbol{\eta},\boldsymbol{\chi}) + \nu_b^{\mu,\,\mathrm{bkg(fake)}}(\boldsymbol{\eta},\boldsymbol{\chi}). \end{aligned}$$

In these expressions, $\nu_b^{e,\,\mathrm{sig}}$ and $\nu_b^{\mu,\,\mathrm{sig}}$ represent the signal yields that involve $\tau^- \to e^-\bar{\nu}_e\nu_\tau$ or $\tau^- \to \mu^-\bar{\nu}_\mu\nu_\tau$ decays on the signal side and $\tau^+ \to h^+ n\pi^0\bar{\nu}_\tau$ on the tag side. The background

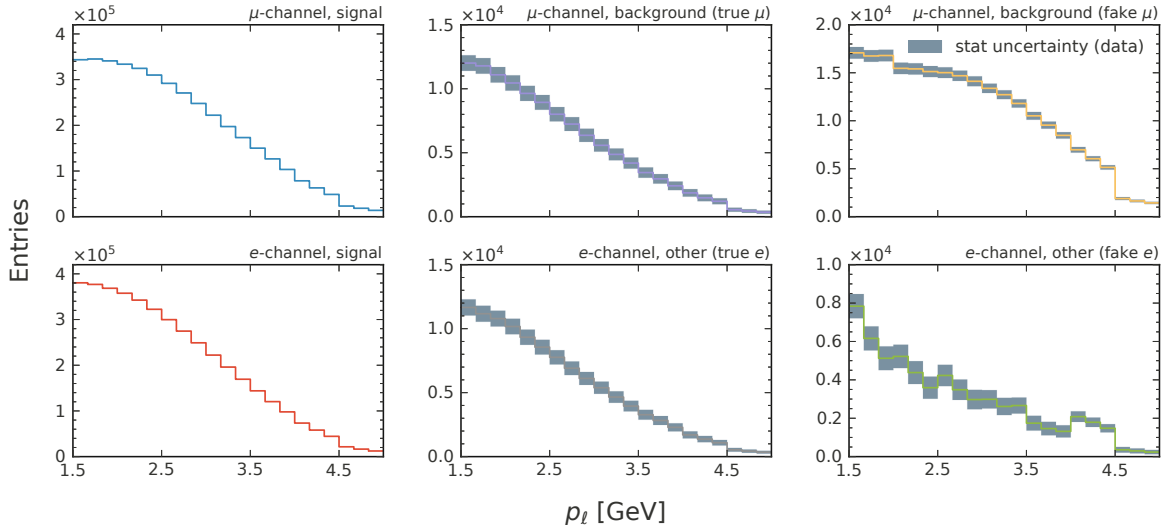


Figure 4.29: The six templates that are used for the fit, scaled to the target luminosity. The templates for the muon channel are in the first row, and the ones for the electron channel are in the second row. One template describes the signal components (first column), one the background with correctly identified leptons (second column) and one the background with all other particles that are misidentified as leptons (third column). The three templates for each channel are summed up and simultaneously fit the data.

components are described by $\nu_b^{e,\,\mathrm{bkg(true)}}$ and $\nu_b^{\mu,\,\mathrm{bkg(true)}}$, which correspond to background events with a correctly identified lepton on the signal side, and $\nu_b^{e,\,\mathrm{bkg(fake)}}$ and $\nu_b^{\mu,\,\mathrm{bkg(fake)}}$, which represent all other background events. The factor κ_e is a multiplicative rate modifier that affects the signal yields of the e and μ templates simultaneously. This enables a direct estimation of R_μ from the fit by only including it as a modifier of the signal template in the muon channel. The factor $\kappa_{e/\mu}^{\mathrm{gen}} \equiv \mathcal{B}_e^{\mathrm{gen}}/\mathcal{B}_\mu^{\mathrm{gen}}$ is a constant that ensures the correct normalisation of R_μ, with the branching fractions

$$\mathcal{B}_\mu^{\mathrm{gen}} \equiv \mathcal{B}^{\mathrm{gen}}[\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau] = 0.173937$$

and

$$\mathcal{B}_e^{\mathrm{gen}} \equiv \mathcal{B}^{\mathrm{gen}}[\tau^- \to e^- \bar{\nu}_e \nu_\tau] = 0.178175$$

used in the generator. The parameter of interest (R_μ) is then estimated along with the nuisance parameters by optimising the fit function 4.17 using the observed data.

4.6 Systematic uncertainties

Systematic uncertainties are included as modifications to the event counts of the templates through nuisance parameters. The resulting variations of the templates should reflect the effects of each systematic source on the lepton momentum distribution. In `pyhf`, such variations can be defined with so-called modifiers. These modifiers add either a multiplicative or an additive term to the nominal event rates of the associated samples. An overview of the available modifiers, the type of modification, and the corresponding constraint term and input is shown in Table 4.19. For the histosys and normsys modifiers, the nuisance parameter

Table 4.19: The relevant modifiers and corresponding constraint terms that can be used to build a `pyhf` model, adapted from [154]. Modifying the event rates is accomplished either with multiplicative (κ) or additive (Δ) terms, which are parametrised by the nuisance parameters. If applicable, the constraint size of these nuisance parameters can be specified with an additional input, shown in the last column.

Name	Modification	Constraint Term c_χ	Input
shapesys	$\kappa_{scb}(\gamma_b) = \gamma_b$	$\prod_b \text{Pois}\left(r_b = \sigma_b^{-2} \mid \rho_b = \sigma_b^{-2}\gamma_b\right)$	σ_b
histosys	$\Delta_{scb}(\alpha) =$ $f_p(\alpha \mid \Delta_{scb,\alpha=-1}, \Delta_{scb,\alpha=1})$	$\text{Gaus}(a = 0 \mid \alpha, \sigma = 1)$	$\Delta_{scb,\alpha=\pm1}$
normsys	$\kappa_{scb}(\alpha) =$ $g_p(\alpha \mid \kappa_{scb,\alpha=-1}, \kappa_{scb,\alpha=1})$	$\text{Gaus}(a = 0 \mid \alpha, \sigma = 1)$	$\kappa_{scb,\alpha=\pm1}$
staterror	$\kappa_{scb}(\gamma_b) = \gamma_b$	$\prod_b \text{Gaus}(a_{\gamma_b} = 1 \mid \gamma_b, \delta_b)$	$\delta_b^2 = \sum_s \delta_{sb}^2$
lumi	$\kappa_{scb}(\lambda) = \lambda$	$\text{Gaus}(l = \lambda_0 \mid \lambda, \sigma_\lambda)$	$\lambda_0, \sigma_\lambda$
normfactor	$\kappa_{scb}(\mu_b) = \mu_b$		

interpolates between a given up and down variation of the histograms ($\Delta_{scb,\alpha=\pm1}$) or the normalisation ($\kappa_{scb,\alpha=\pm1}$). The interpolating functions are f_p and g_p, respectively. Different types of interpolation are possible, but the method chosen for the present work is piecewise linear interpolation [153]. This means that the values of the function are linearly interpolated between the nominal and the up or down variation. An example of this for the histosys modifier is shown in Figure 4.30, where a nominal histogram is shown together with the up and down variation. The piecewise linear interpolation with a value of $\alpha = 0.7$ is overlayed, which is for each bin at 70 % of the distance between the nominal and the up variation. The type of modifier

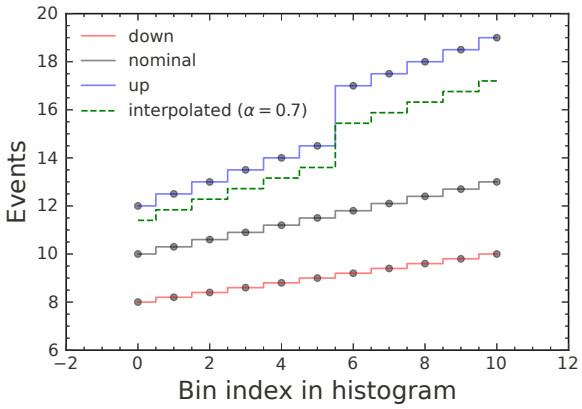


Figure 4.30: An example of the interpolating function f_p for the histosys modifier. The green dashed line is the output of the function with $\alpha = 0.7$ when, together with the nominal histogram, the shown down and up-variations are provided.

that will be used for the different systematic effects mainly depends on the correlation between the template bins. For example, the shapesys modifier adds one independent multiplicative nuisance parameter for each sample bin. This allows for changes in the shape of the distribution

described by this sample, which is why it is often referred to as an *uncorrelated shape* modifier. In contrast, the histosys modifier only allocates a single nuisance parameter that interpolates between the nominal event rates of the sample and the provided up and down variations. This is fully correlated between the bins, so it is often referred to as a *correlated shape* modifier. In `pyhf`, the same nuisance parameter can also be assigned to different templates, which is essential to correlate a systematic variation across multiple samples or channels. Since nuisance parameters are always fully independent by construction, the available modifiers can only be used to model either fully correlated or completely independent effects. However, arbitrary correlation across template bins can be modelled using multiple correlated shape modifiers. This method will be used for the most dominant systematic uncertainty, lepton identification, which is described in further detail in Section 4.6.2.

4.6.1 Treatment of template variations

Split of variations into normalisation and shape components

For better fit stability, it is recommended to split the variations of the templates into normalisation (normsys) and shape (histosys) components, which are fully correlated. The normalisation component is calculated with

$$\kappa_{\text{norm}}^{\text{up/dn}} = \sum_{b \in \text{bins}} \frac{n_b^{\text{up/dn}}}{n_b^{\text{nom}}}, \tag{4.19}$$

where $n_b^{\text{up/dn}}$ are the up or down variations and n_b^{nom} are the nominal values of the templates in bin b. The shape component is calculated with

$$n_{b,\,\text{shape}}^{\text{up/dn}} = \frac{n_b^{\text{up/dn}}}{\kappa_{\text{norm}}}. \tag{4.20}$$

This split into two components will always be realised whenever a correlated shape modifier is used. In `pyhf`, the nuisance parameter associated with the two components is shared between modifiers if given the same name (α in Table 4.19). Doing this split is mathematically equivalent to just using a single histosys component but can lead to improved fit behaviour. The number of nuisance parameters allocated by the normsys and shapesys components is still one.

Modifier veto

In cases where the modifications to the templates are minimal and the variation of the nuisance parameters has a negligible effect on the likelihood, it is better not to include the parameters since they might cause instabilities in the fit. To handle such cases, certain conditions for not including systematics in the model are defined. For the correlated shape variations, both of the following conditions must be met:

- The p-value of a χ^2 test that compares the shape variation to the nominal histogram is larger than 0.99 *or* the largest difference between the shape variation and the nominal histogram is smaller than 10 % of the smallest expected statistical uncertainty in data.
- The difference between the normalisation component from unity is smaller than 10 % of the smallest relative expected statistical uncertainty in data.

The expected statistical uncertainty in data is calculated as

$$\sigma^{\text{stat}}_{\text{exp},\,cb} = \sqrt{\sum_{s \in \text{samples}} \nu^0_{scb}}, \tag{4.21}$$

with ν^0_{scb} being the nominal event rates in MC. The smallest expected statistical uncertainty in data is always taken from all bins of the corresponding channel, which are affected by the shape variation. If the modifier is shared for multiple samples, it has to be vetoed in all the templates to be omitted.

For the uncorrelated shape variations, the veto is applied bin-wise if the following condition is met:

- The uncertainty is smaller than 10 % of the expected statistical uncertainty in data for the corresponding bin.

In the following, the expected relative statistical uncertainty in data will be indicated as hatched areas in the figures that show the variations of the templates, for example, in Figure 4.33. The legend will show the smallest relative expected statistical uncertainty in data for the corresponding template. Furthermore, if the variation is a correlated shape, the p-values of the χ^2 test and the difference of the normalisation component from one will be shown separately for up and down variations. The plot also indicates whether the individual variations meet the veto conditions (when shown in red) or not (when shown in green). Only if a systematic effect is vetoed for all the samples is the template modification not included in the model, which is also shown on the plots with a checkbox.

For uncorrelated shape modifiers, the legend shows the largest relative variation of the corresponding template. Since the veto, in this case, is applied bin-wise, whether the modification is included is shown by the colour of the individual variations: red when the veto applies and green when it fails.

Symmetrisation

All template variations with individual up and down components are symmetrised before including them in the model. The symmetrisation is done by keeping the difference between the up and down variations constant and only introducing a shift to equal their difference from the nominal template. This has only a small impact on most of the covered cases. In cases where only the up variation is shown in the following figures, the down variation is always included as the symmetric version of the up variation.

Smoothing

For some specific cases, the 353QH algorithm [156] is used twice to smooth the histogram variations before including them in the model. Smoothing helps to remove statistical fluctuations in cases where the up and down variations of the histogram are obtained by an independent reprocessing of MC or data samples, only with some parameter changes, or when the variation is based on random sampling. The smoothing is applied only to the relative variations with respect to the nominal histogram, so it only affects the fluctuations in the variation itself. The

Table 4.20: The treatment of the template variations for the different systematic sources. The checkmark indicates when either smoothing or symmetrisation is applied.

	Smoothed	Symmetrised
Photon energy	✓	✓
Momentum correction	✓	✓
π^0 efficiency	✓	✓
Photon efficiency	✓	✓
Particle decay-in-flight	✓	✓
Modelling of ISR	✓	✓
Modelling of FSR	✓	✗
Neural network selection	✓	✗
Lepton identification	✗	✗
Sample normalisation	✗	✗

statistical fluctuations of the templates, which originate from a limited sample size, are not affected by this and are treated separately.

The template variation before symmetrisation or smoothing is shown as a dashed grey line in the following figures. In contrast, the resulting symmetric and smoothed variation is shown as a dashed red line.

Table 4.20 shows whether the systematic variations described in the following were smoothed or symmetrised with the procedures mentioned above.

4.6.2 Lepton identification

To correctly model the correlation of lepton ID uncertainties, the following relation between the covariance matrix C_{ik}^{sys} and the correlated shape variations Γ_{ij} can be used:

$$C_{ik}^{\text{sys}} = \sum_{j=1}^{N_s} \Gamma_{ij}\Gamma_{kj}, \tag{4.22}$$

where Γ_{ij} is a variation of bin i associated to some systematic source j and the sum is over N_s independent systematic sources. If instead any arbitrary covariance matrix C^{sys} is given, it can be decomposed into $N_s = \dim(C^{\text{sys}})$ shape variations via eigendecomposition:

$$C^{\text{sys}} = VUV^T = (V\sqrt{U})(V\sqrt{U})^T = \Gamma\Gamma^T \Longrightarrow \Gamma = V\sqrt{U}. \tag{4.23}$$

Here, V is a matrix with columns corresponding to eigenvectors, and U is a diagonal matrix with the related eigenvalues. Each eigenvector, multiplied by the square root of the corresponding eigenvalue, gives a correlated shape variation that can be implemented with histosys modifiers in `pyhf`. Together, these independent shape variations model the original covariance matrix. This process can be thought of as a coordinate rotation in the hyperdimensional space of a multivariate normal distribution, defined by the original covariance matrix, such that the new covariance matrix is diagonal. In some cases, the eigenvalues might be zero or very close to zero. This will be, for example, the case for strongly correlated uncertainties, which would result in only one dominant shape variation. To avoid unnecessary parameters in the fit, one can define a threshold for the eigenvalues below which the corresponding variation is ignored.

To implement lepton ID uncertainties in the model with this procedure, we first need to obtain the covariance matrix, which describes the correlation between the template bins. The statistical component of the lepton ID corrections is naturally fully independent across all correction bins. However, the correlation of the systematic component is currently unknown, and some conservative assumptions must be made while maintaining a realistic model. The lepton ID corrections are binned in three dimensions, for which the correlation has to be considered:

1. Charge: Positive and negative charged tracks are combined in the measurement. Having fully correlated uncertainties will result in the largest uncertainty on R_μ.
2. Polar angle: Because the polar angle bins are summed over, the uncertainties are also considered fully correlated for this dimension to be most conservative.
3. Momentum: The fit is performed with templates binned in momentum, making it very sensitive to the correlation of lepton ID uncertainties in this dimension. A priori, it is unclear which correlation assumption is most suitable. To find the most conservative option, different cases have to be studied. As shown later in Section 5.3.2, assuming that the uncertainties are fully correlated is the best choice with the current model.

Using these assumptions, the covariance matrix is obtained using a toy approach. For this, 5000 variations of the correction factors are generated with k-dimensional multivariate Gaussian distributions that follow the correlation assumed for the different dimensions. The statistical component of the corrections is described with a random vector,

$$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{\mathbf{stat}}), \tag{4.24}$$

where $\boldsymbol{\Sigma}^{\mathbf{stat}}$ is a diagonal matrix containing the corresponding variances and $\boldsymbol{\mu}$ is the central value of the individual weights. For the systematic component, the random vector

$$\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}^{\mathbf{sys}}), \tag{4.25}$$

is used, where $\boldsymbol{\Sigma}^{\mathbf{sys}}$ contains all the information about the assumed correlations. In the most general case, it can be expressed as

$$\boldsymbol{\Sigma}^{\mathbf{sys}} = \begin{bmatrix} \sigma_{x_1} & & 0 \\ & \sigma_{x_2} & \\ & & \ddots \\ 0 & & \sigma_{x_k} \end{bmatrix} \begin{bmatrix} 1 & \rho_{x_1,x_2} & \cdots & \rho_{x_1,x_k} \\ \rho_{x_2,x_1} & 1 & \cdots & \rho_{x_2,x_k} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{x_k,x_1} & \rho_{x_k,x_2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \sigma_{x_1} & & 0 \\ & \sigma_{x_2} & \\ & & \ddots \\ 0 & & \sigma_{x_k} \end{bmatrix}, \tag{4.26}$$

where σ_{x_i} are the systematic uncertainties corresponding to each bin. If the uncertainty is asymmetric, the larger of the up or down components is used as σ_{x_i}. The correlations are ρ_{x_i,x_j}. With the abovementioned assumptions, this would result in $\rho_{x_i,x_j} = 1$. The covariance matrices will span all momentum/theta/charge bins of a single type of lepton ID corrections. The types of corrections used for this analysis are

- e efficiency: affecting 2 samples in the e-channel: signal and background (true e)
- μ efficiency: affecting 2 samples in the μ-channel: signal and background (true μ)

- $\pi \to e$ mis-id rate: affecting 1 sample in the e-channel: background (fake e)
- $\pi \to \mu$ mis-id rate: affecting 1 sample in the μ-channel: background (fake μ)
- $K \to e$ mis-id rate: affecting 1 sample in the e-channel: background (fake e)
- $K \to \mu$ mis-id rate: affecting 1 sample in the μ-channel: background (fake μ)

and they are considered to be completely independent from each other. A summary of the σ_{x_i} in the region relevant for this analysis is given in Table 4.21. The table shows the overall range of the correction factors, the average correction factors, the overall range of the uncertainties on the correction factors, and the average uncertainty. The numbers given for mis-id rate are for pions, which are misidentified as leptons.

Table 4.21: A summary of the correction factors and associated uncertainties for lepton identification efficiencies and the $\pi \to \ell$ mis-id rates. The range gives the largest and smallest values, while the average is the weighted average over all events, specific for this analysis. The uncertainties are expressed as a percentage of the corresponding correction factors.

	Correction		Uncertainty	
	Range	Average	Range	Average
Electron efficiency	$0.99 - 1.01$	1.00	$0.05\% - 1.35\%$	0.22%
Muon efficiency	$0.87 - 0.99$	0.97	$0.10\% - 6.00\%$	0.38%
$\pi \to e$ mis-id	$1.00 - 13.0$	2.61	$3.54\% - 162\%$	41.4%
$\pi \to \mu$ mis-id	$0.25 - 1.10$	0.82	$3.11\% - 76.9\%$	11.7%

In the end, we obtain a k-dimensional random weight vector,

$$\mathbf{w} = \mathbf{X} + \mathbf{Y}, \tag{4.27}$$

from which the 5000 toy weights are generated. This is repeated for each type of correction, with the matching toy weights applied to the MC sample to obtain 5000 sets of templates, binned in signal momentum with the binning mentioned in 4.5.3. To illustrate this, the distributions of the total number of events in the signal template for the electron channel and the muon channel are shown in Figure 4.31. Also shown as vertical dashed lines are the values obtained when the weights are varied up or down by their uncertainty, adding statistical and systematic components in quadrature. This would correspond to a fully correlated variation (also for the statistical component). The uncertainties of the electron efficiency corrections are dominated by the systematic component, so for the e-channel, the fully correlated variations are well within the distribution obtained from the toys. In contrast, the muon efficiency corrections are dominated by the statistical component, resulting in a much smaller spread of the toys with respect to the up and down variations in the μ-channel. This point will be discussed in more detail later.

Using the set of toy templates, the sample covariance between bin j and bin k can be computed as

$$C_{jk} = \frac{1}{N-1} \sum_{i=1}^{N} \left(x_{ij} - \bar{x}_j\right)\left(x_{ik} - \bar{x}_k\right), \tag{4.28}$$

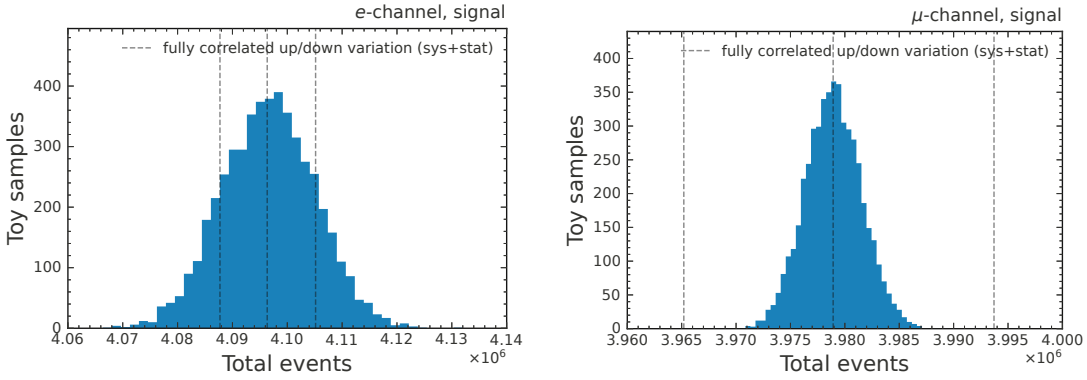


Figure 4.31: The total number of events in the signal template for the electron channel (left) and the muon channel (right), obtained with different generated toy weights of the efficiency correction factors. The dashed lines represent the values obtained when varying the weights up or down by their total uncertainty.

where x_{ij} is the i^{th} toy value of the j^{th} bin, $\bar{x}_j$ denotes the sample mean of the corresponding bin, and the sum is over all N toys. Thus, for the six templates with 21 bins each, we obtain a covariance matrix of dimension $6 \cdot 21 = 126$. However, since there is only one lepton candidate in each event and the correction factors for different particle types are independent, there is a lot of redundant information in the full covariance matrix and most of the correlations are either 0 or 1. When computed with Equation 4.28, they will never be precisely 0 or 1 due to statistical fluctuations and limited numerical precision. Still, the dimensionality can be reduced based on the knowledge of these correlations. In addition, there is no need to explicitly consider all 21 bins for each template since the correction factors themselves are binned in steps of 0.5 GeV, resulting in only 7 bins over the relevant momentum range. Alternatively, as mentioned above, one could remove the redundant variations by setting a threshold on the eigenvalues. Nonetheless, in the present case, it is worthwhile to incorporate these features in advance. This also allows to more easily separate the uncertainties for the different types of correction factors. In the end, only $4 \cdot 7 = 28$ independent shape variations are needed, which can be divided into four types, corresponding to e efficiency ($e^{\pm} \to e^{\pm}$), μ efficiency ($\mu^{\pm} \to \mu^{\pm}$), e mis-id rate ($h^{\pm} \to e^{\pm}$) and μ mis-id rate ($h^{\pm} \to \mu^{\pm}$) systematics. The e and μ efficiency uncertainties are fully correlated for the two templates in each channel with the same lepton type, since they are based on the same toy weights. The correlations obtained from the toys are shown in Figure 4.32. Each 7×7 block shows the correlation in the seven momentum bins for the different correction types. For the e efficiency, the correlations over momentum are close to one, while for the μ efficiency, the correlation matrix is almost entirely diagonal. This difference comes from the underlying uncertainties, which for the μ efficiency are almost entirely dominated by the (uncorrelated) statistical component, while for the e efficiency, the (correlated) systematic component is dominant. The μ efficiency corrections are obtained primarily from the $\mu\mu\gamma$ calibration channel, which has very small systematic uncertainties due to the purity of the channel. The uncertainties of the μ efficiency corrections shown in Figure 3.16 in Section 3.6.6 are almost exclusively statistical, especially for high momentum. The e efficiency corrections are obtained mainly from the Bhabha calibration channel, which has more considerable associated systematic uncertainties due to an unexplained discrepancy between the correction factors obtained with different selections,

which was discussed in Section 3.6.3. The resulting relative variations of the individual

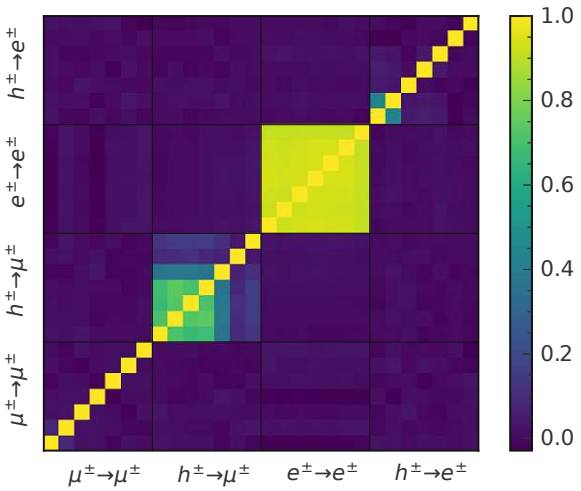


Figure 4.32: The correlation matrix of the uncertainties related to lepton efficiencies ($\mu^\pm \to \mu^\pm$ and $e^\pm \to e^\pm$) and mis-id rates ($h^\pm \to \mu^\pm$ and $h^\pm \to e^\pm$), which is implemented with multiple variations of the templates in the fit model. Each submatrix illustrates the correlation across the momentum bins. The correlations are obtained from the underlying correlations between the correction bins using 5000 toy samples.

templates are shown in Figure 4.33. Each colour denotes an independent histosys variation, which, taken together, models the desired correlation. Only the individual shape variations for the signal and background with the same lepton type are fully correlated for each channel (note that the relative shape variations are precisely the same for the two templates). Since the uncertainties are almost completely independent in the muon channel, each variation only dominates a single bin. On the contrary, one variation dominates everything in the electron channel, while the others are close to zero because of the high correlation. The size of the dominant variation is around 0.2 %, which also corresponds to the uncertainty in the relevant bins of Figure 3.16. The uncertainties associated with fake corrections are generally much larger and have non-trivial correlations due to contributions from both statistical and systematic components of the uncertainties, such as from the tau calibration channel described in Section 3.6.3. The differences in the correlations of the uncertainties of the efficiency corrections are also why the overall impact of electron identification on the uncertainty of R_μ is larger than for muon identification, even though the numbers in Table 4.21 would suggest the opposite. The impact of all different systematic sources on R_μ will be given in Section 4.7.

4.6.3 π^0 efficiency

The π^0 reconstruction efficiency correction compensates for the difference between the efficiency of π^0 reconstruction in data and simulation. The correction factors are measured using decays of D mesons, namely $D^0 \to K^-\pi^+\pi^0$, $D^0 \to K^-\pi^+$ and $B^+ \to \pi^+ D^{*0}$ ($\to \pi^0 D^0$ ($\to K^-\pi^+$)). In addition, they are cross-checked using $\tau^+ \to \pi^+\pi^-\pi^+ n\pi^0\nu_\tau$ decays, with $n > 0$. All these studies were conducted by the Belle II performance group and the resulting correction factors

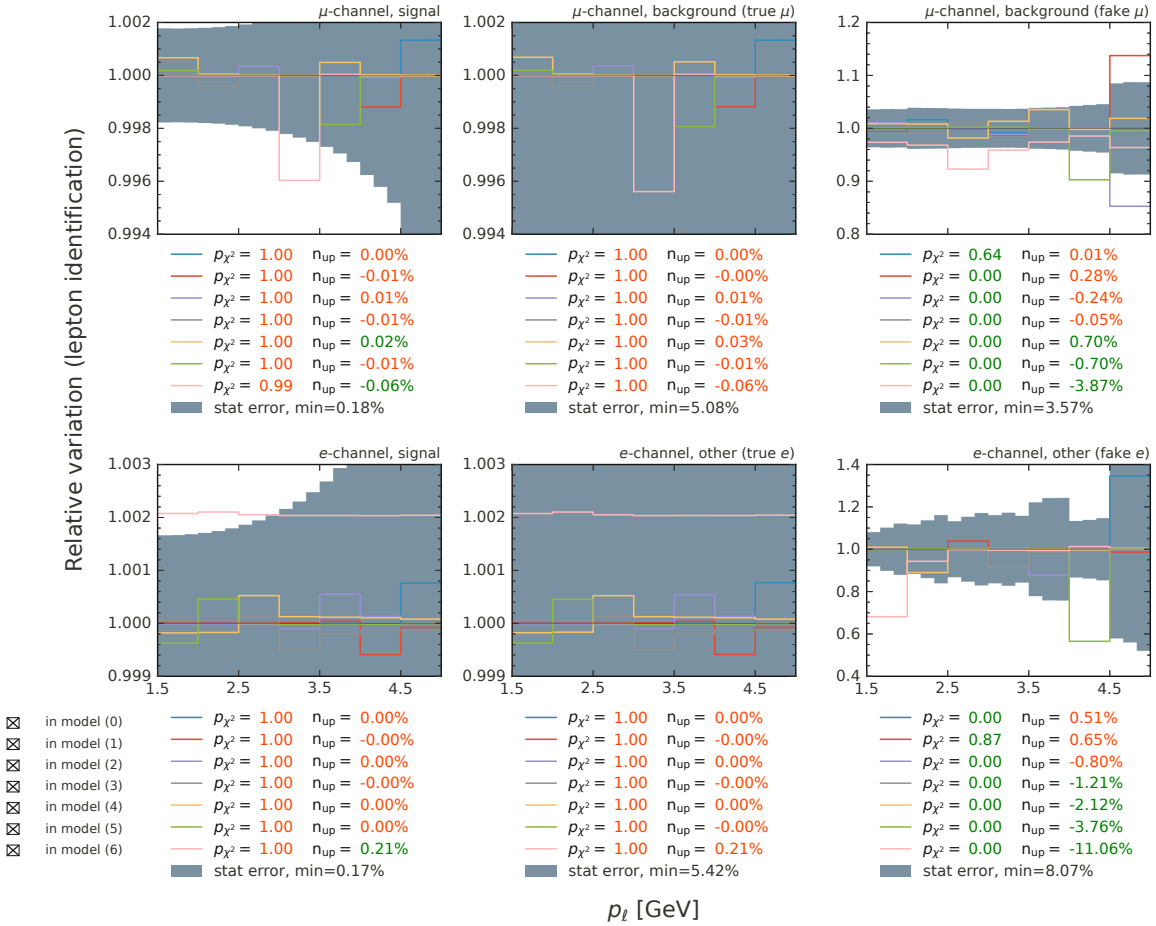


Figure 4.33: The relative shape variations of the templates due to lepton ID uncertainties. Each template is affected by seven independent histosys modifiers, shown in different colours. Together, these shape variations model the desired correlation between the template bins.

are provided to all analysts. The correction factors are measured as a function of the π^0 momentum with three components to the total error for each bin:

- the statistical error,
- the systematic error associated with the corrections from the D meson channel and
- a systematic error obtained from the differences in correction factor between the D meson and tau channels.

The central values of the correction factors, together with statistical and systematic uncertainties, are shown in Table 4.22. The recommendations of the performance group are to take the statistical component as uncorrelated and the systematic components as completely correlated across momentum bins but independent from each other. For simplicity, since this measurement is not very sensitive to the π^0 corrections and the systematic uncertainties are dominant for all the momentum bins, the statistical component will also be considered fully correlated, which will only slightly inflate the size of the overall uncertainties. The π^0

Table 4.22: Correction factors and uncertainties for the π^0 selection used in this analysis.

π^0 momentum [GeV]	Correction	Stat.	Syst. (D meson)	Syst. (channel dep.)
0.2-3.0	1.000	±0.003	±0.037	-
0.2-0.4	0.917	±0.004	±0.049	-
0.4-0.6	0.965	±0.004	±0.036	-
0.6-0.8	0.988	±0.004	±0.079	-
0.8-1.0	1.013	±0.005	±0.058	-
1.0-1.5	1.042	±0.004	±0.045	±0.039
1.5-2.0	1.044	±0.005	±0.041	±0.051
2.0-3.0	1.011	±0.005	±0.040	±0.030

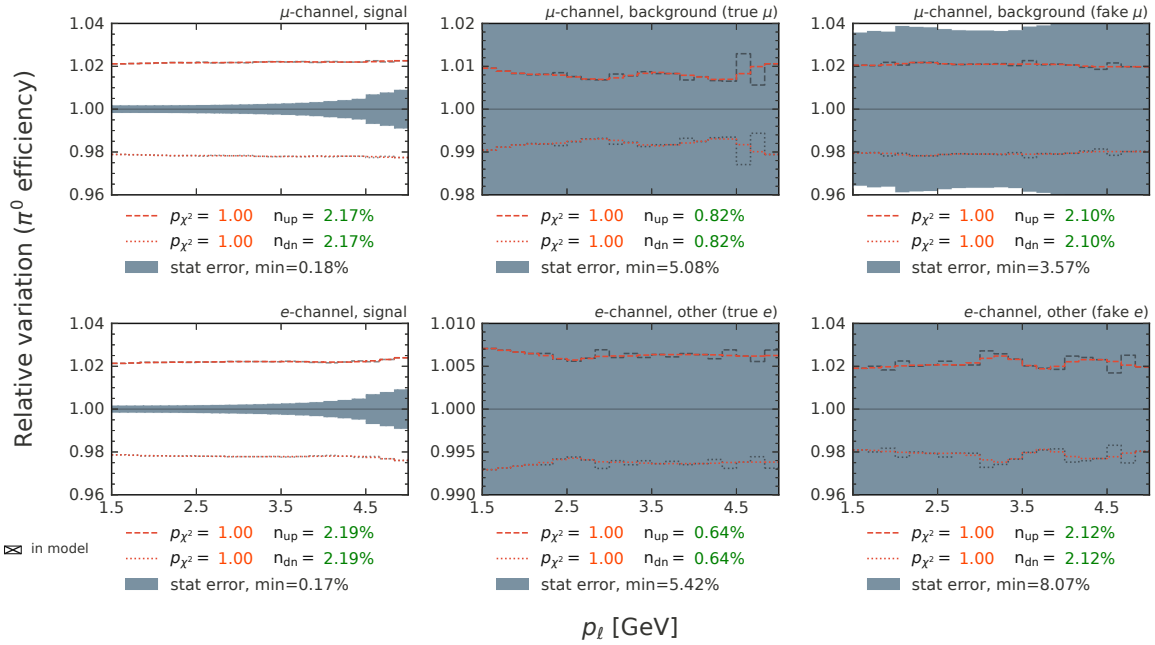


Figure 4.34: The relative variation of the templates due to the π^0 efficiency correction uncertainty. It affects the muon and electron channels similarly, thus cancelling mostly for the determination of R_μ.

corrections are applied by correcting the distributions of the π^0 multiplicity, accounting for all bin migrations. This correction is repeated for the down variation of the correction factors. After applying the tag side π^0 requirement, two samples are obtained, where the sample with the down variation is a subset of the nominal sample. The down variation of the templates is the ratio of these samples, which is taken to be fully correlated, and thus implemented with a histosys modifier. The up variation is implemented as the symmetric version of the down variation. The relative variations of the templates are shown in Figure 4.34.

4.6.4 Trigger

A correction must be applied to the MC templates to account for differences in trigger simulation and data. This correction is applied to each bin and has an associated error, a

systematic uncertainty for the measurement. The triggers used in this analysis are defined in Section 3.4.7, and the nominal trigger selection and the method for estimating trigger efficiencies were described in Section 4.4.3.

In Figure 4.35, the trigger efficiency for the electron and muon channels is shown for data and MC. In data, the trigger efficiency can only be assessed with the reference method ('ref'), as defined in Equation 4.14. In contrast, in MC, also the absolute efficiency without restriction to a reference sample can be computed ('abs'). The difference between the efficiency in MC, calculated with the reference and absolute methods, can then be taken as a systematic uncertainty due to the reference trigger method. Any difference between the two methods can be attributed to not fully independent ECL and CDC triggers. In addition, an uncertainty derived from the change in the data/MC ratio when using an alternative reference sample ('alt ref') is included. For this alternative definition, the `fso` and `syo` bits are used as reference triggers in addition to `ffo` and `fyo` (see Section 3.4.7 for the trigger bit definitions).

The efficiencies computed with the alternative reference events are also shown in the figure and deviate very little from the nominal efficiencies in data. However, there is a significant shift in the MC efficiencies, which is included as an uncertainty on the correction factors. This should reflect the systematic uncertainty of the mismodelling of the reference triggers.

The ratio of the trigger efficiency in data and MC (computed with the reference method) is used as a correction factor for the MC templates. This correction factor is shown in the subplot of Figure 4.35, where the error bars include all the effects mentioned above, which are added in quadrature:

- the statistical uncertainty from data and MC sample size,
- the difference between the trigger efficiency computed in the reference sample and the absolute efficiency in MC,
- the difference between the correction factor when computed in the reference sample and an alternative reference sample (with the addition of `fso` and `syo` trigger lines).

The uncertainty associated with the trigger correction is then assigned as a variation of the templates using 42 histosys modifier. This variation is implemented to be independent across momentum bins and channels, but fully correlated for the individual momentum bins across the three samples in each channel. The relative uncertainty assigned to each bin is shown in Figure 4.36. The most significant contribution to uncertainty is the difference between the correction factor computed with the nominal and alternative reference triggers. An additional fully correlated (normsys) variation is added to the muon channel only, which is based on the difference of the correction factors for the integrated e and μ channels. This should account for the unlikely case that the observed difference in correction factors originates from an unaccounted systematic effect only present in one channel. This difference is very small and amounts to 0.037 %, which is also reassuring in the sense that the trigger correction itself does not have a significant impact on R_μ. Another correlated normsys variation is added as the difference of the integrated trigger correction factor computed using the nominal and the alternative reference triggers, which is implemented fully correlated between the e and the μ-channel. This should account for the fact that the observed differences due to the choice of the reference trigger likely originate from a common source, which would make them correlated across momentum and the channels. Since the origin of this discrepancy is not fully

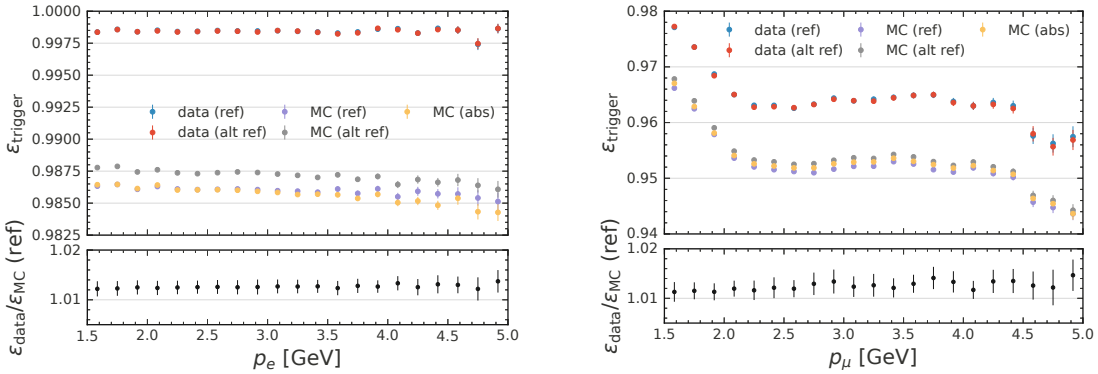


Figure 4.35: The trigger efficiency in data and MC for the e-channel (left) and the μ-channel (right). For MC, the efficiency within the reference sample and the absolute trigger efficiency are shown. The subplots show the data/MC ratio of the trigger efficiency, which is used to correct the MC templates. The error bars on the ratio include the statistical and systematic uncertainties.

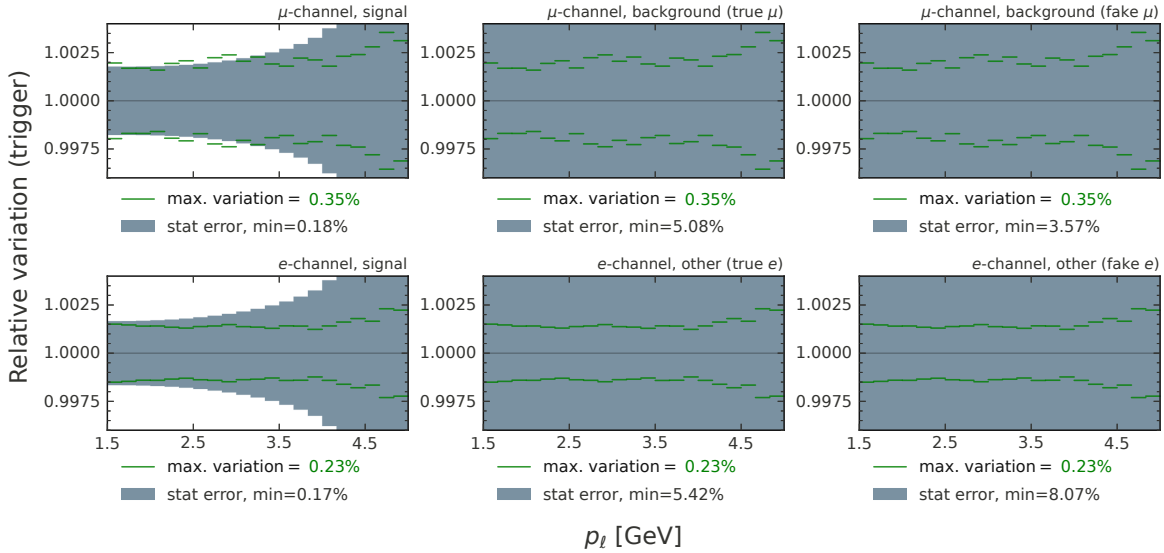


Figure 4.36: The relative uncertainty associated with the trigger correction, assigned independently to each template bin.

understood, the exact correlations are not known, which is why it was included in the model as a fully independent effect, which is the more conservative choice for the resulting uncertainty on R_μ. However, it allows the overall normalisation of the templates not to vary as much as if they were correlated, which is why this variation is included in addition. The integrated difference is 0.158 % in the μ-channel and 0.132 % in the e-channel. This overall normalisation does not have a significant impact on R_μ, since the size of the variation in both channels is very similar. The trigger efficiencies in data and MC, the resulting correction factors, and the quantities used to compute the uncertainties on the correction factor are shown in Table 4.23 for the momentum integrated samples of the μ-channel and the e-channel. The similarity of the differences with the usage of alternative reference triggers indicates that this effect is highly correlated across the channels, and cancels when propagated to R_μ.

Implementing the trigger uncertainty as independent across momentum bins and channels means that the uncertainty associated with the trigger will likely be overestimated. The impact of reducing this uncertainty on the total uncertainty of R_μ is minimal, for which some studies will be shown in Section 5.3.7.

Table 4.23: A summary of quantities related to the trigger correction and associated uncertainties computed for the integrated muon and electron samples, as well as a combination of the two. The trigger efficiencies ε in data and MC are shown together with their statistical uncertainties. The resulting data/MC correction factor κ_{nom} differs only slightly between the μ-channel and the e-channel. Also shown is the resulting statistical uncertainty on the correction factor (both from data and MC) σ_{stat}, the change of the correction factor when using the absolute efficiency in MC (κ_{abs}), and the change of the correction factor when using an alternative reference sample (κ_{alt}).

	μ-channel	e-channel	μ+e-channel
ε (data)	$0.966\,02 \pm 0.000\,10$	$0.998\,45 \pm 0.000\,02$	$0.982\,31 \pm 0.000\,05$
ε (MC)	$0.954\,41 \pm 0.000\,07$	$0.986\,08 \pm 0.000\,04$	$0.969\,52 \pm 0.000\,04$
κ_{nom}	1.012 17	1.012 54	1.013 20
$\sigma_{\text{stat}}/\kappa_{\text{nom}}$	0.013 %	0.004 %	0.007 %
$(\kappa_{\text{nom}} - \kappa_{\text{abs}})/\kappa_{\text{nom}}$	0.072 %	−0.009 %	0.033 %
$(\kappa_{\text{nom}} - \kappa_{\text{alt}})/\kappa_{\text{nom}}$	0.158 %	0.132 %	0.149 %

4.6.5 Modelling of final-state radiation

To account for the mismodelling of FSR in MC, a variation of the templates should be included in the model, which reflects such effects. This could be a non-negligible source of systematic uncertainty since it will affect the electron channel much more than the muon channel. In addition, because FSR changes the momentum of the lepton, an impact on the shape of the distributions is expected. To correctly vary the amount of FSR present in MC, the momenta of all generated FSR photons on the signal hemisphere are saved. For this, FSR photons are defined as primary photons with a τ lepton as an ancestor and no π^0 as a mother. To simulate the effect of a reduced FSR rate, the (generator) momenta of a random fraction of the FSR photons are added to the signal lepton momentum. This is done for 5 % of all photons, which is considered conservative.

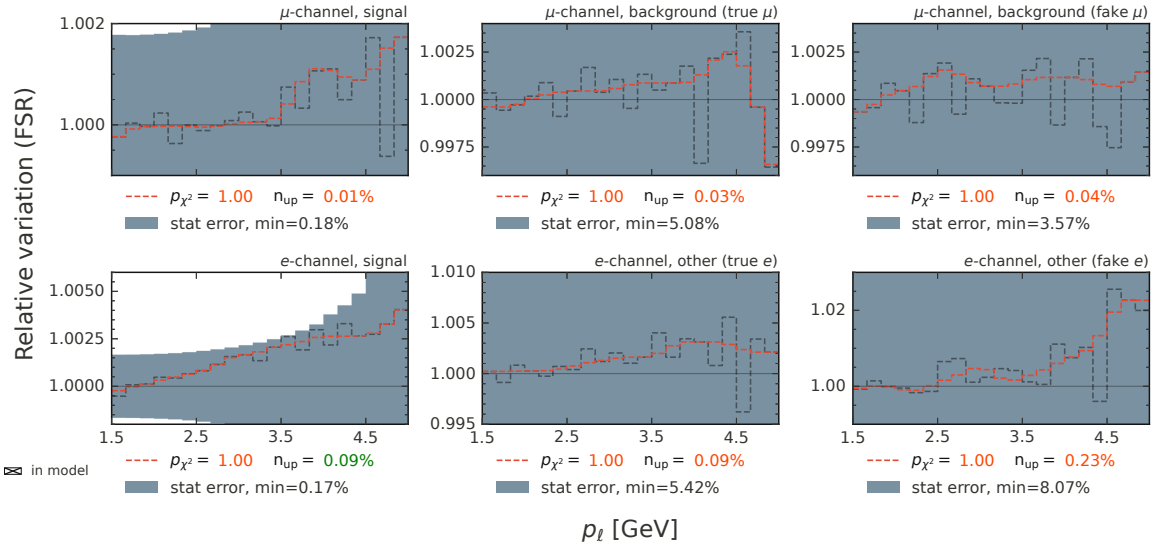


Figure 4.37: The variations of the templates as a result of a change in FSR rate.

The magnitude of this variation is based on studies with a J/ψ sample, using $J/\psi \to e^+e^-$ to analyse the modelling of final-state radiation, as well as $J/\psi \to \mu^+\mu^-$ as a control mode without radiative effects. In the region sensitive to FSR mismodelling, an agreement well within 5 % between data and MC was observed. These studies with the J/ψ sample were carried out by a Belle II colleague.

The simulation program responsible for generating FSR, PHOTOS [128], is also expected to be much more accurate than 5 %. In leptonic tau decays, the agreement with the exact analytical formula of radiative corrections to the decay rate was found to be around 1 % [157].

Unfortunately, adding FSR photons to MC is not so straightforward. However, it is reasonable to assume that the effect of a higher FSR rate would be exactly opposite to that of a lower rate. Based on this assumption, the symmetric variation of the templates will be enough to capture both effects. The resulting variation of the templates is implemented with a correlated shape modifier (histosys) and is shown in Figure 4.37.

4.6.6 Modelling of initial-state radiation

To estimate the effect of mismodelled initial-state radiation (initial-state radiation (ISR)) in MC, events are scaled by

$$w_{\text{ISR}} = (1 \pm 0.05)^{N_{\text{ISR}}}, \tag{4.29}$$

where N_{ISR} is the number of generated photons without a τ lepton as ancestor and with an electron or positron as mother particle in the event. Such varied templates are shown in Figure 4.38 and included in the model with a correlated shape modifier (histosys). The 5 % variation used in Equation 4.29 is a conservative estimate of the size of this effect. Additional checks were performed by varying ISR photons from 5 % to 20 %, and for all cases the impact on R_μ was comparably small (see Section 5.5.2). This is because the effect of ISR essentially cancels due to the similar influence on the templates in the electron and muon channels.

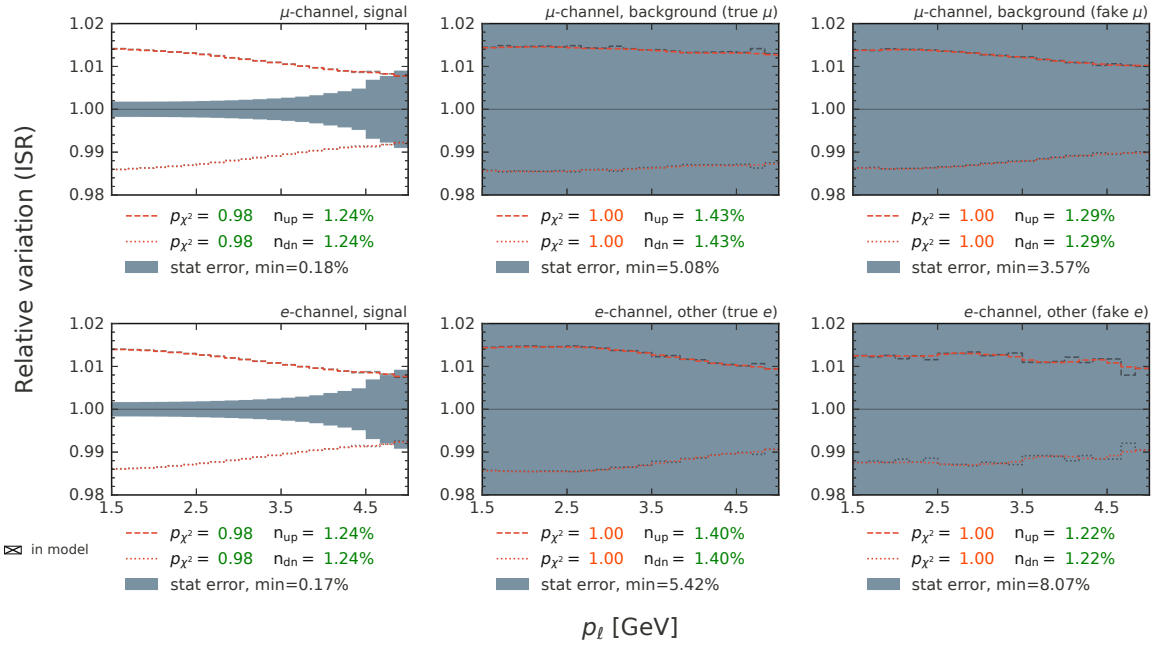


Figure 4.38: The template variations associated with a change in ISR rate.

4.6.7 Particle decay-in-flight

To estimate the effect of mismodelled decay-in-flight rate of particles in MC, events are scaled by

$$w_{\text{decay}} = (1 \pm 0.05)^{N_{\text{decay}}}, \tag{4.30}$$

where N_{decay} is the number of truth-matched tracks that are reconstructed from the decay product of another (long-lived) particle. Such varied templates are shown in Figure 4.39 and included in the model with a correlated shape modifier (histosys). Similar to the ISR variation, the 5 % variation is thought to be conservative and the impact on R_μ will be small due to the similar variation of the signal templates in the electron and muon channels. The main difference between the two channels comes from $\tau \to \pi\nu_\tau$ and $\tau \to \pi\pi^0\nu_\tau$ decays on the signal side. This is because the pion has a much shorter lifetime (2.6×10^{-8} s) compared to the muon (2.2×10^{-6} s), and due to helicity suppression, it decays about 99.97 % of the time into a muon and a muon neutrino [10]. In the MC simulation, these events are still considered to have a true muon, which is why this difference appears in the background template with the true lepton.

4.6.8 Modelling of the momentum distribution

Since the objective is to measure the normalisation ratio of two templates, a systematic effect that affects both channels equally would largely cancel. To account for any such effect that would change the shape of the distribution in a way not covered by one of the included shape variations, additional parameters are added to the fit model. For this, uncertainties are added for each of the 21 bins of the templates. These uncertainties are implemented to be fully correlated across all six templates, but independent between the momentum bins. With `pyhf`, this is done by adding correlated shape modifiers, which are shared for all templates, but

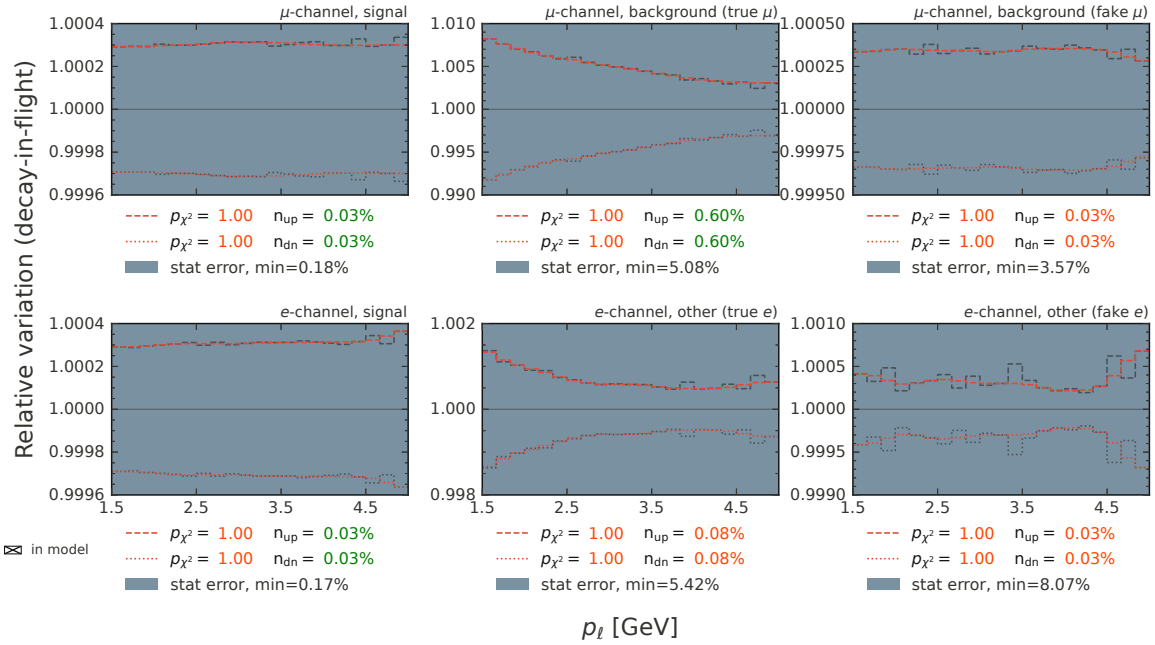


Figure 4.39: The template variations associated with a change in decay-in-flight rate.

the shape is only different from the nominal templates in a single bin. This amounts to one additional nuisance parameter per template bin.

The size of this variation is obtained from the spread of the data/MC ratio in a sample with all selection cuts applied except for the lepton ID requirement on the signal track, previously shown in Section 4.4.8. This greatly increases the background with pions on the signal side, while not being sensitive to any effects of particle identification. The standard deviation of the data/MC ratios computed with the same binning over the lepton momentum, which is used for the fit, amounts to 0.0081. This ratio and the corresponding distributions in data and simulation are shown in Figure 4.40. Using the standard deviation of these data/MC ratios as an estimate to reflect the variation in the shape of the momentum distribution, an uncertainty of 0.81 % is assigned to each momentum bin in the model. This will be referred to as a 'global shape' systematic, which can be thought of as an uncertainty to account for the mismodelling of the momentum distribution.

4.6.9 Sample normalisation

In addition to a correlated shape difference, there could also be some systematic effect that affects the overall normalisation in all bins. This is addressed by assigning an uncertainty on the normalisation of the different samples. This includes effects such as cross-section and branching fraction uncertainties as well as normalisation uncertainties due to background mismodelling.

Taupair normalisation

To obtain a normalisation uncertainty for the taupair process, the same sample that was used for the global shape systematic (Section 4.6.8) is used as a control sample. It has all the

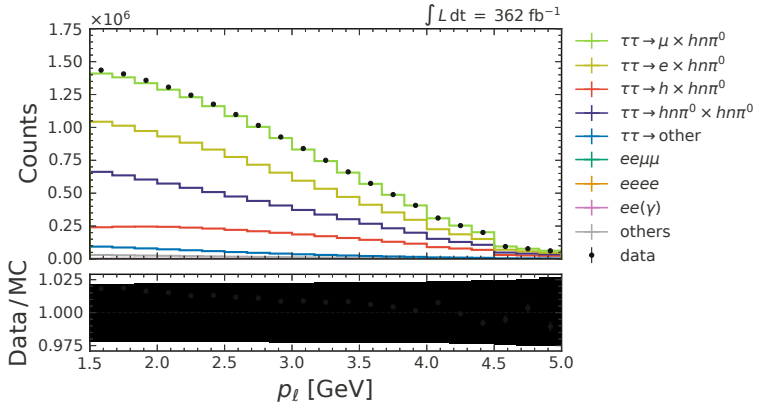


Figure 4.40: The distribution of the signal momentum in data and MC, with the same binning used for the template fit. No lepton ID selection was applied to estimate the systematics associated with the modelling of the momentum distribution. The standard deviation of the data/MC ratios, shown in the lower plot, is used to determine the size of the corresponding uncertainty in the momentum bins.

selection cuts applied except for particle identification on the signal track, which significantly increases the amount of taupair events while keeping background from other processes at a minimum. This sample is used to estimate a normalisation uncertainty without being sensitive to lepton ID uncertainties. The distribution of visible energy in the CM frame was shown in Figure 4.25 in Section 4.4.8, indicating good data/MC agreement. The distribution of signal momentum with the same binning used for the fit was shown in the previous Section, in Figure 4.40. To account for potential normalisation differences originating from non-taupair processes, a simple two-component fit with templates corresponding to taupair and non-taupair events is performed. A 25 % normalisation uncertainty (normsys) is assigned to the non-taupair template and an uncorrelated shape uncertainty (shapesys) of 0.81 % (taken from Section 4.6.8) to the taupair template. This allows for changes in the shape of the distribution, which is already covered by the global shape systematic. The pre-fit distribution is shown in Figure 4.41 on the left. An estimate of the normalisation of the taupair template, which is left entirely unconstrained, is obtained from the fit. The difference of the estimated scaling from unity with the addition of the fit error is used as the uncertainty for the taupair process. If this uncertainty is smaller than the dominant uncertainties known to affect the normalisation of taupair events, they are used instead. In this analysis, this is the uncertainty on the taupair cross-section, 0.33 % [158], and the branching fraction uncertainty of the tag side tau decays, 0.62 % [10]. Their quadrature sum amounts to 0.7 %. For the present case the fit result is 1.0072 ± 0.0021, with the post-fit distribution shown in Figure 4.41 on the right. This amounts to a normalisation uncertainty for the taupair process of 0.92 %, which exceeds the 0.7 % mentioned before. The estimated uncertainty from data is still sensitive to effects originating from the π^0 requirement on the tag side, which is already taken into account in a separate systematic. However, the chosen approach is conservative and at worst the overall normalisation uncertainty is slightly overestimated, which does not significantly impact the uncertainty on R_μ. The data-driven method was chosen to be able to account for larger normalisation differences since otherwise the fit model would not be flexible enough to adapt the scale of the templates to the data level. In particular, this could be the case for some of

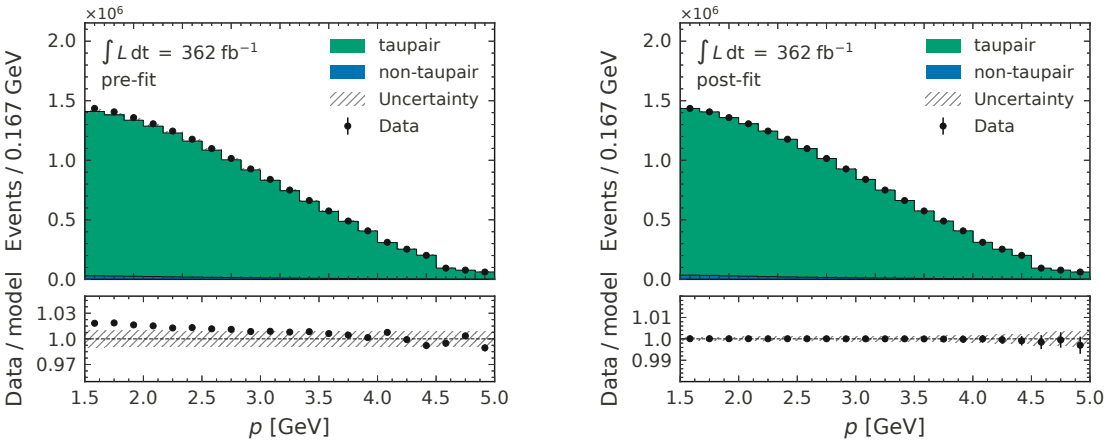


Figure 4.41: The pre-fit (left) and post-fit (right) distributions of the control sample, which is used to verify the normalisation uncertainty of the taupair process.

the sample splits used in Section 5.5 to test the stability of the result, since some of these splits use subsets of data for which no separate corrections are available. However, for this nominal case the observed data/MC difference of the taupair enriched sample is consistent with what was expected.

Taupair with wrong signal side

For taupair events with a misidentified signal side but the correct tag side, the uncertainty of the corresponding tau branching fraction is used as a normalisation uncertainty. This is dominated by $\tau \to \pi\nu$ with 1.03 % and $\tau \to \pi\pi^0(\pi^0)$ with 0.62 %.

Taupair with wrong tag side

As seen in Table 4.16, many events are classified as background because they have the wrong tag side decay, even if there is a leptonic tau decay on the signal side. Many of these 'fake tag' events have a mis-reconstructed K^0_S or K^0_L in the decay, for which an additional uncertainty should be assigned. A conservative uncertainty of 25 % is chosen for the normalisation of these events, as the impact on R_μ is minimal.

Non-taupair events

Similarly, an overall normalisation uncertainty for non-taupair backgrounds is assigned. For this, the different backgrounds are grouped as they only contribute very little to the composition of the total background (see Table 4.16):

- the $ee\tau\tau$ sample
- the $q\bar{q}$ samples: $u\bar{u}$, $d\bar{d}$, $c\bar{c}$, $s\bar{s}$
- the $\ell\ell$ samples: $ee(\gamma)$, $\mu\mu(\gamma)$
- the $ee\ell\ell$ samples: $eeee$, $ee\mu\mu$

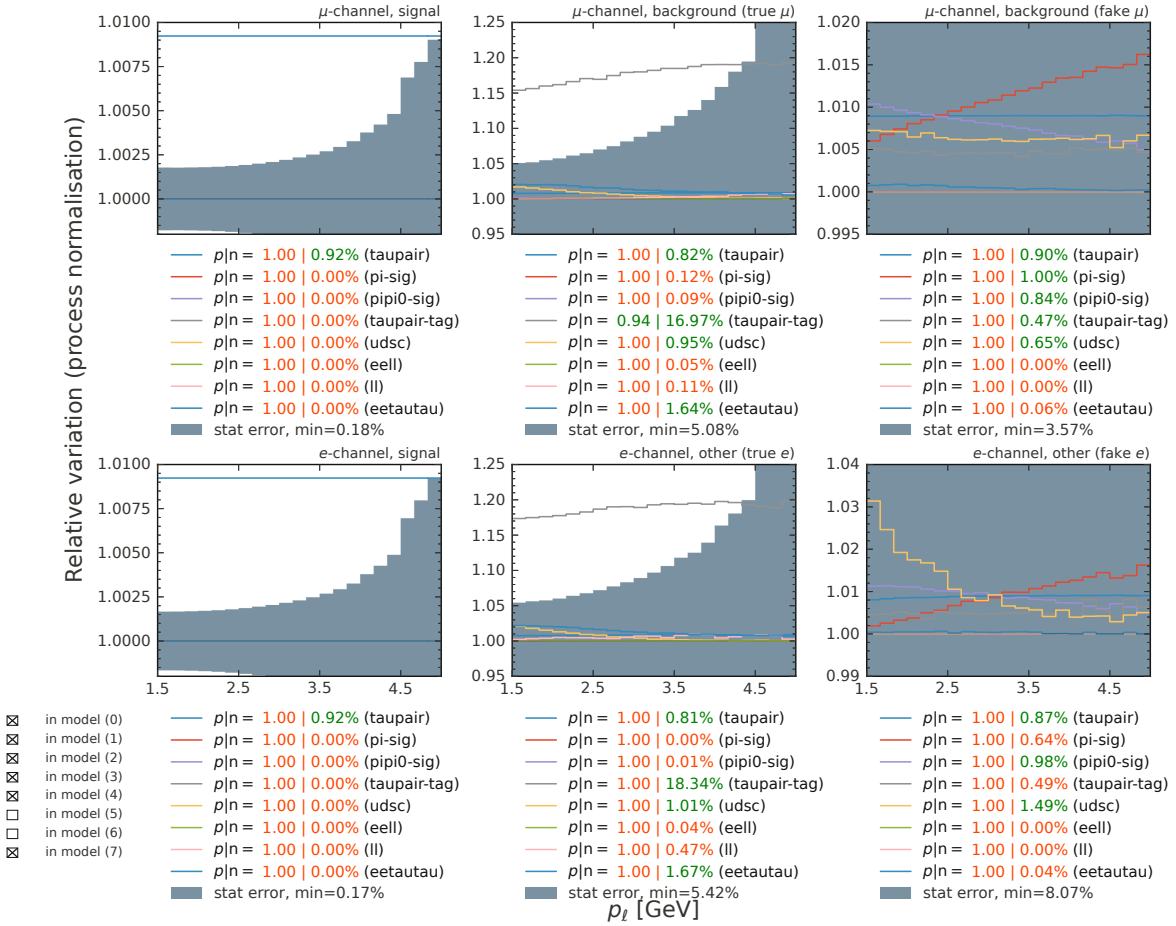


Figure 4.42: The relative variation of the templates due to overall normalisation uncertainties. Each colour shows an independent variation that is fully correlated across all templates.

For each of these groups, independently, a 25 % normalisation uncertainty is assigned, which is based on the initial data/MC discrepancy observed in the very high thrust region before any selections (Figure 4.9a). In particular, the last bin is dominated by $ee(\gamma)$ background and shows a 20 % excess of data/MC. This normalisation uncertainty is conservative and has little impact on the uncertainty on R_μ due to the small fraction of the background that survives the selections.

These normalisation uncertainties are implemented with correlated shape modifiers (histosys). The effect of the normalisation of each of the components mentioned above on the templates is shown in Figure 4.42. Here, 'taupair' refers to the normalisation of all taupair events; 'pi-sig' and 'pipi0-sig' refer to taupair events with the wrong signal side but the correct tag side; 'taupair-tag' refers to taupair events with the wrong tag side; and the other names refer to the non-taupair events.

4.6.10 Luminosity

The luminosity uncertainty is implemented using the lumi modifier from Table 4.19, a normalisation that affects all templates equally. As detailed in Section 3.5.1, the official Belle II

measurement of the offline luminosity with $\Upsilon(4S)$ beam energy is

$$\mathcal{L}_{\text{int}} = 361.654 \pm 0.021\,(\text{stat}) \pm 2.170\,(\text{sys})\,\text{fb}^{-1}.$$

This corresponds to a relative uncertainty of 0.6 %. Like any other global scaling of the templates, the uncertainty on the luminosity has a negligible impact on R_μ. However, contrary to the full cancellation one might expect from writing down the ratio when counting the background subtracted yields, there can still be a small effect on R_μ when performing a fit. This is because a global scaling of the data will not necessarily be absorbed by the nuisance parameter associated with the luminosity. Correlated changes in the normalisation of the templates can also be achieved with a combination of various other systematic effects, which may subsequently have an impact on R_μ. Examples are the taupair normalisation or π^0 efficiency, which are correlated, but not completely equal for all the templates. Which parameters collectively result in similar effects on the templates can be seen from the correlations of the fitted parameters, which are shown in Figure 5.31 in Section 6.2.2 for some selected parameters, including the luminosity nuisance parameter. Adding additional freedom to the fit due to luminosity uncertainty thus slightly increases the overall uncertainty on R_μ, which is, however, a very small effect, and the impact on R_μ is of the order of 0.01 %.

4.6.11 Tracking efficiency

Following the latest recommendation of the tracking performance group in Belle II, a systematic uncertainty for each track of 0.24 % is assigned to MC to account for differences in track reconstruction between simulation and data. This uncertainty is fully correlated across all samples and is included in the model through a normsys modifier. Since the event has two charged tracks, the normalisation uncertainty is 0.48 %. This is based on the assumption that the data/MC ratio of the tracking efficiency does not depend on the type of the particle. The measurement of R_μ would be mainly sensitive to data/MC differences between e and μ. At the current level of precision, no significant difference between electrons and muons is expected, and the used assumption is valid. For future measurements, when other uncertainties are reduced, this might have some impact and should be studied in more detail.

4.6.12 Momentum correction

To correct for the scale of the momentum of charged particle tracks, global correction factors provided centrally by the Belle II tracking performance group are used. The momentum obtained by the tracking algorithm is biased since it is derived from an imperfect magnetic field map in the reconstruction process. This can be checked by studying the peak position of particles with a well-known invariant mass. The present correction to the track momentum is obtained from the study of the D^0 mass peak. The scale factor is computed independently for each run and applied to data. To estimate the systematic uncertainty, the data is reprocessed using the up and down variation of the scale factor, corresponding to $\pm 1\sigma$, where σ is the total uncertainty on the correction factor. Then, the relative variation of the corresponding templates (obtained with the varied scale factor) with respect to the data (with the nominal scale factor) is added as a correlated shape variation on the MC templates. These shape variations are shown in Figure 4.43. The effect is small (~ 0.01 %) and does not meet the criteria to be included in the fit.

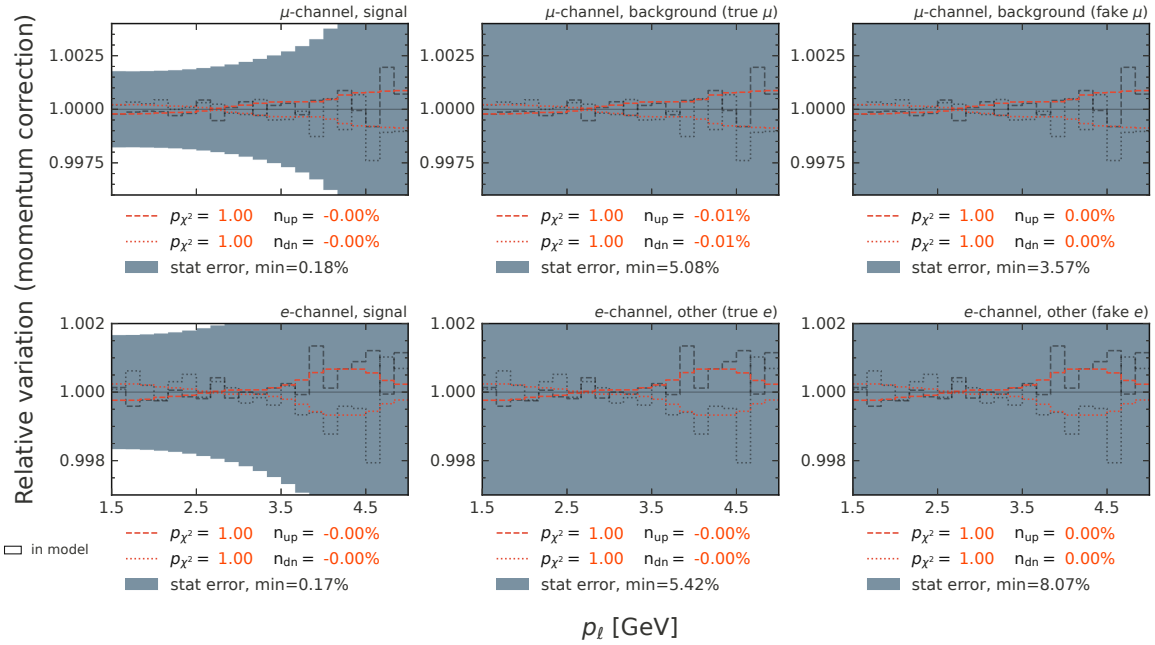


Figure 4.43: The shape variations originating from the varied momentum scale factor in data. The size of the variations is small and does not meet the criteria for having them included in the fit.

4.6.13 Tag side modelling (signal embedding)

A signal embedding technique is used to evaluate the impact of the MC modelling on the tag side. The combined $e+\mu$ sample is used for this, and truth-matched signal events (either $\tau^- \to e^- \bar{\nu}_e \nu_\tau$ or $\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$) are embedded in data and MC. For all these samples, the preselection is applied, except for a looser cut in the signal momentum, namely $0.5\,\text{GeV} < p_\ell < 5.5\,\text{GeV}$. The embedding is done by randomly choosing an event from the true signal events and then replacing the track on the signal side in the sample that is to be embedded. To account for a possible impact of FSR photons, which will be more prominent in events with an electron, the 4-momentum of the closest photon is added to the signal track when its momentum direction is within a cone of 0.1 rad around the momentum of the signal track. The tracks are replaced in quintiles of the signal momentum magnitude (in the CM frame) to choose signal tracks for which the magnitude of the momentum is not too dissimilar from the original track, to preserve the correct kinematic relations. These quinitles are chosen independently for the signal samples and the combined $e + \mu$ sample, and the corresponding bin edges are listed in Table 4.24. Tracks with momentum below 0.8 GeV are also replaced; however, their contribution to the final embedded sample is minimal.

After choosing a signal track for the replacement, it is then rotated so that the momentum direction matches the original momentum direction in the CM frame. As a next step, the signal momentum is boosted from the CM frame to the laboratory frame, and all variables that use information from the signal momentum (in particular event kinematics) are updated to reflect the properties of the replaced signal. Then, all analysis selections are applied again, and four embedded samples are obtained, namely:

Table 4.24: The bin edges for the quintiles in which the embedding is performed, defined over the momentum of the signal track in the CM frame.

	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}
$e + \mu$ sample	0.8	1.289	1.839	2.443	3.182	6.0
$\tau^- \to e^- \bar{\nu}_e \nu_\tau$ sample	0.8	1.271	1.8	2.388	3.113	6.0
$\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$ sample	0.8	1.326	1.886	2.488	3.215	6.0

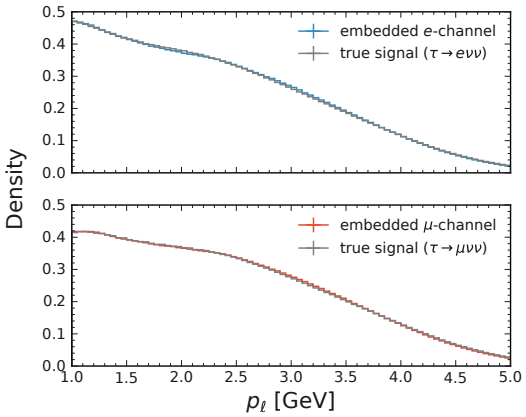


Figure 4.44: A comparison of the embedded samples in MC to the original distributions of $\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$ and $\tau^- \to e^- \bar{\nu}_e \nu_\tau$ which were used for the embedding. The figure shows the momentum of the signal track, with a looser selection than what is used in the analysis. It extends up to 1 GeV at its lower end and does not include the veto of the lepton ID correction bins in which no mis-id rate corrections are available.

- $\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$ embedded in MC,
- $\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$ embedded in data,
- $\tau^- \to e^- \bar{\nu}_e \nu_\tau$ embedded in MC,
- $\tau^- \to e^- \bar{\nu}_e \nu_\tau$ embedded in data.

Before comparing data and MC with these embedded samples, it is important to check how well the embedding worked by comparing the distributions of the embedded MC samples to the distributions of the original signal events used for the embedding. This comparison is shown in Figure 4.44 and Figure 4.45 with the embedded e-channel in blue and the embedded μ-channel in red. In Figure 4.44 all analysis selection except for the cut on the ANN output and a looser cut on the signal momentum was applied. In case of the embedded samples, the selection was imposed on the updated variables. The figure shows that even from 1.0 GeV to 1.5 GeV, where there are the largest differences between the e and the μ samples, the momentum distribution of the signal track is well reproduced. The comparisons shown in Figure 4.45 are with the same momentum cut as for the analysis but still without the cut on the ANN output. Overall, the kinematics of the signal events, as well as the ANN output, are well reproduced in the embedded samples. The dips in the polar angle distribution of the signal (4.45b) stem from artefacts introduced by the lepton ID selections. The dip at 90° can also be seen in the e-efficiencies shown in Figure 4.15 and originates from the gap in the ECL detector. Also,

the drop of the μ-efficiencies towards the endcaps can be seen. Both effects appear in the embedded samples because it consists of both the electron and the muon channel, with a relative size of approximately 50 %. The true signal also has the lepton ID cut applied, but since it is truth-matched MC, only one lepton type is present.

Using the embedded samples, the data/MC agreement within the electron and muon samples can now be compared without being influenced by any effects of lepton universality. Because the same signal sample is used to embed in data and MC, any observed discrepancies come from the tag side and will largely cancel when measuring R_μ. Some distributions are shown in Figure 4.46, where the embedded e sample is on the left and the embedded μ sample is on the right. It should be noted that for all these plots, the hatched uncertainty band in the ratio only includes lepton ID uncertainties.

While the data/MC agreement is primarily flat and within the given uncertainties, there are regions with significant discrepancies. As long as these discrepancies appear equally in the e and μ samples, they are not concerning for the present measurement. Most interesting is the difference between the discrepancies in the two samples. The momentum distribution of the signal lepton is shown in Figure 4.47 The data/MC ratio follows a very similar shape in both samples and deviates not more than $\sim 2.5\,\%$ from 1. The shape of the data/MC ratio resembles what was previously observed when estimating the shape uncertainty to account for the mismodelling of the momentum distribution, shown in Figure 4.40.

The mismatch of the differences is obtained by computing the ratio of the data/MC ratios. This double ratio is shown in Figure 4.48, where the blue markers are the result of not adding the momentum of the FSR candidate to the signal track and the red markers are obtained when these photons are taken into account in the replacement. The difference between both double ratios is small and does not impact the size of the estimated uncertainty. The double ratio is again shown in Figure 4.49, using the coarser lepton ID binning in steps of 0.5 GeV. The reason for using a coarser binning here is that the ratio is affected by statistical uncertainties, shown as error bars in the plot. The coarse bins reduce these uncertainties and help to detect systematic deviations from 1. The observed differences are consistent with unity within the statistical uncertainties. Since it is not possible to reliably separate any statistical fluctuations from real systematic effects, a conservative approach is chosen to extract uncertainties: First, the ratios are fit with polynomials of degree 0 to 2. Then, uncertainty bands for each polynomial are computed, and the largest difference from 1 of any uncertainty band is taken as the size of the uncertainty. The fitted polynomials with uncertainty bands are shown in different colours in Figure 4.49, and the resulting symmetric envelope is shown in grey. This procedure is conservative, accepting the possibility of inflated uncertainties so as not to risk underestimating them. The reason why this is done only up to second order is that even though higher order polynomials might as well give a good description of the ratio in terms of a χ^2 test, the LOOCV [159] score increases with higher-order polynomials. This indicates an overfitting of statistical fluctuations, especially since the χ^2 value no longer decreases for the third-order polynomial. Figure 4.50 shows the fits, including a third-order polynomial.

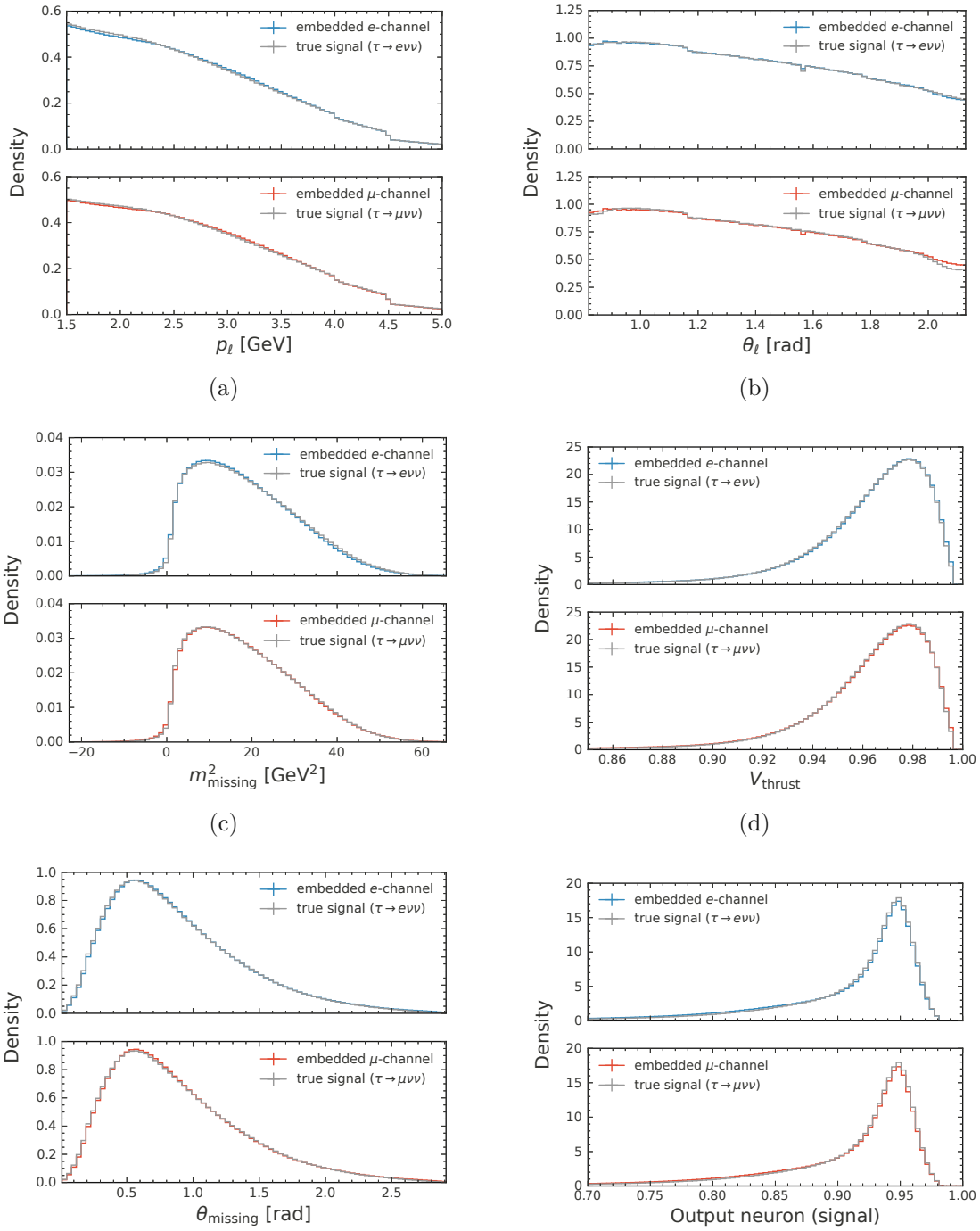


Figure 4.45: A comparison of the embedded samples in MC to the original distributions of $\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$ and $\tau^- \to e^- \bar{\nu}_e \nu_\tau$ which were used for the embedding. The figure shows the distributions of signal track variables (a, b), event variables (c, d, e) and the ANN output (f).

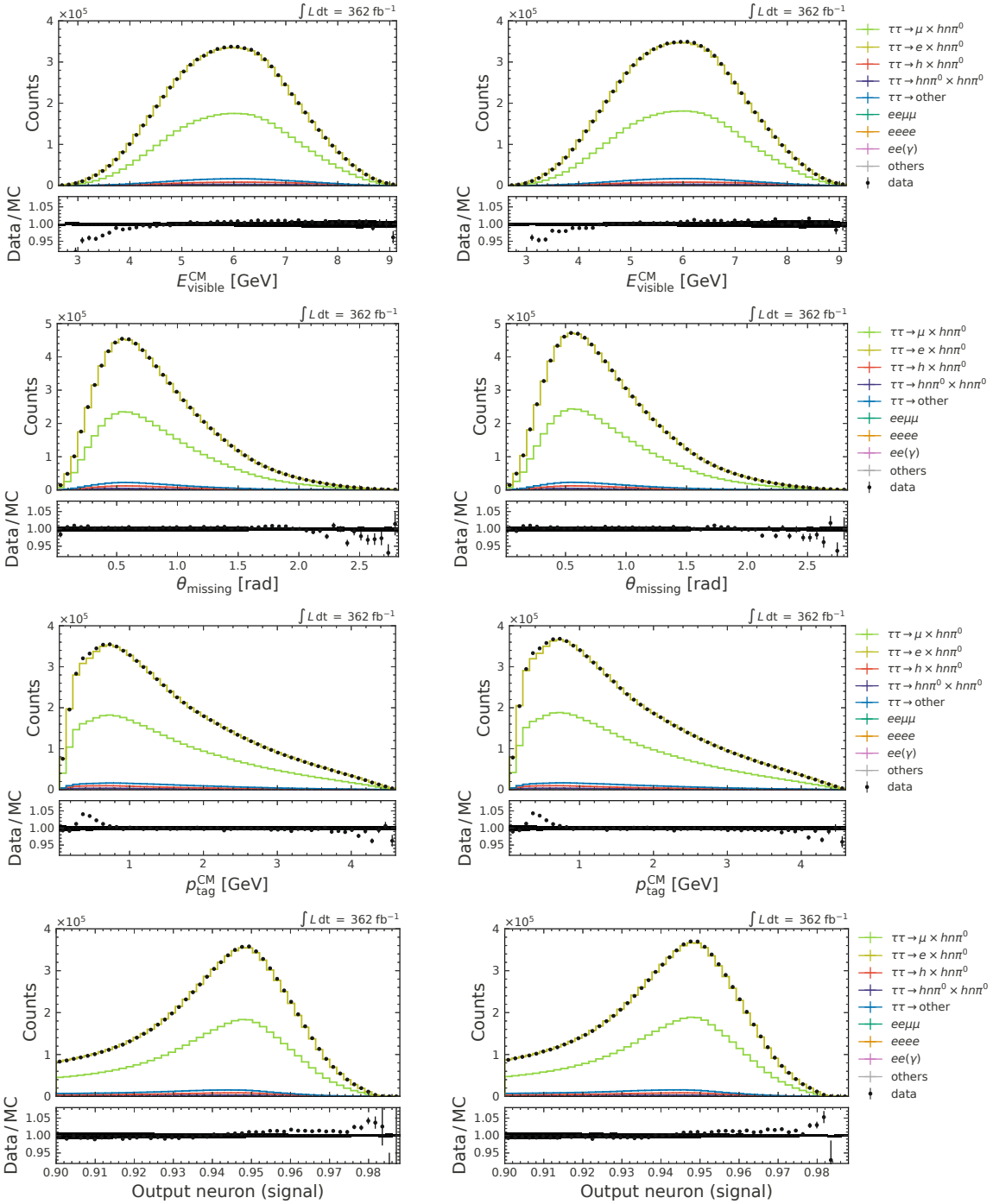


Figure 4.46: Data/MC comparison of the embedded samples for different variables. The e sample is shown on the left and the μ sample on the right. An ANN cut of 0.9 is applied, and the hatched uncertainty bands only include lepton ID uncertainties.

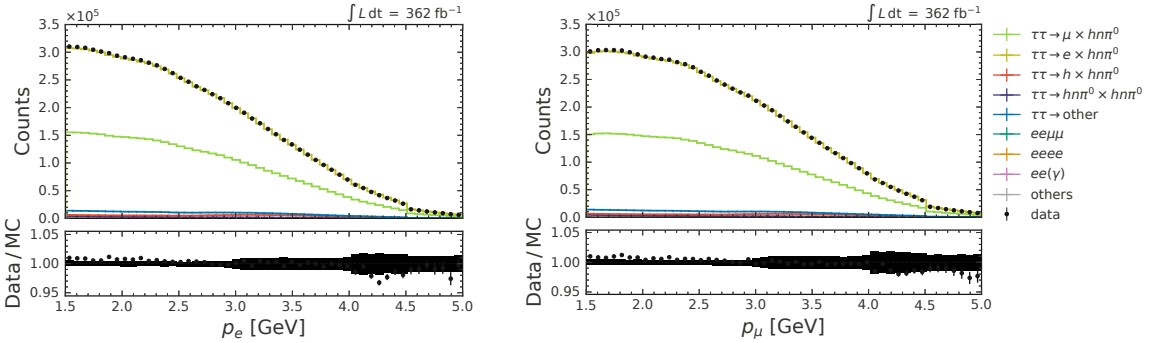


Figure 4.47: Data/MC comparison of the signal lepton momentum in the embedded e sample (left) and the embedded μ sample (right) with an ANN cut of 0.9 applied. The hatched uncertainty bands only include lepton ID uncertainties.

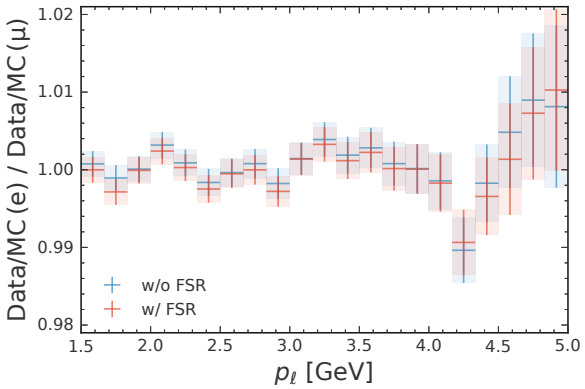


Figure 4.48: The ratio between the data/MC ratios from the embedded e and μ samples, in bins of the lepton candidate momentum. For the blue makers, the FSR candidates on the signal side are not taken into account, while for the red markers, the closest photons are added to the respective signal momenta when their direction is within a cone of 0.1 rad around the signal momentum direction.

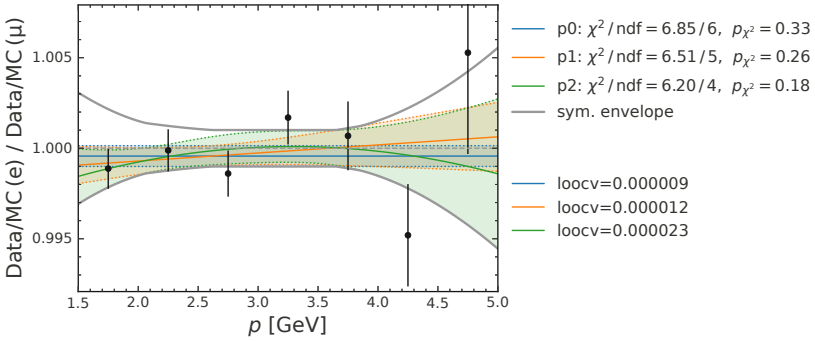


Figure 4.49: Polynomials up to degree 2, fitted to the ratio of data/MC ratios from the embedded samples. The uncertainty bands of the polynomials and the resultant symmetric envelope are overlayed. The legend gives the p-values of the χ^2 test and the Leave-One-Out Cross-Validation (LOOCV) score for the individual polynomials.

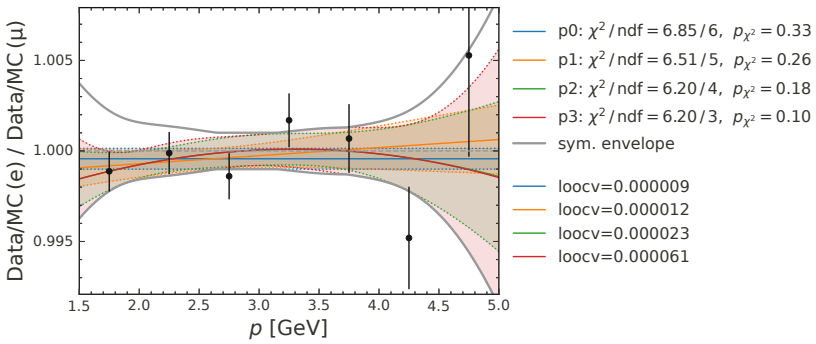


Figure 4.50: Polynomials up to degree 3, fitted to the ratio of data/MC ratios from the embedded samples. The uncertainty bands of the polynomials and the resultant symmetric envelope are overlayed. The legend gives the p-values of the χ^2 test and the LOOCV score for the individual polynomials.

Using the symmetric envelope in Figure 4.49, a per-bin uncertainty is defined for the templates, where the size of the uncertainty corresponds to the value of the envelope in the centre of the bin. This uncertainty is implemented with an uncorrelated shape (shapesys) modifier. Figure 4.51 shows the relative size for each bin. The uncertainty is only assigned to the e and μ signal templates and is vetoed for the background templates due to the small size. This is consistent with the intended application since the embedded samples only reflect the signal decays and are used to estimate the bias due to the tag side mismodelling, which arises from the selection process.

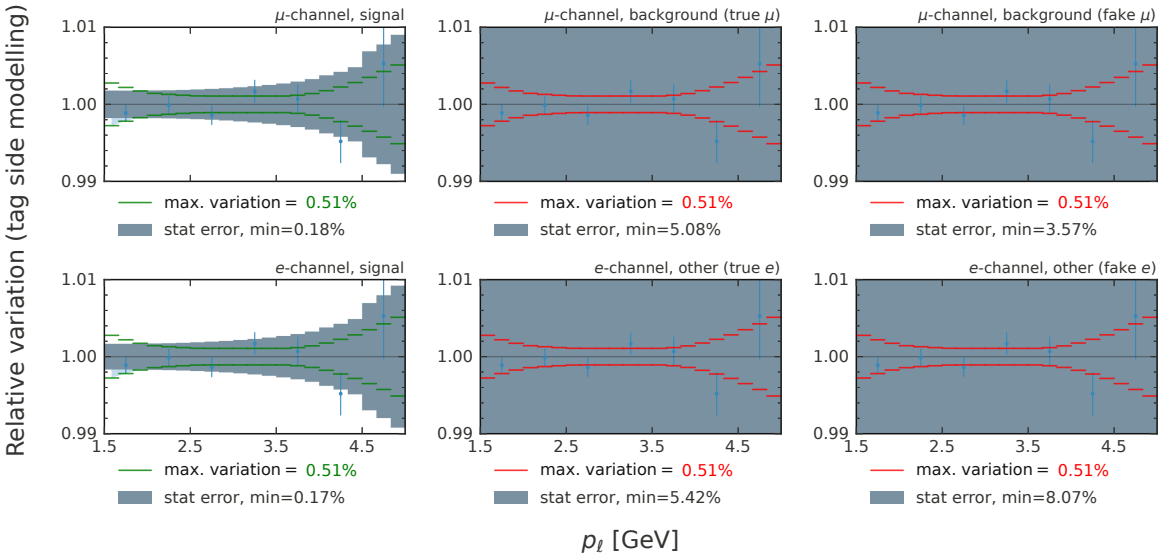


Figure 4.51: The relative uncertainty on each template bin, obtained from the symmetric envelopes of the polynomials fitted to the blue markers. The blue markers show the ratio of the data/MC ratios of the embedded e and μ samples.

Neural network selection

In addition to using the embedded samples to derive an uncorrelated uncertainty for the different template bins, they can also be used to derive a fully correlated variation across all templates, reflecting the observed data/MC discrepancies. Since most of these discrepancies will already be covered by other systematics or are just statistical fluctuations, the data/MC ratio of, for example, Figure 4.47 is not taken as a shape variation. However, most of these effects can be cancelled by computing the ANN efficiency,

$$\varepsilon_{\mathrm{NN}} = \frac{N_{\mathrm{pass}}}{N_{\mathrm{pass}} + N_{\mathrm{fail}}}, \tag{4.31}$$

where N_{pass} are the number of events passing the ANN selection and N_{fail} are the number of events removed by the cut on the ANN output. This efficiency can be computed in every momentum bin and individually for the e sample and the μ sample. Then, the ratio of the ANN efficiency in data over the ANN efficiency in MC can be used to reflect the change in the data/MC ratio before and after applying the ANN cut. Since the data/MC difference in the low ANN output region comes from mismodelled backgrounds, adding this as a shape variation of the templates is not desirable, since almost all of these backgrounds are removed after the cut on the ANN output is applied. Instead, the efficiency ratio with respect to a cut of 0.8 is computed, where the signal becomes dominant. The data/MC ratios of the efficiency ratios are shown in Figure 4.52. This is taken to be a fully correlated shape variation across all templates and is implemented with the histosys modifier. This complements shape variations equal for the e and μ channels but independent between the momentum bins, already covered by the global shape systematic described in Section 4.6.8.

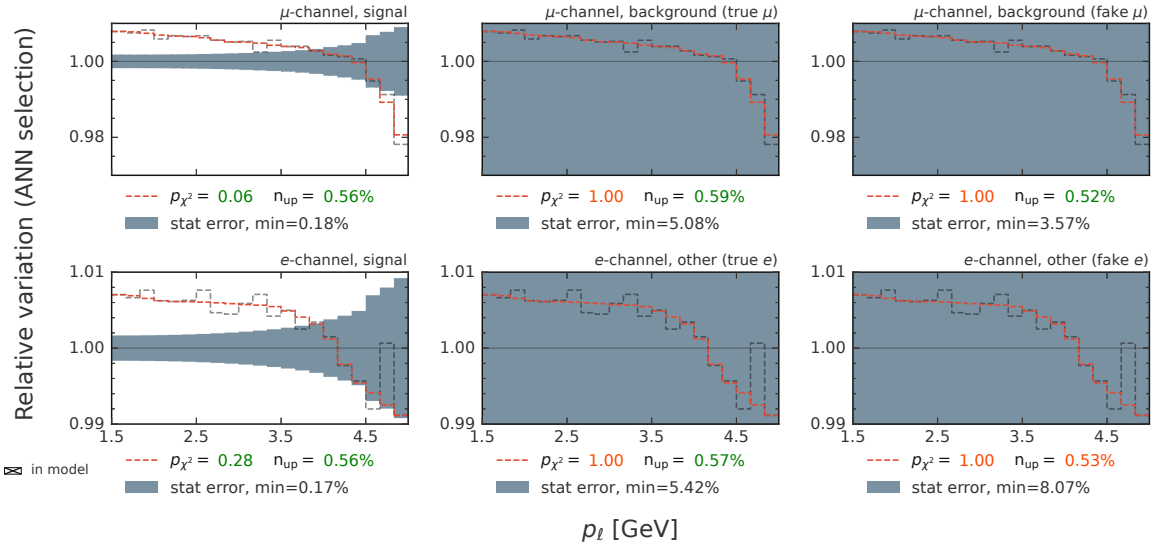


Figure 4.52: The correlated shape variations of the templates, resulting from the data/MC ratio of the ANN efficiency.

4.6.14 Detector misalignment

The alignment of the Belle II tracking system is described by about 60 000 parameters, covering the local alignment of sensors and wires up to the relative alignment of sub-detectors. The

alignment parameters are determined from cosmic ray tracks and various data samples in the calibration procedure of Belle II. More details on the alignment procedure are reported in [160, 161]. To obtain an estimate of the effect due to residual misalignment on the determination of R_μ, additional MC samples of $e^+e^- \to \tau^+\tau^-$ have been produced, using nominal alignment parameters as well as the following different alignment configurations:

- Charm misalignment (v0): Difference from data in the day-to-day alignment of the vertex detectors and the CDC.
- Charm misalignment (v1): Residual misalignment after vertex detector and CDC alignment.
- Misalignment prompt processing: Difference between prompt alignment and reprocessing alignment.
- Misalignment VXD CDC wires: Residual misalignment in the iteration of complete global alignment (from zeros), including CDC wires.

Each sample corresponds to an integrated luminosity of $100\,\mathrm{fb}^{-1}$ and for each alignment configuration, the beam spot position was recalibrated. In all samples, the same events from generation and simulation have been used as input, and the different misalignment configurations have only been applied in the reconstruction step. This is to suppress statistical fluctuations that would originate from independently generated samples. The misalignment samples were generated by a Belle II colleague.

Figure 4.53 shows the distribution of signal momentum in bins of 1 GeV width with the different alignment configurations. The ratio plot shows the ratio of the special alignment configurations to the nominal alignment configuration. Following the recommendations regarding alignment systematics, the variation from 'charm misalignment (v0)' is used to estimate the statistical uncertainty associated with misalignment and the most extensive bin-wise variation of 'misalignment prompt processing', 'misalignment VXD CDC wires' and 'charm misalignment (v1)' to estimate the systematic uncertainty. Then, the statistical and systematic components are summed in quadrature, as shown on the right side of Figure 4.53. To be conservative, the most extensive variation of the four bins is used and applied as an uncorrelated shape variation to each bin of the templates. The resulting uncertainty is 0.038 %, which is still very small. To not unnecessarily increase the number of nuisance parameters, this uncertainty is added in quadrature to the per-bin uncertainty attributed to the tag side modelling, obtained from the embedding studies shown in Section 4.6.13.

4.6.15 Dependence on the generator model

The modelling of the decay $\tau^- \to \pi^-\pi^0\nu_\tau$ in the TAUOLA generator [124] has been implemented from the measurement of the hadronic spectrum carried out by the Belle collaboration [162]. In total, 11 parameters describe the unfolded spectrum $\mathrm{d}(\pi\pi^0)/\mathrm{d}s$, with intermediate resonances $\rho(770)$, $\rho'(1450)$ and $\rho''(1750)$ parametrised using the Gounaris-Sakurai model [163]. The generator provides two parameter configurations, from which two independent taupair data sets have been produced to estimate the impact on R_μ that comes from the parametrisations. The first parametrisation uses values determined from a fit to the $M_{\pi\pi^0}$ distribution when the normalisation factor $|F_\pi(0)|^2$ is set equal to unity as expected for a Breit-Wigner shape of the

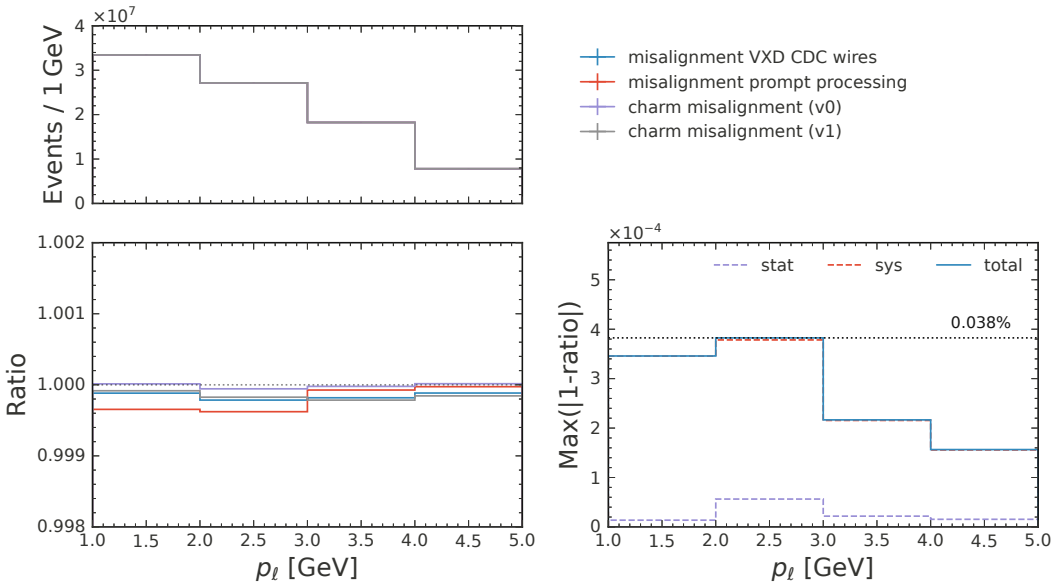


Figure 4.53: Estimation of the misalignment systematic with different alignment configurations. The upper left plot shows the distribution of signal momentum in bins of 1 GeV width with the different configurations. The ratio plot on the lower left shows the ratio of the special alignment configurations to the nominal alignment configuration. The lower right plot shows the resulting statistical and systematic uncertainty that is attributed to the misalignment, computed based on the deviation of the ratio from unity.

resonances and the obtained sample will be referred to as 'MC1'. For the second parametrisation, all parameters were left floating to account for a possible deviation from the form of the fitting function, and the corresponding sample will be referred to as 'MC2'. Figure 4.54 shows the squared invariant mass on the tag side for the two-parameter configurations, separated according to the charge of the decay $\tau^- \to \pi^-\pi^0\nu_\tau$. Here, only events passing all the selections are shown. The different parametrisations are expected to cause only minor differences in the tails of the squared invariant mass distribution of $\pi^-\pi^0$. In the ratio in Figure 4.54, no significant differences are observed, except for $M^2_{\text{tag}} < 0.5\,\text{GeV}^2$, where the 'MC1' sample has slightly more events than the 'MC2' sample. These differences affect only a few events.

In Figure 4.55, the different models are compared in bins of signal momentum. In the upper left figure, the total number of entries is shown. In the lower left, the ratio of the two samples is shown, with the positive and negative samples combined. In the lower right, the difference between this ratio and unity is shown, which is around 0.1 %. The deviations from one of the ratios are consistent with statistical fluctuations, and no significant shift is observed. This also aligns with the expectation that the $\tau^- \to \pi^-\pi^0\nu_\tau$ decay is well modelled, especially the peak of the $\pi^-\pi^0$ invariant mass. Thus, based on this study, it was decided not to add an additional systematic uncertainty. The tag side modelling is expected to affect the e and μ channels equally and will have little impact on R_μ. If some mismodelling still affects the two channels differently, it would already be taken into account by the signal embedding procedure described in Section 4.6.13.

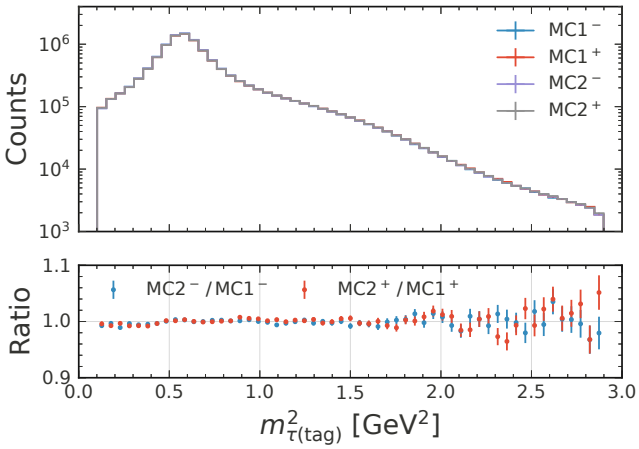


Figure 4.54: The squared invariant mass on the tag side, comparing the different parameter configurations of the decay $\tau^- \to \pi^-\pi^0\nu_\tau$. The samples are additionally split according to the charge of the τ-lepton.

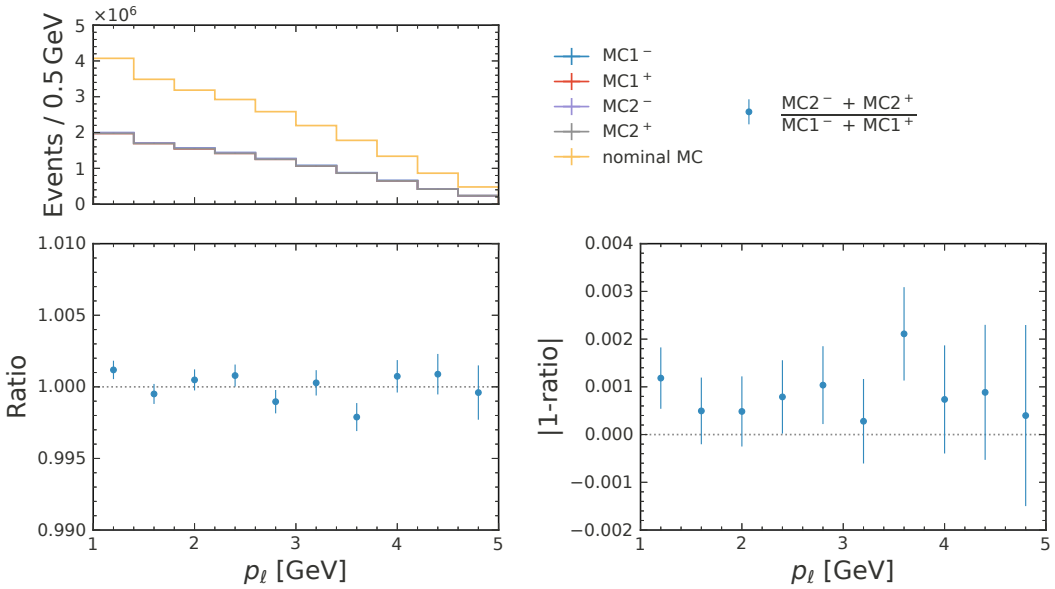


Figure 4.55: A comparison of the samples generated with different settings in TAUOLA, binned in signal momentum. The upper left figure shows the yields after the analysis selections for the different configurations and the nominal MC sample. The ratio of the two configurations is shown on the lower left, with positive and negative charges combined On the lower right, the deviation of the ratio from unity is shown. The difference between the samples is well within their statistical uncertainties.

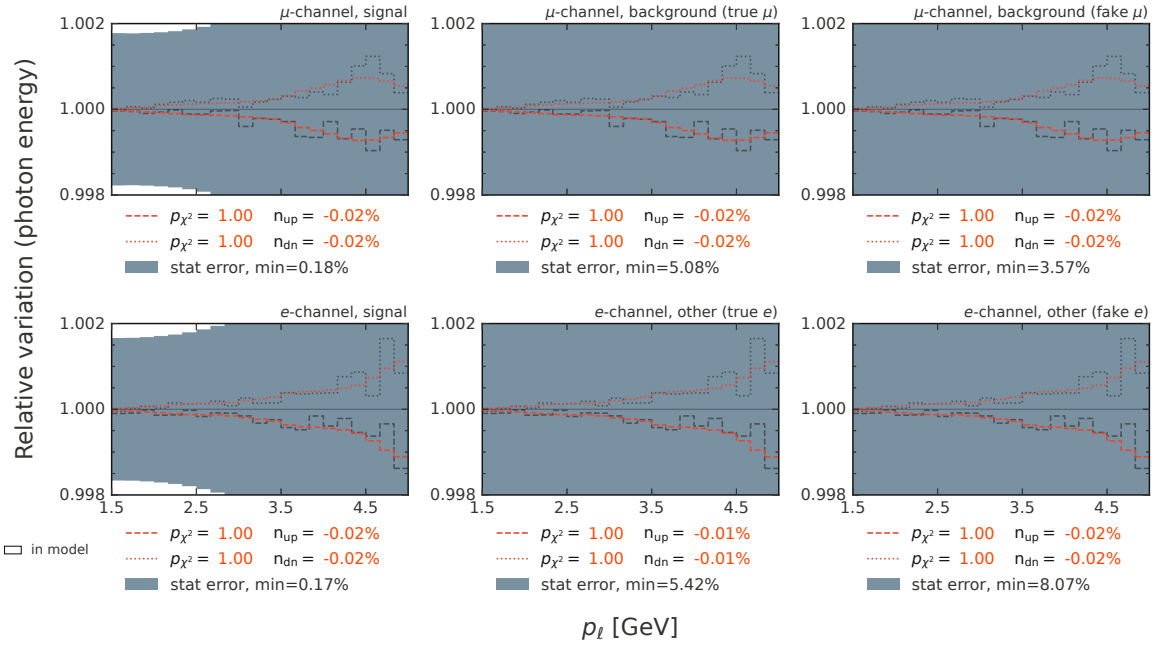


Figure 4.56: The relative template variations due to systematics associated with the photon energy correction. The variations are at the order of 0.01 % and can be neglected.

4.6.16 Photon energy

The reconstructed photon energy can differ from the actual energy of the photon, so it needs to be calibrated. This is also called the photon energy bias. A correction factor is obtained to account for differences between data and simulation, which is then applied to data. The correction factors are provided by the Belle II neutrals performance group, by studying the invariant mass of two photons in the decays $\pi^0 \to \gamma\gamma$ and $\eta \to \gamma\gamma$. They are binned in photon energy and azimuthal and polar angle of its momentum direction. Similarly to the momentum scale correction (see 4.6.12), a systematic up and down variation of the energy bias is used to obtain the impact on the selected events. Two additional data samples with varied energy bias corrections are produced and the relative difference of the resulting templates is computed with respect to the nominal one. This relative difference can be included as a shape variation on the MC templates. It is shown in Figure 4.56. The variation of the templates would only be at the order of 0.01 % and does not satisfy the requirement to be included in the model.

4.6.17 Photon efficiency

To correct for photon efficiency in MC, correction weights provided centrally by the Belle II neutrals performance group are used. These weights are given in bins of momentum, polar angle, and photon energy, and they have an associated statistical and systematic uncertainty.

There are multiple options for how to use these weights to correct the photon efficiency in MC:

1. Randomly removing photon candidates according to the weights: This means throwing away photon candidates at the reconstruction level for which the associated weight is smaller than a random number sampled between 0 and 1. This will correctly account

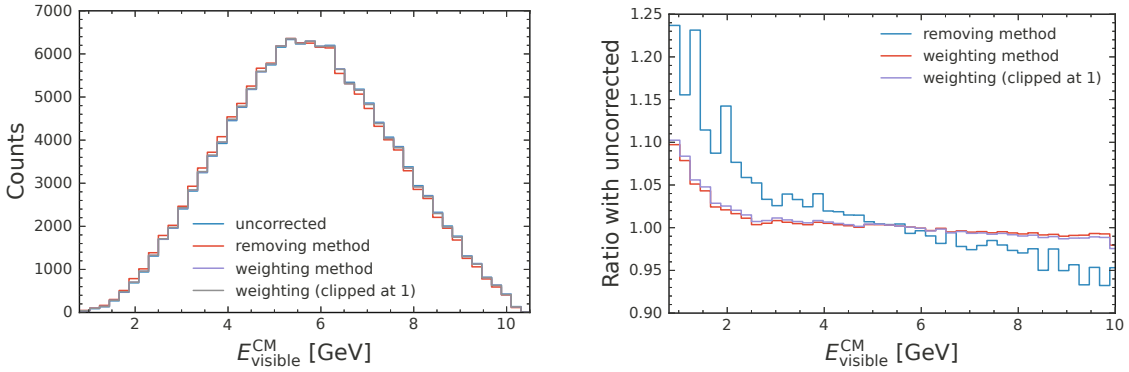
for bin migration in variables that use the photon energies (e.g. the visible energy) and produce the correct photon multiplicity. One disadvantage of this method is that it cannot deal with correction weights larger than 1 (photons cannot be added to the photon list). Another downside is that when estimating systematic uncertainties due to these corrections, the whole MC sample needs to be reproduced multiple times, removing photons with weights that vary up and down.

2. Weighting the events: A much more convenient method would be to weight the events with a weight factor obtained as the product of the photon efficiency weights of all photons in the event. This can be done after the reconstruction step and would only require storing the weights of all photons or directly their product. However, this method will reduce the total number of events in MC (if most photon efficiency weights are less than 1). This does not make sense, as the whole event is not necessarily discarded because some photons are undetected. The number of events can only change when there is a bin migration due to the photon efficiency correction that would result in the event being cut away by some analysis selection.

3. Weighting the events and dividing by their mean: One solution to the issue of losing events would be to force the number of events to remain the same by reweighing the weights so that their mean is 1. This method seems a little arbitrary and will give different results depending on the cuts applied before this reweighing. If it is done after some analysis selection, the bin migration of events 'out of the selection' is ignored since the total number of events stays constant. It should be applied before any selections, and even then it is unclear if it correctly reproduces the bin migrations in event-based variables.

Methods 1 (removing method) and 3 (weighting method) were compared with a small MC sample to check if they would give similar results. The result can be seen in Figure 4.57, where the distributions of visible energy in the CM for the uncorrected MC and the different methods are shown. In addition, results for the weighting method where larger efficiency weights are clipped to unity are shown to see the impact of neglecting them. In Figure 4.57b, the ratio of the distributions obtained with the different methods with respect to the uncorrected distribution is shown. One can see that applying the photon efficiency corrections (with the majority of weights smaller than one) results in a shift of the visible energy towards smaller values, which is expected. The shift is much less pronounced for the weighting method than for the removing method. This led to the conclusion that the removing method should be used to get correct distributions of event kinematic variables if photon efficiency corrections larger than one can be ignored. This seems feasible in the present case, since the difference between the weighting method using all weights and the weighting method where weights are clipped to unity is negligible.

Having settled on a method to correct photon efficiency, the associated systematic is implemented in the model with a correlated shape variation. The variations are obtained from the ratio of the varied MC samples (with photons removed using the up and down variations of the efficiency weights) to the nominal MC sample. This ratio is shown in Figure 4.58 for each of the six templates. A sizeable shift in the momentum distribution is observed due to the uncertainty of the photon efficiency corrections, which cannot be neglected.



(a) The distribution of visible energy for the different methods.

(b) The ratio of the corrected distributions with respect to the uncorrected distribution.

Figure 4.57: A comparison of different methods to correct photon efficiency in MC, overlayed distributions (a) and ratio (b). Compared to the removal method, the shift towards lower visible energy is much less pronounced for the weighing method.

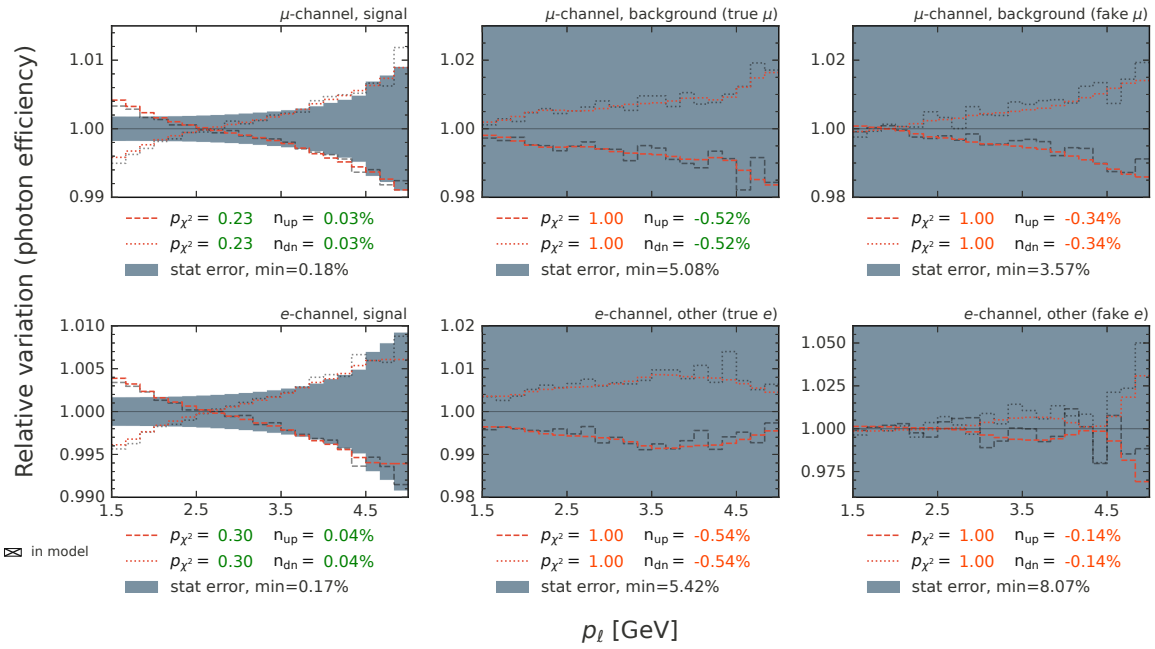


Figure 4.58: The template variations associated with the uncertainty of the photon efficiency correction.

4.7 Impact of systematic effects

The contribution of each systematic effect to the overall uncertainty on R_μ is obtained using a toy approach, where 2000 instances of simulated data (toys) are sampled from the fit model for each systematic. In every toy sample, only auxiliary data belonging to the nuisance parameters which model the systematic are altered based on the related constraint terms. The standard deviation from the resulting distribution of the estimated R_μ values is taken as the impact of the individual systematic effects. Table 4.25 lists the systematic sources described in the previous section together with their contribution to the relative uncertainty on R_μ. The table also shows how many parameters are needed to model each systematic, as well as which modifiers were used. The stars next to the parameter numbers indicate that these systematics have shared parameters. The small detector misalignment uncertainty was added in quadrature to tag side modelling for each of the 42 bins individually, which is possible since both effects are independent and fully uncorrelated across the momentum bins. The number of nuisance parameters generally depends on the correlation of the respective systematic uncertainties. In total, 194 nuisance parameters are included in the fit model. The dominant source of uncertainty for R_μ stems from lepton identification, which contributes 0.32 %, with the next largest uncertainty amounting to 0.10 %, which is associated with the trigger. The total relative systematic uncertainty is 0.37 %, which corresponds to an absolute systematic of 0.0036 when multiplied by the SM prediction of R_μ of 0.9726. Together with the quoted statistical uncertainty of 0.07 %, this results in an overall relative uncertainty of 0.38 %, corresponding to a total absolute uncertainty of 0.0037. This aligns with the uncertainty derived from a single fit with Asimov data, an artificial dataset that matches the expected yields from simulation. The statistical component of the uncertainty, as well as the Asimov fit, is described in more detail in Section 5.3.

Ranking plot

Another way to estimate the impact of the systematic sources on R_μ is with a so-called ranking plot. This is a ranking of all the individual nuisance parameters according to their effect on the parameter of interest. The influence on R_μ is obtained by varying the nuisance parameter θ up and down by one standard deviation from the best-fit value $\hat{\theta}$. Here, one can distinguish between the pre-fit impact and the post-fit impact, depending on whether the variation is done with the uncertainty provided in the constraint term of the nuisance parameter, $\Delta\theta$, or whether the post-fit uncertainty $\Delta\hat{\theta}$ of the nuisance parameter is used. The fit is then repeated with the nuisance parameter in question fixed to the varied value, and the change of the fitted value of R_μ relative to the nominal fit result, ΔR_μ, is quoted as the impact. This is shown in Figure 4.59 for each nuisance parameter and independently for the up and down variations. In addition, this plot also shows the pull values, which will be discussed in more detail in Section 5.3.5. The names of the nuisance parameters shown in Figure 4.59 and the systematic with which they are associated are described in more detail in Table 4.26.

Table 4.25: The relative expected uncertainty on R_μ, categorised according to various systematic sources. The uncertainties were obtained with toy samples, where for each systematic type 2000 samples were used. Also shown are the number of parameters needed to model each systematic and what modifiers were used. The stars next to the parameter numbers associated with tag side modelling and detector misalignment denote that the parameters are shared. The total relative uncertainty, including both statistical and systematic errors, is 0.38 %, which corresponds to an absolute uncertainty of 0.0037.

	Uncertainty [%]	Parameters	Modifier
Charged-particle identification	0.32	28	histosys
- Electron identification	- 0.22	7	histosys
- Muon misidentification	- 0.19	7	histosys
- Electron misidentification	- 0.12	7	histosys
- Muon identification	- 0.05	7	histosys
Trigger	0.10	42	histosys
		2	normsys
Modelling of FSR	0.08	1	histosys
Normalisation of individual samples	0.07	8	normsys
Size of the simulated samples	0.06	42	staterror
Modelling of the momentum distribution	0.06	21	histosys
Tag side modelling (signal embedding)	0.05	42*	shapesys
π^0 efficiency	0.02	1	histosys
Particle decay-in-flight	0.02	1	histosys
Modelling of ISR	0.01	1	histosys
Luminosity	0.01	1	lumi
Tracking efficiency	0.01	1	normsys
Detector misalignment	< 0.01	42*	shapesys
Photon efficiency	< 0.01	1	histosys
ANN efficiency	< 0.01	1	histosys
Photon energy	< 0.01	0	histosys
Momentum correction	< 0.01	0	histosys
Total systematic uncertainty	0.37		
Statistical uncertainty	0.07		
Total	0.38	193	

Table 4.26: The names and descriptions of the nuisance parameters defined in the `pyhf` model. If a range is given, multiple nuisance parameters are assigned, and the number is added to the end of the name. The reference points to the section where the associated systematic effect was introduced.

Name	Range	Description	Section
`R_mu`	N/A	the parameter of interest	4.5.4
`BF_e`	N/A	norm for $\tau \to e\nu_e\nu_\tau$ (κ_e)	4.5.4
`gen_BF_scaling`	N/A	scaling for R_μ ($\kappa^{\text{gen}}_{e/\mu}$)	4.5.4
`lumi`	N/A	luminosity	4.6.10
`pi0`	N/A	π^0 efficiency	4.6.3
`tracking_efficiency`	N/A	tracking efficiency	4.6.11
`FSR`	N/A	FSR modelling	4.6.5
`ISR`	N/A	ISR modelling	4.6.6
`decay`	N/A	decay-in-flight	4.6.7
`gammaEfficiency`	N/A	photon efficiency	4.6.17
`nn_efficiency`	N/A	ANN selection	4.6.13
`normalisation_mu_sig_x_bkg`	N/A	wrong tag $\tau \to \mu\nu_\mu\nu_\tau$	
`normalisation_e_sig_x_bkg`	N/A	wrong tag $\tau \to e\nu_e\nu_\tau$	
`normalisation`	00-07	sample normalisation	4.6.9
`trigger_alt_ref`	N/A	trigger normsys	4.6.4
`trigger_diff`	N/A	trigger normsys	4.6.4
`trigger_correction_e_channel`	00-20	trigger shapesys	4.6.4
`trigger_correction_mu_channel`	00-20	trigger shapesys	4.6.4
`lid_p_part_10_e_channel_e`	00-06	e efficiency	4.6.2
`lid_p_part_10_e_channel_fake`	00-06	$h \to e$ mis-id rate	4.6.2
`lid_p_part_10_mu_channel_mu`	00-06	μ efficiency	4.6.2
`lid_p_part_10_mu_channel_fake`	00-06	$h \to \mu$ mis-id rate	4.6.2
`global_shape`	00-20	modelling of momentum	4.6.8
`selection_e_channel_e_sig_e`	00-20	tag side modelling (e)	4.6.13
`selection_mu_channel_mu_sig_mu`	00-20	tag side modelling (μ)	4.6.13
`mc_stat_error_e_channel`	[00-20]	MC sample size	4.19
`mc_stat_error_mu_channel`	[00-20]	MC sample size	4.19

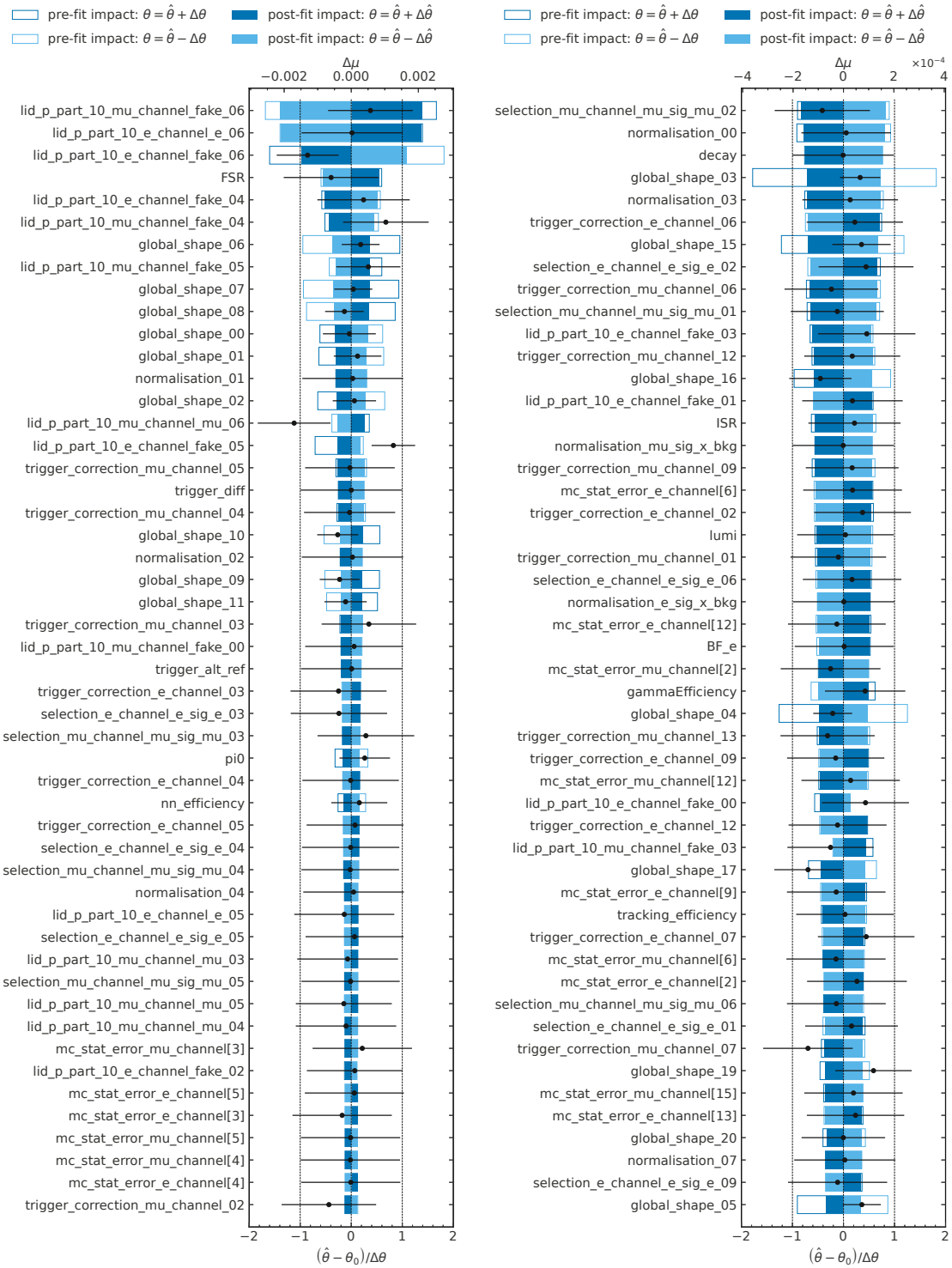


Figure 4.59: Ranking plot showing the impact of varying the different nuisance parameters within their associate uncertainties on R_μ. For the pre-fit impact $\Delta\theta$ denotes the uncertainty provided in the constraint term of the nuisance parameter, while for the post-fit impact, $\Delta\hat{\theta}$ is the uncertainty as obtained from the fit. In addition, the pull values and associated uncertainties of the nuisance parameters are overlayed with black markers.

CHAPTER 5

Validation studies

5.1 Introduction

This chapter will show several studies that test the validity of the estimated systematic effects and the fit result. First, in Section 5.2, a control sample using $e^+e^- \to (\tau^+ \to e^+\nu_e\bar{\nu}_\tau)(\tau^- \to \mu^-\nu_\tau\bar{\nu}_\mu)$ events will be studied. Then, in Section 5.3, the template fit and the properties of the estimator will be studied in more detail. In Section 5.4, the shape of the simulated distributions will be compared with what is observed in the data. Finally, in Section 5.5, the consistency of the obtained estimate of R_μ from the fit will be studied for different data subsets and other analysis conditions.

All these studies will be realised without exposure to the actual fit result or any information that could indicate it. This is achieved by adding a random (but reproducible) unknown shift Δ to R_μ in the likelihood function by doing the replacement

$$R_\mu \to R_\mu^{\text{shifted}} + \Delta. \tag{5.1}$$

This leaves the uncertainties and the fit behaviour invariant, but introduces a random offset to the fitted parameter of interest. For Δ, a random number with a fixed seed was generated, sampled uniformly between $-0.02 \cdot R_\mu^{\text{SM}}$ and $0.02 \cdot R_\mu^{\text{SM}}$. The random shift will only be removed for the final unblinding of the measurement. In some cases, data and MC will be compared in the individual channels, where any discrepancy in the overall normalisation could provide hints on the value to be measured for R_μ. This is avoided by artificially scaling the MC so that the total number of events matches the observed events in the data.

5.2 Control channel

To validate the modelling of the distributions associated with the signal side without being sensitive to hints of lepton universality violation, taupair events with both leptonic tau decays in a single event, that is $\tau \to e\nu_e\nu_\tau$ in one hemisphere and $\tau \to \mu\nu_\mu\nu_\tau$ in the other, are used. In the following, this will be called the $e \times \mu$ channel, which is used as a control sample. Any potential issue with the signal side would also show up in this channel. In particular, the validity of lepton ID corrections and associated uncertainties can be tested.

The advantage of using this sample is that most non-taupair backgrounds can be rejected without explicitly reconstructing any neutral particles. Thus, no associated systematics have to be considered. Moreover, no information about R_μ can be inferred from this sample, since it is only sensitive to the product of the leptonic branching fractions rather than their ratio. Even if an overall normalisation difference is observed between data and MC, this would not indicate an increase or decrease of one of the two branching fractions since the current uncertainty on these branching fractions is 0.28 % ($\tau \to e\nu_e\nu_\tau$) and 0.29 % ($\tau \to \mu\nu_\mu\nu_\tau$) [10]. In contrast, other effects that change the overall normalisation are much less precisely known. It is easiest to see this when looking at the expected number of observed events:

$$N_{obs} = 2 \cdot \sigma_{ee\to\tau\tau} \cdot \mathcal{L} \cdot \mathcal{B}(\tau \to e\nu_e\nu_\tau) \cdot \mathcal{B}(\tau \to \mu\nu_\mu\nu_\tau) \cdot \varepsilon_{\text{sig}}. \tag{5.2}$$

The systematic uncertainty associated with the luminosity $\mathcal{L}$ is around 0.6 %, more than a factor of two larger than the uncertainty on the branching fractions. This ensures a minimal observer-expectancy effect when looking at this control sample.

5.2.1 Preselection

To select a clean sample of $\tau \to e\nu_e\nu_\tau$ and $\tau \to \mu\nu_\mu\nu_\tau$, a similar selection strategy is implemented as for the primary measurement. Initially, to select the electron and muon candidate, the same cut on the lepton ID variables as used in the primary sample is applied ($P_e > 0.5$ BDT-based and $P_\mu > 0.9$ likelihood (LH)-based). First, a cut-based preselection is optimised with the annealing technique described in Section 4.4.5. Before optimising the preselection, the following loose cuts on the track kinematics and thrust value are applied:

$$\begin{aligned} 0.85 &\leq V_{\text{thrust}} \\ 0.4\,\text{rad} &\leq \theta_\ell < 2.6\,\text{rad} \\ 1.0\,\text{GeV} &\leq p_\ell < 5.0\,\text{GeV} \end{aligned}$$

Here, θ_ℓ and p_ℓ denote the polar angle and momentum of the lepton candidates in the laboratory frame, which in this case refers to the electron and the muon candidate in both hemispheres. The obtained ROC curve is shown in Figure 5.1, with a black cross indicating the chosen selection. The exact cuts of this preselection are:

$$\begin{aligned} 3.53\ \text{GeV}^2 &\leq m^2_{\text{missing}} < 74.87\ \text{GeV}^2 \\ 0.25\,\text{rad} &\leq \theta^{\text{CM}}_{\text{missing}} < 2.87\,\text{rad} \\ 0.15\,\text{GeV} &\leq p^{\text{CM}}_{t,\text{missing}} < 4.00\,\text{GeV} \\ 0.16\ \text{GeV}^2 &\leq \mu^2_{\text{max (e)}} < 2.67\ \text{GeV}^2 \\ 0.29\ \text{GeV}^2 &\leq \mu^2_{\text{max }(\mu)} < 2.58\ \text{GeV}^2 \\ 1.02\,\text{GeV} &\leq p^{\text{CM}}_{t,\text{max}} < 4.83\,\text{GeV} \\ 0.01\,\text{GeV} &\leq p^{\text{thrust frame}}_{t,\text{min}} < 1.21\,\text{GeV} \end{aligned}$$

The variables m^2_{missing}, $\theta^{\text{CM}}_{\text{missing}}$, $p^{\text{CM}}_{t,\text{missing}}$, $p^{\text{CM}}_{t,\text{max}}$ and $p^{\text{thrust frame}}_{t,\text{min}}$ have the same meaning as defined in Table 4.10. The features $\mu^2_{\text{max (e)}}$ and $\mu^2_{\text{max }(\mu)}$ are the largest possible missing mass squared in the tau decay in the respective hemisphere. They are derived from the expression for the missing mass

$$m^2_{\text{missing}(\tau)} = (p_\tau - p_{\tau,\text{rec}})^2 = m^2_\tau + m^2_{\tau,\text{rec}} - 2E_{\text{beam}}E_{\tau,\text{rec}} + 2\cos\theta_\tau\,|\vec{p}_{\tau,\text{rec}}|\,\sqrt{E^2_{\text{beam}} - m^2_\tau}, \tag{5.3}$$

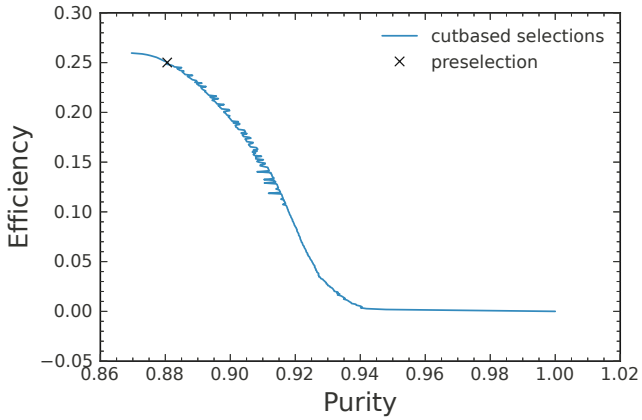


Figure 5.1: A ROC curve obtained by optimising cuts for maximum purity with fixed efficiencies. A black cross marks the set of cuts chosen for the preselection of the $e \times \mu$ sample.

Table 5.1: The most prominent processes after applying the preselection to the $e \times \mu$ sample. The signal efficiency is 25.05 %.

Process	Fraction
$ee \to (\tau \to e\nu_e\nu_\tau)(\tau \to \mu\nu_\mu\nu_\tau)$	88.04 %
$ee \to (\tau \to e\nu_e\nu_\tau)(\tau \to hn\pi^0\nu_\tau)$	5.39 %
$ee \to ee\mu\mu$	2.5 %
$ee \to (\tau \to e\nu_e\nu_\tau)(\tau \to h\nu_\tau)$	2.29 %
$ee \to \tau\tau$ (other)	1.15 %
$ee \to ee\tau\tau$	0.31 %
$ee \to c\bar{c}$	0.12 %
$ee \to ee(\gamma)$	0.1 %

where p_τ is the true (unknown) tau 4-momentum, $p_{\tau,\mathrm{rec}}$ is the reconstructed 4-momentum from the visible decay products and θ_τ is the angle between the corresponding 3-momenta. In this equation, E_τ is set to the beam energy $E_{\mathrm{beam}} = \frac{\sqrt{s}}{2}$, and only quantities in the CM frame are used. Now everything on the right-hand side of Equation 5.3 is known, except for θ_τ. To obtain $\mu^2_{\mathrm{max\,(e)}}$ and $\mu^2_{\mathrm{max}\,(\mu)}$, $\theta_\tau = 0$ is used and thus the largest possible value for $m^2_{\mathrm{missing}(\tau)}$ is obtained.

This preselection only removes a small fraction of background events, since the sample was already relatively clean due to the unlikely combination of an electron and a muon in low-multiplicity backgrounds. In Figure 5.2, all the variables used in the preselection are shown, and the applied selection is indicated with red arrows. All the selections, except for the one on the shown variable, have been applied to show the distributions. The remaining processes after this selection and the most dominant taupair backgrounds are shown in Table 5.1.

5.2.2 Multivariate selection

As a next step, an ANN was trained in the same way as described in Section 4.4.7. As input features, the same variables as for the preselection are used, with the addition of $\theta^{\mathrm{CM}}_{\mathrm{min,missing}}$,

Table 5.2: The most prominent processes after applying all the selections for the $e \times \mu$ sample. The signal efficiency is 8.9 %.

process	fraction
$ee \to (\tau \to e\nu_e\nu_\tau)(\tau \to \mu\nu_\mu\nu_\tau)$	90.9%
$ee \to (\tau \to e\nu_e\nu_\tau)(\tau \to h\nu_\tau + n\pi^0)$	4.5%
$ee \to (\tau \to e\nu_e\nu_\tau)(\tau \to h\nu_\tau))$	2.5%
$ee \to ee\mu\mu$	1%
$ee \to \tau\tau$ (other)	0.8%
$ee \to ee\tau\tau$	0.17%
$ee \to c\bar{c}$	0.05%

for which the definition was given in Equation 4.1. For training, all the cuts mentioned in the previous section are applied, except $p_\ell > 1\,\text{GeV}$, which is loosened to $p_\ell > 0.5\,\text{GeV}$.

In Figure 5.3, the ROC curve for the ANN is shown, both with lepton ID corrections applied and without. The training itself was carried out with the lepton ID corrections in place to reflect the conditions in data more accurately. Figure 5.4 shows the distribution of the output neuron for classifying the signal. The ANN is beneficial to remove the remaining Bhabha events and a significant fraction of the $ee\mu\mu$ background. A loose cut at 0.75 on the output of the ANN is sufficient to remove a large portion of the backgrounds while maintaining most of the signal events. In Figure 5.4, the signal efficiency and purity corresponding to this cut are indicated with a black cross.

In addition to the selection on the ANN output, the kinematics of the lepton candidates are restricted to the same region used in the primary analysis:

$$\begin{aligned} 1.0\,\text{GeV} &\leq p_\ell &&< 5.0\,\text{GeV} \\ 0.82\,\text{rad} &\leq \theta_\ell &&< 2.13\,\text{rad} \end{aligned}$$

After all these selections, the composition of the remaining events is shown in Table 5.2.

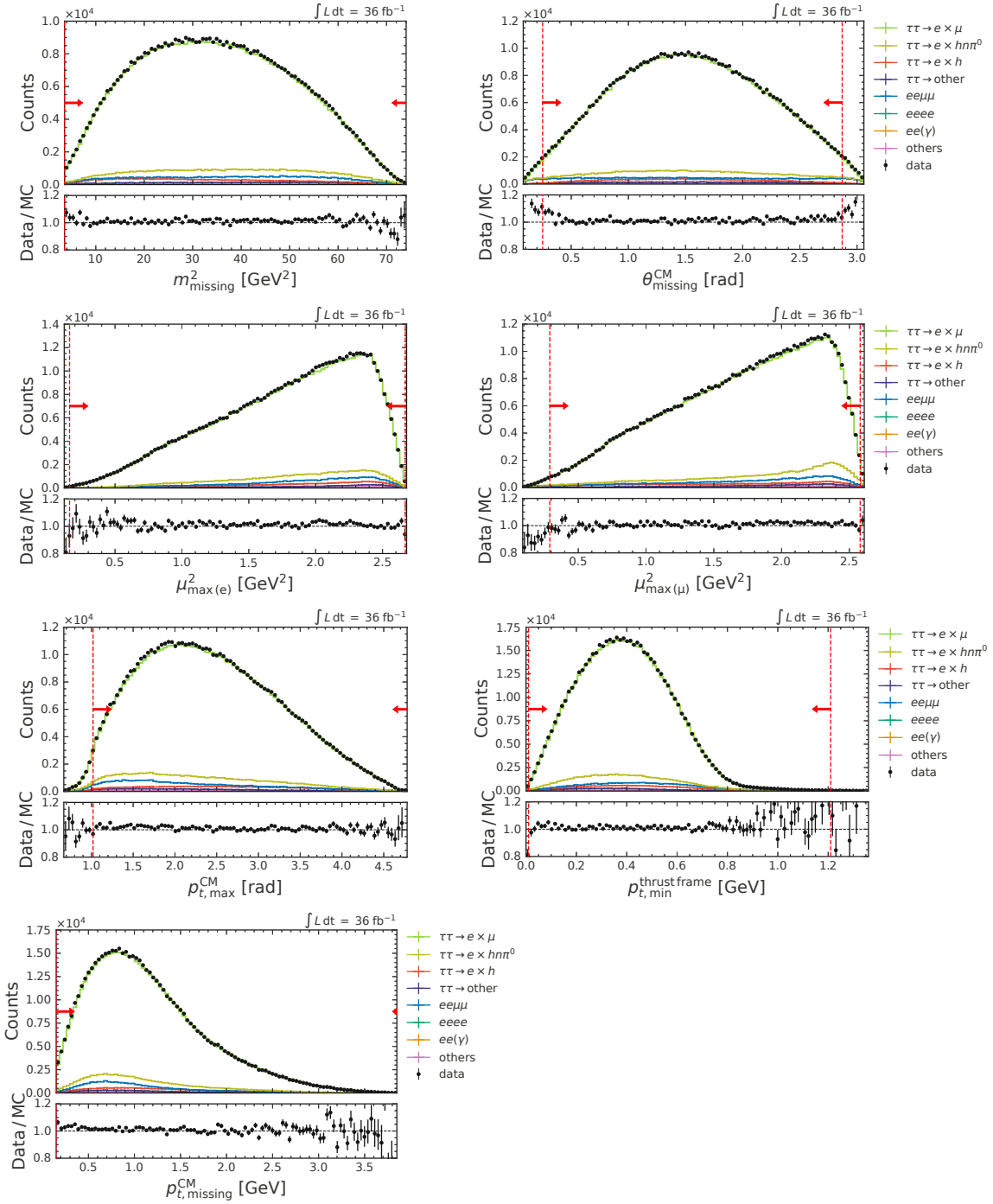


Figure 5.2: The distributions for the variables used to optimise the preselection of the $e \times \mu$ sample. All other selections, except for the one on the shown variable, have been applied.

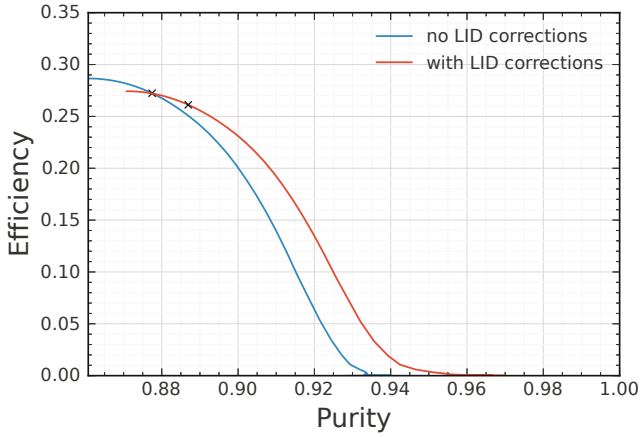


Figure 5.3: The ROC curve of the neural network for the $e \times \mu$ control sample.

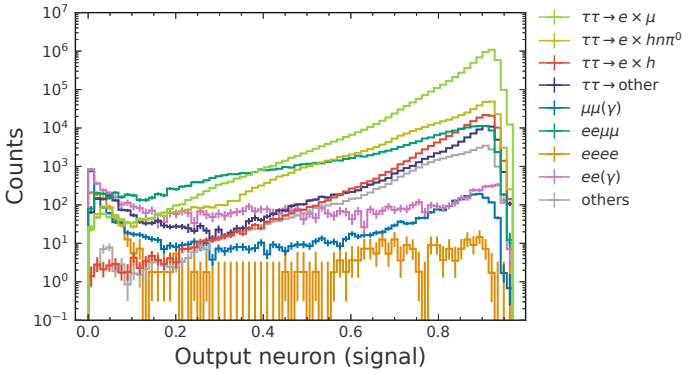


Figure 5.4: The neural network output for the $e \times \mu$ control sample.

5.2.3 Trigger

For the $e \times \mu$ sample, the same trigger lines as for the main channels (Section 4.4.3) are used with the same method to estimate systematics associated with the trigger correction (Section 4.6.4). Figure 5.5 shows the trigger efficiencies and resultant correction factors in bins of the lepton candidate's momentum. The overall efficiency for this sample in the data is greater than 99.5 %. As for the main channels, a $\sim$1 % higher efficiency in data than in TSIM is observed. There is a slight momentum dependence of the trigger efficiency, especially at lower momentum for the electron candidate. This is expected since the dominant trigger line is `hie`, which relies on the energy deposited in ECL by the electron.

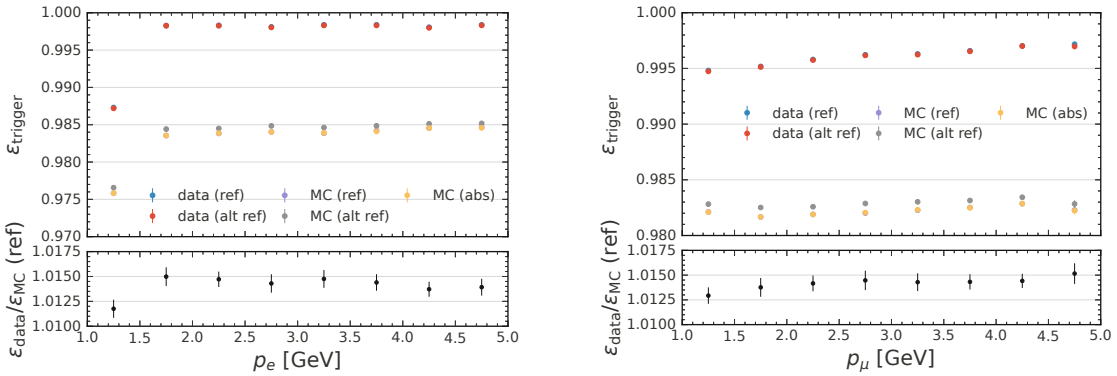


Figure 5.5: Trigger efficiencies and data/MC correction factors with associated systematics for the $e \times \mu$ control sample. The efficiency is shown in momentum bins for the electron candidate (left) and muon candidate (right).

5.2.4 Comparison with data

Good agreement is observed when comparing data and MC of some selected variables, as shown in Figure 5.6. In this figure and in the following, the pull value shown in the lower plot is defined as

$$\frac{N_{\text{data}} - N_{\text{MC}}}{\sqrt{\sigma_{\text{stat}} + \sigma_{\text{sys}}}}, \tag{5.4}$$

where σ_{sys} comprises the systematic uncertainties for lepton ID, trigger and luminosity only. All other systematic effects should have a negligible impact on this sample. Although no information on R_μ measurement can be acquired from looking at the data in this sample, the overall normalisation is still blinded in these figures by normalising the yield in MC to the observed yield in data. This is because a future measurement can be conducted using the $e \times \mu$ channel to obtain a measurement of the absolute branching fractions.

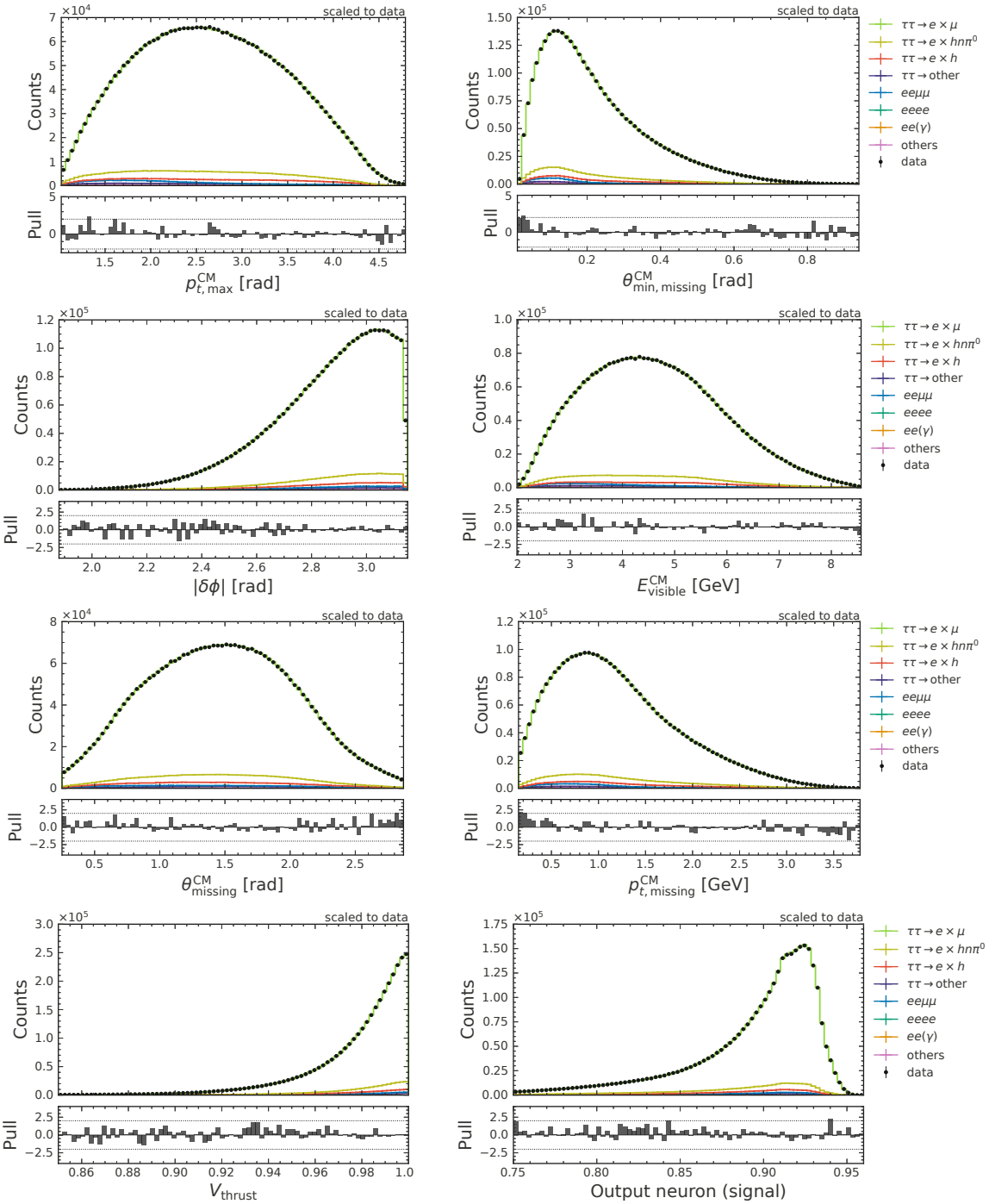


Figure 5.6: Selected distributions of the $e \times \mu$ control sample, comparing data and MC. The pull values are computed considering statistical and systematic uncertainties from lepton ID, trigger and luminosity.

5.2.5 Lepton ID effects

Since a crucial aspect of this measurement are uncertainties associated with lepton ID corrections, it is important to validate them by comparing data to MC in bins of momentum (shown in Figure 5.7) and polar angle (shown in Figure 5.8) of the lepton candidates. The respective other dimension is projected onto the indicated variable when showing the distributions for the electron and muon candidates.

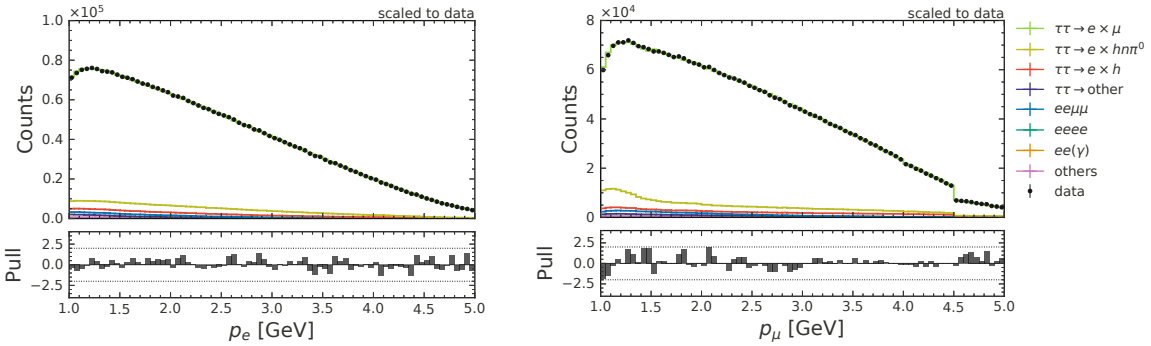


Figure 5.7: A comparison of data and MC in the $e \times \mu$ control sample. On the left-hand side, the momentum of the electron candidate is shown, and on the right-hand side, the momentum of the muon candidate is shown.

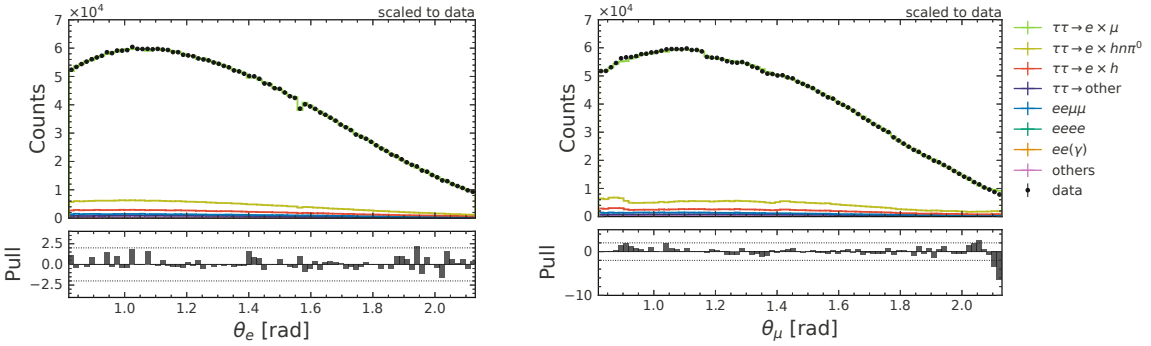


Figure 5.8: A comparison of data and MC in the $e \times \mu$ control sample. On the left-hand side, the polar angle of the electron candidate is shown, and on the right-hand side, the polar angle of the muon candidate is shown.

The momentum distribution shows good agreement for the $e^\pm$ candidate. Also for the $\mu^\pm$ candidate the agreement is good, except for a slight drop in data events at the momentum bins closest to 1 GeV. Similarly, for the polar angle distributions of the $e^\pm$ candidate there is good agreement, whereas for the $\mu^\pm$ candidate an even more significant drop in data at a large polar angle is observed. Both of these drops are independent and most likely come from lepton identification inefficiencies, which are different in data compared to simulation. These efficiency drops also exist in MC, as shown in Section 4.4.6 (Figure 4.15). The lepton ID corrections correct for such inefficiencies but are only provided integrated over larger bins in momentum and theta. Thus, some discrepancies are still observed if they affect only a part of the bin. However, as long as the kinematics of the lepton in the calibration channel used to derive the

corrections is similar to that of the lepton in this analysis, these discrepancies will cancel when integrated over the whole bin. In Figure 5.9 and Figure 5.10, the data/MC agreement is shown in all the individual momentum and polar angle bins used for the lepton efficiency corrections, for the electron candidate and the muon candidate, respectively. The first eight bins represent the lowest theta bin (0.82 rad to 0.92 rad), while the last eight bins show the highest theta bin (2.04 rad to 2.13 rad). Good agreement between data and the corrected MC is observed in the individual bins, with a slight data deficit for the highest polar angle bin ranging from 2.04 rad to 2.13 rad for the muon candidate (bin number greater than 80 in Figure 5.10). Since this deficit is within 2σ of the quoted uncertainties, it is still acceptable. This check indicates that at the level of precision of the control sample, the lepton ID corrections appropriately amend the mismodelling in simulation. While the measurement will be performed over the integrated polar angle bins and the variation of the correction factors within these bins is acceptable, the momentum bins are split into sub-bins, as this dimension is used for defining the templates. This is the main reason why the analysis is restricted to leptons with momentum larger than 1.5 GeV, since for the lepton ID momentum bin ranging from 1 GeV to 1.5 GeV, this division could be problematic because of the variations of the muon efficiency and mis-id rate corrections within the bin.

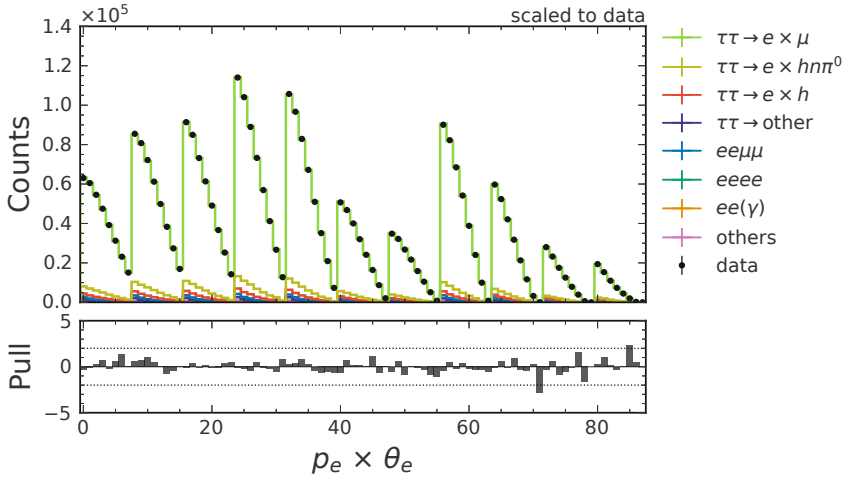


Figure 5.9: A comparison of data and MC in the $e \times \mu$ control sample. The binning is defined over the momentum and polar angle of the electron candidate. The first eight bins correspond to the low theta region, and the last eight correspond to the high theta region. The bins reflect the binning of the lepton ID corrections.

5.2.6 Fit to data

Previously, when the distributions of the lepton candidates were shown, the corrections and uncertainties of the respective other track were integrated. This should have a similar impact for each bin, but especially the polar angles of the two tracks will not be independent. It is not feasible to completely disentangle the lepton ID corrections for the electron candidate from the muon candidate. Neither is it possible to disentangle efficiency from mis-id rate uncertainties. However, it is possible to perform a fit that correctly models these uncertainties and accounts for all correlations. The fit results will show whether the model can describe the data or whether the systematic uncertainties cannot account for some significant discrepancies.

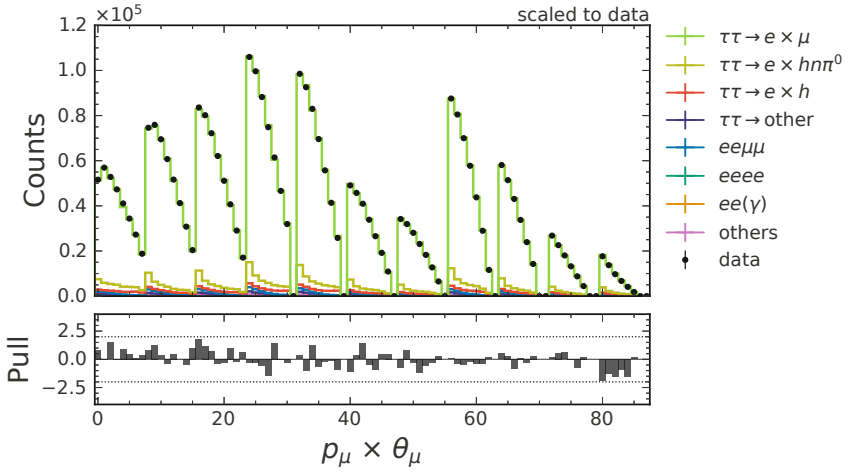


Figure 5.10: A comparison of data and MC in the $e \times \mu$ control sample. The binning is defined over the momentum and polar angle of the muon candidate. The first eight bins correspond to the low theta region, and the last eight correspond to the high theta region. The bins reflect the binning of the lepton ID corrections.

The degree to which this validates the corrections is limited by the statistical precision of the $e \times \mu$ sample.

For the fit, the same method as for the main channels is used: A binned maximum likelihood fit with `pyhf`, where the template is now defined over two dimensions of the momenta of the two tracks. Following the binning of the lepton ID corrections, each momentum dimension consists of eight bins, ranging from 1 GeV to 5 GeV in steps of 0.5 GeV. This amounts to a total of 64 bins. The fit uses four templates based on the truth-matched particle type in MC:

- $e \times \mu$
- $e \times \neg\mu$
- $\neg e \times \mu$
- $\neg e \times \neg\mu$,

where $\neg e$ and $\neg\mu$ denote all particle types except electron and muon respectively. The pre-fit templates are shown in Figure 5.11, where the first eight bins correspond to the lowest and the last eight bins to the highest electron momentum bin.

To incorporate the lepton ID uncertainties, the same method as described in Section 4.6.2 is used. The resulting correlation matrix is shown in Figure 5.12. The four blocks indicated by the labels are the fit templates, and each block is a 64×64 matrix, showing the correlation between the individual bins. Since there are now two lepton candidates in the event that have to be considered simultaneously, the lepton ID uncertainties of each template are correlated with at least two other templates. A much larger number of nuisance parameters are needed to account for these correlations than for the primary measurement. From this matrix's eigendecomposition (see Equation 4.23), $64 \cdot 4 = 256$ independent shape variations are obtained. As mentioned in Section 4.6.2, not all of them are needed to correctly model the correlations

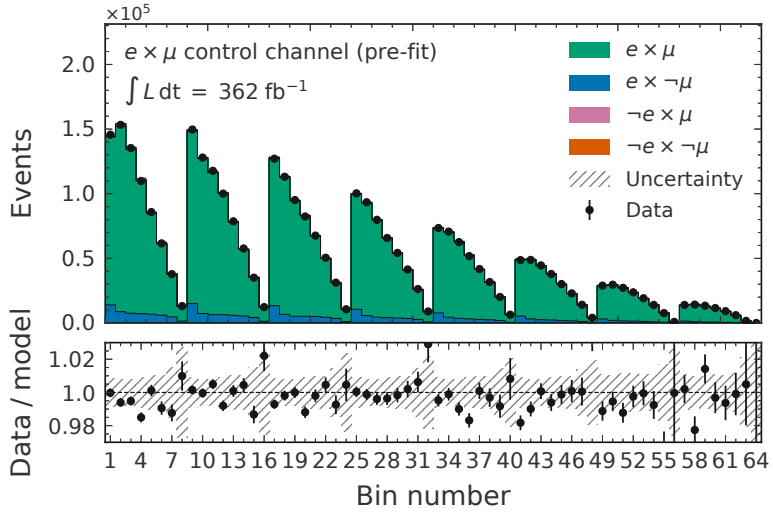
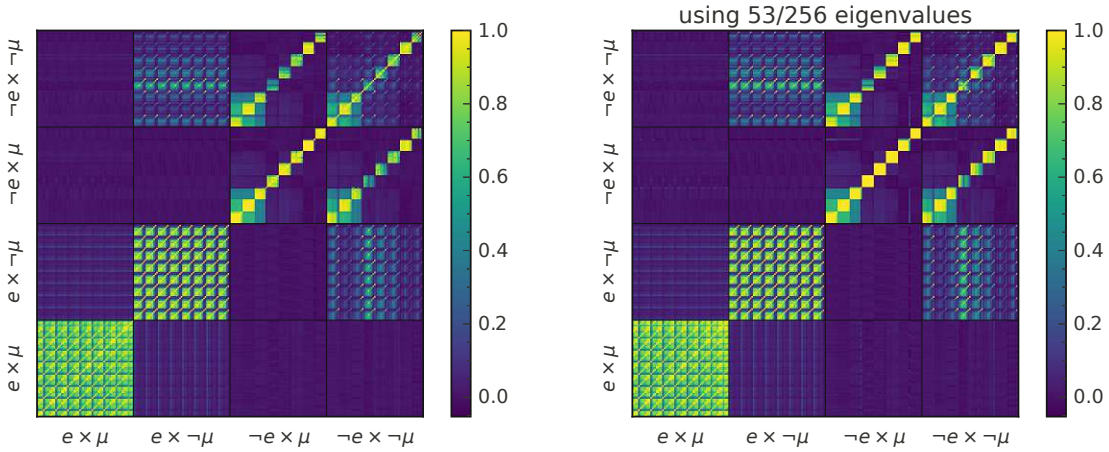


Figure 5.11: The pre-fit templates for the control sample fit. The templates are defined over the two dimensions of the lepton candidate momenta.

since many will hardly affect the templates. Setting a threshold on the eigenvalues of 250, the total number of variations can be reduced to 54. The original correlation matrix is shown in Figure 5.12a, and the correlation obtained when only using 54 eigenvalues is shown in Figure 5.12b. The original matrix is correctly reproduced by the 54 shape variations. Thus, only 54 nuisance parameters are used in the model to describe the lepton ID uncertainties and their correlation.



(a) The original correlation obtained from the lepton ID correction tables.

(b) The correlation obtained when only 53 shape variations, corresponding to the largest eigenvalues, are used.

Figure 5.12: The correlation matrix of the lepton ID uncertainties. The four blocks indicated by the labels are the fit templates, and each block is a 64 × 64 matrix, showing the correlation between the individual bins and templates.

The post-fit templates are shown in Figure 5.13. The model agrees reasonably well with the data. A goodness-of-fit test gives a p-value of 0.41, indicating good compatibility between the

post-fit model and the observations.

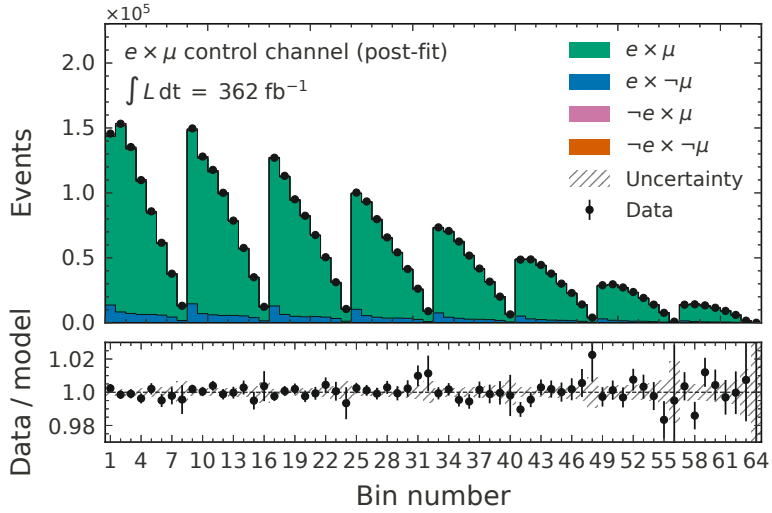


Figure 5.13: The post-fit templates for the control sample fit. The templates are defined over the two dimensions of the lepton candidate momenta. The p-value of a goodness-of-fit test is 0.23.

Figure 5.14 shows the pull plot for all the nuisance parameters used in the fit. The pull values are calculated as $(\hat{\theta} - \theta_0)/\Delta\theta$, where $\hat{\theta}$ is the MLE of the nuisance parameter, θ_0 is the expected value, and $\Delta\theta$ is the width of the constraint term c_χ in Table 4.19. The error bars shown together with the pull values reflect the post-fit uncertainties of the nuisance parameters. It is reassuring to see that none of the nuisance parameters pulls away much more than 1σ of the constraints from the nominal values. If extreme pull values were observed, it would indicate that some of the systematics are underestimated. Unfortunately, it cannot be concluded from this check alone that the uncertainties are not underestimated - they can only be checked up to a certain precision. Below that, the sensitivity is lost since other effects, such as the statistical uncertainty, will compensate for discrepancies.

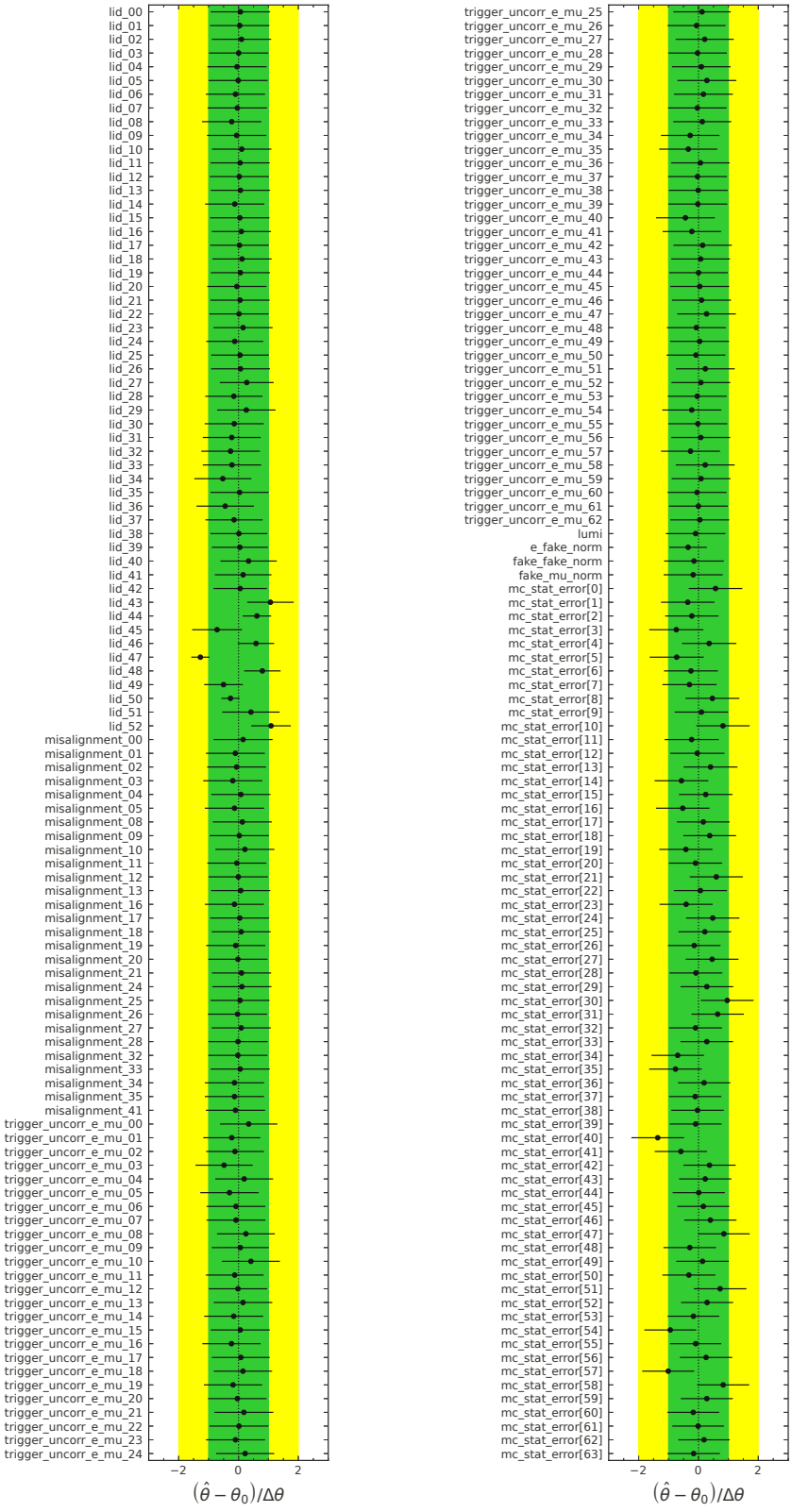


Figure 5.14: A pull plot showing all nuisance parameters used to fit the $e \times \mu$ sample with a two-dimensional template defined over the electron and muon candidate momentum.

5.2.7 3-channel fit

An additional check that can be realised with the $e \times \mu$ sample is to perform a combined fit with the two main channels. From this 3-channel fit, the absolute branching fractions of $\tau \to e\nu_e\nu_\tau$ ($\mathcal{B}_e$) and $\tau \to \mu\nu_\mu\nu_\tau$ ($\mathcal{B}_\mu$) can be extracted and R_μ can be obtained as their ratio. To obtain the correct uncertainty on R_μ, it is important to take into account the correlation of the uncertainties of the fitted branching fractions. Given that there are no incompatibilities with the implemented systematic effects and everything was done correctly, the obtained value of R_μ should be similar to the one extracted directly from the 2-channel fit.

While this will primarily serve as a cross-check for the present analysis, the expected precision on the measured branching fractions is similar to the best measurements from older experiments, making it also attractive as a standalone measurement in the future. In principle, the absolute branching fractions can also be measured only from the 2-channel fit. The advantage of adding the $e \times \mu$ control sample is that it constrains the product $\mathcal{B}_e \times \mathcal{B}_\mu$, which, together with $R_\mu = \mathcal{B}_\mu/\mathcal{B}_e$, will significantly reduce the uncertainty on the branching fractions. Because of the branching fraction product, normalisation uncertainties that affect the $e \times \mu$ channel will be reduced by 50 % for the individual branching fractions. In the e and μ channels, the precision of the absolute branching fractions is limited by the uncertainty coming from the tag side, which cancels for R_μ but not for the individual branching fractions.

A relatively coarse binning of the $e \times \mu$ control sample is chosen since the primary interest is normalisation, which will be dominated by the uncertainty of luminosity and tracking efficiency, independently of the number of bins. Similarly to what was done in Section 5.2.6, a two-dimensional template of the momenta of the electron and muon candidate is used. The leptons are separated into higher 1.5 GeV to 3 GeV and lower 3 GeV to 5 GeV momentum, which results in a total number of 4 bins. Then, three templates based on the generated tau decays in MC are defined:

1. $\tau \to e\nu_e\nu_\tau \times \tau \to \mu\nu_\mu\nu_\tau$
2. $\tau \to e\nu_e\nu_\tau \times \tau \to h\nu_\tau(\geq 0\,\pi^0)$
3. other: all other tau decays and non-taupair backgrounds

The normalisation of the first template is affected by both fit parameters ($\mathcal{B}_e$ and $\mathcal{B}_\mu$), which are included in the model as normfactor modifiers. The second template is only modified by the parameter $\mathcal{B}_e$, and a normalisation uncertainty of 2.5 % is assigned based on the studies with the $\tau \to hn\pi^0\nu_\tau$ tag in the main channels. The third template only contains about 2 % of the events, and a conservative uncertainty of 25 % is assigned to its normalisation since the backgrounds were not studied in detail. This third template is dominated by the $ee \to ee\mu\mu$ process (see Table 5.2). All templates are affected by the luminosity uncertainty (0.63 %) and the tracking uncertainty (0.48 %). The first two templates have an additional normalisation uncertainty of 0.33 %, which originates from the uncertainty of the taupair cross-section (0.919 ± 0.003 nb) [158].

To simultaneously fit the three channels, it is essential to include the correct correlation of the systematic variations in the model. For most relevant uncertainties (such as tracking, luminosity, or taupair cross-section), the correlation is 100 %. The systematic with the

most complicated correlation between the channels is lepton ID. Here, the templates in the $e \times \mu$ channel are simultaneously correlated with electron and muon efficiency or mis-id rate uncertainties. Fortunately, the same toy approach as before (described in Section 4.6.2) can be used to obtain the correlation between the individual bins and templates. To have the correct correlation, the generated toy weights must be the same as what was used for the e and μ channels, which can be achieved by setting the same seed. The resulting correlation matrix obtained from the toy sample is shown in Figure 5.15. Each entry in this matrix shows the correlation between two bins, and the labels correspond to the different templates:

- $\mu^{\pm} \to \mu^{\pm}$: $\mu^{\pm}$ templates in the μ-channel
- $h^{\pm} \to \mu^{\pm}$: $h^{\pm}$ template in the μ-channel
- $e^{\pm} \to e^{\pm}$: $e^{\pm}$ templates in the e-channel
- $h^{\pm} \to e^{\pm}$: $h^{\pm}$ template in the e-channel
- $e^{\pm} \times \mu^{\pm}$: first template in the $(e \times \mu)$-channel
- $e^{\pm} \times h^{\pm}$: second template in the $(e \times \mu)$-channel
- other: third template in the $(e \times \mu)$-channel

This correlation is then modelled with 40 independent modifications of the templates, using the same method as previously in Section 4.6.2.

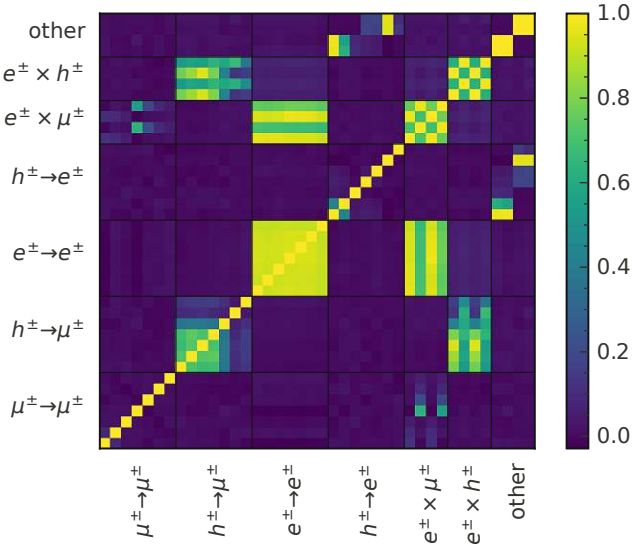


Figure 5.15: The correlation between the bins and templates of the e-channel, the μ-channel and the $(e \times \mu)$-channel, obtained with 5000 toy weights.

For all other systematics, the same procedure as for the e and μ channels is used. The resulting contributions of the individual systematics to the overall uncertainty are shown in Table 5.3. These uncertainties are estimated with 2000 toys for each systematic, where the sampled data reflect the variations from this systematic effect in the same way as it was described in

Table 5.3: The expected uncertainties from the 3-channel fit, attributed to different systematic sources. The uncertainties were obtained with toy samples, where for each systematic type 2000 samples were used. Also shown is the correlation obtained from the toy samples and the resulting uncertainty on R_μ when taking the ratio of the fitted branching fractions. The uncertainties are given as a percentage of the expected nominal values of $\mathcal{B}_e$, $\mathcal{B}_\mu$ and R_μ.

	$\mathcal{B}_{\tau^\pm\to e^\pm\nu\bar{\nu}}$ [%]	$\mathcal{B}_{\tau^\pm\to \mu^\pm\nu\bar{\nu}}$ [%]	correlation	R_μ [%]
Charged-particle identification	0.22	0.22	0.06	0.30
Trigger	0.07	0.09	0.31	0.10
Modelling of FSR	0.05	0.04	< 0.01	0.08
Normalisation of individual samples	0.21	0.22	0.94	0.08
Size of the simulated samples	0.06	0.07	0.57	0.06
Modelling of the momentum distribution	0.03	0.04	< 0.01	0.06
Tag side modelling (signal embedding)	0.03	0.04	< 0.01	0.05
Particle decay-in-flight	< 0.01	< 0.01	< 0.01	0.02
Luminosity	0.29	0.31	1.00	0.01
Modelling of ISR	< 0.01	0.01	1.00	0.01
Tracking efficiency	0.24	0.25	1.00	0.01
Neural network selection	0.01	< 0.01	0.99	< 0.01
Detector misalignment	0.03	0.03	0.97	< 0.01
Photon efficiency	0.01	< 0.01	1.00	< 0.01
$e \times \mu$ sample selection	< 0.01	< 0.01	1.00	< 0.01
π^0 efficiency	0.05	0.05	1.00	< 0.01
Photon energy	< 0.01	< 0.01	1.00	< 0.01
Momentum correction	< 0.01	< 0.01	1.00	< 0.01
Total systematic uncertainty	0.52	0.55	0.77	0.36
Statistical uncertainty	0.04	0.05	-0.32	0.07
Total	0.52	0.55	0.76	0.37

Section 4.7. Using these toy samples, the correlation between the fitted branching fractions and the contribution of the systematic to the uncertainty on their ratio R_μ can be obtained directly, as shown in the table. The uncertainty on $\mathcal{B}_e$ is 0.000 92 and on $\mathcal{B}_\mu$ 0.000 95, with a correlation of 0.76. The resultant total uncertainties (from summing the individual components in quadrature) and correlation are consistent with what is obtained from a single fit.

The three channels and their templates are shown in Figure 5.16, both pre-fit (left) and post-fit (right). The shown data is Asimov data, to not unblind the absolute branching fraction measurement by revealing the data/MC agreement in all three channels.

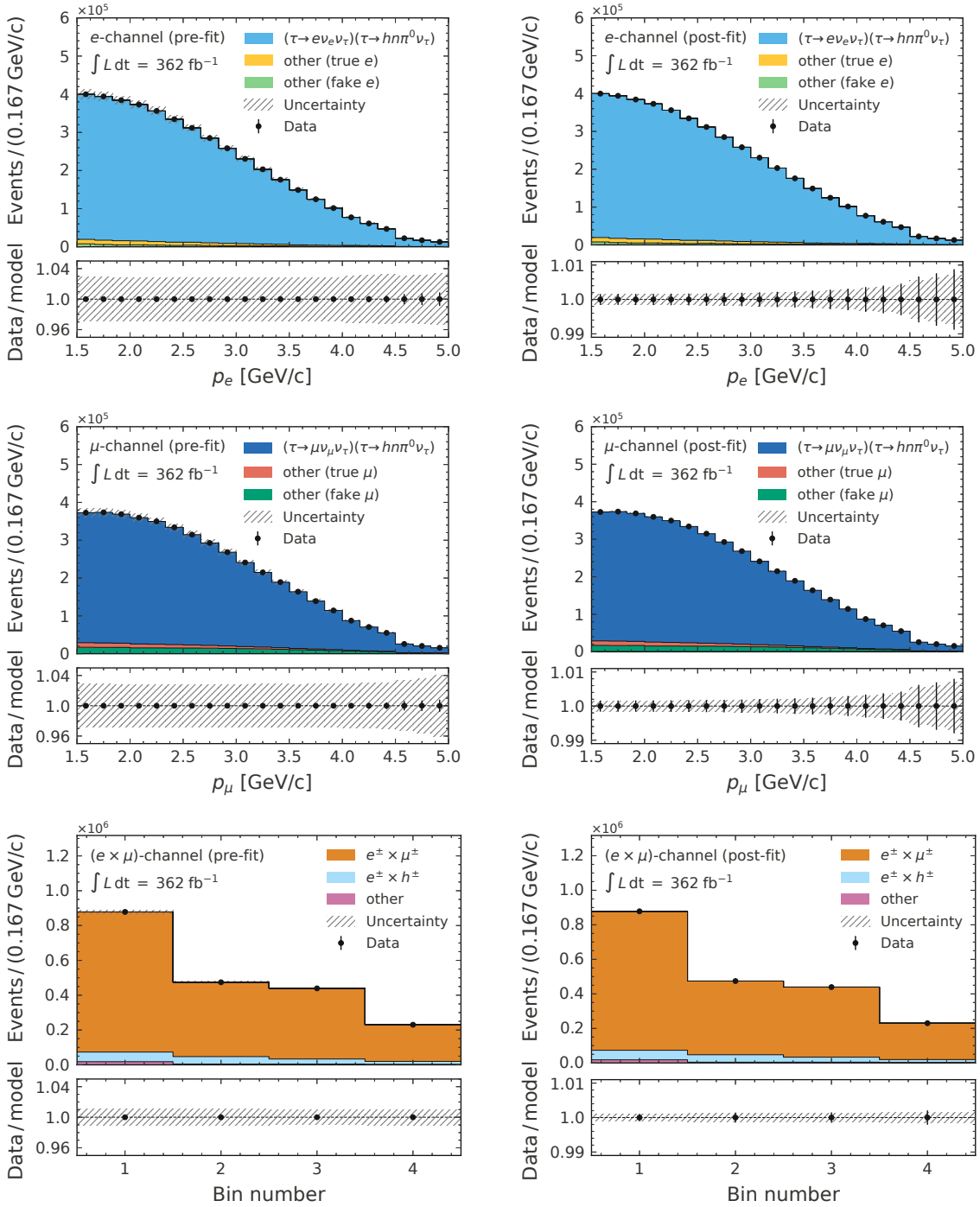


Figure 5.16: The three channels and their templates used to extract the absolute branching fractions, pre-fit (left) and post-fit (right), using Asimov data.

When performing the fit with actual data, an unknown random shift (from $-2\,\%$ to $2\,\%$) was included in the likelihood function for both parameters $\mathcal{B}_e$ and $\mathcal{B}_\mu$, keeping the measurement blinded (this means that the normalisation of the signal templates is shifted by this amount). This random shift corresponds to the same shift used in the 2-channel fit, which allows for a comparison of the fitted values of R_μ directly with the actual data. This difference can be calculated without risking unblinding the measurements. Taking the ratio of the branching fractions, the difference between the central value of R_μ from the 3-channel fit and the value of R_μ directly obtained from the 2-channel fit is less than $0.02\,\%$, which indicates excellent consistency. The slight difference can be attributed to the information gained on some nuisance parameters with the addition of the third channel. This can also be seen from the uncertainties, where the total uncertainty on R_μ with the 3-channel fit with 0.0036 is slightly less than with the 2-channel fit with 0.0037. This is mainly due to a slight reduction in the systematic uncertainty associated with lepton ID, from $0.32\,\%$ to $0.3\,\%$. One could even consider including more control channels sensitive to lepton ID effects to decrease the associated uncertainty on R_μ.

5.2.8 Conclusion of control channel studies

No significant discrepancies were found when comparing the $e \times \mu$ sample with data. The uncertainties incorporated in the analysis can account for the observed differences.

An important observation from the $e \times \mu$ sample is the drop in efficiency at the lower end of the muon momentum spectrum and the high polar angle region. However, since the leptons in the main channels have restricted kinematics from 1.5 GeV to 5.0 GeV, the drop at low momentum is not a concern. The high polar angle region requires further investigation, and additional checks will be carried out in Section 5.4 and Section 5.5.5.

There was no clear bias or underestimation of the lepton ID uncertainties. However, due to the limited statistical precision of the control sample, it may not be sensitive enough to validate the lepton efficiency uncertainties to the required level for the measurement. Despite this, the agreement within the accessible precision is crucial for moving on to the next steps since it already indicates that the signal side is well-modelled. Still, it is necessary to perform additional validations and checks on the main channels, while remaining blind to the results. These checks include, among other things, stability checks to ensure that lepton ID does not bias the measurement, which can be found in Section 5.5.3.

Finally, the fit using all three channels simultaneously showed good consistency with the results of the 2-channel fit, indicating that the included systematic effects are compatible. It also shows that there is no problem with the fitting method itself.

5.3 Fit validation

In this section, several studies to validate the fit procedure and study the properties of the estimator are shown. Most of these checks use simulated sets of observables sampled from the fit model, which will be referred to as toy samples.

5.3.1 Linearity checks

2-channel fit

As a first check, a toy study is performed with the nominal model for different assumed values of R_μ. Seven values ranging from 0.965 to 0.985 are tested, which is approximately $\pm 1\,\%$ around the SM value of 0.972 56. For each of these assumed values, 10 000 toys are sampled and the pull value

$$\text{pull} = \frac{R_\mu^{\text{fit}} - R_\mu^{\text{gen}}}{\sigma_{R_\mu}^{\text{fit}}} \tag{5.5}$$

is computed for each of them. Here, R_μ^{fit} refers to the estimated value of R_μ, R_μ^{gen} to the generated value, and $\sigma_{R_\mu}^{\text{fit}}$ to the estimate of the uncertainty of R_μ. One of the resulting distributions of pull values is shown in Figure 5.17 (with $R_\mu^{\text{gen}} = 0.98$). Then the mean ($\mu$) and standard deviation (σ) are extracted by fitting a Gaussian distribution to the obtained values.

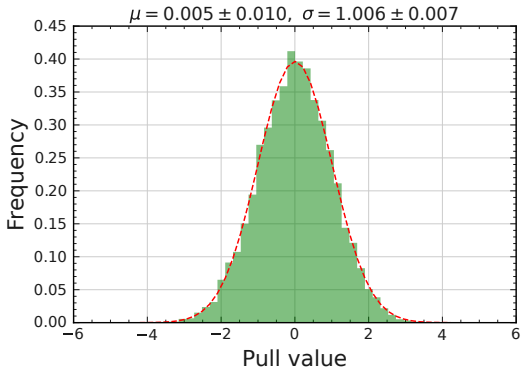


Figure 5.17: The pull distribution obtained with 10 000 toys, assuming $R_\mu^{\text{gen}} = 0.98$.

For all the tested cases, it was found that the mean is consistent with 0 and the standard deviation with 1. The estimator is unbiased and the uncertainties of R_μ are correctly estimated. The fitted σ of the pull distributions is shown in Figure 5.18.

The fit bias, obtained by fitting the distribution of

$$\text{bias} = R_\mu^{\text{fit}} - R_\mu^{\text{gen}} \tag{5.6}$$

and extracting its mean, is shown in Figure 5.19 for each of the assumed values.

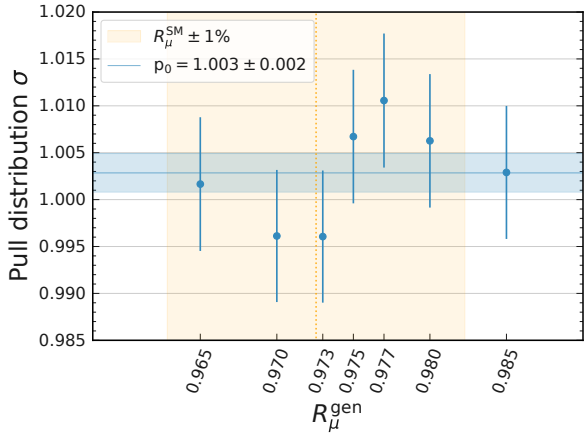


Figure 5.18: The fitted standard deviations of the pull distributions for different values of R_μ^{gen}. The result of a fit with a polynomial of degree 0 is overlayed. All values are consistent with 1.

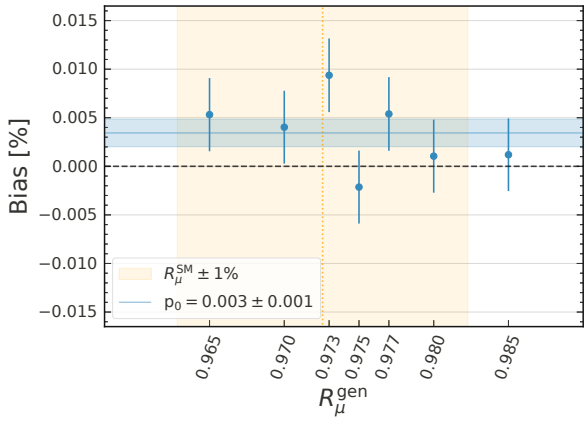
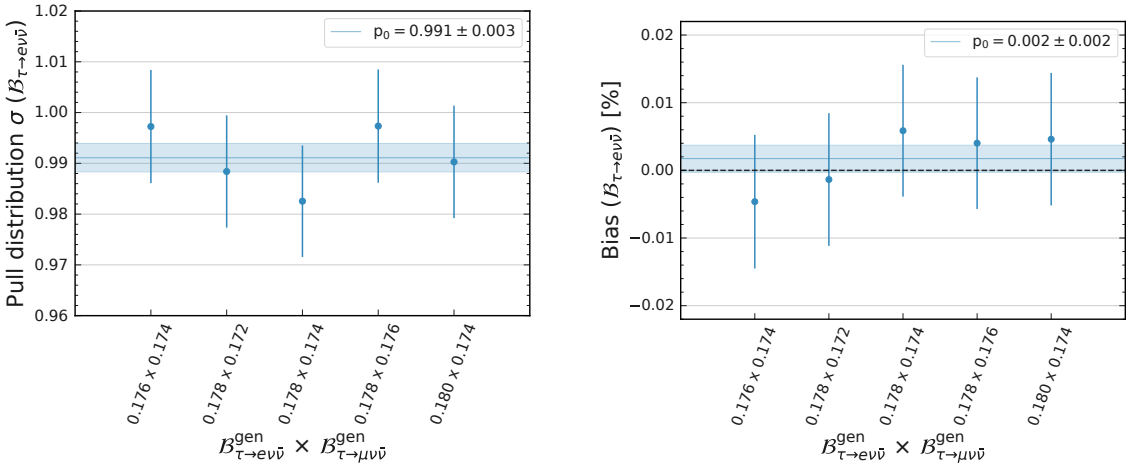


Figure 5.19: The fit bias obtained from the distributions of toys for different values of R_μ^{gen}, given as a percentage of the generated value. The result of a fit with a polynomial of degree 0 is overlayed. All values are consistent with 0.

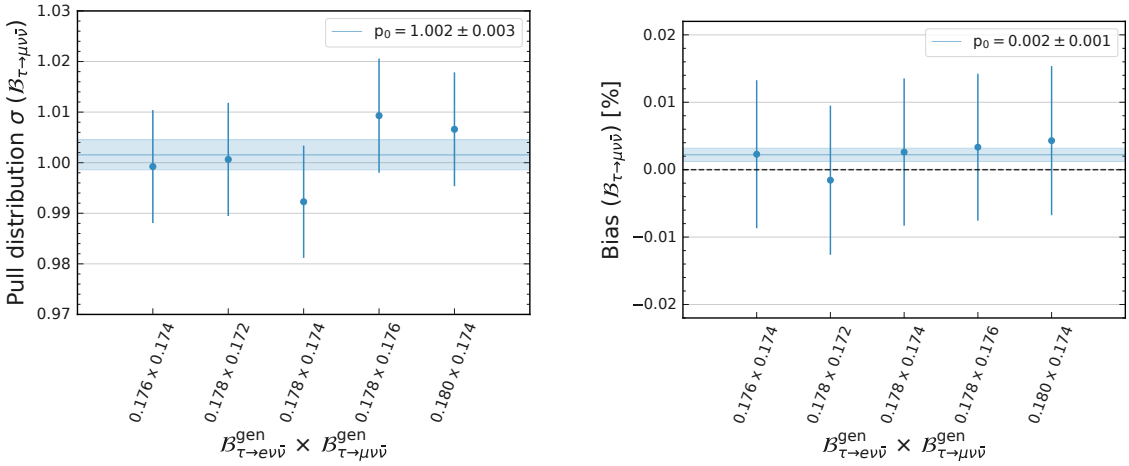
3-channel fit

In this section, some of the studies described in Section 5.3.1 are repeated for the 3-channel fit used to estimate the absolute branching fractions. For this, 4000 toy samples and five combinations of assumed $\tau \to e\nu_e\nu_\tau$ and $\tau \to \mu\nu_\mu\nu_\tau$ branching fractions around the current PDG values are used. The results are shown in Figures 5.20 and 5.21.



(a) σ of the fitted pull distributions (b) relative bias

Figure 5.20: The results of toy studies with the 3-channel fit, with different assumed values for $\mathcal{B}(\tau \to e\nu_e\nu_\tau)$ and $\mathcal{B}(\tau \to \mu\nu_\mu\nu_\tau)$.



(a) σ of the fitted pull distributions (b) relative bias

Figure 5.21: The results of toy studies with the 3-channel fit, with different assumed values for $\mathcal{B}(\tau \to e\nu_e\nu_\tau)$ and $\mathcal{B}(\tau \to \mu\nu_\mu\nu_\tau)$.

5.3.2 Lepton ID correlation

As mentioned in Section 4.6.2, the exact correlation of the systematic component of the lepton ID uncertainties between the individual correction bins is unknown. Therefore, some assumptions about the correlation in the charge, polar angle, and momentum dimensions must be made. While fully correlated uncertainties for charge and polar angle can be justified as being the most conservative, this cannot be taken for granted for the momentum dimension. To test the impact of the correlation between momentum bins, 5000 toys that follow correlation A are sampled and then fitted with a model that assumes correlation B. Correlations from 0 to 1.0 are tested in steps of 0.25, so A $\in \{0, 0.25, 0.5, 0.75, 1.0\}$ and B $\in \{0, 0.25, 0.5, 0.75, 1.0\}$. This results in a total number of 25 possible combinations. The relative uncertainty on R_μ

obtained from the toys for each combination is shown in Figure 5.22. It ranges from 0.27 to 0.39 and is most prominent in the fully correlated case.

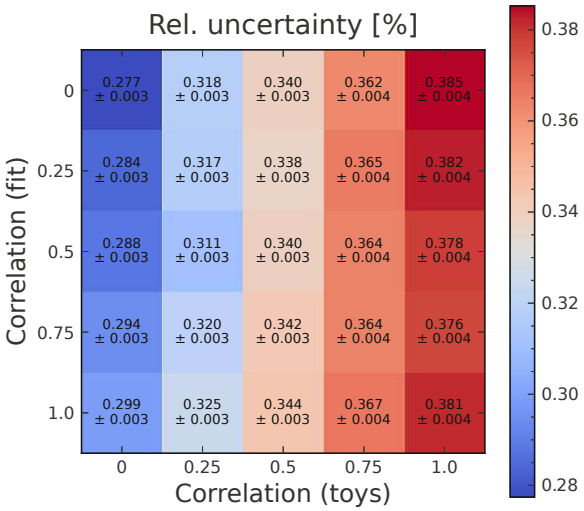
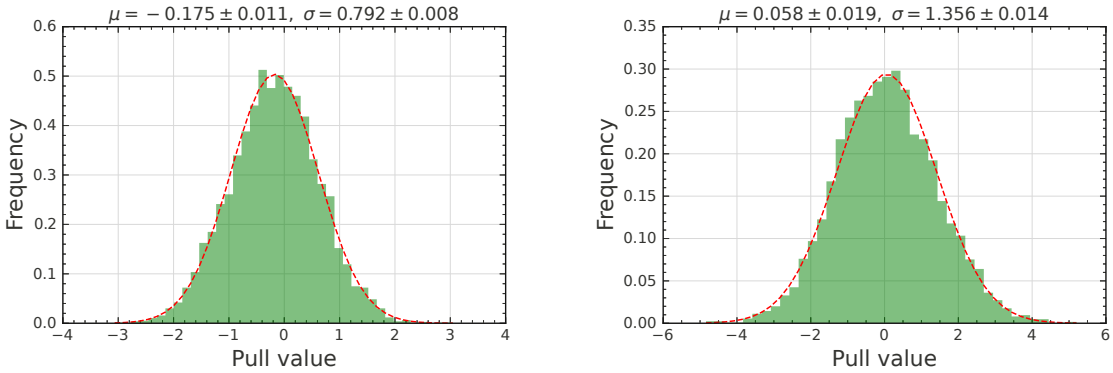


Figure 5.22: The relative uncertainty on R_μ for different combinations of lepton ID systematic correlation in the toys and assumed in the fit.

Since the fit model will not necessarily assume the actual correlation, the uncertainty estimate on R_μ may be wrong. To be conservative, the correlation where uncertainties are not underestimated for any case should be used. This information is obtained from the σ of the fitted pull distribution of toys. Two such pull distributions are shown in Figure 5.23, where on the left-hand side the correlation in the toys is 0 and the correlation in the fit model is 1.0, and on the right-hand side it is vice versa. For the first case, the pull σ is 0.794 ± 0.028, meaning that the uncertainty on R_μ is overestimated, while for the second case, the pull σ is 1.337 ± 0.047, meaning that it is underestimated. The results from this toy study for all the



(a) Toys without correlation, fitted with a model that assumes 100 % correlation.

(b) Toys with 100 % correlation, fitted with a model that assumes no correlation.

Figure 5.23: Pull distributions to study the estimator when the correlation in the model is different from the underlying correlation.

possible combinations are shown in Figure 5.24. Assuming a correlation of 1.0, the uncertainty

is not underestimated in any tested case. Thus, this correlation is used in the model.

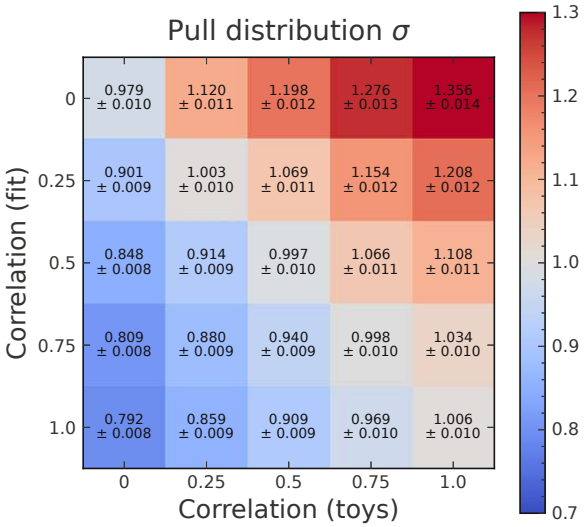


Figure 5.24: The fitted σ of the pull distributions for different combinations of lepton ID systematic correlation in the toys and assumed in the fit.

5.3.3 ANN cut dependence

Figure 5.25 shows the expected uncertainty of R_μ as a function of the cut on the ANN output. With a loose cut, the uncertainty increases due to the larger background. Instead, when cutting too tight, the signal efficiency becomes very small, and the statistical uncertainty becomes dominant. A cut between 0.85 and 0.95 results in the best precision. Figure 5.26 shows the fit bias for different ANN output selections. No bias is seen in any of the cases.

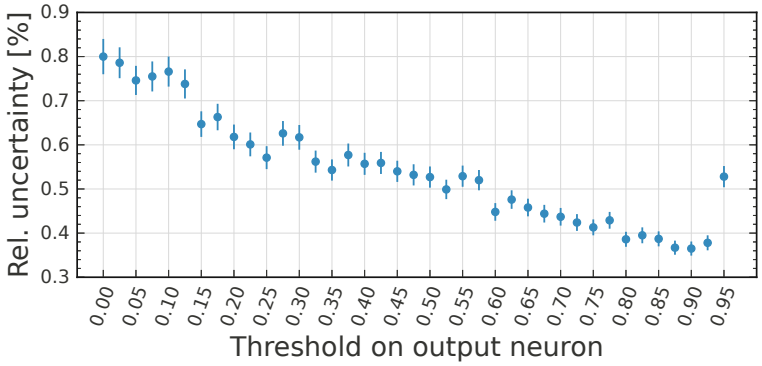


Figure 5.25: The relative uncertainty on R_μ for different cuts on the ANN output. A cut between 0.85 and 0.9 gives the best sensitivity.

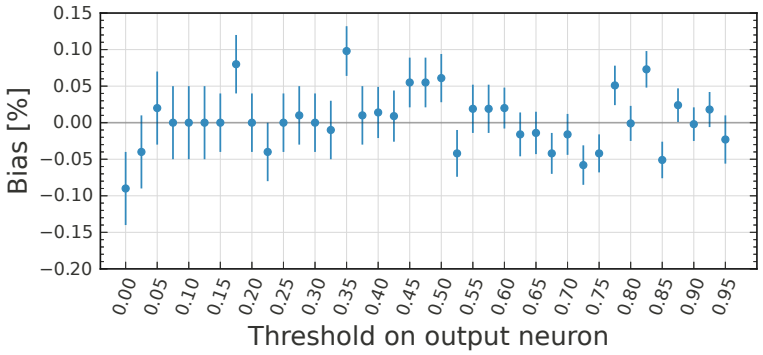


Figure 5.26: The fit bias relative to R_μ for different cuts on the ANN output. No significant bias is observed.

5.3.4 Likelihood profiles

This section shows profile likelihood scans of the parameters of interest. The scans are performed with the random shifts included in the likelihood functions, to not unblind the measurements. In the following figures, the axis labels only show the difference relative to the best-fit value. Figure 5.27 shows the scan for R_μ when using real data, while Figure 5.28a and Figure 5.28b show the scans for $\mathcal{B}(\tau \to e\nu_e\nu_\tau)$ and $\mathcal{B}(\tau \to \mu\nu_\mu\nu_\tau)$, respectively, when using Asimov data. The likelihood profile corresponds to what is obtained with the Gaussian approximation, which is assumed by the HESSE method of MINUIT [71] to calculate the uncertainties of the fitted parameters. This Gaussian approximation and its width σ, which is the uncertainty obtained from the fit, are also shown in the figures. Also shown are likelihood scans with all nuisance parameters fixed to their MLEs and only the parameters of interest left free. This is used to estimate the statistical component of the total uncertainty. The likelihood is Gaussian in this case as well.

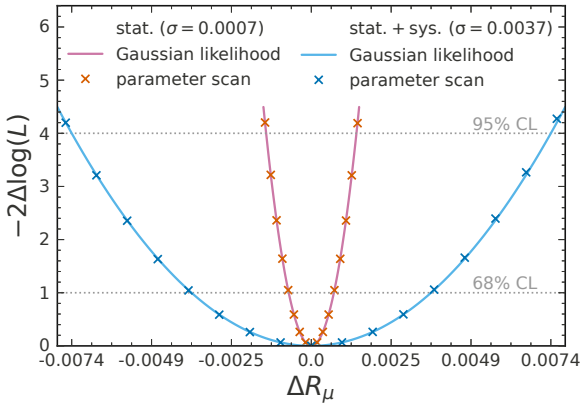
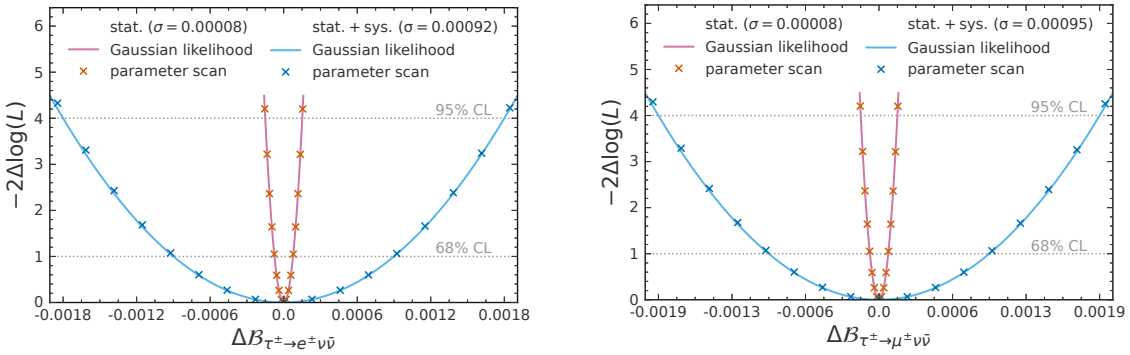


Figure 5.27: The likelihood profile of R_μ obtained from the fit with two channels. The crosses mark the scanned values, while the solid line indicates the values of a Gaussian likelihood with a width σ corresponding to the uncertainty obtained from the fit. The red curve is obtained from a fit with all nuisance parameters fixed to their best-fit values, indicating the statistical component of the total uncertainty. The blue curve corresponds to the nominal fit, with an uncertainty that includes both statistical and systematic effects.



(a) The likelihood profile of $\mathcal{B}(\tau \to e\nu_e\nu_\tau)$, obtained from the fit with three channels.

(b) The likelihood profile of $\mathcal{B}(\tau \to \mu\nu_\mu\nu_\tau)$, obtained from the fit with three channels.

Figure 5.28: The crosses mark the scanned values, while the solid line indicates the values of a Gaussian likelihood with a width σ corresponding to the uncertainty obtained from the fit. The red curve is obtained from a fit with all nuisance parameters fixed to their best-fit values, indicating the statistical component of the total uncertainty. The blue curve corresponds to the nominal fit, with an uncertainty that includes both statistical and systematic effects.

In Figure 5.29, two-dimensional contours of the likelihood profile are shown, scanning both $\mathcal{B}(\tau \to e\nu_e\nu_\tau)$ and $\mathcal{B}(\tau \to \mu\nu_\mu\nu_\tau)$ using Asimov data and the fit with the three channels. These contours represent a confidence region corresponding to a CL of 0.68, 0.95, and 0.99. The contours show the correlation between the two parameters, which is equivalent to 0.76. This correlation results in a substantial reduction in systematic uncertainties when calculating $R_\mu = \frac{\mathcal{B}(\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau)}{\mathcal{B}(\tau^- \to e^- \bar{\nu}_e \nu_\tau)}$.

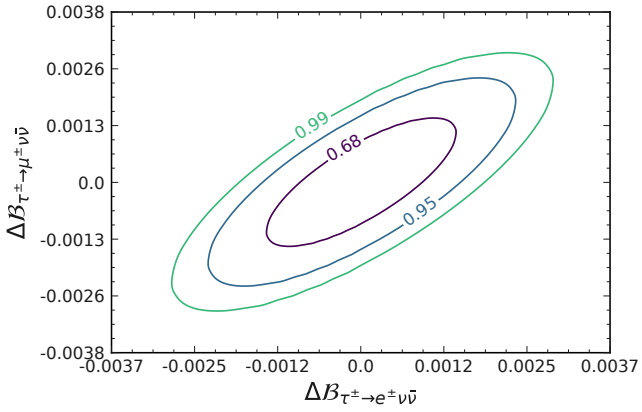


Figure 5.29: Contour plot of the two-dimensional likelihood profile when scanning both $\mathcal{B}(\tau \to \mu\nu_\mu\nu_\tau)$ and $\mathcal{B}(\tau \to e\nu_e\nu_\tau)$. The shown contours represent a confidence region corresponding to a CL of 0.68, 0.95, and 0.99 and the correlation between the parameters is 0.76.

5.3.5 Pull plot

The pulls of the nuisance parameters for the fit with the two channels are shown in Figure 5.30. These were obtained using the model with the random shift in the likelihood function for R_μ, but this does not affect the fit results for the nuisance parameters. The pull values are calculated as

$$\text{pull value} = \frac{\hat{\theta} - \theta_0}{\Delta\theta}, \tag{5.7}$$

where $\hat{\theta}$ is the MLE of the nuisance parameter, θ_0 is the expected value and $\Delta\theta$ is the width of the constraint term c_χ in Table 4.19. In Figure 5.30, the error bars associated with the pull values reflect the post-fit uncertainties of the nuisance parameters. The names of the nuisance parameters were already broken down in Table 4.26. No strong pulls of any nuisance parameters away from their expected values are observed. Only one parameter has a pull value slightly more than $\Delta\theta$ away from the expected. This is the parameter `lid_p_part_10_mu_channel_mu_06`, which models the muon efficiency uncertainty in the momentum bin ranging from 3 GeV to 3.5 GeV, which is the last entry in the legend of Figure 4.33 (the pink component). This means that for the templates with true muons in the μ-channel, the nominal event rates are reduced by about 0.5 %. This is not concerning since a 1σ effect can be anticipated, and the uncertainties related to muon efficiency corrections are solely driven by their statistical components, which is especially high in this specific bin. This can also be seen from Figure 3.16 in Section 3.6.6. One might expect more nuisance parameters to show an effect larger than 1σ (32 % of 194, so around 62) if they are normally distributed. The reason for this is that, in some cases, the provided uncertainties are chosen very conservatively, which allows the fit model to follow the data very well (see also Section 5.3.7, where a goodness-of-fit test is discussed). This impacts the pulls of all nuisance parameters since it will often be preferable (in terms of the change in the likelihood value) to vary the shape of the templates using these underconstrained nuisance parameters. The extreme case would be unconstrained nuisance parameters that could freely change the shape of all templates. Then, all the pull values would be 0, and no information about the nuisance parameters could be gained from the fit (the post-fit uncertainty would be $\Delta\theta$). Another indication of having conservative constraints is the size of the error bars in Figure 5.30. They are close to 1 when the fit cannot further constrain the corresponding nuisance parameter compared to what was initially provided. In contrast, when they are smaller than 1, the fit provides enough information to obtain a more precise parameter estimate. This could be either because the fit is very sensitive to the effect modelled by the nuisance parameter or, especially when the pull values are close to 0, the uncertainties were overestimated. In Figure 5.30, this is particularly striking for the global shape parameters, indicating that the value obtained in Section 4.6.8 for the corresponding uncertainty was very conservative. This is expected to have little impact on the fitted value of R_μ and its uncertainties. This will be shown explicitly and discussed further in Section 5.3.7.

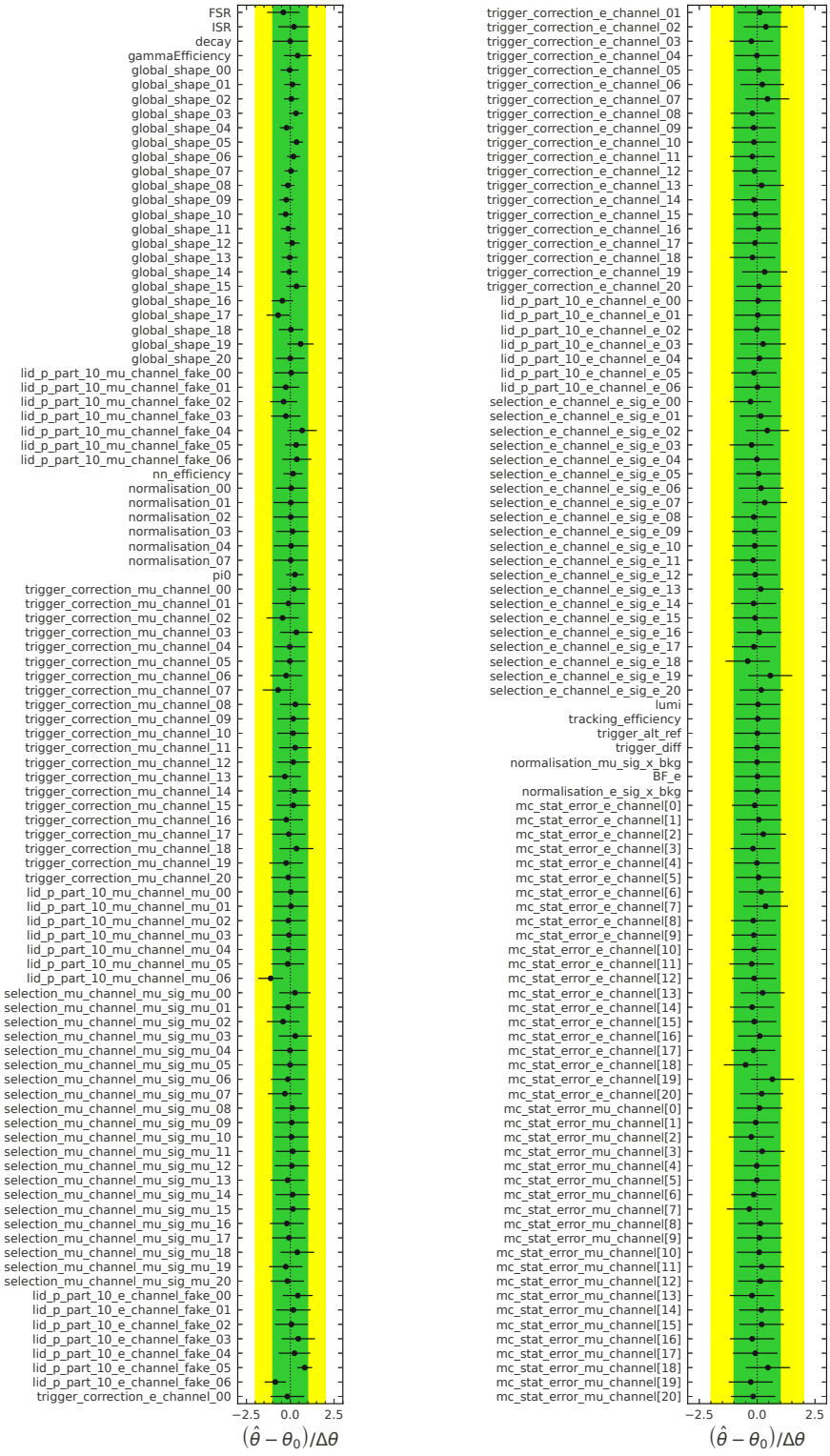


Figure 5.30: The pull plot for the 2-channel fit, showing all nuisance parameters with names as described in Table 4.26. The black marker indicates the pull value, and the error bar shows the post-fit uncertainties of the nuisance parameters. None of the parameters has a pull value that deviates significantly from zero.

5.3.6 Correlations

In Figures 5.31, 5.32 and 5.33, the correlations between some selected fit parameters are shown. It is not feasible to show all correlations between the 195 parameters, but the most important ones are included.

Figure 5.31 shows the parameters that have at least a correlation above 0.25 or below −0.25 with any other parameter. As already mentioned in Section 4.6.10, the nuisance parameter associated with scaling the luminosity is anti-correlated to the parameter that models the π^0 efficiency correction uncertainty and some of the normalisation effects, which also results in a correlation of −0.03 with R_μ.

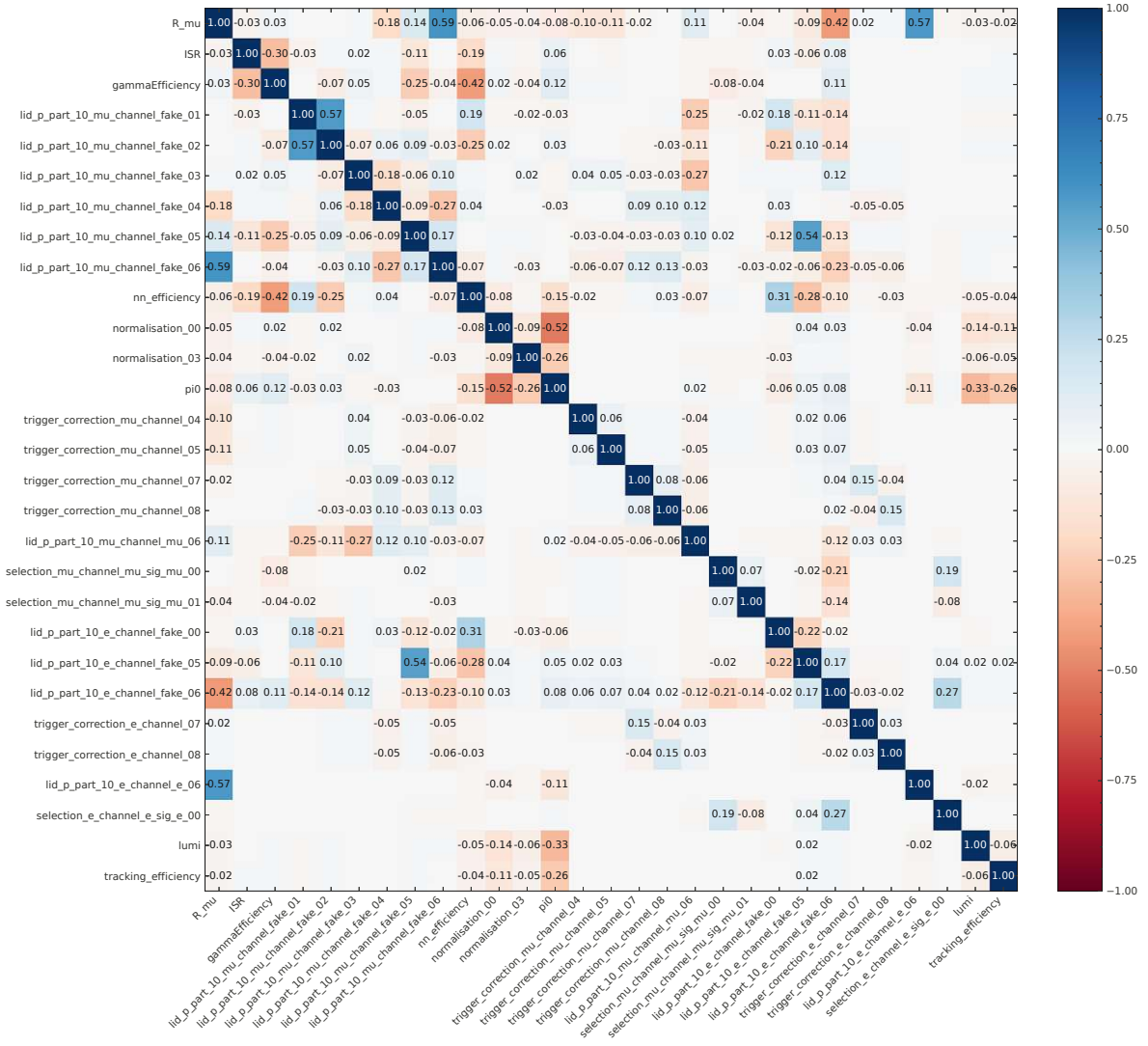


Figure 5.31: The correlation of some selected fit parameters with correlations above 0.25 or below −0.25 with any other parameter.

Figure 5.32 shows the correlation between the global shape nuisance parameters, where one parameter scales the yields in one momentum bin, correlated for all templates. Here, a peculiar correlation structure arises, with more significant correlations in blocks of three momentum

bins and larger blocks in multiples of three. This arises from the fact that lepton ID corrections and associated uncertainties are provided in steps of 0.5 GeV. This makes the variations of the templates due to lepton ID effects correlated in steps of three templates bins, as it was shown in Figure 4.33, which in turn results in a higher correlation between the global shape parameters. Some additional structure originating from the lepton ID correlations, previously shown in Figure 4.32, can also be seen. For example, the correlation of the mis-id rate correction uncertainties ($h^{\pm} \to \mu^{\pm}$) over the first 15 momentum bins is clearly visible.

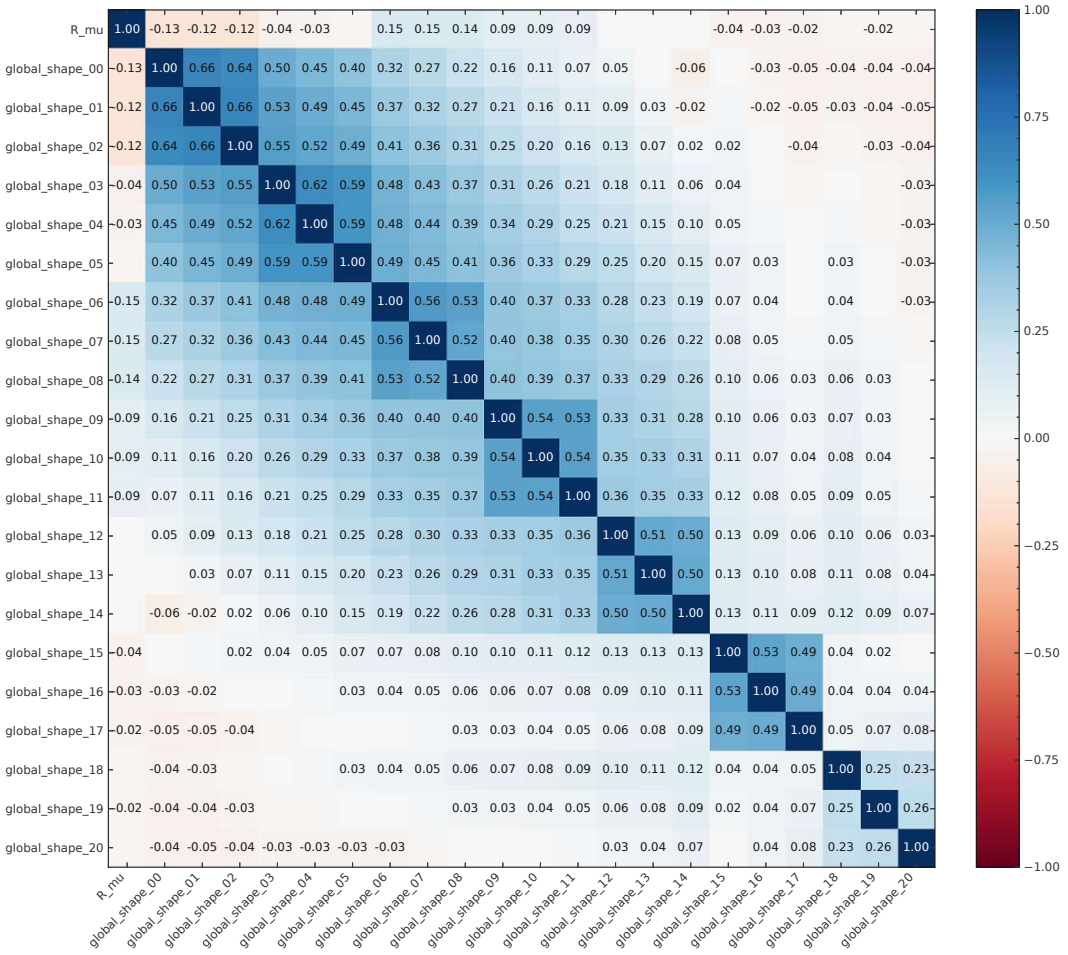


Figure 5.32: The correlation between the global shape nuisance parameters.

Figure 5.33 shows the parameters most correlated with R_μ, which generally indicates a more significant impact on the corresponding systematic effects on the result. Unsurprisingly, these are primarily parameters associated with lepton ID correction uncertainties but also FSR, global shape, and the normalisation applied to the μ-channel due to the difference between the integrated trigger efficiency correction factors in the e and μ-channels (`trigger_diff`). It is important to note that the correlation of, for example, the global shape parameters with R_μ originates from their correlation with the lepton ID parameters. Similarly to the luminosity uncertainty that does not completely cancel for R_μ, the global shape uncertainty also contributes to the overall uncertainty on R_μ, simply because it makes the model more flexible. However, one has to keep in mind that the size of this effect strongly depends on the

other systematic effects included in the model and the correlation between them.

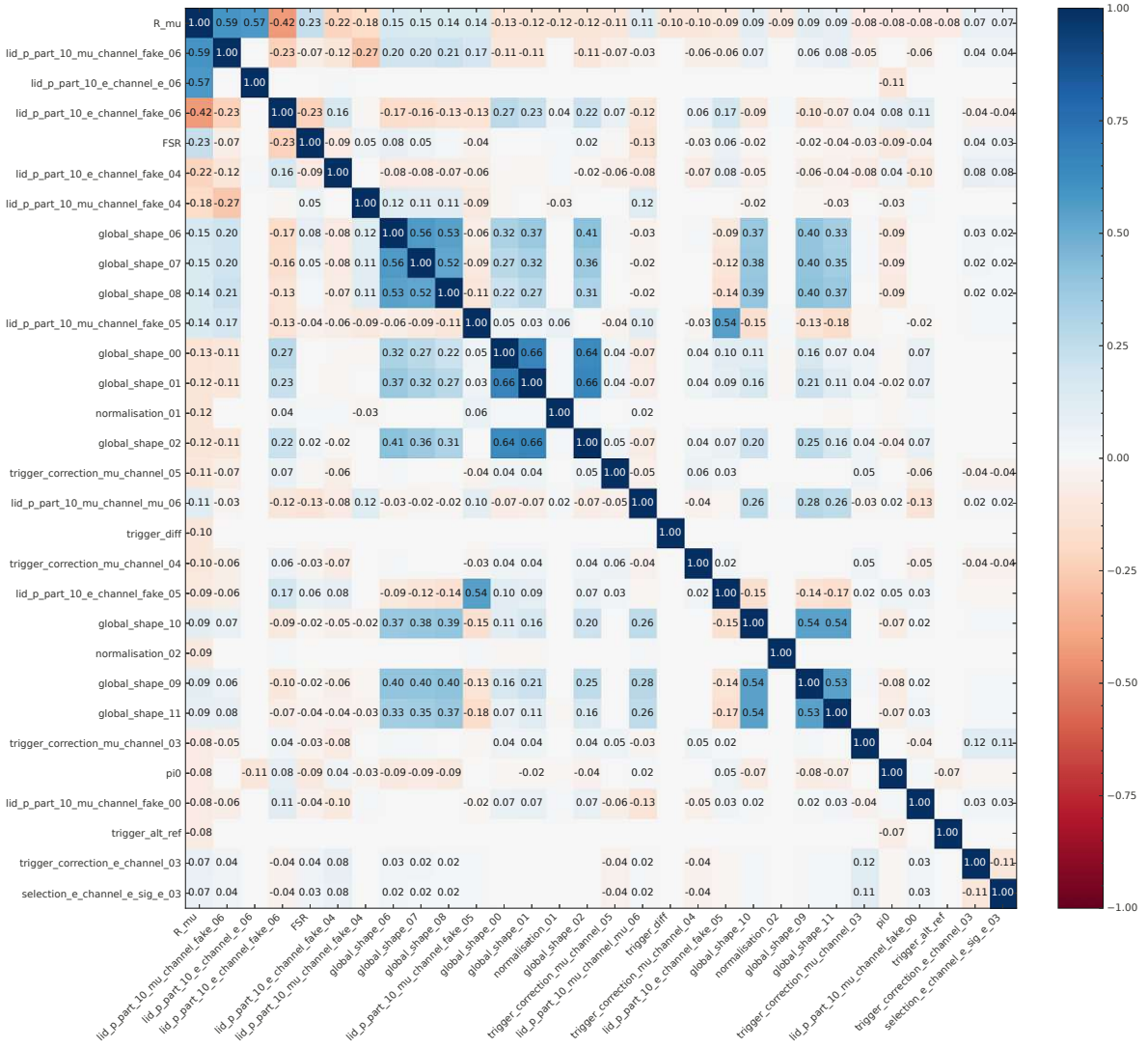


Figure 5.33: The correlation of some selected fit parameters highly correlated with R_μ.

5.3.7 Goodness-of-fit

A goodness-of-fit test is performed based on the difference between the maximum of the likelihood function and the likelihood value using a 'saturated model', referring to the model which follows data perfectly [151], which can be achieved by setting all parameters equal to the data in the likelihood function. The goodness-of-fit test is then based on the assumption that the saturated model L_{sat} reflects the true underlying model, in which case $-2\log(L/L_{\text{sat}})$ follows a χ^2 distribution with as many degrees of freedom as there are free parameters in the fit, which is just one in the present case (see Section 2.5.2). Using the maximum of the likelihood function L, a p-value of 0.9987 is obtained, indicating an overestimation of uncertainties that allow for the variations of the template yields. Since, for many systematic effects, a rather conservative estimate was used, this was to be expected. However, this does not necessarily

imply that the uncertainty on R_μ is inflated since the systematic variations that give the most freedom to the template shapes are subdominant for R_μ because they are correlated across the channels and largely cancel for the ratio of branching fractions.

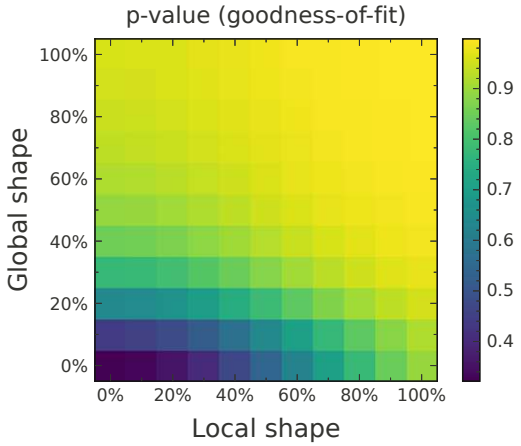


Figure 5.34: The p-value from a goodness-of-fit test for different sizes of the uncertainties associated with the independent shape variations. 'Local shape' refers to independent uncertainties across the templates, while 'Global shape' refers to uncertainties correlated across templates. The p-value ranges from about 0.32 % to 99.87 %.

The reason that the model can fit the data very well is mainly attributed to the systematics that impact the shape of the templates independently across momentum, meaning that for each momentum bin, there is a nuisance parameter associated with it. Two types of parameters can be distinguished: one that is entirely independent between the templates (trigger and tag side modelling) and another that is correlated between templates, which are mainly the parameters associated with modelling the momentum distribution. To test the impact of reducing the uncertainties associated with these effects, the fit was repeated with the size of the constraints of both types varying between 0 % and 100 % of the original size. It was found that the p-value indeed decreases when tighter constraints are used. This is shown in Figure 5.34, with 'Local shape' referring to the fully independent uncertainties and 'Global shape' referring to the ones correlated across templates. The p-value decreases to about 0.32 % when both effects are reduced to 0 %. At the same time, the central value of R_μ is very stable and its uncertainty is reduced only slightly. This can be seen in Figure 5.35: On the left, the shift of the estimated value of R_μ relative to the nominal measurement is shown, while the corresponding uncertainty is shown on the right. The most significant shifts of R_μ are approximately 4×10^{-4} when only one of the two effects is reduced while the other is kept at the nominal level. When both effects are reduced to 0 %, the uncertainty is reduced by approximately 10 %. Interestingly, in that case, the central value is hardly affected.

The impact on R_μ is more clearly visible in Figure 5.36. It shows the measured values for all cases together with their respective uncertainties. The horizontal dashed line indicates the value obtained from the nominal fit, while the dotted lines represent the corresponding $\pm 1\sigma$ uncertainty range. The estimates of R_μ differ only slightly from the nominal value, sometimes being larger and sometimes smaller. However, their error bars are almost always within the $\pm 1\sigma$ range of the nominal fit. Given that for some of these cases the uncertainty of R_μ will be underestimated, the nominal measurement is favoured given the lack of knowledge

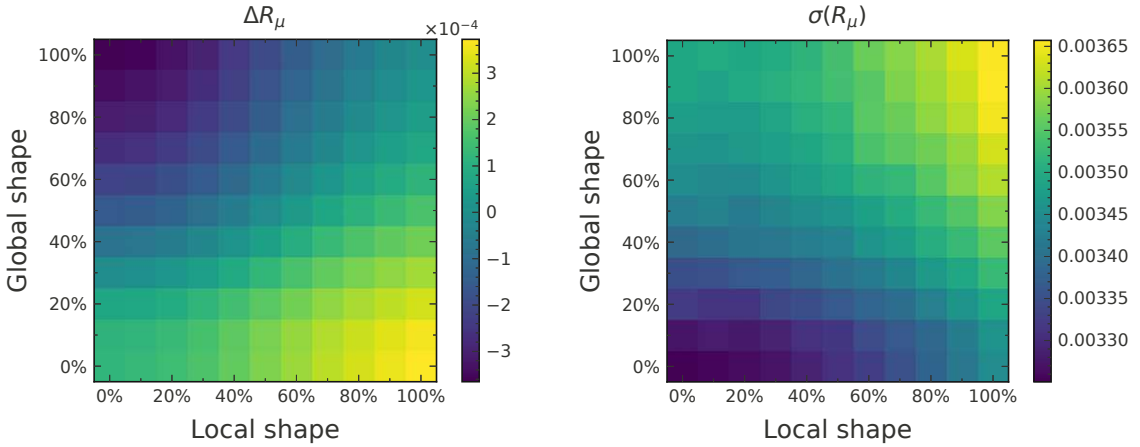


Figure 5.35: The difference between the fitted values of R_μ and the nominal result (left) and the associated uncertainties (right) for different sizes of the implemented independent shape variations.

about the actual uncertainties and their correlation. However, the uncertainty of the nominal measurement could potentially exceed 68.3 % coverage, which is still preferred considering the alternative. In the end, there is no indication that the conservative choices of the uncertainties significantly impact the result, and it remains valid.

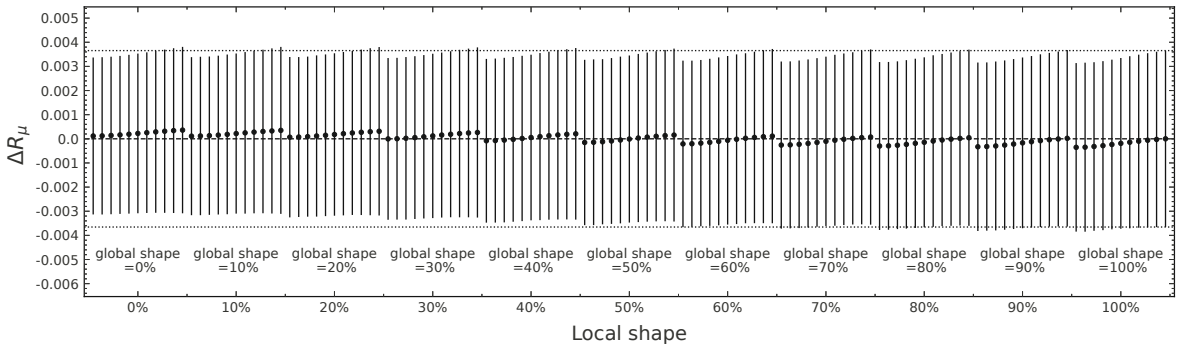


Figure 5.36: The measurements of R_μ for different sizes of the uncertainties associated with the independent shape variations. The y-axis shows the difference with respect to the nominal result, indicated by the horizontal dashed line. The horizontal dotted lines indicate the $\pm 1\sigma$ uncertainty range from the nominal result.

5.4 Data validation

In Figure 5.37, the data/MC agreement in the electron channel (left) and the muon channel (right) is shown for the missing momentum, the thrust, the visible energy in the CM, and the output of the ANN. The entire dataset is used, but to not be sensitive to differences in normalisation that would unblind R_μ, the MC distributions are scaled to match the total number of data events. There are no concerning discrepancies observed between data and

MC. In particular, the observed discrepancies are equally present in the e-channel and the μ-channel.

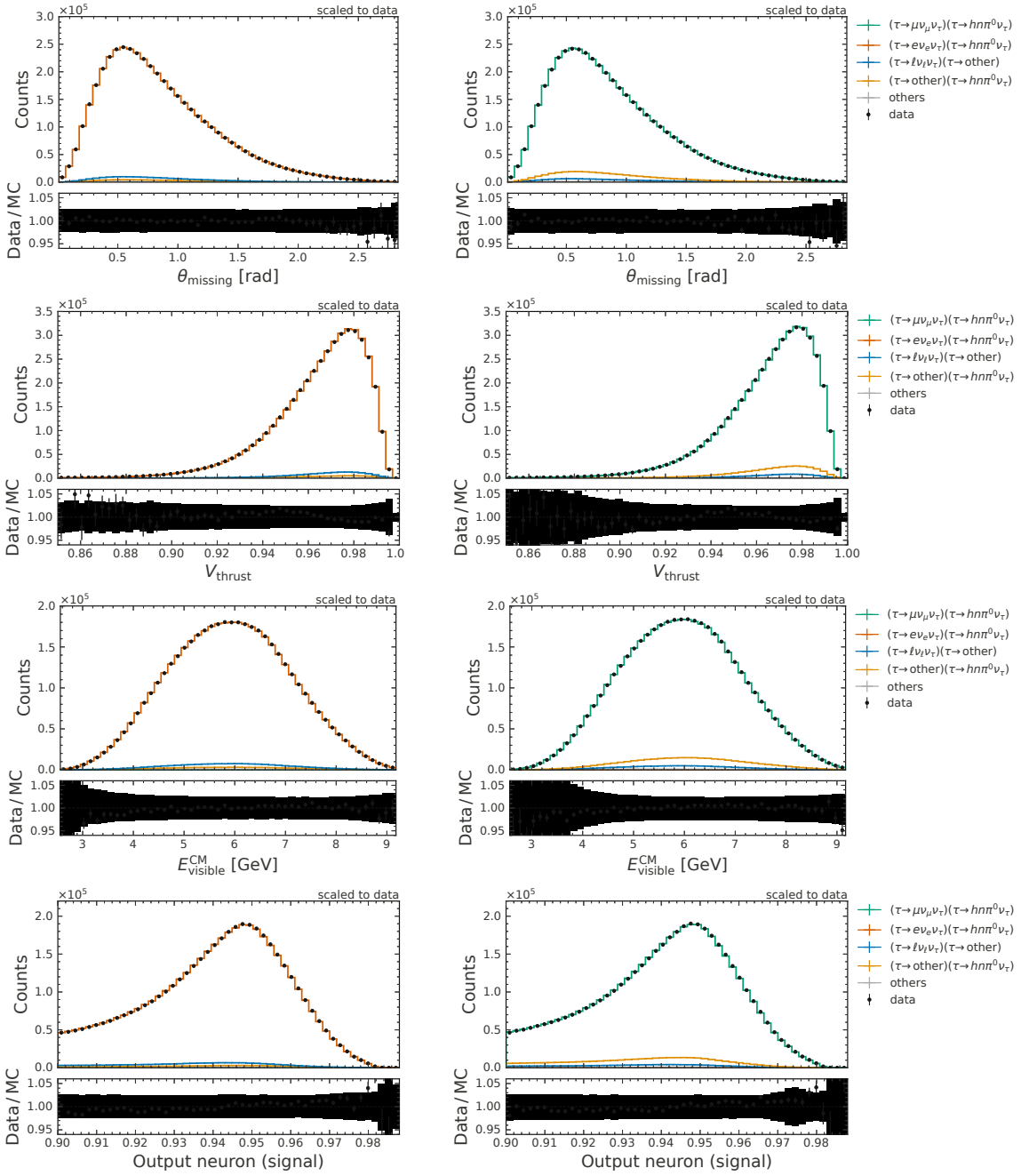


Figure 5.37: The data/MC agreement in the electron channel (left) and the muon channel (right) for the missing momentum, the thrust, the visible energy in the CM and the output of the ANN.

In Figure 5.38, the data/MC agreement in the electron channel (left) and the muon channel (right) is shown for the lepton candidate momentum and polar angle. The data/MC ratio in the momentum distribution is flat for the muon channel, but a slight trend can be observed

in the electron channel, with more data for lower electron momenta and less data for higher electron momenta. This is presumably related to photon efficiency corrections associated with FSR photons, which the included systematic effects cover. Looking at the data/MC ratio in the polar angle, some effect is observed in the muon channel for high theta. As discussed in Section 5.2.5, this originates from changes in muon efficiency and mis-id rates towards the end of the KLM barrel, which is not well reproduced in MC. Provided that the polar angle distribution in the calibration channels used to derive the muon efficiency corrections is similar to the polar angle distribution in the measurement, those discrepancies are corrected when integrated over the whole correction bin. This can be seen when comparing data and MC with the binning in which the corrections are provided, shown in Figure 5.39. It is observed that the overall size of the data/MC differences in the last bin is reduced. Additionally, an overall trend is observed in the data/MC ratio of the polar angle distributions, with less data toward the low and high theta regions. Although it is unclear what is causing this effect, it probably has a common origin for the e-channel and the μ-channel (such as the tag side) since it is present in both channels. This makes the measurement of R_μ insensitive to it.

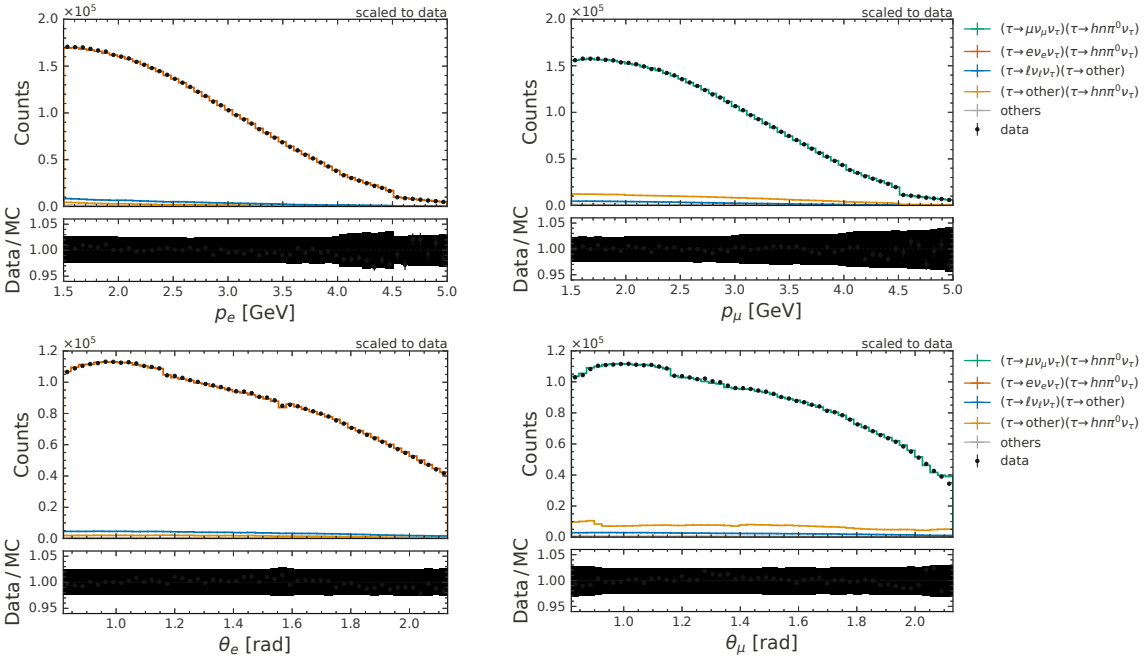


Figure 5.38: The data/MC agreement in the electron channel (left) and the muon channel (right) for the momentum and polar angle of the lepton candidate.

5.5 Stability checks

In this section, several checks are performed to test the stability of the fit result for different conditions. Only the difference between the nominal estimate and the estimate obtained with the various checks will be compared. This was done to avoid being influenced by the observer-expectancy effect by exposure to the actual value. In addition, to avoid doing this accidentally, the estimates of the parameter of interest will be obtained with the random shift included in the likelihood, which is the same for all checks and thus cancels for the difference. Furthermore, all the checks presented in this section were first performed using only a random

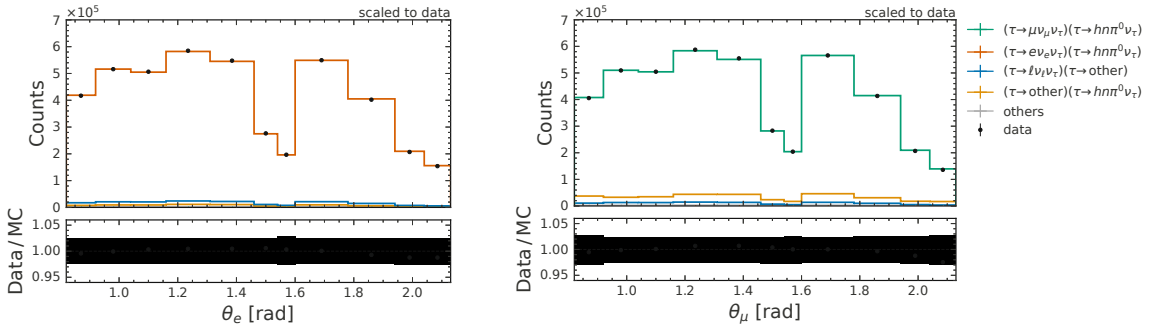


Figure 5.39: The data/MC agreement in the electron channel (left) and the muon channel (right) for the polar angle of the lepton candidate, following the binning used for the electron and muon efficiency corrections.

10 % of the data, to ensure that there are no major issues before proceeding with the entire dataset.

In the following figures, $\Delta R_\mu \equiv R_\mu^{\text{shifted}}\,(\text{check}) - R_\mu^{\text{shifted}}\,(\text{nominal})$ will be shown, where $R_\mu^{\text{shifted}}\,(\text{check})$ is the estimated value for the indicated check and $R_\mu^{\text{shifted}}\,(\text{nominal})$ is the estimate with the nominal conditions. The error bar will indicate the total uncertainty of R_μ^{shifted}, as obtained from the fit (combined statistical and systematic uncertainty). In addition, horizontal error caps indicate the part of the total uncertainty that originates from lepton ID uncertainties, while transparent solid error bars indicate statistical uncertainties.

5.5.1 Split samples (independent)

The stability of R_μ for different sample splits is shown in Figure 5.40. The nominal result is compared to results obtained with samples of restricted lepton momentum (larger and smaller than 3 GeV, as well as split at 2.5 GeV and 3.5 GeV), polar angle (larger and smaller than 1.31), negative and positive charge, and with the polar angle of the missing momentum vector larger and smaller than 0.8. Some of these splits do not have any events in common and are therefore statistically independent, as indicated by the vertical dashed lines in Figure 5.40. Although for those the statistical uncertainties are independent, the systematic components could still be correlated.

The only split that is incompatible when only statistical uncertainties are considered is the split into low and high momentum, separated at 3 GeV. However, since momentum is used as the fit dimension, the correlations of the most dominant systematic uncertainties in the fit model are already known. In particular, the mis-id rate uncertainties are independent for high and low momentum, since for high momentum, their statistical component dominates due to a low number of events in the relevant calibration samples, such as it is the case for the tau channel using $\tau^\pm \to 3\pi^\pm\nu$ decays, which was briefly discussed in Section 3.6.3. In addition, as discussed in Section 3.6.6, the $\mu\mu\gamma$ calibration channel requires that at least one of the muon candidates has momentum below 3 GeV, which results in increased statistical uncertainties just above 3 GeV. Furthermore, the BDT which is used for electron identification was trained independently for momenta below and above 3 GeV which can give rise to independent systematic effects. No trend is observed when performing a different split in momentum, such as those separated at 2.5 GeV and 3.5 GeV.

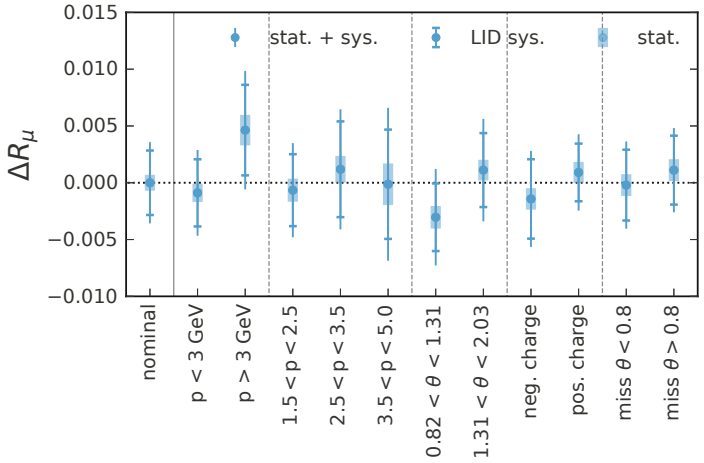


Figure 5.40: Stability of R_μ for statistically independent split samples, indicated by the vertical dashed lines. The horizontal error caps indicate the part of the total uncertainty originating from lepton ID uncertainties, while transparent solid error bars indicate the statistical uncertainty.

5.5.2 Split samples (overlapping)

The stability of R_μ for different samples that partially contain the same events is shown in Figure 5.41. Because the samples overlap, the statistical uncertainties are not independent and similar values are expected to be measured for R_μ. The nominal value is compared to the values obtained with different cuts on P_e and P_μ, to a fit with different assumptions on the correlation of the systematic components of the lepton ID uncertainties (0.5 and 0.0, contrary to the nominal value of 1.0), to a fit with increased luminosity uncertainty from 0.6 % to 10 %, to fits with different bin sizes, where the number of splits of the 0.5 GeV lepton ID bins were varied from 1 to 4, and to fits with varying the amount of ISR photons by 10 % to 20 % (contrary to the nominal value of 5 %). Very good consistency is observed with the nominal value for all checks, and the checks with increased luminosity and ISR uncertainties do not show larger uncertainties on R_μ, which is an important check of the expectation that they cancel. The results with different lepton ID correlations show smaller uncertainties, which is expected from the studies shown in Section 5.3.2. The checks with different lepton ID requirements are further discussed in the following sections.

5.5.3 Lepton identification requirements

The stability of R_μ for different tightness of selections on P_e is shown in Figure 5.42, additionally also split into the different subsets already shown in the previous sections. The compared cuts on P_e are 0.5, 0.9, 0.95, and 0.99, since the lepton ID corrections were only available for these values. While slight variations are observed for some subsets, this is well within the uncertainties, which are dominated by the electron efficiency uncertainty for the tighter cuts. The uncertainty increases significantly with tighter cuts on P_e, indicative of the larger deviations from the nominal value when applying tighter cuts. This originates from a discrepancy in the Bhabha efficiency correction factors between different events selected by the HLT, which are more significant for tighter cuts and for which a systematic uncertainty is assigned (see Section 3.6.3). The cut at 0.5 results in the smallest uncertainties and is most robust to

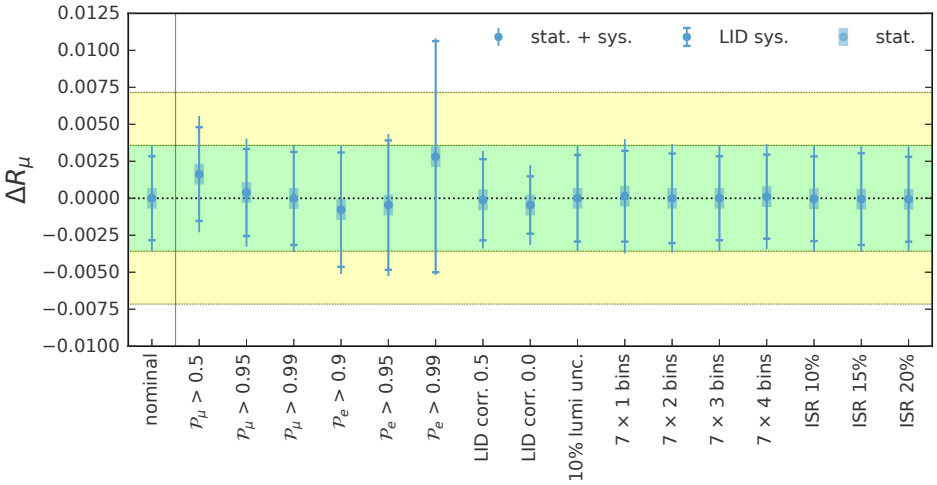


Figure 5.41: The stability of R_μ for different partially overlapping samples. The horizontal error caps indicate the part of the total uncertainty originating from lepton ID uncertainties, while transparent solid error bars indicate the statistical uncertainty.

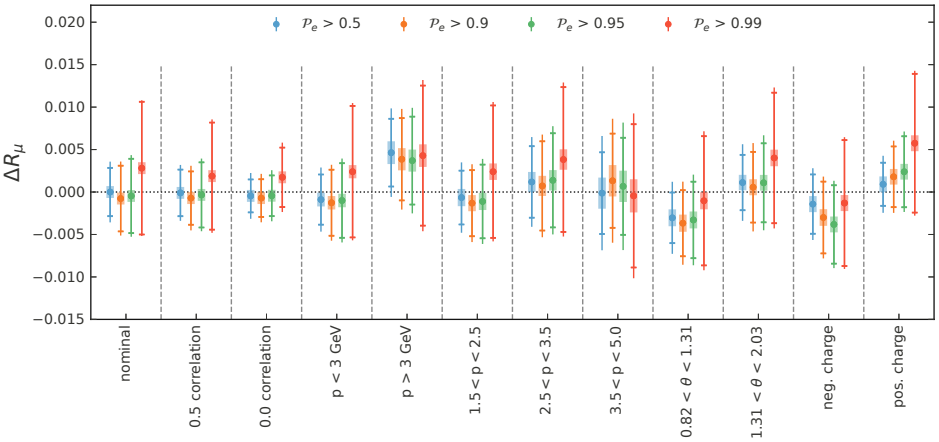


Figure 5.42: Stability checks with different selections on P_e. The horizontal error caps indicate the part of the total uncertainty originating from lepton ID uncertainties, while transparent solid error bars indicate the statistical uncertainty.

systematic effects, which is why it was chosen in the first place. Figure 5.43 shows the compatibility for each of the obtained R_μ values with the one obtained with the nominal cut on P_e (0.5). It gives the consistency for different assumptions on the correlation of the electron efficiency uncertainties between the different working points, shown as coloured markers. The assumed correlation ranges from 0.0 to 1.0, since the true correlation is unknown and cannot be easily obtained. For this check, all values are consistent with any assumed correlation of the uncertainties.

In Figure 5.44, the same checks with different cuts on P_μ are shown, which show very stable behaviour except for slightly larger values with a cut at 0.5. Figure 5.45 shows the compatibility of each of the obtained R_μ values with the one obtained with the nominal cut on P_μ (0.9). The compatibility depends on the correlation of the muon efficiency and mis-id rate uncertainties

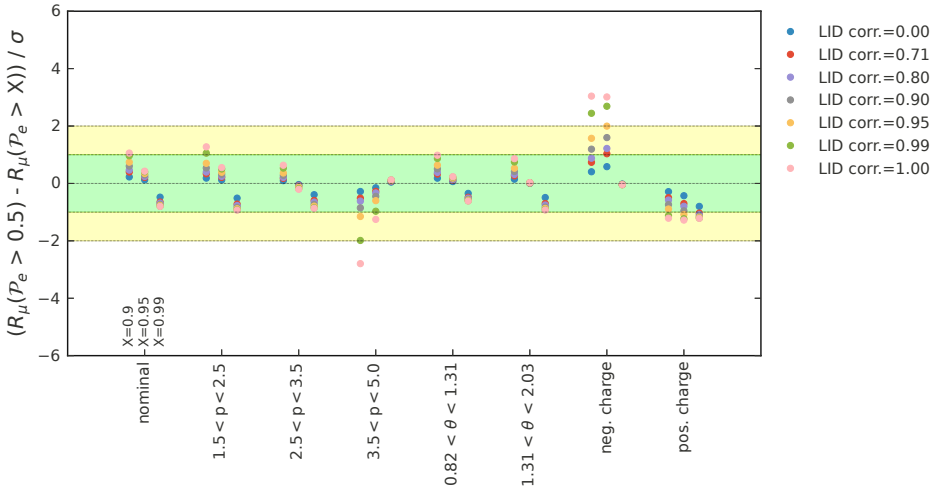


Figure 5.43: The compatibility between the observed values when changing the cut on P_e. The statistical uncertainty on R_μ is taken to be fully correlated, while different correlations are tested for the lepton ID uncertainty.

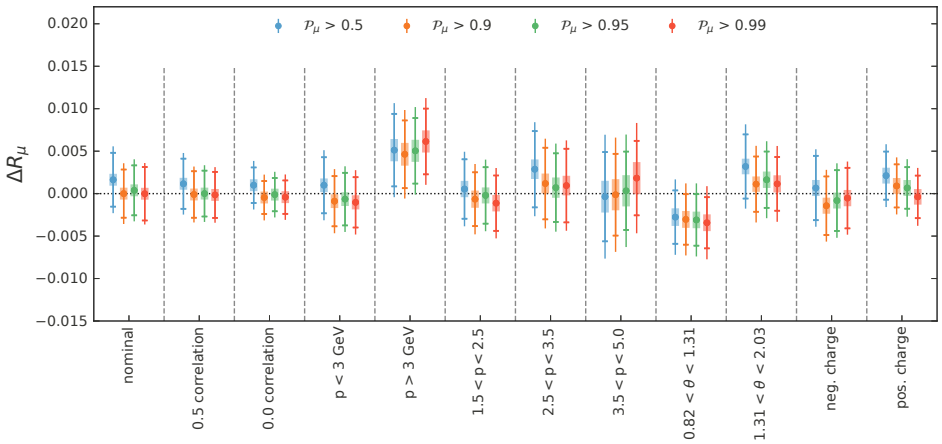


Figure 5.44: Stability checks with different selections on P_μ. The horizontal error caps indicate the part of the total uncertainty that comes from lepton ID uncertainties.

between the different cut values. If the uncertainties are highly correlated, there is some tension between the obtained values, but a correlation below 0.9 would already make them compatible. In the following, the correlation of the statistical uncertainty of the corrections will be estimated.

5.5.4 Correlations of muon identification uncertainties

Using the $e^+e^- \to \mu^+\mu^-\gamma$ sample which was used to calibrate the muon efficiency corrections, the correlations of the statistical component of the muon identification uncertainties between different cuts on P_μ can be estimated. This is done with an integrated sample that restricts the probe muon to the same momentum and polar angle region used in the analysis. First, the integrated muon identification efficiency is computed in data and MC for different cuts on P_μ, shown in Figure 5.46a. Then, the ratio of the obtained efficiencies is taken while propagating

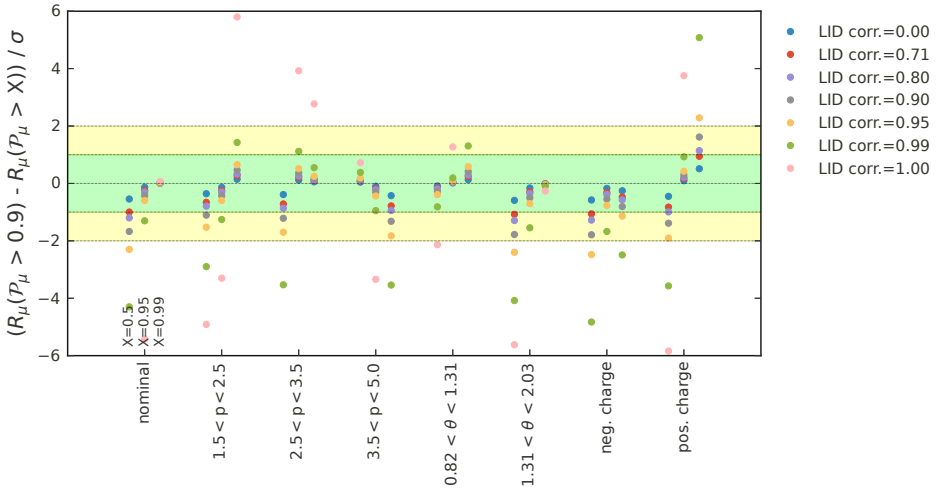
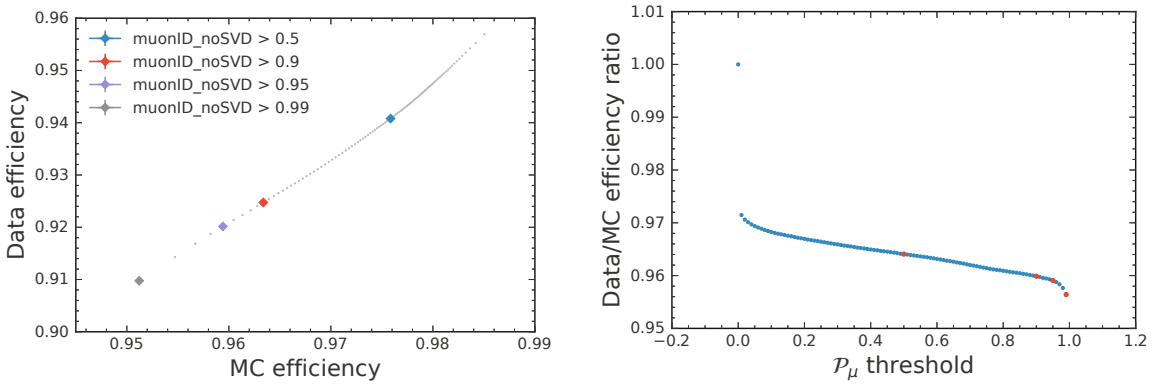


Figure 5.45: The compatibility between the observed values when changing the cut on P_μ. The statistical uncertainty on R_μ is taken to be fully correlated, while different correlations are tested for the lepton ID uncertainty.

the statistical uncertainty and keeping information about the correlations. The resulting ratio is shown in Figure 5.46b. In Figure 5.47, the correlation of the uncertainties between any two



(a) (b)

Figure 5.46: The muon identification efficiencies in data and MC (a), as well as their ratio (b) for different selections on P_μ, in order to estimate the correlations of the associated statistical uncertainties.

cuts on P_μ is shown. As expected, for two close cut values, the correlation is almost one, while very different cuts give a lower correlation. Ultimately, the only correlations of interest are between the cut values $0.5, 0.9, 0.95,$ and 0.99, which are shown in Table 5.4.

The same study is carried out with the pion sample that is used to derive $\pi \to \mu$ mis-id rate correction factors ($\tau^\pm \to 3\pi^\pm\nu$). The efficiencies in data and MC and their ratio are shown in Figure 5.48. In Figure 5.49, the correlation is shown for different cuts on P_μ, and in Table 5.5, the correlation is shown for the relevant working points. The correlation is similar to the one obtained for the muon efficiency corrections.

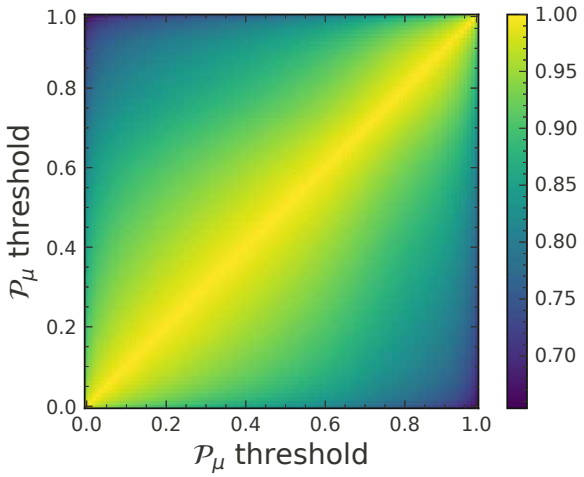


Figure 5.47: The correlation of the statistical uncertainty of the muon efficiency correction factor between different cuts on P_μ.

Table 5.4: Correlations of the statistical component of the muon efficiency uncertainty between different cuts on P_μ, computed using the $\mu\mu\gamma$ calibration channel.

	0.5	0.9	0.95	0.99
0.5	1	0.869	0.839	0.781
0.9	0.869	1	0.965	0.899
0.95	0.839	0.965	1	0.931
0.99	0.781	0.899	0.931	1

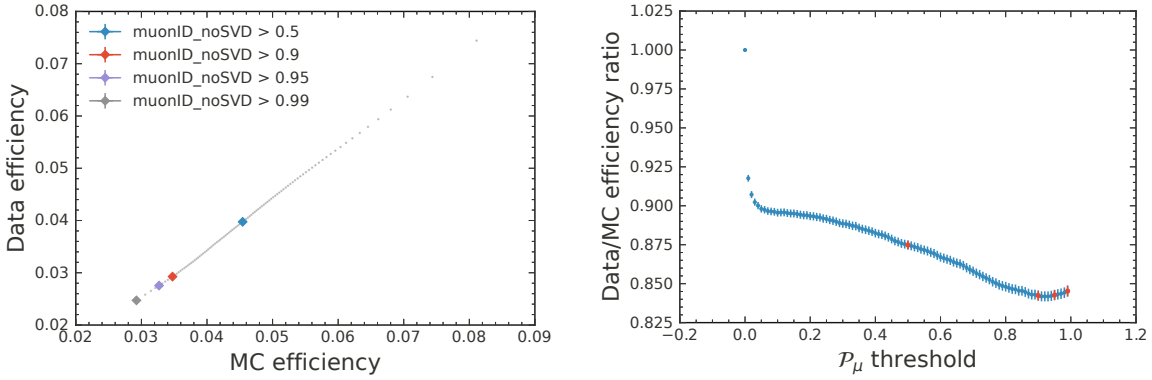
Table 5.5: Correlations of the statistical component of the $\pi \to \mu$ mis-id rate uncertainty between different cuts on P_μ, computed using the $\tau^\pm \to 3\pi^\pm\nu$ calibration channel.

	0.5	0.9	0.95	0.99
0.5	1	0.86	0.834	0.788
0.9	0.86	1	0.969	0.916
0.95	0.834	0.969	1	0.945
0.99	0.788	0.916	0.945	1

Now, the compatibility between the checks shown in Figure 5.44 can be computed using the obtained correlations. This is indicated in Figure 5.50 with a black marker. With this correlation, all checks are consistent with each other within 2σ, suggesting that the small observed differences arise only due to statistical fluctuations.

5.5.5 Polar angle splits

In this section, the impact on R_μ is checked when measured with the lepton in different θ regions. Figure 5.51 shows the stability in four coarse bins, with edges coinciding with edges of the lepton ID correction bins. In Figure 5.52, a denser theta binning is used, corresponding to the lepton ID correction bins defined in Section 3.6.4. For both cases, the observed variation



(a) (b)

Figure 5.48: The $\pi \to \mu$ mis-id rate in data and MC (a), as well as their ratio (b) for different selections on P_μ, in order to estimate the correlations of the associated statistical uncertainties.

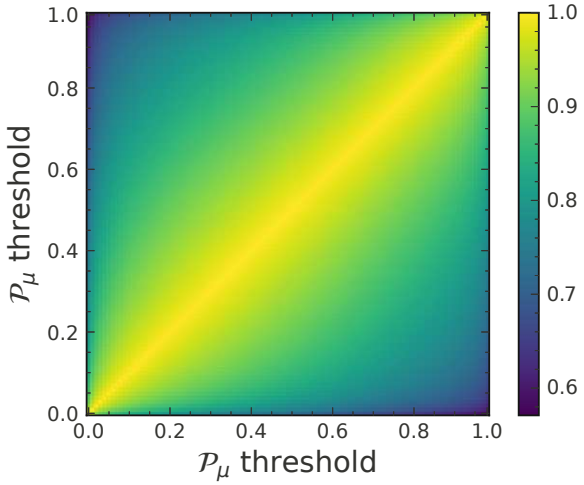


Figure 5.49: The correlation of the statistical uncertainty of the $\pi \to \mu$ correction factor between different cuts on P_μ.

is consistent with statistical fluctuations. A χ^2/ndf of 13.4/10 (p-value: 0.2) is obtained for the dense theta split when only considering the statistical uncertainties. One reason for the observed shift in the last bin could be that it is the region in which the muon efficiency and mis-id rate correction factor changes within the bin, as previously observed. This is also reflected by the larger systematic uncertainty of R_μ when measured in that bin, which originates from lepton ID uncertainties, but also from the normalisation uncertainty (introduced in Section 4.6.9) due to larger data/MC discrepancies in that bin which can also be seen from Figure 5.39 in Section 5.4. However, as is apparent from the same figure, the number of events with the lepton candidate in the range $2.04 < \theta < 2.13$ is small.

5.5.6 ANN cut stability

The stability of R_μ between different cuts on the neural network output is shown in Figure 5.53. The nominal cut on the ANN output is 0.9. The measured value is very stable, with a slight

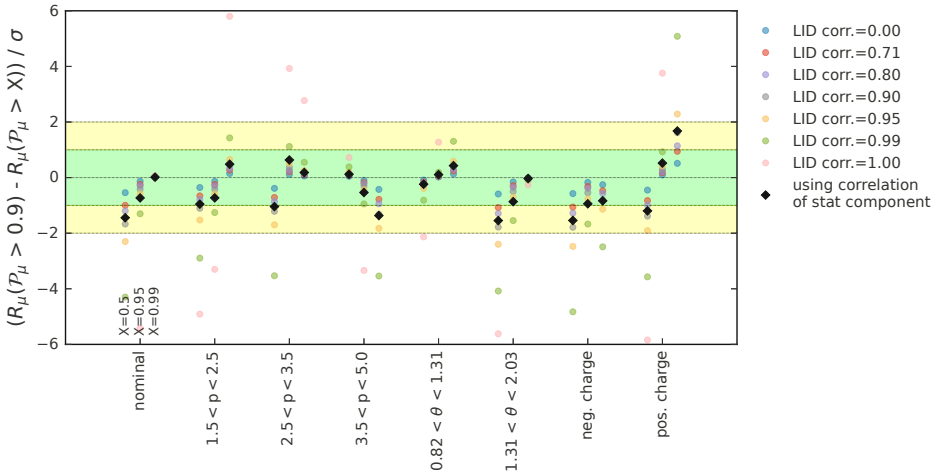


Figure 5.50: The compatibility between the observed values when changing the cut on P_μ. The statistical uncertainty on R_μ is taken to be fully correlated, while different correlations are tested for the lepton ID uncertainty. The black marker indicates the correlation obtained for the muon efficiency uncertainty.

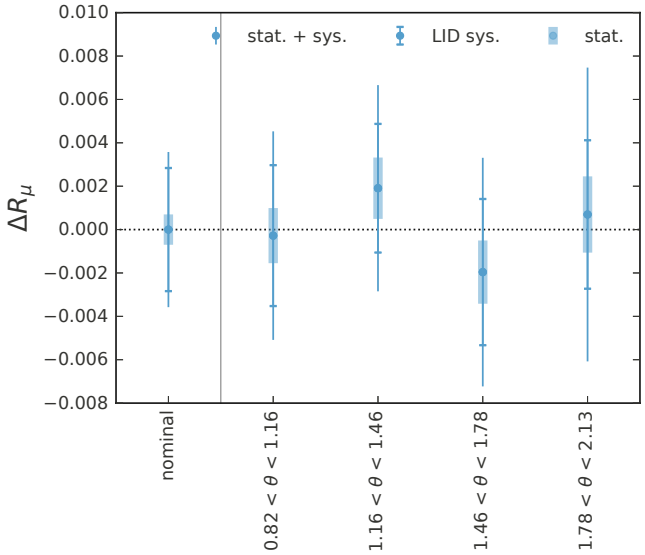


Figure 5.51: The stability of R_μ in four different θ regions, with edges coinciding with edges of the lepton ID correction bins.

increase for tighter cuts on the output. This can be explained by the fact that a significant fraction of signal events are removed, resulting in a substantial statistical uncertainty. For cuts larger than 0.96, there are too many empty template bins, which causes the fit to fail. In Figure 5.53b, the change in total uncertainty is shown as a function of the cut. It shows that the cut at 0.9 is indeed close to optimal. For looser cuts, the uncertainty increases due to larger backgrounds, whereas for tighter cuts, a steep increase in the statistical uncertainty is observed.

The hypothesis that the observed shift for tighter cuts is due to a statistical fluctuation can be

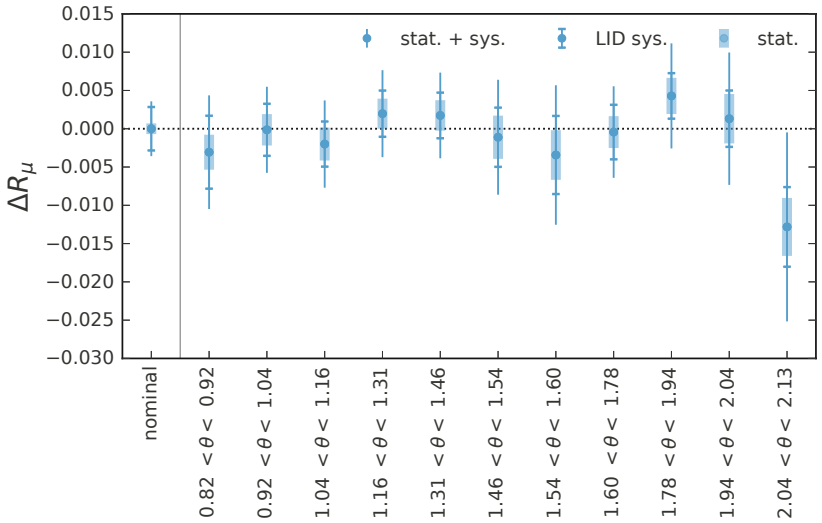
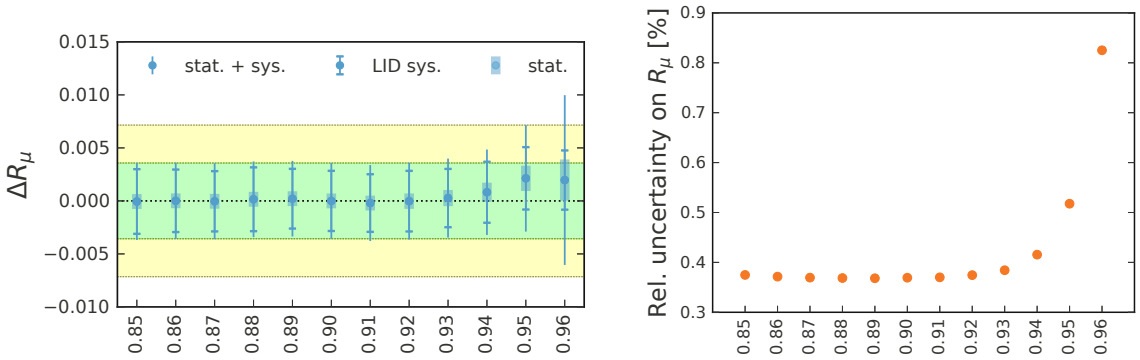


Figure 5.52: The stability of R_μ in eleven different θ regions, corresponding to the lepton ID correction bins.

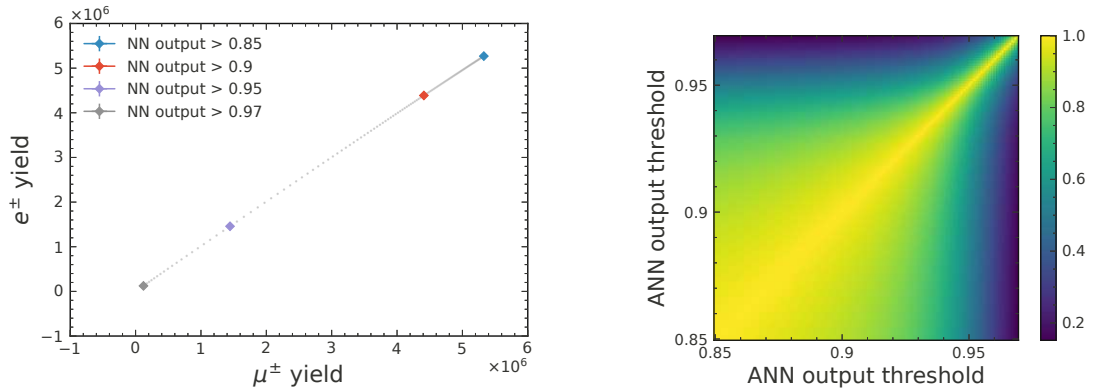


(a) (b)

Figure 5.53: Stability checks with different cuts on the neural network output (a) and the change in the overall uncertainty (b).

tested by computing the correlation of the statistical component of the R_μ values between different cuts on the ANN output. This is similar to what was done for the efficiency corrections in Section 5.5.4. In Figure 5.54a, the electron and muon channel yields in data are shown for different cuts on the ANN output. Taking the ratio of the yields and propagating the statistical uncertainties results in a reasonable estimate of the statistical uncertainty on R_μ. Considering that tighter cuts always select a subset of the looser cuts, the correlation of the uncertainties can be obtained, shown in Figure 5.54b. This matches the expectation that close cuts for lower ANN outputs result in highly correlated statistical uncertainties, whereas a tight cut and a loose cut have largely independent uncertainties.

Using this correlation for the statistical components, the compatibility with different assumptions on the correlation of the systematic component can be checked. This is shown in Figure 5.55 when comparing neighbouring cuts and in Figure 5.56 when comparing always to the nominal cut of 0.9. In all the cases, the measured values are compatible, even when



(a) The yields of the electron and muon channels in data, for different cuts on the ANN output.

(b) The correlation of the statistical uncertainty on R_μ between different cuts of the ANN output.

Figure 5.54: The yields for the e-channel and the μ-channel as a function of the ANN cut, and the such computed correlation between different cuts of the ANN output.

making the most conservative assumption on the correlation of the systematic uncertainties.

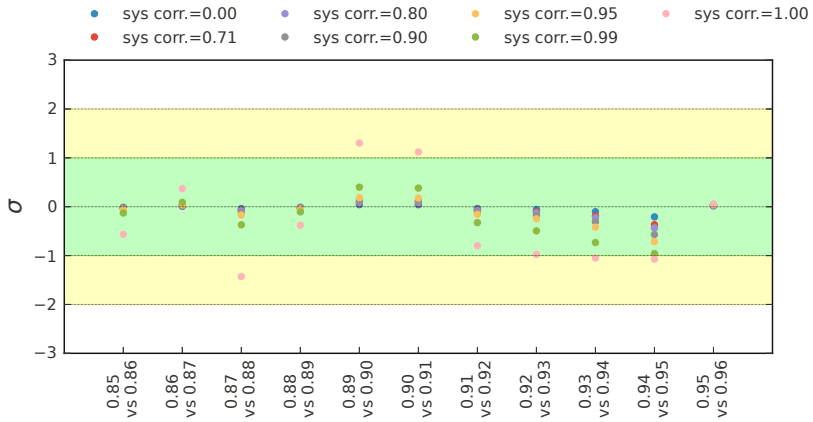


Figure 5.55: The compatibility between the obtained values of R_μ for neighbouring cuts on the ANN output, where the correct correlation was used for the statistical uncertainties. For the correlation of the systematic uncertainties, different values are compared.

5.5.7 Time dependence

Figure 5.57 shows the stability of R_μ for different experiments (see Section 3.5.1 for the corresponding time periods and luminosities). Here, one has to keep in mind that some trend can be expected due to a time dependence of lepton ID efficiencies, and thus also associated correction factors (the conditions in MC do not change). This time dependence mainly occurs because of an increase in beam backgrounds in the later data-taking periods, which were recorded with higher instantaneous luminosity. The data/MC correction factors that were also applied here are only available integrated for the entire dataset. Thus, the purpose of this check is only to ensure that there are no sudden changes in the measured value, which could indicate an issue with some part of the data. In Figure 5.57 only a slight trend towards higher

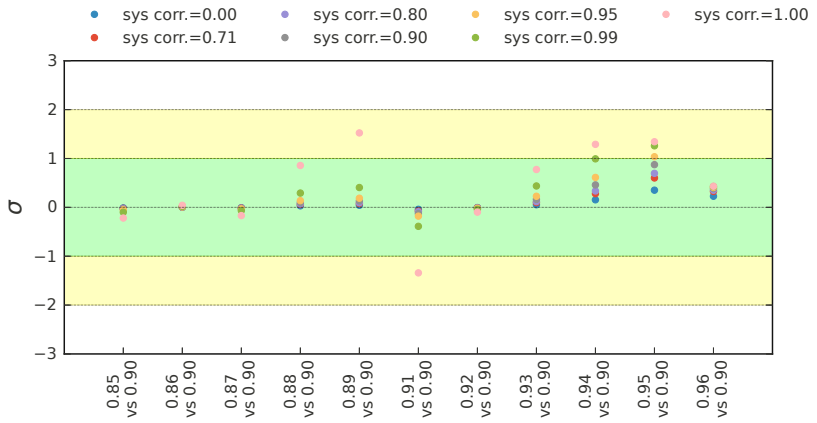


Figure 5.56: The compatibility between the nominal measurement of R_μ and the results obtained with different cuts on the ANN output, where the correct correlation was used for the statistical uncertainties. For the correlation of the systematic uncertainties, different values are compared.

values is observed for later data. This is insignificant and even consistent with statistical fluctuations. When only statistical uncertainties are taken into account, the reduced χ^2/ndf is 7.975/4, corresponding to a p-value of 0.092.

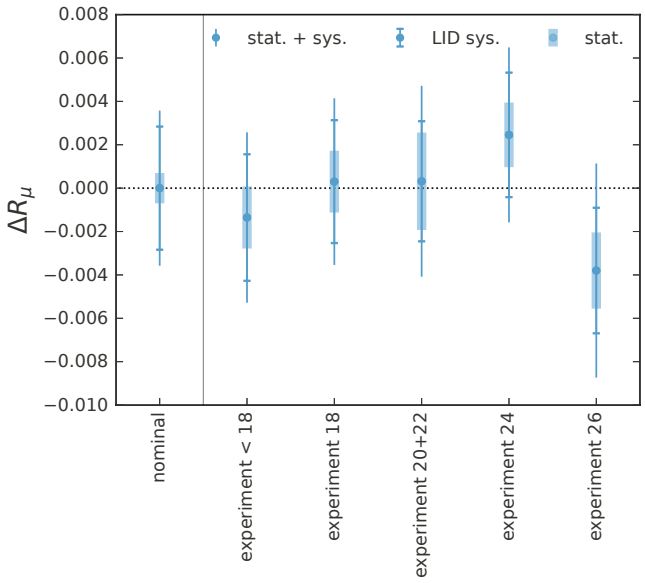


Figure 5.57: The measured value of R_μ for different experiments. A slight trend can be explained by the time dependence of correction factors.

CHAPTER 6

Discussion of the results

6.1 Introduction

This chapter presents the measurement results, their interpretations, limitations, and prospects. All studies presented in the previous sections were carried out while being blind to R_μ and related quantities that could be used to infer the value that will be measured in the data. The fit could be validated and studied without unblinding R_μ by adding a random shift of R_μ to the likelihood function, as mentioned in Section 5.1. The results obtained for the pulls of the nuisance parameters or the correlation of the fitted parameters are equivalent to the fit without the random shift. After ensuring that all checks and fit properties are satisfactory, the next step towards fully unblinding the analysis was to measure R_μ with 10 % of the available data, randomly sampled from the entire dataset. This partial unblinding is presented in Section 6.2.1, followed by the full unblinding in Section 6.2.2 after observing no issues. In Section 6.3, the results are compared with previous experiments; in Section 6.4, the observations are interpreted, and in Section 6.5, some of the limitations with the current measurement are addressed. Finally, Section 6.6 gives a conclusion and outlook.

All numerical results presented in this chapter are rounded to four decimal places, while intermediate calculations were realised using all digits.

6.2 Fit results

6.2.1 Partial unblinding

As a first step of unblinding, only 10 % of the data are used, randomly sampled from the whole data-taking period. The impact of using only 10 % of the available data increases the statistical uncertainty. While for 100 % a statistical (total) relative uncertainty of 0.07 % (0.38 %) can be expected, for 10 % this uncertainty on R_μ becomes 0.22 % (0.44 %). Other options for this partial unblinding that were considered were to use only 1 %, 2 %, or 5 % of the data, resulting in a total precision on R_μ of 0.8 %, 0.62 % or 0.5 %, respectively. While only a part of the data was used to perform the fit, the evaluated systematic uncertainties and variations of the templates were the same as for the measurement with all of the data. The pre-fit distributions

of the electron and muon channels are shown in Figure 6.1, together with the distribution of the random 10 % of the data. The data/MC ratio in the electron channel is consistently higher than in the muon channel, indicating a measured value of R_μ lower than the SM expectation. The post-fit distributions of the electron and muon channels are shown in Figure 6.2.

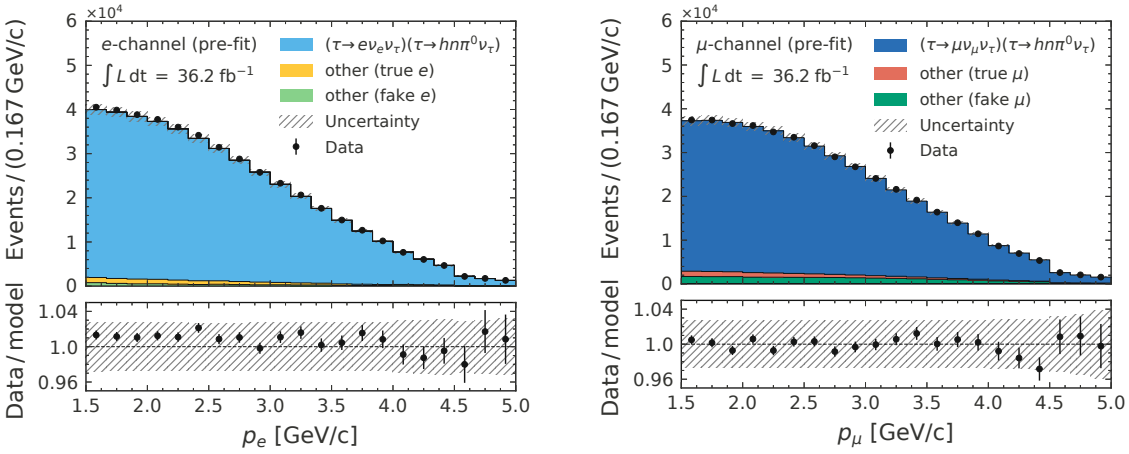


Figure 6.1: The pre-fit momentum distributions for lepton candidates in the e-channel (left) and μ-channel (right), compared to the observed data, using only 10 % of the total available data. The bottom part of the figures displays the ratio of the observed data to the predicted event counts from the fit model. The shaded region represents changes in the event counts based on the systematic variations incorporated in the fit model.

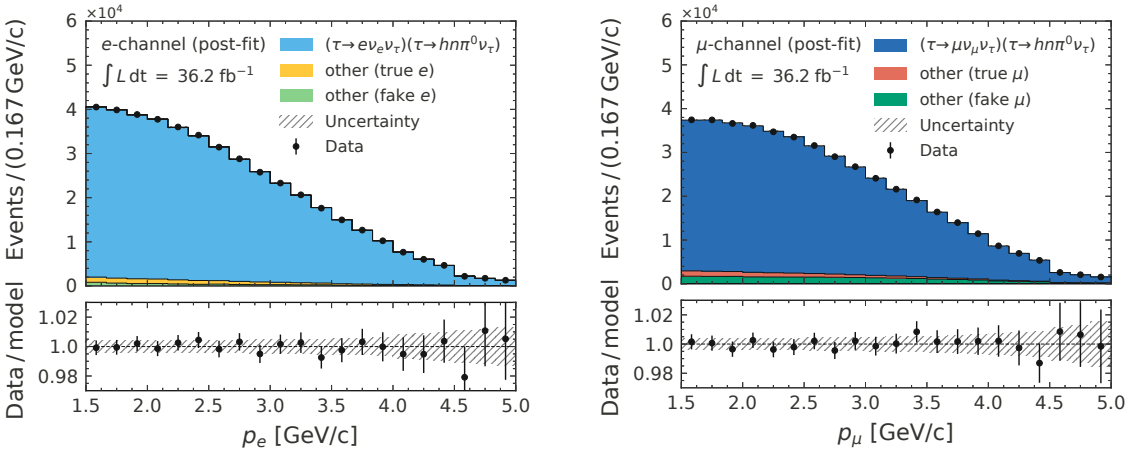


Figure 6.2: The fitted distributions for lepton candidates in the e-channel (left) and μ-channel (right), compared to the observed data, using only 10 % of the total available data. The bottom part of the figures displays the ratio of the observed data to the fitted number of events. The shaded region represents changes in the fitted event counts based on the systematic variations incorporated in the fit model after being constrained by the observed data.

A value of R_μ of 0.9652 ± 0.0043 is obtained from the fit, with a statistical component of the uncertainty of 0.0021 and a systematic component of 0.0038. The measured value is below the SM expectation but still consistent at the level of $1.72\,\sigma$. In Figure 6.3, the likelihood profile of R_μ for the partially unblinded result is shown, both for the fit with all nuisance

parameters floating (blue curve) and for a fit with all nuisance parameters fixed to their best-fit values (red curve), indicating the statistical uncertainty associated with the estimated value of R_μ. The partially unblinded distributions and results indicate no unexpected issues with the measurement.

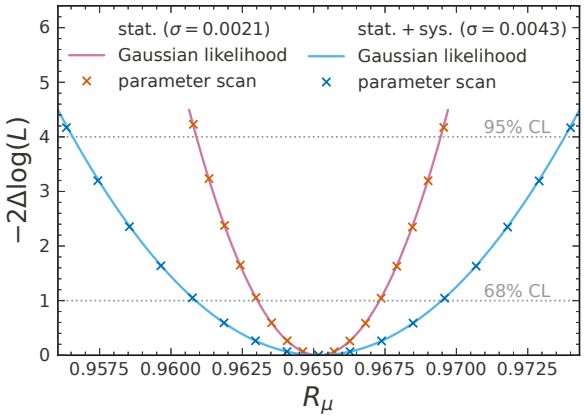


Figure 6.3: The likelihood profile for R_μ using 10 % of the data. The blue curve is obtained with the nominal fit, where all nuisance parameters are kept floating. In contrast, the red curve is obtained with all nuisance parameters fixed to their best-fit values, indicating the statistical uncertainty associated with the estimated value of R_μ.

6.2.2 Full unblinding

Using the full available dataset corresponding to 362 fb^{-1}, 4 358 376 events are observed in the electron channel, and 4 371 737 events are observed in the muon channel. A value of R_μ of 0.9675 ± 0.0037 is measured, which is consistent with the SM expectation of 0.9726 within $1.38\,\sigma$. Separated in terms of statistical and systematic uncertainties, the measured value for the ratio of branching fractions is

$$R_\mu = \frac{\mathcal{B}(\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau)}{\mathcal{B}(\tau^- \to e^- \bar{\nu}_e \nu_\tau)} = 0.9675 \pm 0.0007 \pm 0.0036. \tag{6.1}$$

When this value is compared with the result obtained using 10 % of the data, the partially unblinded result indicates a downward fluctuation in the sampled data consistent with the estimated statistical uncertainty. The pre-fit distributions of the electron and muon channels using the entire data sample are shown in Figure 6.4. As for the partially unblinded results, there is slightly less data in the muon channel than in the electron channel, compared to their respective expectations from the MC templates. While the data/MC ratio in the muon channel is relatively flat, an almost linear trend is observed in the electron channel, with more data than expected for lower electron momentum. As mentioned in Section 5.4 when comparing the shapes of the distributions, this is likely related to differences in photon reconstruction efficiency between data and MC and the presence of more FSR photons in the e-channel. This effect is covered by the systematic uncertainties included in the fit model. Figure 6.5 displays the momentum distributions for electron and muon candidates after fitting the model to the observed data. The shaded regions in the bottom part of the figures represent changes in the fitted event counts based on the systematic variations incorporated in the fit model after being

constrained by the observed data. To obtain this, the nuisance parameters are varied according to their post-fit uncertainties while also considering their correlations. As expected from the goodness-of-fit test with a p-value close to 1, already discussed in Section 5.3.7, the fitted model matches the observed yields very closely. The measured values of the nuisance parameters and their uncertainties can be inferred from the pull plot already shown in Section 5.3.5. There are no significant pulls of any nuisance parameter with respect to its expected value. The correlations of some of the fitted nuisance parameters were shown in Section 5.3.6.

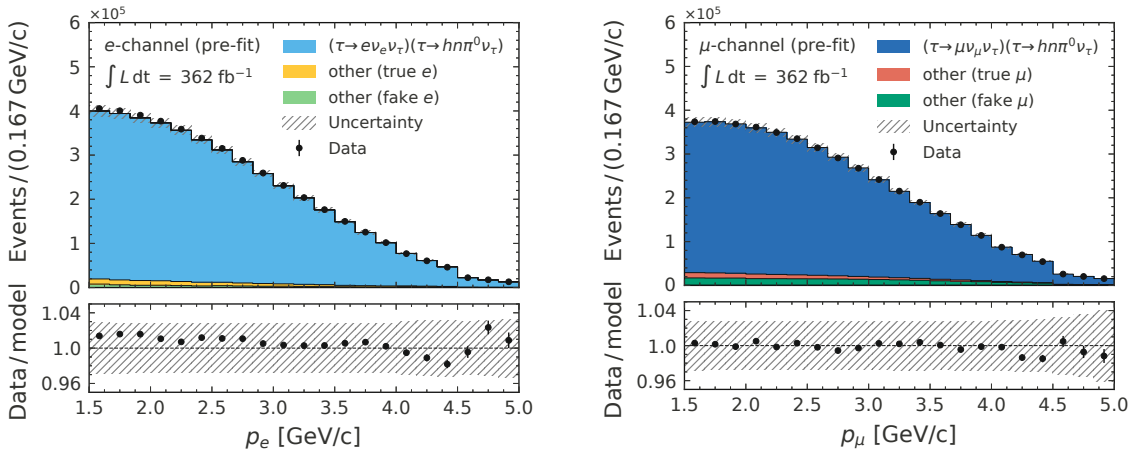


Figure 6.4: The pre-fit momentum distributions for lepton candidates in the e-channel (left) and μ-channel (right), compared to the observed data. The bottom part of the figures displays the ratio of the observed data to the predicted event counts from the fit model. The shaded region represents changes in the event counts based on the systematic variations incorporated in the fit model.

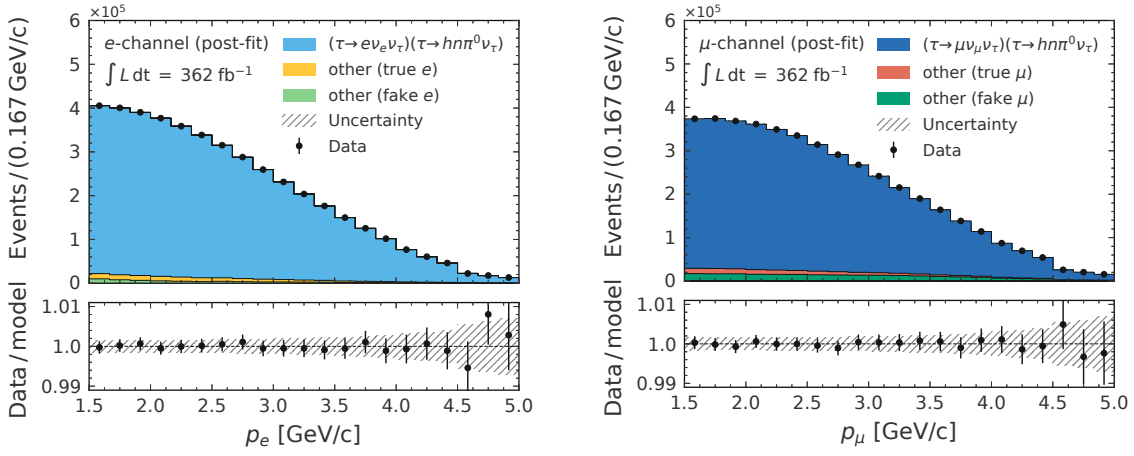


Figure 6.5: The fitted distributions for lepton candidates in the e-channel (left) and μ-channel (right), compared to the observed data. The bottom part of the figures displays the ratio of the observed data to the fitted number of events. The shaded region represents changes in the fitted event counts based on the systematic variations incorporated in the fit model after being constrained by the observed data.

Figure 6.6 shows the likelihood profile of R_μ, demonstrating that it closely follows a Gaussian

distribution. The blue curve shows the profile obtained with all nuisance parameters floating, with the curvature at the best-fit value resulting in the total uncertainty of 0.0037. The red curve shows the profile obtained with all nuisance parameters fixed to their best-fit values, indicating the statistical uncertainty associated with the estimated value of R_μ, which is 0.0007.

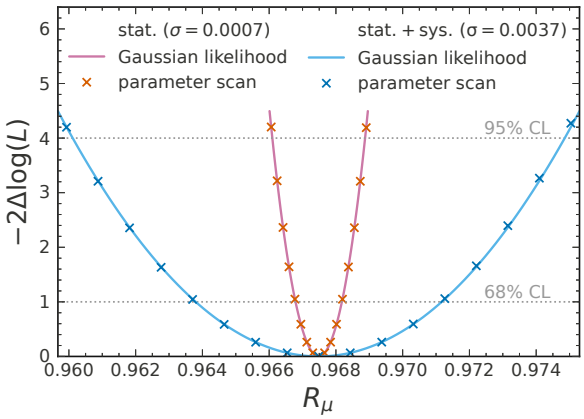


Figure 6.6: The likelihood profile for R_μ after full unblinding. The blue curve is obtained with the nominal fit, where all nuisance parameters are kept floating. In contrast, the red curve is obtained with all nuisance parameters fixed to their best-fit values, indicating the statistical uncertainty associated with the estimated value of R_μ.

6.3 Comparisons with previous studies

In Figure 6.7, the comparison of the unblinded value to previous B-factory measurements and the global fit result by HFLAV [13] is shown. On the left, the values are compared for R_μ and on the right for the corresponding values of $|g_\mu/g_e|_\tau$. The measured value of R_μ is translated to $|g_\mu/g_e|_\tau$ using Equation 1.2, resulting in

$$\left|\frac{g_\mu}{g_e}\right|_\tau = 0.9974 \pm 0.0019. \tag{6.2}$$

All values are consistent with the expectations of the SM. The HFLAV fit contains the measurements by CLEO, *BABAR*, and many more results of individual measurements of tau branching fractions, but not yet this Belle II result. While previous measurements were slightly above the SM value, the present result is slightly below. This is expected to bring the global average closer to the SM value.

6.3.1 Absolute branching fractions

Although the absolute branching fractions were not measured in the present work, one can compare the expected precision from the three-channel fit described in Section 5.2.7 with previous measurements of the $\tau^- \to e^-\bar{\nu}_e\nu_\tau$ and $\tau^- \to \mu^-\bar{\nu}_\mu\nu_\tau$ branching fractions. This comparison is shown in Figure 6.8, where the Belle II value is obtained from a fit with Asimov data. The central value corresponds to what was used as the branching fraction in the generator, as listed in Table 4.11. The separation of the total uncertainty into statistical and systematic

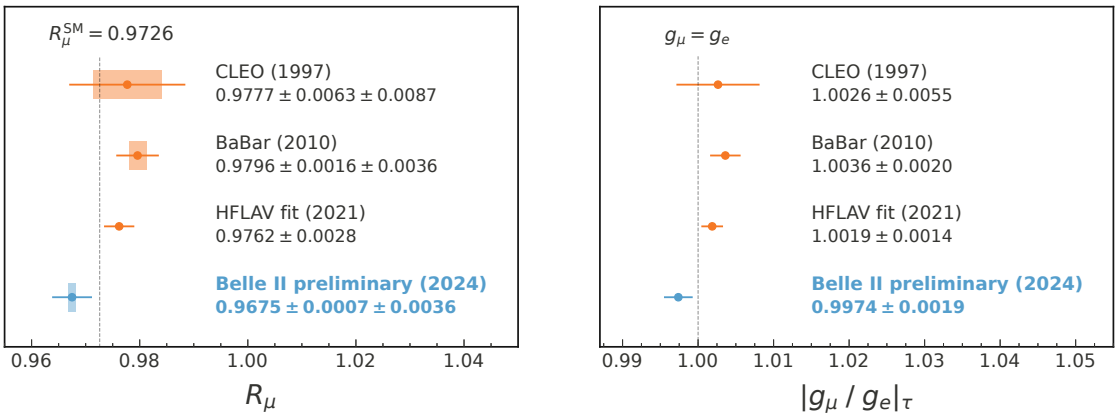


Figure 6.7: A comparison of the present result to earlier measurements of R_μ (left) and $|g_\mu/g_e|_\tau$ (right) conducted by B-factories [11, 12] along with the global fit result from HFLAV [13]. The error bars show the total uncertainties of the respective quantities. In addition, in the figure on the left, the shaded regions indicate the statistical uncertainties. The vertical lines mark the values as predicted by the SM.

components is based on the likelihood profile scans presented in Figure 5.28 in Section 5.3.4. Comparing the uncertainties, the current analysis would already be competitive with the older measurements from LEP experiments. Only the ALEPH result [164] is significantly more precise. The Belle II result could contribute to the global average and would be an essential measurement to further reduce the precision of the branching fractions, which are also a necessary input for e-τ and μ-τ LFU tests, as described in Section 2.4.2.

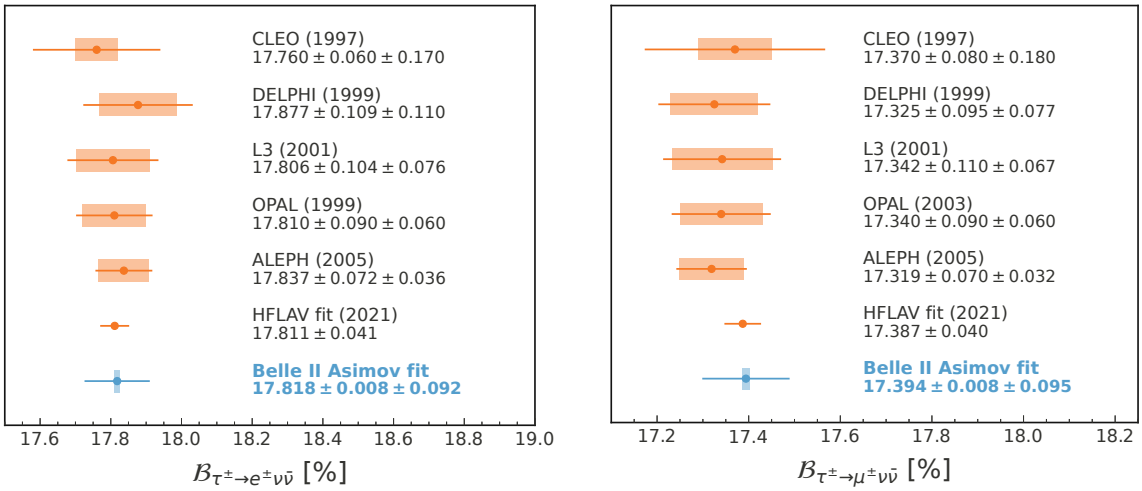


Figure 6.8: Previous individual measurements of $\mathcal{B}(\tau^- \to e^- \bar{\nu}_e \nu_\tau)$ (left) and $\mathcal{B}(\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau)$ (right) by CLEO [11] and the LEP experiments [164–168] along with the global fit result from HFLAV [13]. The Belle II entry shows the preliminary sensitivity achieved with the three-channel fit described in this work. The fit is performed using Asimov data, resulting in a central value equivalent to the branching fractions set in TAUOLA. The error bars show the total uncertainties, while the shaded regions indicate the statistical uncertainties.

6.4 Interpretation of findings

From the results of the present work, a CI at 68.27 % CL can be constructed using the profile likelihood ratio outlined in Section 2.5.2. The resulting interval spans a range from 0.9955 to 0.9993 over $|g_\mu/g_e|_\tau$. In the frequentist interpretation, this means that when the measurement is repeated many times, in 68.27 % of the cases, the constructed CI will cover the actual value of $|g_\mu/g_e|_\tau$.

Testing the SM hypothesis of $|g_\mu/g_e|_\tau = 1$ using the profile likelihood ratio as a test statistic (see Section 2.5.2), a p-value of 0.166 is obtained. This indicates that, under the assumption that the SM hypothesis is true, in about 16.6 % of the outcomes from repeated measurements, the measured value is less consistent with the SM than the observed value from this measurement. Based on this result, the SM hypothesis cannot be ruled out. At the same time, it is important to acknowledge that no hypothesis can be confirmed based on statistical inference.

These interpretations are well summarised in Figure 6.9, where the p-value associated with the profile likelihood scan from Figure 6.6 is shown, assuming that negative two times the logarithm of the profile likelihood ratio follows a χ^2 distribution with one degree of freedom. The scan values were translated from R_μ to $|g_\mu/g_e|_\tau$ using Equation 1.2. The CI corresponding to some CL can be constructed by finding the values of $|g_\mu/g_e|_\tau$ where the p-value corresponds to $1 - \text{CL}$. At the same time, the p-value of testing the SM hypothesis can be directly read from the value at $|g_\mu/g_e|_\tau = 1$.

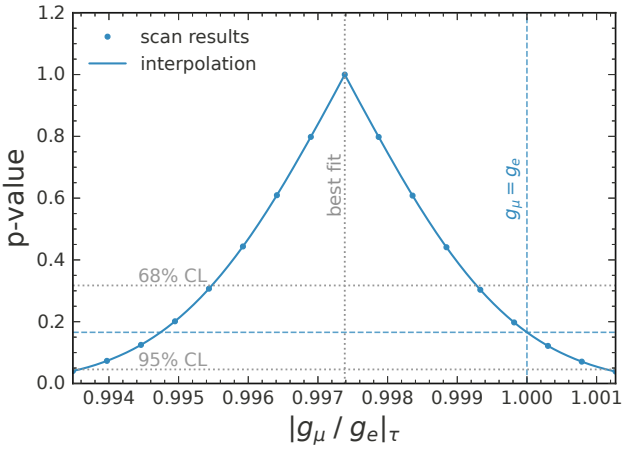


Figure 6.9: The p-values associated with the profile likelihood scan of R_μ, translated to $|g_\mu/g_e|_\tau$. The point estimate from the fit gives a p-value of one and is highlighted with a vertical dotted line. The CI corresponding to some CL can be constructed by finding the values of $|g_\mu/g_e|_\tau$ where the p-value corresponds to $1 - \text{CL}$. This is indicated for a CL of 68.27 % and 95.45 % with horizontal dotted lines. The p-value of testing the SM hypothesis can be directly read from the value at $|g_\mu/g_e|_\tau = 1$ and is indicated with a dashed line.

The tests in Section 5.3.7 suggest that the constructed confidence intervals may be slightly overcovering. This can be attributed to the conservative choice of some uncertainties due to the limited knowledge of the actual values. Furthermore, the assumed correlations between the systematic variations of the templates are conservative, as the true correlation is often unknown and difficult to determine. Making a conclusive statement about overcoverage without

knowing the actual uncertainties and correlations is difficult. It is possible that reducing the uncertainties without considering the correct correlations would introduce a bias and cause undercoverage. Given these considerations, it is preferable to risk overcoverage. The studies with reduced uncertainties indicate that the effect is, in any case, not greater than 10 %, and the interpretation of the results remains unaffected.

To see the impact of this result on the global average, one can compare the average of previous measurements quoted in the PDG to a combination of prior measurements with the result of this work. The PDG averaged the results of ARGUS [169], CLEO [11] and *BABAR* [12] and obtained a value of 0.979 ± 0.004 [10]. Combining these three results with the present result, a value for R_μ of 0.9735 ± 0.0026 is obtained. Here, completely independent uncertainties between all the measurements were assumed, which might not be the case but is good enough for this naive comparison. This average perfectly agrees with the SM prediction of 0.9726. Moreover, translated to $|g_\mu/g_e|_\tau$, this corresponds to 1.0005 ± 0.0013, which is of similar precision as the currently most precise LFU test using pion decays, which is $g_e/g_\mu = 0.9996 \pm 0.0012$ [14].

6.5 Addressing limitations

The dominant sources of uncertainties in this measurement arise from the identification of leptons (0.32 %), followed by uncertainties associated with the trigger (0.10 %); see also the summary of all systematic effects shown in Table 4.25. These sources are particularly challenging because they can affect the signal of $\tau^- \to e^- \bar{\nu}_e \nu_\tau$ and $\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$ differently, leading to non-cancellation when measuring R_μ.

Extensive work has been done to study lepton identification uncertainties, which is reassuring in terms of their validity. However, there is still room for improvement, particularly concerning electron identification uncertainties. These uncertainties are dominated by their systematic components, which will not scale with acquiring a larger dataset. To address this, dedicated studies must be performed to explain the observed discrepancies with the Bhabha calibration channel, shown in Section 3.6.3, which are currently added as systematic uncertainties. Once the source of these differences is understood, it is reasonable to assume that the systematics will be reduced, possibly by introducing an additional dimension in which the corrections are binned.

Another realistic way to improve the measurement is to introduce a π^0 veto on the signal side. As discussed in Section 4.4.2, this is currently not included because data/MC differences in the reconstruction of false π^0 candidates can significantly impact the measurement of R_μ. A correction factor and associated uncertainties must be obtained through a detailed study using control channels to address this. While this could add to the overall uncertainty on R_μ, it could potentially outweigh the increased uncertainties associated with mis-id rate uncertainties originating from the larger contamination of pions on the signal side. With the π^0 veto, the fraction of $\tau \to \pi\pi^0\nu$ decays on the signal side would be approximately one-third compared to the present case, in both the e and μ channels, as shown in Table 6.1. For this, a simple requirement of having no π^0 candidates on the signal side using the π^0 selection from Section 4.4.1 was imposed, but one could also think of having a more refined veto. For reference, the contribution of $\tau \to \pi\nu$ decays on the signal side is also listed, which is unaffected by the π^0 veto. The impact of these reduced pion backgrounds on the systematic uncertainty due to muon mis-id rate corrections would be a reduction from currently 0.19 % to 0.11 %.

Similarly, the systematic uncertainty due to electron mis-id rate corrections would be reduced from 0.12 % to 0.05 %.

Table 6.1: The fraction of tau decays involving pions on the signal side for the electron and the muon channel. The current case ('without veto') is compared to introducing the requirement that no π^0 candidates are on the signal side ('with veto'). The number of $\tau \to \pi\nu$ decays in the sample is unaffected by the veto.

Decay	e-channel	μ-channel
$\tau \to \pi\pi^0\nu$ (without veto)	0.9 %	3 %
$\tau \to \pi\pi^0\nu$ (with veto)	0.3 %	1 %
$\tau \to \pi\nu$	0.36 %	2.2 %

The uncertainty associated with the trigger correction can be reduced with detailed studies on why the correction factor varies for different choices of the reference trigger. Although the currently included uncertainties are very conservative, the overall impact on the total uncertainty of R_μ remains subdominant compared to lepton identification. When the uncertainty associated with lepton ID corrections can be reduced, it will be more critical to better understand the systematics associated with the trigger. In particular, the correlation between the uncertainties in the e-channel and the μ-channel is important to understand. One way to improve the analysis is to use additional trigger lines, such as newer CDC-based triggers using Neuro 3D tracks. A promising candidate for this is the single-track trigger, which can trigger on individual tracks. By doing this, the trigger efficiency can increase and reach close to 100 % for both channels, making the measurement less sensitive to trigger systematics altogether.

Other sources of systematics that, at some point, could become a limiting factor are the modelling of FSR and the modelling of the background. Currently, both effects are included using rather conservative estimates. Understanding these effects in more detail would involve extensive studies, possibly only beneficial after reaching a sub-permille precision on R_μ.

The measurement of the absolute branching fractions is sensitive to various systematic effects, particularly those that impact the overall normalisation of the templates. Not as much attention has been paid to these effects in the present work since they cancel for the R_μ measurement. Although the $e \times \mu$ channel was intended only as a control channel, the prospects of using it to precisely measure the absolute branching fractions are promising. While it is relatively clean compared to other channels, particularly regarding uncertainties related to neutral reconstruction, it still needs to be studied in more detail as additional effects may play a role. Currently, the most significant systematics are associated with luminosity, tracking efficiency, lepton identification, and sample normalisation, as listed in Table 5.3. In addition, the effect of the trigger should be carefully studied since systematic shifts from estimating the trigger efficiency with a reference sample would not cancel, as it is the case for the R_μ measurement. When writing this, the official luminosity measurement of Belle II was not yet finalised for the present dataset and is still subject to some change in the central value and associated uncertainty. While this has little impact on the measurement of R_μ, a change in the luminosity would directly impact the measurement of the absolute branching fractions, which is another reason why it was not pursued for publication together with the test of LFU.

6.6 Conclusion and outlook

This work presented a test of lepton universality by measuring the ratio of the leptonic branching fractions of the tau lepton. Using $362\,\mathrm{fb}^{-1}$ of Belle II data, the measured value

$$R_\mu = \frac{\mathcal{B}(\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau)}{\mathcal{B}(\tau^- \to e^- \bar{\nu}_e \nu_\tau)} = 0.9675 \pm 0.0007 \pm 0.0036$$

is in agreement with the SM expectation of 0.9726. The uncertainty is split into a statistical and a systematic component, with the latter being dominant. The limiting systematic source for this measurement is associated with the identification of the charged leptons, which comprises about 86 % of the total systematic uncertainty. Extensive studies were conducted to estimate and reduce the systematic effects of this measurement and to confirm its validity. When translated to the ratio of the weak couplings, the result is

$$\left|\frac{g_\mu}{g_e}\right|_\tau = 0.9974 \pm 0.0019,$$

which at the moment is the most accurate test of lepton universality using tau decays obtained from a single experiment. A journal paper describing this analysis has been submitted to the Journal of High Energy Physics (JHEP), and the full statistical model will be published on HEPData to simplify combinations with past and future measurements.

The future for this measurement looks promising, and even more precise tests of lepton universality can be expected. At the time of writing this, another measurement by the Belle II collaboration is being prepared using the 3x1 topology of taupair events, following a similar approach as outlined in this work. This measurement could exceed the precision achieved with the current measurement for several reasons: The main improvement with respect to this analysis would be the addition of a π^0 veto, which significantly reduces uncertainties due to mis-id rate uncertainties, as discussed in Section 6.5. The challenge will be reliably estimating the correction due to the difference between data and simulation in reconstructing fake π^0 candidates and the systematic uncertainties associated with applying this veto. Another advantage is that the trigger is less of a concern for the 3x1 topology since additional track-based triggers that require the presence of three tracks in the event can be used. Studies by colleagues have shown that using a selection of ECL and CDC-based triggers already results in an efficiency close to 100 % for this topology, which significantly reduces the impact of the trigger correction and associated uncertainties. Furthermore, selecting taupair events with the 3x1 topology generally results in lower backgrounds, which can reduce the uncertainties associated with it.

The combination of the 3x1 measurement with this 1x1 measurement can be achieved using the `pyhf` library. A combined fit using four channels rather than two can be performed, in which the correlations of the systematic effects common to both measurements must be taken into account. Since both measurements are expected to be dominated by lepton identification systematics, the gain from this combination will be limited. However, it will be an important confirmation of the present measurement.

One step further would be to measure the absolute branching fractions by adding a fifth channel, as outlined in this thesis (Section 5.2.7). The measurement of the absolute branching fractions (and the measurement of R_μ) could be further improved by adding even more channels

to the fit, such as those discussed in Section 4.3. The most promising candidate for this would be events with purely leptonic decays for both tau leptons.

In the more distant future, the best way to reduce the uncertainties of this measurement is to perform more detailed studies on lepton identification and the associated systematic effects, with the calibration channels used to derive the corrections discussed in Section 3.6. It will be especially important to obtain the correlation of the systematic uncertainties between the correction bins, which is currently unknown. Knowing these correlations could already have a significant impact when using the current set of corrections, as it was shown in Section 5.3.2 or as indicated by Figure 4.3 in Section 4.3. One could even consider including these calibration channels in the fit model and directly obtaining the correction factors, associated uncertainties, and their correlation with the fit. Furthermore, new calibration channels can be studied to obtain a more robust estimation of the lepton identification corrections and better understand the various systematic effects.

With these improvements, Belle II could test lepton universality with an accuracy below one per mille, resulting in the most precise test of lepton universality altogether, surpassing even other methods, such as using pion decays. Moreover, Belle II can perform precise measurements of the τ mass and lifetime, which, together with the leptonic tau branching fractions, provides a powerful test of e-τ and μ-τ universality. Other experiments are also expected to improve their precision in the coming years, and exciting results can be anticipated. New colliders, such as FCC-ee [170, 171], will also be able to perform this measurement and help to further enhance our understanding of fundamental physics. Precision measurements with tau leptons are and will be an essential aspect of this, as they provide a unique way to test the Standard Model and search for new physics beyond it.

Additional work

During the time of my PhD, I was actively involved in several journal publications that are not quite suitable for inclusion within the main content of this dissertation. For completeness, they are listed here, together with a brief mention of my own contributions.

F. Abudinén et al. „Punzi-loss: a non-differentiable metric approximation for sensitivity optimisation in the search for new particles“. In: *The European Physical Journal C* 82.2 (Feb. 2022). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-022-10070-0

In close collaboration with *H. Haigh* and *G. Inguglia*.

Primary contributions:

- idea of optimising the Punzi figure of merit with neural networks
- initial development in the search for a Z' boson [142, pp. 71 sqq.]
- PyTorch implementation: *feichtip/punzinet*. Version v1.0.0. Feb. 2022. DOI: 10.5281/zenodo.5992328

Andreas Crivellin, Claudio Andrea Manzari, Wolfgang Altmannshofer, Gianluca Inguglia, Paul Feichtinger, and Jorge Martin Camalich. „Toward excluding a light Z' explanation of $b \to s\ell^+\ell^-$“. In: *Physical Review D* 106.3 (3 Aug. 2022), p. L031703. ISSN: 2470-0029. DOI: 10.1103/PhysRevD.106.L031703

Primary contributions:

- recast of a search for a Z' boson with early Belle II data [172]
- recalculation of upper limits when assuming sizeable Z' widths

---

I. Adachi et al. „Search for an Invisible Z' in a Final State with Two Muons and Missing Energy at Belle II“. in: *Physical Review Letters* 130.23 (23 June 2023), p. 231801. ISSN: 1079-7114. DOI: 10.1103/PhysRevLett.130.231801

In close collaboration with *H. Haigh*, *G. Inguglia*, *M. Campajola* and *E. Graziani*.

Primary contributions:

- preceding studies of event selection and sensitivity [142]
- implementation of the fitting procedure and statistical interpretation of the results

List of Figures

List of Tables

Acronyms

ANN artificial neural network 22–24, 63–65, 69, 70, 76, 79, 87–95, 123, 125–127, 129, 138, 142, 143, 163, 164, 172, 173, 181, 183–185, 200–204, 206–208

ARICH Aerogel Ring Imaging Cherenkov counter 34, 35, 44, 48

basf2 Belle II analysis software framework 37, 38, 43

BCE binary cross-entropy 23

BDT boosted decision tree 45, 81, 141, 175

CDC Central Drift Chamber 33–35, 37–39, 44, 45, 48, 70, 72, 75, 76, 112, 130, 194, 195, 199, 201

CI confidence interval 21, 192, 209

CL confidence level 21, 165, 192, 206, 209

CM centre-of-mass 2, 3, 28, 36, 41, 43, 46, 48, 49, 55, 60, 72, 77–80, 89, 93, 95, 118, 122, 123, 134, 142, 172, 173, 201, 202, 207, 210, 212

DAQ data acquisition 32, 38

ECL Electromagnetic Calorimeter 36–39, 44, 45, 48, 49, 51, 53, 63, 70–72, 74–76, 112, 123, 146, 195, 199, 201, 210

FOM figure of merit 82, 96, 212

FPGA Field Programmable Gate Array 38

FSR final-state radiation 72–74, 114, 115, 122, 124, 127, 138, 169, 174, 188, 194, 203, 204, 211

HEP high energy physics 19, 20, 25

HER high energy storage ring 29, 31

HFLAV Heavy Flavor Averaging Group 2, 81, 190, 191, 209

HLT high level trigger 38, 39, 44, 48, 49, 176

HVP hadronic vacuum polarization 14

IP interaction point 25, 29, 31, 36, 38, 72

IR interaction region 31

ISR initial-state radiation 115, 116, 138, 176, 203

KLM K_L - Muon detector viii, 37, 39, 44, 45, 48, 54, 55, 70, 85, 174, 199

L1 level 1 38, 39, 43, 44, 48, 49, 51

lepton ID lepton identification 44, 46, 53, 54, 62, 64, 68–70, 73, 84, 85, 91, 93–100, 105, 106, 110, 117, 118, 123, 124, 126, 127, 140, 141, 143, 146–151, 155, 158, 161–163, 169, 175–184, 194, 200–208, 210, 212

LER low energy storage ring 29, 31

LFU lepton flavour universality 1, 2, 13, 14, 17–19, 191, 193, 194, 210

LH likelihood 141

LOOCV Leave-One-Out Cross-Validation 124, 127, 128, 204

LR learning rate 24

LS1 long shutdown one 41, 42, 60, 199

MC Monte Carlo 42, 44, 46–48, 54, 55, 60, 62, 64, 68, 73, 74, 76–79, 84–86, 88–90, 93–96, 98, 99, 104, 107, 111–116, 118, 121–125, 129, 130, 132–135, 138, 140, 141, 146–150, 154, 172–174, 178, 179, 181, 184, 188, 201–205, 207, 211, 212

mDST mini Data Summary Tables 38

mis-id rate misidentification rate 44–48, 54, 55, 57, 59, 69, 86, 100, 107–109, 123, 138, 149, 155, 169, 174, 175, 177, 179–181, 193–195, 199, 200, 203, 207, 212, 213

ML maximum likelihood 20, 98

MLE maximum likelihood estimate 21, 152, 164, 166

MLP multilayer perceptron 22

PDF probability density function 44

PDG Particle Data Group 2, 18, 27, 42, 81, 82, 193, 199, 210, 211

PID particle identification 44, 46–49, 53, 73, 88–90, 199, 202

PMT photomultiplier tube 35

PSD pulse shape discrimination 37

PXD Pixel detector 32, 33, 39, 41, 199

QED quantum electrodynamics 10

ROC receiver operating characteristic 64, 65, 81–83, 89, 91, 141–143, 200–202, 205

RPC Resistive Plate Chamber 37

SM Standard Model 1, 2, 5–14, 18, 19, 42, 136, 187, 188, 190–193, 195, 198, 209, 210

SVD Silicon Vertex detector 32, 33, 199

TOP Time-of-propagation counter 34, 35, 44, 48, 199

TSIM trigger simulation 43, 75, 76, 146, 201

Bibliography

[1] M. L. Perl et al. „Evidence for Anomalous Lepton Production in $e^+ - e^-$ Annihilation“. In: *Physical Review Letters* 35 (22 Dec. 1975), pp. 1489–1492. DOI: 10.1103/PhysRevLett.35.1489.

[2] M.L. Perl et al. „Properties of anomalous $e\mu$ events produced in $e^+ - e^-$ annihilation“. In: *Physics Letters B* 63.4 (1976), pp. 466–470. ISSN: 0370-2693. DOI: 10.1016/0370-2693(76)90399-3.

[3] M.L. Perl et al. „Properties of the proposed τ charged lepton“. In: *Physics Letters B* 70.4 (1977), pp. 487–490. ISSN: 0370-2693. DOI: 10.1016/0370-2693(77)90421-X.

[4] A. Stahl. *Physics with Tau Leptons*. Springer Tracts in Modern Physics. Springer Berlin Heidelberg, 2000, pp. 1–317. DOI: 10.1007/BFb0109618.

[5] J. P. Lees et al. „Evidence for an Excess of $\overline{B} \to D^{(*)}\tau^-\overline{\nu}_\tau$ Decays“. In: *Physical Review Letters* 109 (10 Sept. 2012), p. 101802. DOI: 10.1103/PhysRevLett.109.101802.

[6] R. Aaij et al. „Measurement of the Ratios of Branching Fractions $\mathcal{R}(D^*)$ and $\mathcal{R}(D^0)$“. In: *Physical Review Letters* 131 (11 Sept. 2023), p. 111802. DOI: 10.1103/PhysRevLett.131.111802.

[7] R. Aaij et al. „Test of lepton universality in beauty-quark decays“. In: *Nature Physics* 18.3 (Mar. 1, 2022), pp. 277–282. ISSN: 1745-2481. DOI: 10.1038/s41567-021-01478-8.

[8] Antonio Pich. „Precision tau physics“. In: *Progress in Particle and Nuclear Physics* 75 (2014), pp. 41–85. ISSN: 0146-6410. DOI: 10.1016/j.ppnp.2013.11.002.

[9] Yung-Su Tsai. „Decay Correlations of Heavy Leptons in $e^+ + e^- \to l^+ + l^-$“. In: *Physical Review D* 4 (9 Nov. 1971), pp. 2821–2837. DOI: 10.1103/PhysRevD.4.2821.

[10] R. L. Workman et al. „Review of Particle Physics“. In: *PTEP* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097.

[11] A. Anastassov et al. „Experimental test of lepton universality in tau decay“. In: *Physical Review D* 55 (1997). [Erratum: Phys.Rev.D 58, 119904 (1998)], pp. 2559–2576. DOI: 10.1103/PhysRevD.55.2559.

[12] Bernard Aubert et al. „Measurements of Charged Current Lepton Universality and $|V_{us}|$ using Tau Lepton Decays to $e^-\bar{\nu}_e\nu_\tau$, $\mu^-\bar{\nu}_\mu\nu_\tau$, $\pi^-\nu_\tau$, and $K^-\nu_\tau$“. In: *Physical Review Letters* 105 (2010), p. 051602. DOI: 10.1103/PhysRevLett.105.051602.

[13] Yasmine Sara Amhis et al. „Averages of b-hadron, c-hadron, and τ-lepton properties as of 2021“. In: *Physical Review D* 107.5 (2023). updated results and plots available at https://hflav.web.cern.ch/, p. 052008. DOI: 10.1103/PhysRevD.107.052008.

[14] A. Aguilar-Arevalo et al. „Improved Measurement of the $\pi \to e\nu$ Branching Ratio“. In: *Physical Review Letters* 115.7 (2015), p. 071801. DOI: 10.1103/PhysRevLett.115.071801.

[15] ATLAS Collaboration. *Precise test of lepton flavour universality in W-boson decays into muons and electrons in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. 2024. DOI: 10.48550/arxiv.2403.02133.

[16] David J Griffiths. *Introduction to elementary particles; 2nd rev. version*. Physics textbook. New York, NY: Wiley, 2008. URL: https://cds.cern.ch/record/111880.

[17] Matthew D. Schwartz. *Quantum Field Theory and the Standard Model*. Cambridge University Press, 2013.

[18] Palash B. Pal. *An Introductory Course of Particle Physics*. Taylor & Francis, July 2014, pp. 1–777. ISBN: 9780429068515. DOI: 10.1201/b17199.

[19] Herwig Schopper, ed. *Particle Physics Reference Library: Volume 1: Theory and Experiments*. Cham: Springer Nature, 2020. ISBN: 978-3-030-38207-0. DOI: 10.1007/978-3-030-38207-0.

[20] Jiri Horejsi. *Fundamentals of electroweak theory*. 2022. DOI: 10.48550/arxiv.2210.04526.

[21] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, 2013. ISBN: 978-1-107-03426-6. DOI: 10.1017/CBO9781139525367.

[22] C. S. Wu et al. „Experimental Test of Parity Conservation in Beta Decay“. In: *Physical Review* 105 (4 Feb. 1957), pp. 1413–1415. DOI: 10.1103/PhysRev.105.1413.

[23] Sheldon L. Glashow. „Partial-symmetries of weak interactions“. In: *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: 10.1016/0029-5582(61)90469-2.

[24] P.W. Higgs. „Broken symmetries, massless particles and gauge fields“. In: *Physics Letters B* 12.2 (1964), pp. 132–133. ISSN: 0031-9163. DOI: 10.1016/0031-9163(64)91136-9.

[25] Peter W. Higgs. „Broken Symmetries and the Masses of Gauge Bosons“. In: *Physical Review Letters* 13 (16 Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.

[26] F. Englert and R. Brout. „Broken Symmetry and the Mass of Gauge Vector Mesons“. In: *Physical Review Letters* 13 (9 Aug. 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.

[27] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. „Global Conservation Laws and Massless Particles“. In: *Physical Review Letters* 13 (20 Nov. 1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585.

[28] K. Ackerstaff et al. „Measurement of the strong coupling constant α_s and the vector and axial-vector spectral functions in hadronic tau decays“. In: *The European Physical Journal C* 7.4 (Mar. 1999), pp. 571–593. ISSN: 1434-6052. DOI: 10.1007/s100529901061.

[29] Jens Erler. „Calculation of the QED coupling alpha (M(Z)) in the modified minimal subtraction scheme“. In: *Physical Review D* 59.5 (Feb. 1999). ISSN: 1089-4918. DOI: 10.1103/physrevd.59.054008.

[30] P. Achard et al. „Measurement of the running of the electromagnetic coupling at large momentum-transfer at LEP“. In: *Physics Letters B* 623.1–2 (Sept. 2005), pp. 26–36. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2005.07.052.

[31] Ponor. *Standard Model – All Feynman diagram vertices*. This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License. To view a copy of this license, visit https://creativecommons.org/licenses/by-sa/4.0/. June 4, 2021. URL: https://commons.wikimedia.org/w/index.php?curid=106285709 (visited on 03/2024).

[32] Jack Lindon. „Particle Collider Probes of Dark Energy, Dark Matter and Generic Beyond Standard Model Signatures in Events With an Energetic Jet and Large Missing Transverse Momentum Using the ATLAS Detector at the LHC“. Presented 30 Oct 2020. Birmingham U., 2020. URL: https://cds.cern.ch/record/2746537.

[33] Izaak Neutelings. *SM particles*. This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License. To view a copy of this license, visit https://creativecommons.org/licenses/by-sa/4.0/. 2024. URL: https://tikz.net/sm_particles/ (visited on 03/2024).

[34] Laurent Canetti, Marco Drewes, and Mikhail Shaposhnikov. „Matter and antimatter in the universe“. In: *New Journal of Physics* 14.9 (Sept. 2012), p. 095012. DOI: 10.1088/1367-2630/14/9/095012.

[35] C. Lazzeroni et al. „Precision Measurement of the Ratio of the Charged Kaon Leptonic Decay Rates“. In: *Physics Letters B* 719 (2013), pp. 326–336. DOI: 10.1016/j.physletb.2013.01.037.

[36] F. Ambrosino et al. „Precise measurement of $\Gamma(K \to e\nu(\gamma))/\Gamma(K \to \mu\nu(\gamma))$ and study of $K \to e\nu\gamma$“. In: *The European Physical Journal C* 64 (2009). [Erratum: Eur.Phys.J. 65, 703 (2010)], pp. 627–636. DOI: 10.1140/epjc/s10052-009-1217-6.

[37] G. Caria et al. „Measurement of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ with a semileptonic tagging method“. In: *Physical Review Letters* 124.16 (2020), p. 161803. DOI: 10.1103/PhysRevLett.124.161803.

[38] Belle II Collaboration et al. *A test of lepton flavor universality with a measurement of $R(D^*)$ using hadronic B tagging at the Belle II experiment*. 2024. DOI: 10.48550/arxiv.2401.02840.

[39] R. Aaij et al. „Test of lepton flavor universality using $B^0 \to D^{*-}\tau^+\nu_\tau$ decays with hadronic τ channels“. In: *Physical Review D* 108 (1 July 2023), p. 012018. DOI: 10.1103/PhysRevD.108.012018.

[40] R. Aaij et al. „Test of Lepton Universality in $b \to s\ell^+\ell^-$ Decays“. In: *Physical Review Letters* 131 (5 Aug. 2023), p. 051803. DOI: 10.1103/PhysRevLett.131.051803.

[41] G. Aad et al. „Test of the universality of τ and μ lepton couplings in W-boson decays with the ATLAS detector“. In: *Nature Physics* 17.7 (2021), pp. 813–818. DOI: 10.1038/s41567-021-01236-w.

[42] Armen Tumasyan et al. „Precision measurement of the W boson decay branching fractions in proton-proton collisions at $\sqrt{s} = 13$ TeV“. In: *Physical Review D* 105.7 (2022), p. 072008. DOI: 10.1103/PhysRevD.105.072008.

[43] D. P. Aguillard et al. „Measurement of the Positive Muon Anomalous Magnetic Moment to 0.20 ppm“. In: *Physical Review Letters* 131 (16 Oct. 2023), p. 161802. DOI: 10.1103/PhysRevLett.131.161802.

[44] T. Aoyama et al. „The anomalous magnetic moment of the muon in the Standard Model“. In: *Physics Reports* 887 (2020). The anomalous magnetic moment of the muon in the Standard Model, pp. 1–166. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2020.07.006.

[45] Sz. Borsanyi et al. „Leading hadronic contribution to the muon magnetic moment from lattice QCD“. In: *Nature* 593.7857 (May 1, 2021), pp. 51–55. ISSN: 1476-4687. DOI: 10.1038/s41586-021-03418-1.

[46] M. Cè et al. „Window observable for the hadronic vacuum polarization contribution to the muon $g-2$ from lattice QCD“. In: *Physical Review D* 106 (11 Dec. 2022), p. 114502. DOI: 10.1103/PhysRevD.106.114502.

[47] Svjetlana Fajfer et al. „Implications of Lepton Flavor Universality Violations in B Decays“. In: *Physical Review Letters* 109 (16 Oct. 2012), p. 161801. DOI: 10.1103/PhysRevLett.109.161801.

[48] R. Alonso, B. Grinstein, and J. Martin Camalich. „Lepton universality violation with lepton flavor conservation in B-meson decays“. In: *Journal of High Energy Physics* 2015.10 (Oct. 2015). ISSN: 1029-8479. DOI: 10.1007/jhep10(2015)184.

[49] Wolfgang Altmannshofer and Itay Yavin. „Predictions for lepton flavor universality violation in rare B decays in models with gauged $L_\mu - L_\tau$“. In: *Physical Review D* 92 (7 Oct. 2015), p. 075022. DOI: 10.1103/PhysRevD.92.075022.

[50] Wolfgang Altmannshofer, Peter Stangl, and David M. Straub. „Interpreting hints for lepton flavor universality violation“. In: *Physical Review D* 96 (5 Sept. 2017), p. 055008. DOI: 10.1103/PhysRevD.96.055008.

[51] Jernej F. Kamenik, Yotam Soreq, and Jure Zupan. „Lepton flavor universality violation without new sources of quark flavor violation“. In: *Physical Review D* 97 (3 Feb. 2018), p. 035002. DOI: 10.1103/PhysRevD.97.035002.

[52] P N Thu et al. „Lepton universality violation in the minimal flipped 331 model“. In: *Progress of Theoretical and Experimental Physics* 2023.12 (Nov. 2023), 123B01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptad135.

[53] Maria-Teresa Dova, John Swain, and Lucas Taylor. „Anomalous charged current couplings of the tau and implications for tau compositeness and two-Higgs-doublet models“. In: *Physical Review D* 58 (1 May 1998), p. 015005. DOI: 10.1103/PhysRevD.58.015005.

[54] Martin Jung, Antonio Pich, and Paula Tuzón. „Charged-Higgs phenomenology in the aligned two-Higgs-doublet model“. In: *Journal of High Energy Physics* 2010.11 (2010), p. 3. ISSN: 1029-8479. DOI: 10.1007/JHEP11(2010)003.

[55] Andreas Crivellin et al. „Searching for lepton flavor universality violation and collider signals from a singly charged scalar singlet“. In: *Physical Review D* 103 (7 Apr. 2021), p. 073002. DOI: 10.1103/PhysRevD.103.073002.

[56] Wolfgang Altmannshofer et al. „Lepton flavor violating Z' explanation of the muon anomalous magnetic moment“. In: *Physics Letters B* 762 (2016), pp. 389–398. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2016.09.046.

[57] Douglas Bryman et al. „Testing Lepton Flavor Universality with Pion, Kaon, Tau, and Beta Decays“. In: *Annual Review of Nuclear and Particle Science* 72.Volume 72, 2022 (2022), pp. 69–91. ISSN: 1545-4134. DOI: 10.1146/annurev-nucl-110121-051223.

[58] A. Pich. „Tau lepton physics: theory overview“. In: *Nuclear Physics B - Proceedings Supplements* 55.3 (1997). vi55c, pp. 3–22. ISSN: 0920-5632. DOI: 10.1016/S0920-5632(97)00195-3.

[59] Izaak Neutelings. *SM decay piechart*. This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License. To view a copy of this license, visit https://creativecommons.org/licenses/by-sa/4.0/. 2024. URL: https://tikz.net/sm_decay_piechart/ (visited on 03/2024).

[60] M. M. Giannini. *Muon decay and neutrino mass*. 2022. DOI: 10.48550/arxiv.2207.02717.

[61] Florian Scheck. „Muon Physics“. In: *Phys. Rept.* 44 (1978), p. 187. DOI: 10.1016/0370-1573(78)90014-5.

[62] T. D. Lee and C. N. Yang. „Possible Nonlocal Effects in μ Decay“. In: *Physical Review* 108 (6 Dec. 1957), pp. 1611–1614. DOI: 10.1103/PhysRev.108.1611.

[63] Andrea Ferroglia et al. „Contributions of the W-boson propagator to the μ and τ leptonic decay rates“. In: *Physical Review D* 88 (3 Aug. 2013), p. 033012. DOI: 10.1103/PhysRevD.88.033012.

[64] M. Fael, L. Mercolli, and M. Passera. „W-propagator corrections to μ and τ leptonic decays“. In: *Physical Review D* 88 (9 Nov. 2013), p. 093011. DOI: 10.1103/PhysRevD.88.093011.

[65] William J. Marciano and A. Sirlin. „Electroweak radiative corrections to τ decay“. In: *Physical Review Letters* 61 (16 1988), pp. 1815–1818. DOI: 10.1103/PhysRevLett.61.1815.

[66] I. Adachi et al. „Measurement of the τ-lepton mass with the Belle II experiment“. In: *Physical Review D* 108 (3 Aug. 2023), p. 032006. DOI: 10.1103/PhysRevD.108.032006.

[67] „Information and Likelihood Theory: A Basis for Model Selection and Inference“. In: *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Ed. by Kenneth P. Burnham and David R. Anderson. New York, NY: Springer New York, 2002, pp. 49–97. ISBN: 978-0-387-22456-5. DOI: 10.1007/978-0-387-22456-5_2.

[68] Erich L. Lehmann and George Casella. *Theory of Point Estimation*. Second. New York, NY, USA: Springer-Verlag, 1998.

[69] C. Radhakrishna Rao. „Information and the Accuracy Attainable in the Estimation of Statistical Parameters“. In: *Breakthroughs in Statistics: Foundations and Basic Theory*. Ed. by Samuel Kotz and Norman L. Johnson. New York, NY: Springer New York, 1992, pp. 235–247. ISBN: 978-1-4612-0919-5. DOI: 10.1007/978-1-4612-0919-5_16.

[70] Harald Cramér. *Mathematical methods of statistics*. Vol. 26. Princeton University Press, 1999.

[71] F. James. *MINUIT: Function Minimization and Error Analysis Reference Manual*. CERN Program Library Long Writeups. 1998. URL: https://cds.cern.ch/record/2296388.

[72] Olaf Behnke and Lorenzo Moneta. „Parameter Estimation“. In: *Data Analysis in High Energy Physics*. John Wiley & Sons, Ltd. Chap. 2, pp. 27–73. ISBN: 9783527653416. DOI: 10.1002/9783527653416.ch2.

[73] Jerzy Neyman and Egon Sharpe Pearson. „On the problem of the most efficient tests of statistical hypotheses“. In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 231.694-706 (1933), pp. 289–337. DOI: 10.1098/rsta.1933.0009.

[74] S. S. Wilks. „The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses“. In: *Annals of Mathematical Statistics* 9.1 (Mar. 1938), pp. 60–62. DOI: 10.1214/aoms/1177732360.

[75] Glen Cowan et al. „Asymptotic formulae for likelihood-based tests of new physics“. In: *The European Physical Journal C* 71.2 (Feb. 9, 2011), p. 1554. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-011-1554-0.

[76] Abraham Wald. „Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large“. In: *Transactions of the American Mathematical Society* 54.3 (1943), pp. 426–482. ISSN: 0002-9947. URL: http://www.jstor.org/stable/1990256.

[77] J. Neyman. „Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability“. In: *Philosophical Transactions of the Royal Society of London Series A* 236.767 (Aug. 1937), pp. 333–380. DOI: 10.1098/rsta.1937.0005.

[78] Grégory Schott. „Hypothesis Testing“. In: *Data Analysis in High Energy Physics*. John Wiley & Sons, Ltd. Chap. 3, pp. 75–105. ISBN: 9783527653416. DOI: 10.1002/9783527653416.ch3.

[79] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. „Optimization by Simulated Annealing“. In: *Science* 220.4598 (1983), pp. 671–680. DOI: 10.1126/science.220.4598.671.

[80] Nicholas Metropolis et al. „Equation of State Calculations by Fast Computing Machines“. In: *The Journal of Chemical Physics* 21.6 (June 1953), pp. 1087–1092. ISSN: 0021-9606. DOI: 10.1063/1.1699114.

[81] W. K. Hastings. „Monte Carlo sampling methods using Markov chains and their applications“. In: *Biometrika* 57.1 (Apr. 1970), pp. 97–109. ISSN: 0006-3444. DOI: 10.1093/biomet/57.1.97.

[82] James L. McClelland, David E. Rumelhart, and PDP Research Group. *Parallel Distributed Processing, Volume 2: Explorations in the Microstructure of Cognition: Psychological and Biological Models*. The MIT Press, July 1986. ISBN: 9780262291262. DOI: 10.7551/mitpress/5237.001.0001.

[83] David E. Rumelhart, James L. McClelland, and PDP Research Group. *Parallel Distributed Processing, Volume 1: Explorations in the Microstructure of Cognition: Foundations*. The MIT Press, July 1986. ISBN: 9780262291408. DOI: 10.7551/mitpress/5236.001.0001.

[84] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. „Learning Internal Representations by Error Propagation“. In: *Parallel Distributed Processing, Volume 1: Explorations in the Microstructure of Cognition: Foundations*. The MIT Press, July 1986. ISBN: 9780262291408. DOI: 10.7551/mitpress/5236.003.0012.

[85] R. J. Williams. „The Logic of Activation Functions“. In: *Parallel Distributed Processing, Volume 1: Explorations in the Microstructure of Cognition: Foundations*. The MIT Press, July 1986. ISBN: 9780262291408. DOI: 10.7551/mitpress/5236.003.0014.

[86] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. URL: http://www.deeplearningbook.org.

[87] Yoshua Bengio et al. „A neural probabilistic language model“. In: *Journal of Machine Learning Research* 3 (Mar. 2003), pp. 1137–1155. ISSN: 1532-4435. URL: https://www.jmlr.org/papers/v3/bengio03a.html.

[88] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. „Learning representations by back-propagating errors“. In: *Nature* 323.6088 (Oct. 1, 1986), pp. 533–536. ISSN: 1476-4687. DOI: 10.1038/323533a0.

[89] Atilim Gunes Baydin et al. „Automatic Differentiation in Machine Learning: a Survey“. In: *Journal of Machine Learning Research* 18.153 (2018), pp. 1–43. URL: http://jmlr.org/papers/v18/17-468.html.

[90] Léon Bottou and Olivier Bousquet. „The Tradeoffs of Large Scale Learning“. In: *Advances in Neural Information Processing Systems*. Ed. by J. Platt et al. Vol. 20. Curran Associates, Inc., 2007. URL: https://proceedings.neurips.cc/paper_files/paper/2007/file/0d3180d672e08b4c5312dcdafdf6ef36-Paper.pdf.

[91] Geoffrey E. Hinton. „A Practical Guide to Training Restricted Boltzmann Machines“. In: *Neural Networks: Tricks of the Trade: Second Edition*. Ed. by Grégoire Montavon, Geneviève B. Orr, and Klaus-Robert Müller. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 599–619. ISBN: 978-3-642-35289-8. DOI: 10.1007/978-3-642-35289-8_32.

[92] Yukiyoshi Ohnishi et al. „Accelerator design at SuperKEKB“. In: *Progress of Theoretical and Experimental Physics* 2013.3 (Mar. 2013), 03A011. ISSN: 2050-3911. DOI: 10.1093/ptep/pts083.

[93] H. Bethe. „Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie“. In: *Annalen der Physik* 397.3 (1930), pp. 325–400. DOI: 10.1002/andp.19303970303.

[94] M. Stanley Livingston and H. A. Bethe. „Nuclear Physics C. Nuclear Dynamics, Experimental“. In: *Reviews of Modern Physics* 9 (3 July 1937), pp. 245–390. DOI: 10.1103/RevModPhys.9.245.

[95] Yung-Su Tsai. „Pair production and bremsstrahlung of charged leptons“. In: *Reviews of Modern Physics* 46 (4 Oct. 1974), pp. 815–851. DOI: 10.1103/RevModPhys.46.815.

[96] Claude Leroy and Pier-Giorgio Rancoita. „Physics of cascading shower generation and propagation in matter: principles of high-energy, ultrahigh-energy and compensating calorimetry“. In: *Reports on Progress in Physics* 63.4 (Apr. 2000), p. 505. DOI: 10.1088/0034-4885/63/4/202.

[97] Kazunori Akai, Kazuro Furukawa, and Haruyo Koiso. „SuperKEKB collider“. In: *Nucl. Instrum. Meth.* A907 (2018), p. 188. DOI: 10.1016/j.nima.2018.08.017.

[98] T. Abe et al. *Belle II Technical Design Report*. 2010. DOI: 10.48550/arxiv.1011.0352.

[99] SuperB Collaboration. „SuperB: A High-Luminosity Asymmetric e^+ e^- Super Flavor Factory. Conceptual Design Report“. In: *arXiv* (2007). DOI: 10.48550/arxiv.0709.0451.

[100] E Kou et al. „The Belle II Physics Book“. In: *Progress of Theoretical and Experimental Physics* 2019.12 (Dec. 2019). ISSN: 2050-3911. DOI: 10.1093/ptep/ptz106.

[101] T. Aushev et al. „Physics at Super B Factory“. In: *arXiv* (2010). DOI: 10.48550/arxiv.1002.5012.

[102] K. Adamczyk et al. „The design, construction, operation and performance of the Belle II silicon vertex detector“. In: *Journal of Instrumentation* 17.11 (Nov. 2022), P11042. DOI: 10.1088/1748-0221/17/11/P11042.

[103] Katsuro Nakamura et al. „The Belle II SVD detector“. In: *Proceedings of The 25th International workshop on vertex detectors — PoS(Vertex 2016)*. Vol. 287. 2017, p. 012. DOI: 10.22323/1.287.0012.

[104] Christian Wolfgang Fabjan and Herwig Schopper, eds. *Particle Physics Reference Library: Volume 2: Detectors for Particles and Radiation*. Cham: Springer Nature, 2020. ISBN: 978-3-030-35318-6. DOI: 10.1007/978-3-030-35318-6.

[105] Valerio Bertacchi et al. „Track finding at Belle II“. In: *Computer Physics Communications* 259 (Feb. 2021), p. 107610. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2020.107610.

[106] Nils 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: Karlsruhe Institute of Technology (KIT), 2015. URL: https://publish.etp.kit.edu/record/21234.

[107] Umberto Tamponi. „The TOP counter of Belle II: Status and first results“. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 952 (2020). 10th International Workshop on Ring Imaging Cherenkov Detectors (RICH 2018), p. 162208. ISSN: 0168-9002. DOI: 10.1016/j.nima.2019.05.049.

[108] Dmitri Kotchetkov et al. „Front-end electronic readout system for the Belle II imaging Time-Of-Propagation detector“. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 941 (2019), p. 162342. ISSN: 0168-9002. DOI: 10.1016/j.nima.2019.162342.

[109] H Kichimi et al. „KEKB beam collision stability at the picosecond timing and micron position resolution as observed with the Belle detector“. In: *Journal of Instrumentation* 5.03 (Mar. 2010), P03011. DOI: 10.1088/1748-0221/5/03/P03011.

[110] R. Pestotnik et al. „The aerogel Ring Imaging Cherenkov system at the Belle II spectrometer“. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 876 (2017). The 9th international workshop on Ring Imaging Cherenkov Detectors (RICH2016), pp. 265–268. ISSN: 0168-9002. DOI: 10.1016/j.nima.2017.04.043.

[111] T. Iijima et al. „A novel type of proximity focusing RICH counter with multiple refractive index aerogel radiator“. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 548.3 (2005), pp. 383–390. ISSN: 0168-9002. DOI: 10.1016/j.nima.2005.05.030.

[112] Kenkichi Miyabayashi. „Belle electromagnetic calorimeter“. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 494.1 (2002). Proceedings of the 8th International Conference on Instrumentatio n for Colliding Beam Physics, pp. 298–302. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(02)01483-3.

[113] T. Kuhr et al. „The Belle II Core Software“. In: *Comput. Softw. Big Sci.* 3.1 (2019), p. 1. DOI: 10.1007/s41781-018-0017-9.

[114] *Belle II Analysis Software Framework (basf2)*. https://doi.org/10.5281/zenodo.5574115. DOI: 10.5281/zenodo.5574115.

[115] A. J. Bevan et al. „The physics of the *B* factories“. In: *Eur. Phys. J.* C74 (2014), p. 3026. DOI: 10.1140/epjc/s10052-014-3026-9.

[116] *basf2 light-2305-korat documentation*. Belle II software documentation of light-2305-korat. URL: https://software.belle2.org/light-2305-korat/sphinx/index.html (visited on 03/2024).

[117] Sebastian Skambraks et al. „A 3D track finder for the Belle II CDC L1 trigger“. In: *Journal of Physics: Conference Series* 1525.1 (Apr. 2020), p. 012102. DOI: 10.1088/1742-6596/1525/1/012102.

[118] S. Bähr et al. *The Neural Network First-Level Hardware Track Trigger of the Belle II Experiment*. 2024. DOI: 10.48550/arxiv.2402.14962.

[119] *Total recorded integrated luminosity up to March 2024*. Belle II Luminosity web page. URL: https://confluence.desy.de/display/BI/Belle+II+Luminosity (visited on 03/2024).

[120] F. Abudinén et al. „Measurement of the integrated luminosity of the Phase 2 data of the Belle II experiment“. In: *Chinese Physics C* 44.2 (Jan. 2020), p. 021001. DOI: 10.1088/1674-1137/44/2/021001.

[121] B. Aubert et al. „Measurement of the total width, the electronic width, and the mass of the $\Upsilon(10580)$ resonance“. In: *Physical Review D* 72 (3 Aug. 2005), p. 032005. DOI: 10.1103/PhysRevD.72.032005.

[122] S. Jadach, B. F. L. Ward, and Z. Was. „The precision Monte Carlo event generator KK for two-fermion final states in e^+e^- collisions“. In: *Computer Physics Communications* 130 (2000), p. 260. DOI: 10.1016/S0010-4655(00)00048-5.

[123] S. Jadach, B. F. L. Ward, and Z. Was. „Coherent exclusive exponentiation for precision Monte Carlo calculations“. In: *Physical Review* D 63 (2001), p. 113009. DOI: 10.1103/PhysRevD.63.113009.

[124] Sw. Banerjee et al. „Monte Carlo Event Generator updates, for τ pair events at Belle II energies“. In: *16th International Workshop on Tau Lepton Physics*. Nov. 2021. DOI: 10.48550/arxiv.2111.05914.

[125] O. Shekhovtsova et al. „Resonance chiral Lagrangian currents and τ decay Monte Carlo“. In: *Physical Review* D 86 (2012), p. 113008. DOI: 10.1103/PhysRevD.86.113008.

[126] M. Chrzaszcz et al. „TAUOLA of τ lepton decays—framework for hadronic currents, matrix elements and anomalous decays“. In: *Computer Physics Communications* 232 (2018), pp. 220–236. DOI: 10.1016/j.cpc.2018.05.017. URL: https://arxiv.org/abs/1609.04617.

[127] I. M. Nugent et al. „Resonance chiral Lagrangian currents and experimental data for $\tau^- \to \pi^-\pi^-\pi^+\nu_\tau$“. In: *Physical Review* D 88 (2013), p. 093012. DOI: 10.1103/PhysRevD.88.093012.

[128] Elisabetta Barberio, Bob van Eijk, and Zbigniew Was. „PHOTOS: A universal Monte Carlo for QED radiative corrections in decays“. In: *Computer Physics Communications* 66 (1991), p. 115. DOI: 10.1016/0010-4655(91)90012-A.

[129] Torbjörn Sjöstrand et al. „An Introduction to PYTHIA 8.2“. In: *Computer Physics Communications* 191 (2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024.

[130] G. Balossini et al. „Matching perturbative and parton shower corrections to Bhabha process at flavour factories“. In: *Nuclear Physics B* B 758 (2006), pp. 227–253. DOI: 10.1016/j.nuclphysb.2006.09.022.

[131] G. Balossini et al. „Photon pair production at flavour factories with per mille accuracy“. In: *Physics Letters B* B 663 (2008), pp. 209–213. DOI: 10.1016/j.physletb.2008.04.007.

[132] C. M. Carloni Calame et al. „The BABAYAGA event generator“. In: *Nucl. Phys.* **B** *Proc. Suppl.* 131 (2004). Ed. by Marco Incagli and G. Venanzoni, pp. 48–55. DOI: 10.1016/j.nuclphysbps.2004.02.008.

[133] C. M. Carloni Calame. „An improved parton shower algorithm in QED“. In: *Physics Letters B* B 520 (2001), pp. 16–24. DOI: 10.1016/S0370-2693(01)01108-X.

[134] C. M. Carloni Calame et al. „Large angle Bhabha scattering and luminosity at flavor factories“. In: *Nuclear Physics B* B 584 (2000), pp. 459–479. DOI: 10.1016/S0550-3213(00)00356-4.

[135] F.A. Berends, P.H. Daverveldt, and R. Kleiss. „Radiative corrections to the process $e^+e^- \to e^+e^-\mu^+\mu^-$“. In: *Nuclear Physics B* B 253 (1985), pp. 421–440. DOI: 10.1016/0550-3213(85)90540-1.

[136] F.A. Berends, P.H. Daverveldt, and R. Kleiss. „Complete lowest-order calculations for four-lepton final states in electron-positron collisions“. In: *Nuclear Physics B* B 253 (1985), pp. 441–463. DOI: 10.1016/0550-3213(85)90541-3.

[137] F.A Berends, P.H Daverveldt, and R Kleiss. „Monte Carlo simulation of two-photon processes: II: Complete lowest order calculations for four-lepton production processes in electron-positron collisions“. In: *Computer Physics Communications* 40.2 (1986), pp. 285–307. DOI: 10.1016/0010-4655(86)90115-3.

[138] S. Uehara. „TREPS: A Monte-Carlo Event Generator for Two-photon Processes at e^+e^- Colliders using an Equivalent Photon Approximation“. In: *arXiv* (July 1996). DOI: 10.48550/arxiv.1310.0157.

[139] S. Agostinelli et al. „GEANT4: A simulation toolkit“. In: *Nucl.Instrum.Meth.* A506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.

[140] Thomas Bayes. „LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S“. In: *Philosophical Transactions of the Royal Society of London* 53 (1763), pp. 370–418. DOI: 10.1098/rstl.1763.0053.

[141] Marco Milesi, Justin Tan, and Phillip Urquijo. „Lepton identification in Belle II using observables from the electromagnetic calorimeter and precision trackers“. In: *EPJ Web of Conferences* 245 (2020), p. 06023. DOI: 10.1051/epjconf/202024506023.

[142] Paul Feichtinger. „Search for an invisibly decaying Z' boson and study of particle identification at the Belle II experiment“. MA thesis. Wien: Technische Universität Wien, 2021. DOI: 10.34726/hss.2021.84843.

[143] S. Brandt et al. „The Principal axis of jets. An Attempt to analyze high-energy collisions as two-body processes“. In: *Physics Letters B* 12 (1964), pp. 57–61. DOI: 10.1016/0031-9163(64)91176-X.

[144] Edward Farhi. „A QCD Test for Jets“. In: *Physical Review Letters* 39 (1977), pp. 1587–1588. DOI: 10.1103/PhysRevLett.39.1587.

[145] Paul Feichtinger. *Selanneal python package*. Version 0.0.3. Github: https://github.com/feichtip/selanneal/. URL: https://pypi.org/project/selanneal/ (visited on 03/2024).

[146] P. A. Zyla et al. „Review of Particle Physics“. In: *PTEP* 2020.8 (2020), p. 083C01. DOI: 10.1093/ptep/ptaa104.

[147] Leo Breiman. In: *Machine Learning* 45.1 (2001), pp. 5–32. ISSN: 0885-6125. DOI: 10.1023/a:1010933404324.

[148] *Permutation feature importance*. scikit-learn web page. URL: https://scikit-learn.org/stable/modules/permutation_importance.html (visited on 03/2024).

[149] F. Pedregosa et al. „Scikit-learn: Machine Learning in Python“. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[150] *Permutation Importance vs Random Forest Feature Importance (MDI)*. scikit-learn web page. URL: https://scikit-learn.org/stable/auto_examples/inspection/plot_permutation_importance.html (visited on 03/2024).

[151] Steve Baker and Robert D. Cousins. „Clarification of the use of CHI-square and likelihood functions in fits to histograms“. In: *Nuclear Instruments and Methods in Physics Research* 221.2 (1984), pp. 437–442. ISSN: 0167-5087. DOI: 10.1016/0167-5087(84)90016-4.

[152] Lukas Heinrich, Matthew Feickert, and Giordon Stark. *scikit-hep/pyhf*. Version v0.7.1. Apr. 2023. DOI: 10.5281/zenodo.7807148.

[153] Kyle Cranmer et al. *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*. June 2012. URL: https://cds.cern.ch/record/1456844/files/CERN-OPEN-2012-016.pdf.

[154] pyhf developers. *Introduction to pyhf*. 2024. URL: https://pyhf.readthedocs.io/en/v0.7.6/intro.html (visited on 03/2024).

[155] Tadeáš Bilka. „Time-dependent CP violation $B^0 \to \eta_c K^0_S$ decays at the Belle experiment, Alignment of the Belle II detector". PhD thesis. Charles U., 2022.

[156] J H Friedman. *Data analysis techniques for high energy particle physics*. 1974. DOI: 10.5170/CERN-1974-023.271.

[157] E. Richter-Was. „QED bremsstrahlung in semileptonic B and leptonic tau decays". In: *Physics Letters B* 303.1 (1993), pp. 163–169. ISSN: 0370-2693. DOI: 10.1016/0370-2693(93)90062-M.

[158] Swagato Banerjee et al. „Tau and muon pair production cross sections in electron-positron annihilations at $\sqrt{s} = 10.58$ GeV". In: *Physical Review D* 77 (5 Mar. 2008), p. 054012. DOI: 10.1103/PhysRevD.77.054012.

[159] Annette M. Molinaro, Richard Simon, and Ruth M. Pfeiffer. „Prediction error estimation: a comparison of resampling methods". In: *Bioinformatics* 21.15 (May 2005), pp. 3301–3307. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/bti499.

[160] Tadeas Bilka et al. „Alignment for the first precision measurements at Belle II". In: *EPJ Web of Conferences* 245 (2020). Ed. by C. Doglioni et al., p. 02023. DOI: 10.1051/epjconf/202024502023.

[161] Tadeas Bilka et al. „Simultaneous Global and Local Alignment of the Belle II Tracking Detectors". In: *EPJ Web of Conferences* 251 (2021), p. 03028. DOI: 10.1051/epjconf/202125103028.

[162] M. Fujikawa et al. „High-statistics study of the $\tau^- \to \pi^-\pi^0\nu_\tau$ decay". In: *Physical Review D* 78 (7 Oct. 2008), p. 072006. DOI: 10.1103/PhysRevD.78.072006.

[163] Erik Bartos et al. „Generalized Gounaris-Sakurai formula and $\rho^0(770)$, $\rho^0(1450)$ and $\rho^0(1700)$ masses and widths". In: *Proceedings of The European Physical Society Conference on High Energy Physics — PoS(EPS-HEP2017)*. Vol. 314. 2018, p. 730. DOI: 10.22323/1.314.0730.

[164] S. Schael et al. „Branching ratios and spectral functions of tau decays: Final ALEPH measurements and physics implications". In: *Phys. Rept.* 421 (2005), pp. 191–284. DOI: 10.1016/j.physrep.2005.06.007.

[165] P. Abreu et al. „Measurements of the leptonic branching fractions of the tau". In: *The European Physical Journal C* 10 (1999), pp. 201–218. DOI: 10.1007/s100520050583.

[166] M. Acciarri et al. „Measurement of the τ branching fractions into leptons". In: *Physics Letters B* 507 (2001), pp. 47–60. DOI: 10.1016/S0370-2693(01)00294-5.

[167] G. Abbiendi et al. „A Measurement of the $\tau^- \to e^-\bar{\nu}_e\nu_\tau$ branching ratio". In: *Physics Letters B* 447 (1999), pp. 134–146. DOI: 10.1016/S0370-2693(98)01553-6. URL: https://arxiv.org/abs/hep-ex/9812017.

[168] G. Abbiendi et al. „A Measurement of the $\tau^- \to \mu^-\bar{\nu}_\mu\nu_\tau$ branching ratios". In: *Physics Letters B* 551 (2003), pp. 35–48. DOI: 10.1016/S0370-2693(02)03020-4.

[169] H. Albrecht et al. „Measurement of exclusive one prong and inclusive three prong branching ratios of the tau lepton“. In: *Zeitschrift für Physik C* 53 (1992), pp. 367–374. DOI: 10.1007/BF01625895.

[170] A. Abada et al. „FCC Physics Opportunities“. In: *The European Physical Journal C* 79.6 (2019), p. 474. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-6904-3.

[171] A. Abada et al. „FCC-ee: The Lepton Collider“. In: *The European Physical Journal Special Topics* 228.2 (2019), pp. 261–623. ISSN: 1951-6401. DOI: 10.1140/epjst/e2019-900045-4.

[172] I. Adachi et al. „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“. In: *Physical Review Letters* 124 (14 Apr. 2020), p. 141801. DOI: 10.1103/PhysRevLett.124.141801.