Department of Physics

Ph.D. Thesis in Physics

Search for an Invisible Boson in Tau Decays with Early Belle II Data and Development of New Analysis Methods for Tau Physics

---

Author: Thomas Kraetzschmar

Thesis Submission Date: August 3, 2022

Technische Universität München
TUM School of Natural Sciences

Search for an Invisible Boson in Tau Decays with Early Belle II Data and Development of New Analysis Methods for Tau Physics

Thomas Michel Gerd Krätzschmar

Vollständiger Abdruck der von der TUM School of Natural Sciences der Technischen Universität München zur Erlangung eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Lorenzo Tancredi

Prüfer*innen der Dissertation:

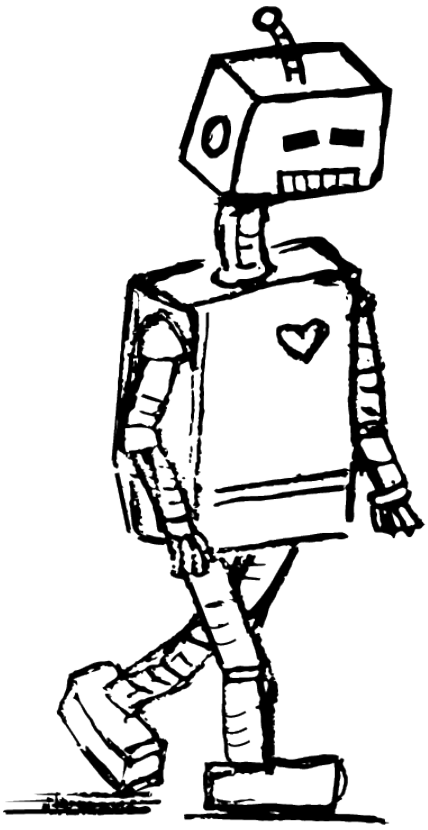
1. Hon.-Prof. Dr. Allen C. Caldwell
2. Prof. Dr. Lukas Heinrich

Die Dissertation wurde am 03.08.2022 bei der Technischen Universität München eingereicht und durch die TUM School of Natural Sciences am 03.11.2022 angenommen.

Dedication

To my father Gerhard Kraetzschmar

I know you would have loved reading this final thesis, I miss my most important critic and the discussions we would have had.
Rest in peace, I know you are watching from above



Acknowledgements

My research effort benefitted from the support of several people I would like to acknowledge gratefully.

First of all, I want to express my deep gratitude to Dr. Frank Simon. He granted me the opportunities, encouragement, and support I needed to pursue my research over the past four years. Frank, I could always approach you when I needed help and received all the support I could have wished for. Thank you!

I also want to thank my thesis advisor Prof. Allen Caldwell who always found the time when I needed his advice – even on short notice. Thank you!

My Future Detector and Belle II colleagues are fantastic people to work with. Regardless of whom I may have bothered with one of my questions, everyone was willing to help me, and it was fun having discussions with you. I am very grateful to Dr. Hendrik Winde, Dr. Thibaud Humair, and Dr. Swathi Sasikumar who were willing to read my thesis and provide me with valuable feedback. Thank you!

Furthermore, I had the great pleasure of working with Dr. Armine Rostomyan, Dr. Petar Rados, and Dr. Francesco Tenchini, whom I met in the DESY-tau group. They provided valuable support and helpful advice when I had questions. I immensely enjoyed our traditional tau-dinners whenever we met in Hamburg or at a B2GM in Tsukuba. Thank you!

Moreover, I want to thank my $\tau \to \ell\alpha$ search analysis team and and all my colleagues at Belle II for their hard work. Thanks to which we managed to pursue our analysis and get ready for publication. Thank you!

Finally, I want to thank my family, who supported me through all this time – the good, the bad, and the awful. My mother always supported me and laid the foundation for my success; I am deeply grateful for this. My father gave me great support and challenged me to do better. I miss you! A special thanks go to my wife, Elisabeth, who showed tremendous patience, understanding, and care. Discussing with her my thesis helped me to deliver the finishing touches. Thank you!

Abstract

The Belle II experiment at SuperKEKB – an asymmetric e^+e^- collider – aims at a total integrated luminosity of $50\,\mathrm{ab}^{-1}$ to pursue a rich program of Standard Model and Beyond the Standard Model physics. Belle II collected about $427.79\,\mathrm{fb}^{-1}$ at the $\Upsilon(4S)$ resonance until the middle of 2021.

This thesis presents a search for the hypothetical lepton-flavor violating process $\tau \to \ell + \alpha$(invisible), where τ is a tau particle that decays, ℓ is a lepton, and α is a Goldstone boson. Several models of beyond-the-Standard-Model physics predict this kind of topology, including Z' and axion-like particle models. With an examined dataset of $63\,\mathrm{fb}^{-1}$, this analysis significantly improves the limit compared to the previous best limit, obtained by ARGUS with an integrated luminosity of $475\,\mathrm{pb}^{-1}$.

Critical elements of the analysis are the event selection, reconstruction, and the interpretation of the overall event kinematics used to search for the two-body decay signature of the signal with an irreducible background of $\tau \to \ell \nu_\tau \nu_\ell$ decays. The event selection is a one-dimensional cut-based strategy. Advanced machine learning algorithms, MLA, were evaluated but did not show substantial gains with the same input parameters; MLAs make controlling systematic uncertainties harder.

A precise knowledge of the tau restframe is required to exploit the two-body kinematics of the signal decay. The analysis uses different strategies to cope with the missing ν_τ kinematic information, including a novel method developed explicitly for this search. This technique, referred to as generalised known kinematics, GKK, fully propagates the probability density function, pdf, of the ν_τ to obtain a pdf for the τ kinematics per event. Beyond the $\tau \to \ell\alpha$ decay, this thesis evaluates the potential of GKK to improve the τ mass measurement.

Upper limits for the branching ratio of $\tau \to \ell\alpha$ for several mass-hypothesis of α, m_α, are obtained with several frequentist approaches. The 95% upper limit on the ratio of branchin ratios,

$$Br(\tau \to \ell\alpha)/Br(\tau \to \ell\nu_\tau\nu_\ell),$$

is in the range of 10^{-2} to 10^{-3}, depending on m_α ranging from 0 to $1.6\,\mathrm{GeV}/c^2$.

A discussion of an electron identification performance study – crucial for the Belle II physics program – complements the physics analysis and its results.

Zusammenfassung

Das Belle II-Experiment am asymmetrischen e^+e^- Teilchenbeschleuniger SuperKEKB zielt auf eine integrierte Gesamtluminosität von 50 ab^{-1} ab, um ein vielseitiges Programm zur Physik des Standardmodells und jenseits des Standardmodells zu verfolgen. Belle II hat bis Mitte 2021 etwa 427,79 fb^{-1} an der $\Upsilon(4S)$-Resonanz gesammelt.

In dieser Arbeit wird die Suche nach dem hypothetischen Zerfall $\tau \to \ell + \alpha$(invisible) vorgestellt, wobei τ ein zerfallendes Tau-Teilchen, ℓ ein Lepton und α ein Goldstone-Boson ist. Mehrere Modelle die das Standard Model erweitern sagen diese Art von Topologie voraus, darunter Z' und axionartige Teilchenmodelle. Mit einem untersuchten Datensatz von 63 fb^{-1} verbessert diese Analyse den wahrscheinlichen Höchstwert der Zerfallsbreite im Vergleich zum bisherigen wahrscheinlichen Höchstwert, der durch ARGUS mit einer integrierten Luminosität von 475 pb^{-1} erzielt wurde, erheblich.

Kritische Elemente der Analyse sind die Datenauswahl, die Rekonstruktion und die Interpretation der gesamten Ereigniskinematik, die für die Suche nach der Zweikörper-Zerfallssignatur des Signals mit einem nicht reduzierbaren Hintergrund aus $\tau \to \ell \nu_\tau \nu_\ell$-Zerfällen verwendet wird. Bei der Datenauswahl handelt es sich um eine eindimensionale, schnittbasierte Strategie. Fortgeschrittene Algorithmen des maschinellen Lernens, MLA, wurden evaluiert, zeigten aber bei denselben Eingabeparametern keine wesentlichen Vorteile; MLAs erschweren die Kontrolle systematischer Unsicherheiten.

Eine genaue Kenntnis des Tau-Ruhesystems ist erforderlich, um die Zweikörperkinematik des Signals zu erkennen. Die Analyse verwendet verschiedene Strategien, um mit der fehlenden kinematischen Information umzugehen, darunter eine neue Methode, die speziell für diese Suche entwickelt wurde. Diese Technik, die als verallgemeinerte bekannte Kinematik, GKK, bezeichnet wird, propagiert die Wahrscheinlichkeitsdichtefunktion, pdf, der ν_τ vollständig, um eine pdf für die τ-Kinematik pro Ereignis zu erhalten. Über den $\tau \to \ell\alpha$-Zerfall hinaus wird in dieser Arbeit das Potenzial von GKK zur Verbesserung der τ-Massenmessung bewertet.

Obergrenzen für die Zerfallsbreite von $\tau \to \ell\alpha$ für verschiedene Massenhypothesen von α, m_α, werden mit verschiedenen frequentistischen Ansätzen ermittelt. Die 95%-Obergrenze für das Verhältnis der Zerfallsbreiten, $Br(\tau \to \ell\alpha)/Br(\tau \to \ell\nu_\tau\nu_\ell)$, liegt im Bereich von 10^{-2} bis 10^{-3}, abhängig von m_α im Bereich von 0 bis 1,6 GeV/c^2.

Eine Diskussion über eine Performance-Studie zur Elektronenidentifikation - entscheidend für das Belle II-Physikprogramm - ergänzt die physikalische Analyse und ihre Ergebnisse.

Contents

Part I

Introduction

The antique philosophers were the first known humans in the European and Mediterranean cultural areas who asked themselves how matter originated. They tried to give explanations based on reason. According to Aristotle [1], the first known Greek philosophers looking for the origin of all things are Thales of Miletus (624–546 B.C.E.), Anaximander (610–545 B.C.E), and Anaximenes of Miletus (d. 528 B.C.E.) of the Milesian school of philosophy [2]. Thales of Miletus is regarded as the founder of natural philosophy and was arguably the first to explain the origin of matter. He believed water to be the primary source [3]. His student Anaximander was critical of this argument. He identified the source of everything to be some unknown boundless source (Greek: “Apeiron") [1]. Anaximenes of Miletus, in turn, argued that air was the source of all things. He is probably also the first one who introduced the concept of material change [2]. Heraclius of Ephesus (540–480 B.C.E) was a critic of the Milesian scholars. He argued that change was an intrinsic property of the world, and he believed fire to be the source of everything [4]. Leucippus (450–370 B.C.E) is arguably the founder of Atomism [5]. Together with his student Democritus (460–370 B.C.E), he developed the antique theory of atoms, which are the smallest constituents of matter [6]. Plato (427–347 B.C.E) lived in the same epoch as Leucippus. He believed in the theory of the four elements – fire, water, air, and earth – which consist of geometric objects [7]. Aristotle (384–322 B.C.E), Plato’s student, added a fifth element, which he called aether. He did not believe in the creation of the universe but instead thought of it as ever-existing and everlasting [8].

For about 2000 years, these ideas were known, favoring the four or five-element theory of Plato and Aristotle. This worldview started to change during the renaissance and the enlightenment. Evidence for the atomic theory started to mount in natural sciences such as chemistry, which found basic building blocks. In 1661 Robert Boyle was one of the first people who developed chemistry from alchemy. He paved the way for introducing chemical elements and found a law that governs the relation of pressure and volume in a temperature-stable gas – today, it is known as Boyle’s law [9]. Daniele Bernoulli showed in 1740 that countless hits of particles could explain the constant pressure of the gas on a container wall. Furthermore, it allowed explaining Boyle’s law [10]. In 1789 Antoine Lavoisier presented a list of chemical elements, substances he said are not dividable by chemical means [11]. John Dalton successfully explained integer numbers for chemical elements in any substance with the Atomic Hypothesis. At the beginning of the 20th century, the scientific community finally accepted the Atomic Hypothesis, when it was able to explain Brownian motion [12] – based on the work of Wiener [13] and Einstein [14, 15] and verified by Perrin [16].

Already before a wide range of the scientific community accepted the Atomic Hypothesis for chemical elements, evidence for a subatomic structure started to mount. In the 19th century, the structure of atoms

became evident. Dmitry Mendelejew was the first who proposed the periodic table known today, and he used it for predicting elements based on gaps in his table[17]. At the end of the 19th century, an acceleration of discoveries started to shed light on the nature of atoms. In 1896 Henri Becquerel reported observing radiation from Uranium [18]. In 1897 Joseph Thomson discovered the electron as part of the Atom [19]. Simultaneously Pierre and Marie Curie worked on radiation that originated from Atoms, coining the term radioactivity [20]. In 1899 Rutherford identified two kinds of radiation, $\alpha-$ and $\beta-$ radiation [21]. He was able to identify $\alpha-$radiation as Helium atoms with two positive charges [22]. Later, Rutherford used α-radiation to probe the structure of matter and found that Atoms must consist of a positively charged atomic nucleus [23]. The structure of the atomic nucleus became evident when Rutherford found the proton in 1919 [24] and James Chadwick the neutron in 1932 [25]. Discoveries of various particles continued, leading Murray Gell-Mann and George Zweig to propose the existence of quarks that make up the proton, neutron, and many of the particles observed until 1964 [26, 27].

Today, we have developed a compelling theory of matter and its interactions, called the Standard Model, and it can explain all effects in the laboratory. Furthermore, the Standard Model can even explain the universe's evolution until several seconds after the Big Bang. However, when we want to explain the universe that developed right after the Big Bang, we cannot explain the processes that resulted in the universe we know today. We must resort to the same means as the antique philosophers, crafting hypotheses based on well-formulated arguments.

With the known processes of the Standard Model, we would expect to have a barren universe – with much less matter – dominated by radiation. In reality, we live in a fertile environment, which allows for a magnitude of matter interactions and the development of human life on earth.

The predicted Standard Model processes produce an equal amount of matter and antimatter and annihilate an almost equal amount. We expect a low survival rate due to minor symmetry-breaking effects; unfortunately, these effects are not significant enough to explain the amount of matter observed in the universe today.

We still speculate on how the universe formed in the very early phases due to our inability to observe processes that enable us to test hypotheses and develop a theory beyond the Standard Model, which could explain the observed dominance of matter in the universe. Due to that, we are now forced to search for evidence of new theories that could bring us a step further towards understanding how the universe came into existence as we observe it today.

The Standard Model of particle physics is the most precise quantum theory describing the basic building blocks of matter and their interac-

tion up to date. The discovery of the Higgs particle spectacularly confirmed it in 2012.

Although the Standard Model is one of the most significant physics achievements yet, it cannot explain several observations in our universe. For example, we cannot incorporate the observed matter-antimatter asymmetry in the universe. We do not know why there is a grouping of particles, what determines the strengths of the four fundamental forces[1], or why we observe charged conjugation and parity, CP, symmetry violations in the weak force but not in another. Also, we observe the effects of a matter content in the universe unexplainable by visible matter to us. This problem is called the Dark Matter problem, and we must find out what Dark Matter is.

Furthermore, the Standard Model describes only the strong, weak, and electromagnetic interaction, ignoring gravitation. The Standard Model does not describe gravity; the best description is General Relativity – a classic field theory. We cannot simply combine these two models, but we must find a quantum description that incorporates gravity and yields new effects – interactions or particles – to verify this description.

Today, particle physicists search for signs leading to the post-Standard Model era and enable us to formulate a unified theory. An important clue comes from the magnetic moment (g-2) measurement of the muon μ[2] [28], which is more than four standard deviations away from the Standard Model prediction. Another one comes from lepton flavor universality measurements from LHCb [29], which deviates more than three standard deviations. New physics models describing these observations will inevitably lead to new particles, probably a boson – if angular momentum conservation holds.

Suppose the boson is of spin $s = 0$. In that case, it could be an axion-like particle that allows for lepton flavor violating processes, potentially explaining the strong CP problem [3] [30]. If on the other hand it is spin $s = 1$, it could be a Z' which emerges from a symmetry group given by the lepton number of the muon μ and tau τ denoted as $U(1)_{L_\mu - L_\tau}$-symmetry [31, 32, 33, 34, 35]. The proposed bosons are good Dark Matter candidates and may explain the matter-antimatter asymmetry via Baryogenesis mechanisms.

B-factories are colliders designed to produce vast amounts of B-mesons. They run at the $\Upsilon(4S)$ resonance corresponding to 10.52 GeV center-of-mass collision-energy, $\sqrt{s}$. To measure time dependant CP vio-

[1] Being electromagnetism, the weak force, the strong force, and gravity

[2] Particle symbols without ± signs indicate all possible charges.

[3] The axion particle emerges as a solution to the problem that no Charged conjugation Parity transformation violation, CP violation, is observed in strong interactions (strong CP problem). Axion-like particles are a class of models constructed like axion particle models. In contrast to the axion, axion-like particles need not solve the strong CP problem.

lation, they operate with asymmetric beam energies. B-factories are the first experiments that found the Standard Model's CP violation.

This type of collider is characterized as the intensity frontier in high energy physics due to its high luminosity. This approach to measurements in particle physics contrasts the high energy frontier, which focuses on achieving the highest possible center-of-mass energy. B-factories could once again pave the way to a subsequent great success in particle physics, as they have done in the case of the Standard Model.

The latest generation of a B-factory is the SuperKEKB collider with the Belle II detector in Tsukuba. The collaboration built those machines to search for physics beyond the Standard Model and started physics runs in 2019.

One way to search for new physics phenomena is precision measurements of Standard Model parameters with B-decays, a check of correspondence of the measured with the predicted values. Another way is to use the enormous amount of e^+e^- collisions and find rare decays not yet discovered, which can only be explained by an expansion of the Standard Model.

Especially in the τ sector, SuperKEKB provides unprecedented opportunities with an estimated end of run data set of more than 45 billion τ-pairs, $\tau^+\tau^-$. With the expected amount of $\tau^+\tau^-$, we can test hypothetical charged lepton flavor violation predictions for various new physics models.

SuperKEKB and Belle II aim to collect $50\,\mathrm{ab}^{-1}$ of data. To do so, SuperKEKB aims for a world record instantaneous luminosity of about $6\times10^{35}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The Belle II collaboration designed the detector to deal with the background environment accompanying the ultra-high luminosity.

Charged lepton flavor violation tests belong to the most anticipated studies in the tau-physics program of SuperKEKB and Belle II. Lepton number conservation is a property of the Standard Model, introduced because no lepton number violating process was observed when the Standard Model was formulated. As we have found hints for deviations from the Standard Model predictions, models including charged lepton number violation are gaining increased attention.

At the early phase of SuperKEKB, we can start testing the Standard Model with the search for $\tau \to \ell\alpha$, with α being an invisible particle. It is a robust probe for models proposing a new boson. The ARGUS collaboration published the last result for the $\tau \to \ell\alpha$ decay, with a data set of about $472\,\mathrm{pb}^{-1}$. We search for a $\tau \to \ell\alpha$ decay in a Belle II dataset of $62.8\,\mathrm{fb}^{-1}$.

This thesis comprises four parts. The first part was the introduction. The second part introduces some aspects of physics at B-factories, both theoretical and experimentally.

The theory chapter first briefly introduces the Stanard Model and some concepts of Quantum Field Theories commonly used in new physics extensions. Afterward, it discusses selected extensions of the Standard Model, which are models that predict the $\tau \to \ell\alpha$ decay. The experimental chapter introduces the SuperKEKB collider and Belle II-detector, discussing relevant performance studies.

The third part describes the $\tau \to \ell\alpha$ search, highlighting studies relevant to this thesis. It contains three chapters, one for the event selection and background suppression, one for the determination of the approximated τ-rest frame, and one chapter which discusses the statistical treatment in our $\tau \to \ell\alpha$ search.

The fourth part gives an outlook and conclusion. The outlook suggests on how the $\tau \to \ell\alpha$ analysis might improve and how the Generalised Known Kinematics method – developed in this thesis – might improve future measurements of the τ-mass. The concludes summerizes this thesis.

The Appendix first provides a chapter that explicitly identifies this author's work in the context of the collaborative effort. The later chapters provide supplementary material.

Part II

Introduction to Physics at *B*-Factories

Chapter 1

Theory

In the last century, physicists developed through the interplay between theory and experiment a compelling description of subatomic processes – the Standard Model. The Standard Model is a tremendous success story because it describes all visible matter in the laboratory.

However, there are still unanswered questions that the Standard Model cannot answer. For example, we do not know all constituents of observed matter in the universe. Furthermore, we do not know how matter came into existence, even the matter we can describe. Also, we are wondering if there is a theory that can unify the description of gravity with the three forces of the Standard Model.

Recent experimental results indicate that there are effects we do not understand yet [28, 29], indicating physics beyond the Standard Model in the lepton sector. In many new physics models, we expect that the coupling of the new physics effect is related to the particle's mass. The tau lepton, τ, is the heaviest lepton we know. It is an excellent probe to search for new physics because its mass allows it to decay hadronically, offering a variety of possible new physics decay modes to study.

Conceptually, a decay of the τ to a lighter lepton, ℓ, and an invisible particle, α, has a very accessible searches signature, due to its two-body spectrum. In the rest frame of the τ we expect this decay, $\tau \to \ell\alpha$, to exhibit a narrow peak on top of a broad three-body spectrum for the ℓ's momentum spectrum. This decay gives strong bounds for popular new physics models. Moreover, it has a relatively weak limit set by the ARGUS collaboration [36].

The following chapter will briefly introduce the Standard Model and its problems, relevant for the $\tau \to \ell\alpha$ search. After that, we will discuss a selection of popular extensions of the Standard Model and important theoretical concepts in the context of the $\tau \to \ell\alpha$ search. At last, we will look at concrete examples of axion-like particle models and charged lepton flavor violating models, which expect the $\tau \to \ell\alpha$ decay.

This chapter is based – if not otherwise stated – on the following textbooks [37, 38, 39, 40]. Please refer to these books for a more in-depth discussion of the Standard Model.

1.1 The Standard Model of Particle Physics

This section introduces the Standard Model with an overview of the fundamental building blocks and forces. First, it introduces quarks and leptons, the fundamental particles of all matter. After that, it discusses the forces the particles interact with and form structures. Finally, it shows how mass is acquired.

Table 1.1: Elementary fermionic particles of the Standard Model, taken from [40]. We denote the particle generation as Gen. and the particle mass as M

	Leptons			Quarks		
Gen.	Particle	Q	M [GeV/c^2]	Particle	Q	M [GeV/c^2]
First	e^-	-1	5.11×10^{-4}	u	+2/3	0.005
	ν_e	0	$< 10^{-9}$	d	-1/3	0.003
Second	μ^-	-1	0.106	c	+2/3	1.3
	ν_μ	0	$< 10^{-9}$	s	-1/3	0.1
Third	τ^-	-1	1.78	t	+2/3	174
	ν_τ	0	$< 10^{-9}$	b	-1/3	4.5

In the Standard Model, there are three particle generations. Each particle in one generation has two "sibling" particles in the other generations, which differ only by mass and the generation grouping, commonly referred to as flavor.

Quarks and leptons are further groupings of the particles within a generation. A generation has two quarks and two leptons, and the two particles within a group form a so-called doublet. The position within a doublet is also a quantum number named isospin.

The up-quark, u, has an electric charge of $\frac{+2}{3}$. The higher flavor correspondents are the charm-quark, c, and the top-quark, t. The isospin counterpart of the u is the down-quark, d, with an electric charge of $\frac{-1}{3}$. The other flavor down-type particles are the strange-quark, s, and the bottom-quark, b.

The electron, e^-, is the first generation charged lepton, and its electric charge of $Q = -1$ historically defines the elementary charge. The electron's higher flavor correspondents are the muon, μ^-, and tau, τ^-, particles.

The electrically neutral neutrino, ν, completes the particle generation. It is fascinating in several aspects. For example, it has a low mass – not yet measured; and its mass eigenstate does not correspond to its flavor eigenstate, as is the case in the quark sector.

A flavor eigenstate corresponds to the charged lepton of a generation, be it an electron neutrino, ν_e; muon neutrino, ν_μ; or a tau neutrino,

ν_τ. The mass eigenstate of the neutrino changes its flavor over time[1] because the flavor and mass eigenstates are not aligned. The elementary particles, grouped into 3 generations, are listed in Table 1.1.

The constituents of visible matter are held together by two forces, the strong and the electromagnetic force. In both cases, the mediators of the forces are massless spin $s = 1$ particles, so-called gauge bosons. The strong force binds quarks into either two quark states, called mesons, or three quark states, called baryons[2]. The formation of quark states follows the law of color neutrality. The strong force has three different charges called colors. Mesons are two bound color and anti-color quarks. Baryons form from three different color charges, color-neutral if combined.

The gauge bosons which mediate the strong force are called gluons. There are eight types of gluons; each one corresponds to the force mediation of a color combination. Gluons also carry color charge, which leads to self-interactions. These self-interactions cause an effect called confinement. Confinement describes that the larger the distance between two particles, the stronger the force, leading to an effective finite length.

The first particle generation forms all stable matter. Here the positively charged u and the negatively charged d form two different baryons. The proton, p, is made out of two u and one d, and the neutron, n, is made out of one u and two d. p and n are the building blocks of atomic nuclei and are called nucleons.

The strong force also binds the color neutral p and n together. The nuclear binding mediators are color neutral light mesons such as π^0, which is a superposition of

$$\left|\pi^0\right\rangle = \frac{1}{\sqrt{2}}\left(|u\bar{u}\rangle - \left|d\bar{d}\right\rangle\right).$$

The outer part of an atom, the so-called electron shell, is populated by electrons e^-, the first generation lepton. Leptons are color neutral, so they are not participating in the strong force.

The electron is bound to the nucleus by the Electro-Magnetic (EM) force, and quantum mechanical rules govern the shell population. The EM force has one charge with two signs, positive and negative. A negative charge is attracted to a positive one and vice versa, whereas a negative charge repulses a negative and a positive charge a positive one. The mass- and chargeless photons, γ, mediate this force. The EM force has an infinite reach but decreases quadratically by distance.

Nature prefers to minimize energy – favoring a neutral EM system, a minimized potential system. The charged components of an atom are e^-

[1] Althought the physics community considers this property part of the Standard Model, it is already an extension that introduced ν masses. Before the observation of neutrino oscillation, they were considered massless.

[2] mesons and baryons are the most common quark state types, but there are more exotic types such as pentaquarks. The reader is referred to textbooks such as [40] or the PDG for further reading on this topic [34].

and $\boldsymbol{p}$; both have the electric charge of one. Due to this, there are as many electrons as protons in an atom. Molecules and more complex systems form because a filled atomic shell is energetically more advantageous.

The strong force overpowers the EM force between the $\boldsymbol{p}$ in the nucleus. At distances of about 1 fm, the strong force is about $\mathcal{O}(10^3)$ more potent than the EM force. Table 1.2 displays the relative strength of forces evaluated at a distance of 1 fm.

Table 1.2: The four known forces of nature with their corresponding mediators, the relative strengths are approximated for a distance of 1 fm. Taken from Reference [40].

Force	Strength	Boson		Spin	Mass [GeV]
Strong	1	Gluon	g	1	0
Electromagnetic	10^{-3}	Photon	$\boldsymbol{\gamma}$	1	0
Weak	10^{-8}	W boson	$\boldsymbol{W^{\pm}}$	1	80.4
		Z boson	$\boldsymbol{Z^0}$	1	91.2
Gravity	10^{-37}	Graviton?	G	2	0

The third column of Table 1.2 is the weak force. Since the weak force is about $\mathcal{O}(5)$ weaker at the low energy scale than the EM force, it is called weak. The weak force, or the weak interaction, enables a dynamic system in which particles can change their quantum state. The most prominent weak interaction is the β-decay.Massive $s = 1$ gauge bosons mediate the weak force. The masses of the weak force's gauge bosons are the reason for the minor effects at low energies. The charged $\boldsymbol{W}$ bosons are the mediators of the β-decays, as is shown in Figure 1.1. A weak interaction with an $\boldsymbol{W}$ is called a charged current. Furthermore, there is the $\boldsymbol{Z^0}$ which mediates weak neutral interactions, so-called neutral currents. Table 1.2 shows the masses of the $\boldsymbol{W}$ and $\boldsymbol{Z^0}$ as well as the coupling strength of the weak force. Standard Model particles acquire mass through the Higgs mechanism, which introduces the Higgs boson, $\boldsymbol{H}$, as a particle into the Standard Model.

Gravity is not part of the Standard Model but plays a vital role in the universe. For example, it enables the fusion processes in the sun as it provides a dense and hot environment enabling the particles to overcome the EM force and fuse into heavier nuclei. The best-known description of gravity is the General Relativity by Einstein. Models which include all four forces in one description are an active research field with various approaches.

1.2 Formal Concepts in Theory

The Standard Model is a so-called Quantum Field Theory (QFT). Quantum Field Theories successfully describe physical concepts, and exten-

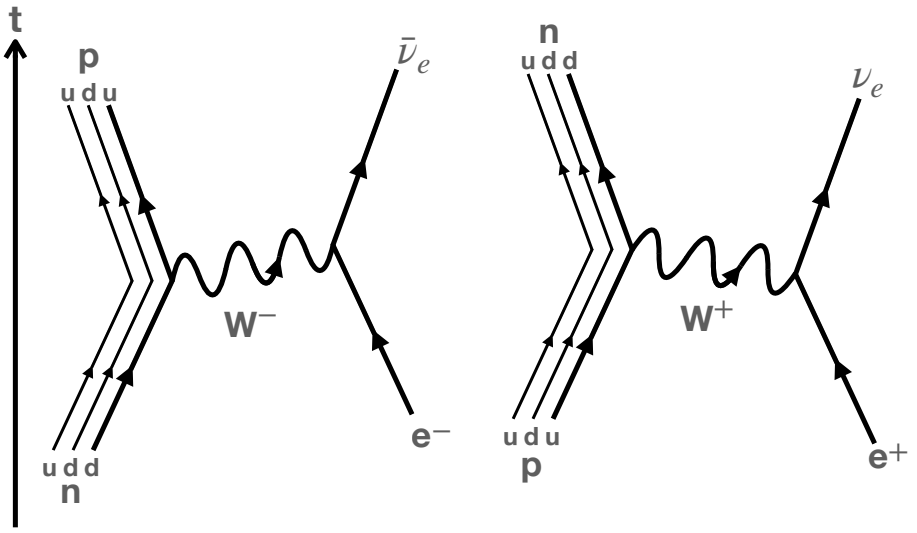


Figure 1.1: Feynman diagrams of the β-decays and inverse β-decays.

sions of the Standard Model use them widely. This section introduces selected concepts of a Quantum Field Theory used in $\tau \to \ell\alpha$ models. First, it introduces the concept of symmetries, showing that symmetries play an essential role in modern physics. After that, it presents the idea of Gauge Theories, which utilize symmetries to describe processes in physics. Finally, it explains the concept of spontaneous symmetry breaking, using the example of the Higgs mechanism, highlighting its importance in many beyond the Stanard Model theories.

1.2.1 Symmetries

In physics, symmetries are fundamental, as they imply conservation laws and provide insights into the dynamics of systems. In the mathematical description modern physics uses, group theory describes symmetries. A group is a set of symmetry operations or transformations, and a representation is an action of a transformation on an object.

If a symmetry exists, it implicates a law of physics invariant under certain transformations. The Lagrangian describing the physical system does not change under a group transformation, and this property means there is a conserved quantity, which is what Noether's theorem states. So, if we identify a new symmetry, we have found a new law in physics. There are two types of symmetries: continuous and discrete symmetries.

Continuous Symmetries

Continuous space-time symmetries give rise to the most fundamental physical laws. The invariance under translation in time leads to energy conservation, the invariance under translation in space leads to momentum conservation, and the invariance under spatial rotation gives rise to angular momentum conservation. These transformations are called

Poincaré transformations, a class of physics laws unchanged under any combination of space-time transformations.

This class of transformations is essential in particle physics as they define so-called gauge groups, used to describe the fundamental forces. Section 1.2.2 introduces the gauge principle.

Discrete Symmetries

The study of discrete symmetries was one of the drivers for developing the Standard Model. It is still one of the most active research topics in particle physics because a violation of these symmetries, or a combination of symmetries, implies a new process or property. The main symmetries are Parity, Charge Conjugation, and Time Reversal.

Parity A parity operation, $\hat{P}$, is the reversal of the spatial coordinates:

$$\hat{P}\vec{x} = -\vec{x}.$$

Parity has eigenstates in particle physics; we know them as particles and antiparticles. Forces that conserve parity, the electromagnetic and strong force, cannot change the parity eigenstate. For example, this property restricts atomic transitions to parity-conserving ones.

Charge Conjugation Charge Conjugation reverses all additive quantum mechanical numbers such as charges, flavor, and lepton number, to name a few. This operation transforms a particle into an antiparticle. Only particles that are their own antiparticles, e.g., photons, γ, and neutral composite particles, are eigenstates of Charge Conjugation.

Time reversal and CPT symmetry Time-reversal reverses all time-like processes such as momentum and angular momentum. That means a process with initial state A and final state B can also go from B to A.

The second law of thermodynamics violates the time-reversal on a macroscopic scale. On particle scales, this violation does not occur. Both electrodynamics and the strong force conserve time-reversal. Only the weak force violates time-reversal in correlation with the violation of Charge Conjugation and Parity, CP. The CPT theorem states that any quantum field theory invariant under Lorentz transformation conserves the combination of Charge Conjugation, Parity, and Time reversal, hence CPT.

Furthermore, it implies that particles and antiparticles are their exact opposite. If CPT holds, every broken discrete symmetry must cancel, e.g., the violation of CP by a time-reversal violation and vice versa.

1.2.2 Gauge Theory

In particle physics, the principle of gauge invariance has emerged to describe fundamental forces.

> A gauge transformation changes variables that leave all measurable quantities unaltered. ([39], p. 112)

A simple example is the force, $\vec{F}$, calculated from the potential energy, U. Here adding a constant c to the potential energy does not alter $\vec{F}$:

$$\vec{F} = -\nabla U = -\nabla(U + c) \tag{1.1}$$

So, if we are looking for a new model, we should find a new conservation law. In quantum mechanics, a conserved quantity is related to a symmetry of the Hamiltonian $\hat{H}$. Here, for each symmetry there is a unitary operator $\hat{U}$, $\hat{U}^\dagger \hat{U} = \mathbf{1}$, which commutes with $\hat{H}$, $[\hat{H}, \hat{U}] = 0$. We mentioned before that a group is a set of transformations. We can build the operator $\hat{U}$ from infinitesimal transformations

$$\hat{U}(\epsilon) = \mathbf{1} + i\epsilon\hat{G}, \tag{1.2}$$

with ϵ being an infinitesimal small parameter and $\hat{G}$ being the transformation generator. As $\hat{U}$ is unitary and $[\hat{H}, \hat{U}] = 0$, it follows that $\hat{G}^\dagger = \hat{G}$ and $[\hat{H}, \hat{G}] = 0$. So, the time evolution of $\hat{G}$ is conserved:

$$\frac{d}{dt}\left\langle \hat{G} \middle| \hat{G} \right\rangle = i\langle [\hat{H}, \hat{G}] \rangle = 0. \tag{1.3}$$

The gauge principle demands one symmetry for each conserved quantity in quantum mechanics. This demand also holds for the Lagrange formalism, used to describe particle interactions. Here the Lagrangian is invariant under a symmetry, equivalent to a conservation law.

We can now restate our search for a new conservation law as a search for a new symmetry as discussed in Section 1.3 in case of the $\boldsymbol{\tau \to \ell\alpha}$ analysis.

The gauge principle ultimately applied in particle physics is local gauge invariance. Local means that a system is locally invariant even if it contains a parameter $\theta(x)$ which can vary almost arbitrarily [3] over space. This property is essential as a local gauge-invariant Lagrangian leads to the requirement of massless gauge bosons.

1.2.3 The Higgs Mechanism - The Idea of Spontaneous Symmetry Breaking

The Higgs mechanism is the most prominent example of spontaneous symmetry breaking. It was introduced due to a problem well-summarised by Goldberg [39]:

> Gauge theories predict massless gauge bosons, but weak mediators have extremely high masses.

[3] It still has to conserve the Jacobi-identity, which is true most of the time.

The solution to this problem was the idea of spontaneous symmetry breaking implemented in the Higgs mechanism, which keeps local gauge symmetry and simultaneously gives mass to the gauge bosons. The Higgs mechanism introduces an additional particle, the Higgs bosons, first observed in 2012 by ATLAS [41] and CMS [42] at the LHC.

The concepts of spontaneous symmetry breaking are also widely applied in new physics models, such as models of the $\tau \to \ell\alpha$ decays because they enable new massive particles. The simplest model is a two-component integer spin field in a U(1) group,

$$\phi = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}, \tag{1.4}$$

which we can interpret as a complex number. A symmetric potential enabling spontaneous symmetry breaking is the so-called Mexican-hat-shaped potential of Figure 1.2. In the ground state of this potential, at $\phi_1 = \phi_2 = 0$, the setting is symmetric and the gauge boson of the model is massless. The symmetry breaks once the system's energy decreases, e.g., the universe cools down. Once the system moves away from the ground state, it falls into the degenerate minimum state of the potential and breaks the symmetry. Figure 1.2 illustrates the case of the Higgs mechanism. The transition is called electroweak phase transition in this case.

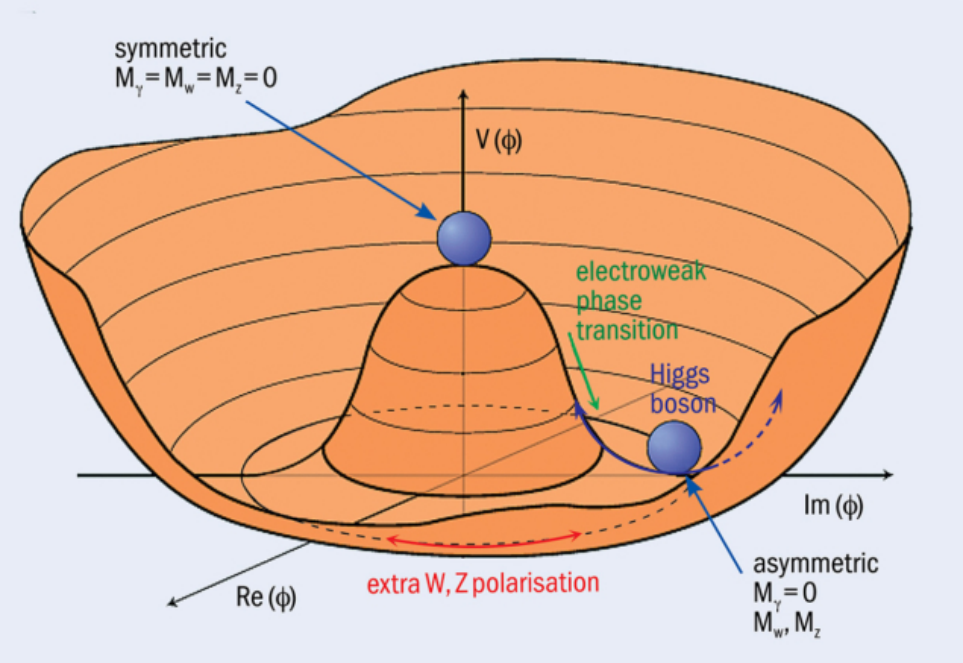


Figure 1.2: The Higgs Potential as described above with the components of ϕ, an U(1) doublet, being interpreted as a real and imaginary part of the field. As indicated at V_0, the gauge bosons are massless in an asymmetric state; once the symmetry is broken – the electroweak phase transition – the gauge bosons obtain mass. Credit: J Ellis/M Neubauer [43]

We can parametrize ϕ such that ϕ_1' describes a massless particle, a so-called Goldstone boson, and ϕ_2' a massive one. The Goldstone boson is a general consequence of this mechanism. We do not observe it in nature, but we choose the gauge in a local gauge theory such that the Goldstone boson disappears.

1.3 Beyond the Standard Model

The Standard Model is an incredible success story in describing physical processes in the laboratory.

In total, there are 26 free parameters in the Standard Model. We cannot explain these 26 parameters from any fundamental reasoning, but we have to accept them. From a theoretical point of view, this is unpleasant, especially since we can identify structures revolving around the families. For example, for the fermion masses on a wide mass range, or that the least required number of families for CP violation is exactly full-filled [38, 40]. There are other theoretical concerns.

The Standard Model interaction strengths vary significantly, but the coupling constants are relatively similar. The Standard Model parameters seem to be fine-tuned such that everything works out and we have a liveable universe. For instance, the strong force is not overpowering. Furthermore, it has no observed CP-violating neutron electric dipole moment, although the Standard Model allows it and shows maximal violation in the weak sector. The Standard Model is not able to describe physics at the Planck Scale, and the Higgs mass was measured much lighter than anticipated from loop corrections [38, 39].

These observations have two possible explanations: The first is that the parameters are accidental; however, this would mean we got fortunate to live in our universe. The second explanation is the existence of an underlying principle enabling us to predict all parameters [39]. For the second explanation, there are also further motivations.

Since the observation of neutrino oscillations, we know that neutrinos have mass. However, the Standard Model predicted mass-less neutrinos because the Higgs mechanism cannot explain the observed neutrino masses without adding a right handed neutrino and introducing a tiny coupling – compared to the other fermions. Furthermore, the Standard Model is not able to explain the baryon asymmetry in the universe [38].

There is numerous cosmological evidence for so-called Dark Matter, which the Standard Model cannot explain. For example, we expect the rotation velocity of stars in spiral galaxies to decrease with increasing radius, but we observe a relatively flat behaviour [44]. Evidence also comes from Galaxy Clusters observations, gravitational lensing, the observations from the Bullet cluster, and the Cosmic Microwave Background [40, 38]. Also, the currently accepted inflation model proposes the existence of dark energy, which is not part of the Standard Model.

There are more fundamental problems and questions unanswered. Can gravity be unified with the other forces [39, 38]? Why is the charge quantized? What is the reason for the number of space-time dimensions [38]?

The case for a theory beyond the Standard Model is thus well motivated. In the following, Section 1.3.1 gives a brief overview of some prominent extensions solving some but not all of the above problems, for example the strong CP problem. Section 1.3.2 discusses lepton flavor violation and presents two crucial model categories, namely the Z' and Axion Like Particles, ALP. At last, Section 1.4 motivates the lepton flavor violating search for an invisible boson in tau decays and presents the idea for the search.

1.3.1 The Strong CP-Problem and Axions

The descriptions follow [38, 39] if not stated otherwise. In general, we should include all terms of the Lagrangian that are compatible with gauge symmetry and renormalizable. In QCD, there exists a so-called θ-term violating CP. We can add

$$\mathcal{L}^{CP}_{QCD} = \theta \times \text{constants} \times \text{Fields} \tag{1.5}$$

to the Standard Model Lagrangian.

The CP-violating term in Equation (1.5) has little impact on effects involving perturbation theory, so it was overlooked for some time until physicists discovered that it has a low-energy effect, leading to an electromagnetic dipole moment of the neutron. The reason for this low-energy effect it that it contributes to the vacuum energy, which is given by

$$E(\theta) \propto m_\pi^2 \frac{m_u m_d}{(m_u + m_d)^2} \cos^2(\theta), \tag{1.6}$$

with the masses of the pion, m_π, the u, m_u, and the d, m_d. Because there are no constraints on the range of θ, one would expect maximal violation. However, when measuring the phase, it is tiny: $\theta < 10^{-10}$. So, the question arises why the phase is so small?

A solution to this problem is the Peccei-Quin-mechanism. Here a new chiral $U(1)_{PQ}$-symmetry is introduced. Equation 1.6 gives the potential, with the minimum at zero. This potential modifies the Lagrangian such that a new particle is introduced – the axion:

$$\mathcal{L}^{A}_{QCD} = (\bar{\theta} - \frac{\phi_A}{f_A}) \times \text{constants} \times \text{Fields} \tag{1.7}$$

ϕ_A is the axion field, f_A is the decay parameter that controls both couplings, and m_A is the axion mass given by the QCD-constraint in [30]:

$$m_A = 5.691(51)\mu eV(\frac{10^{12}\,\text{GeV}}{f_A}). \tag{1.8}$$

This model provides an automated phase cancellation due to the axion field. Furthermore, the new particle introduced is a Dark Matter candidate and in some baryogenesis models it can explain the baryon asymmetry in the universe.

Experimental searches focus on a range of

$$10^9 < f_A \leq 10^{12}\,\text{GeV}.$$

The lower bound emerges due to the demand for sufficiently strong coupling to matter such that it affects the cooling of stars. We can search for this effect in supernovae and red giants. The expected Dark Matter content of the universe gives the upper bound.

1.3.2 Lepton Flavor Violation

Historically, physicists thought lepton flavor would be violated. After discovering the muon in 1937, they regarded it as a heavy electron decaying as

$$\mu \to e + \gamma. \tag{1.9}$$

Physicists were surprised they never found this kind of decay, but instead, a three-body decay involving two invisible particles. Nishijima and Schwinger came up with the two-neutrino hypothesis forbidding lepton flavor violation, LFV. Measurements at the Brookhaven National Laboratory confirmed the neutrino theory in 1962. Ultimately this introduced lepton family conservation in the Standard Model, assuming mass-less ν.

Since discovering neutrino oscillations, we know that the lepton families are not conserved, and neutrinos have mass. It is now proofen that a conserved quantity introduced by experimental constraines is not conserved. Furthermore, because neutrinos show that lepton flavour is not conserved, we also expect charged lepton flavor violation, cLFV. In the Standard Model charged lepton flavor violation is induced by neutrino-oscillations, at a heavily suppressed scale of $\mathcal{O}(10^{-45})$ to $\mathcal{O}(10^{-54})$. There are also many different theories beyond the Standard Model predigting charged lepton flavor violating processes at a considerably higher scale ($\sim \mathcal{O}(10^{-45})$) than the Standard Model. These reasons are a strong motivation to study charged lepton flavor violation again.

Depending on the model, different yields are expected for decays of the type $\ell \to \ell'\gamma$ and $\ell \to \ell'\ell'\ell'$, with a lepton ℓ decaying into a lighter family lepton ℓ' [45]. Some of the proposed models have an enhanced cLFV coupling in the τ-sector, for example Z', or ALPs. Lepton flavor violations studies in the τ-sector are a complementary to collider searches concerned with lepton flavor violating Higgs-models. Furthermore, the τ is a unique probe as it allows to study of many leptonic and semi-leptonic decay modes giving insights into the cLFV couplings between quarks and leptons.

At Belle II, the most prominent decay-modes are $\tau \to 3\mu$, as it has a highly suppressed background and the sensitivity scales with luminosity. The other mode is $\tau \to \mu\gamma$ which has a higher expected branching ratio in many theories beyond the Standard Model than $\tau \to 3\mu$. The challenging part in this mode is the high background. Semi leptonic modes $\tau \to \mu h$, with h being any hadron, are also promising modes.

All named modes are of high theoretical interest but only competitive in the future when Belle II has at least the same amount of data available as its predecessors [46].

Recent results from the measurement of g-2 of the μ [28] and lepton universality [29] have produced hints for physics beyond the Standard Model. Two models have emerged as promising candidates from these results and are discussed in the following.

The Z' Model-Class

The basic idea of this model is that the Standard Model allows extending it with an observed, accidental symmetry [33], without producing anomalies and in agreement with experimental constraints. This new symmetry would give rise to a new gauge boson, the Z', with the properties:

- massive;
- electrically neutral;
- a colour singlet; and
- spin $s = 1$.

It can couple to quarks and leptons both to the right- and left-handed particles, with no constraints on the generations from first principles. The new symmetry must break spontaneously to be well-behaved at high-energy scales. The theories in which a Z' emerges vary, but we can identify four types.

A straightforward extension is adding a U(1) symmetry, which is also the most exciting type for this thesis. The U(1) type allows for various models. A prominent subclass is models with lepton-flavor-dependent charges. These models have anomaly-free solutions and can explain neutrino masses. Other types might incorporate fermion charges or extended Higgs models in a Grand Unified Theory.

Less prominent Z' models use more complicated symmetries; for example, some models extend the electroweak group. Furthermore, some Kaluza-Klein extensions, models proposing extra dimensions, include a series of Z'-boson pairs. At last, the Z' may arise as a composite particle [34].

Lepton Flavour Violating Axions and Axion Like Particles

This discussion follows the review paper of Calibbi, Redigolo, Ziegler, and Zupan [30].

Axions and ALPs are not only well-motivated by the strong CP problem and Dark Matter, but they can also solve the Standard Model flavor puzzle and neutrino-mass problem. The lepton flavor violation can emerge in the classic QCD-axion sector in the so-called Dine, Fischler, Srednicki, and Zhitnitsky (DFSZ)-models. ALP theories loosen the constraint of this QCD-axion sector, for example, the fixed strength-mass-relation. We can view ALPs as a generalization of the axion idea.

Typically, the lepton flavor violation in ALP theories relates to the neutrino sector, explaining the neutrino-mass problem. A widespread assumption is that neutrinos are Majorana particles. For example, there are so-called majoron models in which the ALP couples pred-dominantly to neutrinos through the Yukawa interactions making this type of model also a viable Dark Matter candidate for a weak enough neutrino coupling.

Other attractive ALP models, so-called familons, might not solve the strong CP problem, but they predict the lepton flavor violating couplings related to the neutrino mass texture. This type of model acts on the lepton sector. Other models are closely related to the DFSZ-models of the classic QCD-axion class but may use more complicated symmetry groups than the U(1).

1.4 Search for a New Invisible Boson in Tau Decays

The $\tau \to \ell\alpha$ search is model-independent. We only assume a two-body-decay topology of the τ decay in the search strategy, whereas the Standard Model distribution is a three-body topology. It has a broad distribution of the lepton momentum in the τ-rest frame, given by the weak interaction resulting in Formula B.8. Figure 1.3 shows the resulting distributions for both decays in the τ-rest frame.

It is interesting to discuss the context this search has in theoretical particle physics. The $\tau \to \ell\alpha$ topology is lepton flavor violating due to the missing ν_τ and ν_ℓ. Furthermore, if spin is conserved, α has to be a boson of spin $s = 0$ or $s = 1$.

In principle, various theories could explain $\tau \to \ell\alpha$. For example, we could identify it with a GUT in which lepton and baryon number violations are allowed such that our LFV-decay is possible. This thesis focuses on two classes of theories that are well-motivated and promising candidates.

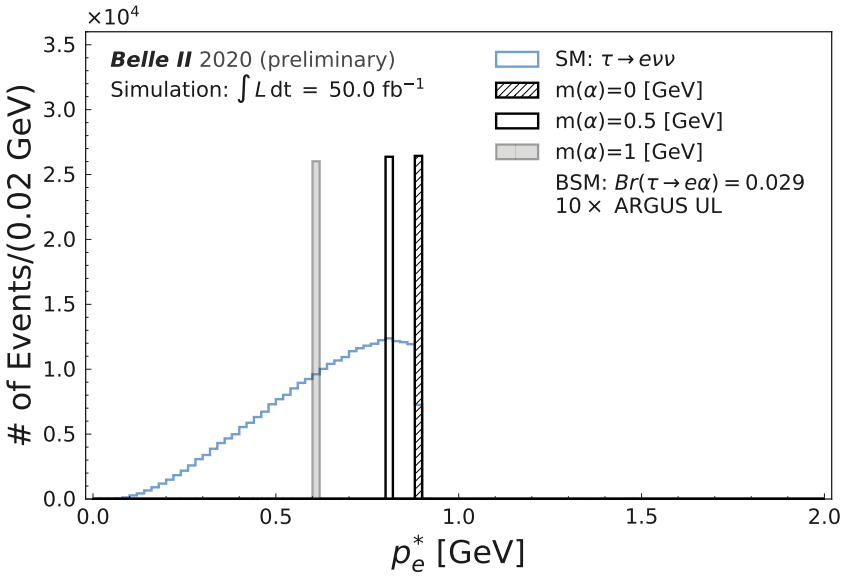


Figure 1.3: Theoretical expected Distributions for a $\tau \to e\nu_\tau\nu_e$ (SM: $\tau \to e\nu\nu$) and a $\tau \to e+\alpha$ decay with different mass-hypothesis for the α (m(α)= ... [GeV] with $c = 1$). We apply the assumed branching ratio for the $\tau \to \ell\alpha$ decay (BSM: $Br(\tau \to e\alpha) = 0.029$) to all mass hypotheses; we chose the branching to be ten times the upper limit provided by the ARGUS collaboration.

1.4.1 $\tau \to \ell\alpha$ in The Context of a Z'

The Z' is a spin $s = 1$ boson, motivated by the lepton flavor symmetry as discussed in Section 1.3.2. One of the primary motivations for Z' models is the muon g-2 discrepancy and its capability to avoid many bounds from cLFV searches [31, 32, 33, 34, 35].

If the hints of lepton universality violation observed by LHCb persist [29], physicists expect that we will find lepton flavor violation in charged leptons [31]. Other motivations for the Z' are that it could explain the large energy gap, a dip in the high-energy neutrino spectrum between 400 TeV and 1 PeV, found in IceCube [47, 48, 49, 50, 51, 52] and resolve the proton radius puzzle [35]. Furthermore, it can account for the neutrino-oscillation pattern, and it is an excellent Dark-Matter portal [33].

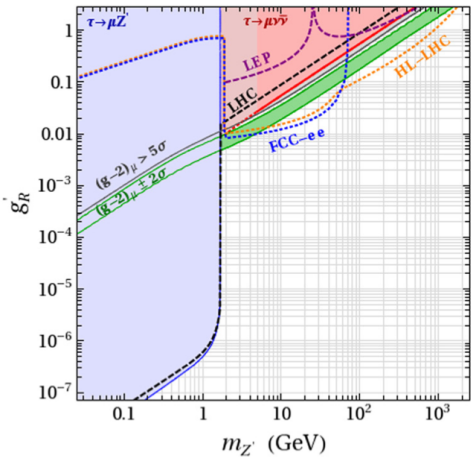
Due to the existing experimental bounds it was found that all models of the type $U(1)_{L_\alpha - L_\beta}$, with α and β being lepton families and $\alpha < \beta$, are ruled out except for $U(1)_{L_\mu - L_\tau}$ [33]. This surviving symmetry has interesting properties.

It is anomaly-free – does not show a strange and unobserved behaviour [32, 33, 35]. It allows to account for neutrino-oscillation patterns [33], has no additional fermions [32], and has a direct coupling of the Z' to the μ. The latter allows to avoid or suppress existing bounds but can explain the g-2 discrepancy of the μ [28, 32].

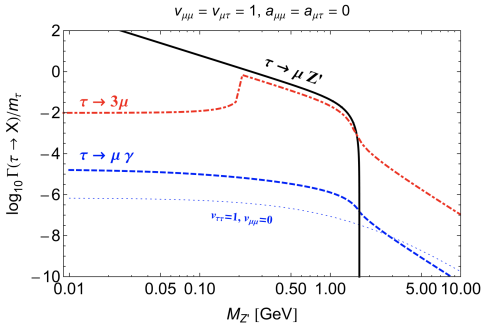
The $U(1)_{L_\mu - L_\tau}$ symmetry can generate Baryogenises through Leptoge-

nesis and can be constructed without right-handed neutrinos – avoiding problems in this field. The symmetry can have a vector or an axial-vector structure, possibly embedded into 2 Higgs-doublets [33]. Due to neutrino bounds, we expect that the mass of the Z' is below the τ mass, enabling the search in τ-decays.

The $\tau \to \ell\alpha$ search can provide strong bounds for Z' models as shown by [31, 35]. The expected bounds are shown in Figure 1.4. Figure 1.4a shows that the $\tau \to \ell\alpha$ search has a high sensitivity for Z' models. The result from ARGUS still exceeds modern constraints obtained at the LHC. In the mass range below $1.6\,\mathrm{GeV}/c^2$, the $\tau \to \ell\alpha$ bound rules out this particular Z' model's parameter space of interest – the one consistent with the observed discrepancy of the measurement of the muon g-2 with the Standard Model expectation. Figure 1.4b indicates that the Z' is still viable, and it shows that the $\tau \to \ell\alpha$ search yields are a complementary bound to other prominent cLFV τ decays searches.



(a) Z' model proposed by [31].



(b) Z' model proposed by [53].

Figure 1.4: The ARGUS results determines the upper limits for several different Z' models. $\tau \to \ell\alpha$ is a sensitive probe in the corresponding parameter space. In Picture (a) $m_{Z'}$ is the Z'-mass, g'_R is the right-handed coupling, the left-handed coupling is $g'_L = g'_R$. The green band is the 2σ-preferred range of the g-2 muon-anomaly. The red region is excluded at 2σ by lepton universality tests in τ-decays. ARGUS excludes the blue region at a 95% Confidence Level (CL). The black dashed line is the CL given by leptonic $W^{\pm}$-decay searches for the Z' at the LHC. The purple dashed line is determined from Z^0 coupling measurements at LEP. Both dashed lines exclude by a 95% CL. The blue and orange dotted lines are sensitivity prospects for Z^0-pole searches at FCC-ee and 3σ sensitivity searches for the $\mu^-\mu^+e^-e^+$ final state at the high-luminosity LHC, respectively. In Picture (b) $M_{Z'}$ is the Z'-mass, m_τ is the τ-mass, $\Gamma(\tau \to X)$ being the decay width to the modes $\mu Z'$, $\mu\mu\mu$ and $\mu\gamma$. Here a vector-like coupling is presented, denoted by $v_{\mu\mu} = v_{\mu\tau} = 1$ for the vector components and $a_{\mu\mu} = a_{\mu\tau} = 1$ for the axial-vector component

1.4.2 $\tau \to \ell\alpha$ in The Context of an Axion Like Particles

The following section is a summary of the $\tau \to \ell\alpha$ prospects of [30], but also summarises motivations from [33, 54, 55, 56].

The axion is a $s = 0$ particle initially motivated by the strong CP problem. Today, it spawned many axion-like particle models, which do not necessarily solve the strong CP problem anymore but share the same idea for the mechanism.

Section 1.3.2 discusses the lepton flavour violating ALPs. Aside from the strong CP problems, ALPs are also well motivated by Dark Matter, the Standard Model Flavour puzzle, and neutrino oscillations.

As in the Z' case, $\tau \to \ell\alpha$ has a role in the searches for ALPs. There are couplings to the τ and other leptons in the QCD-axion and ALP models. The sensitivity for these searches is often limited in the τ-sector due to the lack of data [30].

In the case of the so-called leptonic familons and majorons, the decay $\ell \to \ell'\alpha$ is enhanced and can probe a unique parameter-space – above the μ mass – as indicated in Figure 1.5.

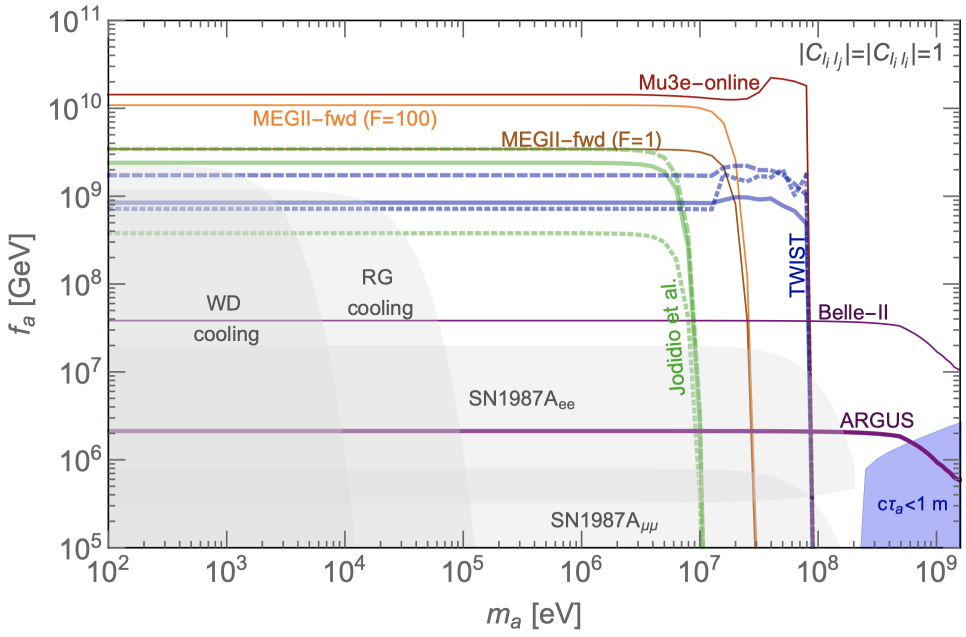


Figure 1.5: ALP parameter space calculated in [30]. f_a is the axion decay parameter and m_a is the axion/ALP-mass, as discussed in Section 1.3.1. The gray shaded regions are astrophysical bounds, the blue shaded region corresponds to a prompt/displaced ALP. All lines except for the purple ones are bounds from searches in the μ-sector. The thick purple line are the current Limits for $\tau \to \ell\alpha$ obtained by ARGUS. The thin purple line is the projection for Belle II.

Chapter 2

SuperKEKB and Belle II

The SuperKEKB collider is an e^+e^- collider built explicitly as the next generation B-Factory. A core feature is its high luminosity. SuperKEKB aims to deliver a 50 times greater data set than its predecessor KEKB.

To achieve this goal, SuperKEKB uses an increased beam current and a new collision scheme with reduced beam size – the nano-beam scheme. Achiving the luminosity goal via increased beam current alone would lead to intolerable beam background rates. The nano-beam scheme reduces the necessary increase in beam current, reducing beam background rates.

The basic idea of the nano-beam scheme is to squeeze the particle bunches at the interaction point to unprecedented small sizes of the order of a few tenths of nanometers. This beam squeezing leads to an increase in interaction probability and instantaneous luminosity.
Figure 2.1 illustrates the difference between the conventional collision scheme used by KEKB and the nano-beam scheme used by SuperKEKB. KEKB's interaction point is much greater compared to SuperKEKB.

The Belle II collaboration, from hereon simply referred to as Belle II, built a new detector based on the general principle used already for Belle. Changes were made to withstand the increased beam background environment and improve the detection performance compared to Belle.

Figure 2.2 illustrates the entire facility of SuperKEKB with the Belle II-detector located at the interaction point of SuperKEKB. The asymmetric beam scheme necessitates two separate storage rings for the positrons and electrons, and the two rings cross at the interaction point. The beams are created and accelerated in the injector linac. The injector linac includes a positron damping ring to improve the beam quality, necessary for the nano-beam scheme and reducing beam backgrounds in the process.

Section 2.1 discusses new techniques SuperKEKB utilizes to increase instantaneous luminosity while simultaneously reducing beam background. Measuring different beam background rates to evaluate the performance is crucial. Section 2.1.1 briefly introduces the CLAWS++ beam monitoring system, which contributes to the beam background reduc-

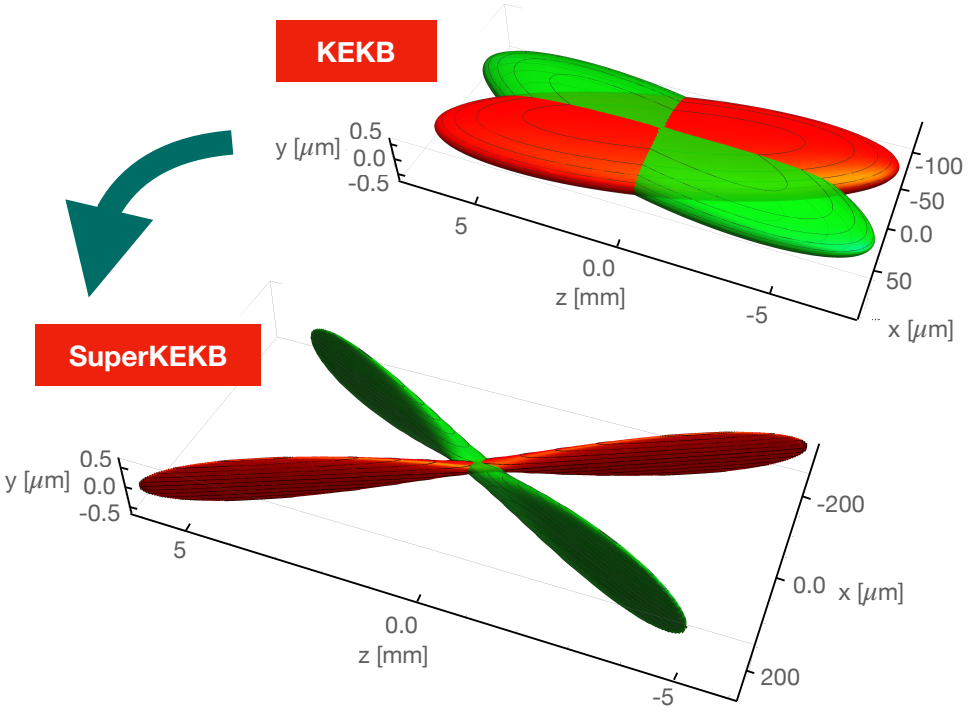


Figure 2.1: Comparison of the KEKB and SuperKEKB Interaction Point [57].

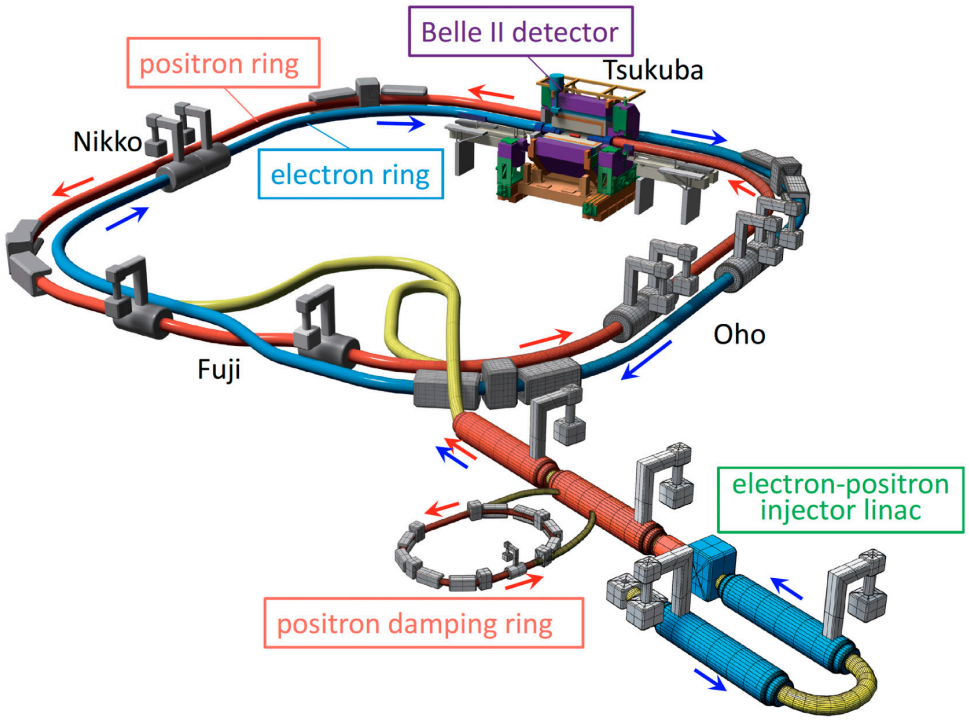


Figure 2.2: The SuperKEKB Collider [58]

tion and is part of this thesis work. As a basis for the discussion of the Belle II-detector, Section 2.2 gives an overview of fundamental principles for particle detectors. Especially the tracking and particle identification capabilities are crucial for the $\tau \to \ell\alpha$-analysis discussed in Part III. Finally, Section 2.4 discusses the electron identification performance, a crucial systematic uncertainty in the $\tau \to \ell\alpha$-analysis. This study is also part part of this thesis work.

If not stated otherwise, this chapter follows the Belle II physics book and technical design report [46, 59].

2.1 SuperKEKB

SuperKEKB is an asymmetric e^+e^--collider running at a center-of-mass of energy of $\sqrt{s} \approxeq 10.58$ GeV. It aims for a world-record instant luminosity of 6×10^{35} cm^{-2}s^{-1}, about 30 times more than its predecessor KEKB. The center-of-mass of energy corresponds to the $\Upsilon(4S)$ resonance, an excited state of the lightest $b\bar{b}$-meson, the $\Upsilon(1S)$. The $\Upsilon(4S)$ decays almost exclusively into a pair of B-mesons, $B\bar{B}$.

SuperKEKB has asymmetric beam characteristics optimized for B-meson physics, especially the measurement of time-dependent-CP violation. SuperKEKB operates two separate storage rings to achieve collisions, one for the e^--beam and one for the e^+-beam.

The High-Energy-Ring, HER, stores the 7 GeV e^--beam. The Low-Energy-Ring, LER, stores the 4 GeV e^+-beam. Both beams are accelerated to their respective energy by a linear accelerator and transferred into their separate storage ring.

SuperKEKB introduces two new technologies – continuous injections and the nano-beam scheme – to achieve its luminosity goal. We shortly recapitulate luminosity to understand the pursued approach.

The instantaneous luminosity $\mathbf{L}$ is the relation of the cross-section, σ, and event rate, $\frac{dN}{dt}$:

$$\mathbf{L} = \frac{1}{\sigma}\frac{dN}{dt}. \tag{2.1}$$

We can interpret the cross-section as the probability for a pair production, e.g. $\Upsilon(4S)$, and the event rate as the number of events N per time t. The idea of the nano-beam scheme is to increase the event rate by increasing the particle density at the interaction point.

Every collider groups particles into particle bunches. Each particle bunch consists of many particles of the same type. Due to their intrinsic charge, particles repel each other within a bunch. There is considerable space between the particles in a bunch; squeezing them to infinite density is impossible.

A three-dimensional Gaussian describes the shape of a bunch, with the beam size for each dimension defined as the respective standard de-

viation, σ_i with $i \in [x, y, z]$ [1].

The bunches are accelerated and collide at the interaction point with other particle bunches. The event rate gives the number of e^+e^- interactions per time. Due to the space in the particle bunches, most particles pass through the other bunch at the interaction point without interacting.

The idea of the nano-beam scheme is to squeeze the particle bunches at the interaction point such that there is a high particle density, leading to an increased event rate. Of course, the particles repel each other again, but they can be forced into a squeezed bunch for a short time at the interaction point before the particle bunch expands.

The nano-beam scheme idea also becomes apparent when rewriting:

$$\mathbf{L} = \frac{N_{e^+} N_{e^-}}{4\pi \sigma_x^* \sigma_y^*} f_c R_L. \tag{2.2}$$

σ_i* is the beam size at the interaction point – assuming the three-dimensional Gaussian shape, N_j is the number of particles in a bunch – with j being the e^- or e^+ bunch respectively, f_c is the frequency of bunch crossings, and R_L is the reduction factor accounting for geometrical efficiencies associated with the finite crossing angle and bunch length. Equation (2.2) shows that $\mathbf{L}$ increases by decreasing the beam-size σ_i* in the x- and y-direction, allowing for an increase in the z-direction, as Figure 2.1 illustrates.

e^+e^--collideres such as SuperKEKB have a beam lifetime of about ten minutes, which means they need to refill their particle bunches every ten minutes. Continuous injection aims to increase the event rate by keeping the bunch currents constant over time.

Traditionally, a ring accelerator first fills the storage rings with many particle bunches. The particles collide after completing the filling and until the beam current falls below a certain threshhold. This operation pattern leads to a dead-time with no collisions when the collider refills with particle bunches.

Continuous injection gets rid of the collision dead-time due to refilling. This filling pattern keeps the number of particle bunches constant and thus increases the averaged beam current over time.A disadvantage of continuous injection is that it increases the beam background. The reason is that an injected particle bunch has a cooldown phase in which it adjusts to the storage ring orbits, and this cooldown phase accompanies a higher potential loss of bunch particles than a cooled particle bunch.

[1] The beam direction defines the Cartesian coordinate system. z is in the direction of the e^--beam, y is the upward direction, and x points outwards of the collider-ring.

2.1.1 Beam Monitoring Detectors: CLAWS++

SuperKEKB has various beam monitoring detectors to steer the collider safely and increase its performance. The beam monitoring detectors are diverse in both setup and purpose. Discussing the CLAWS++ detector system provides multiple examples for beam monitoring applications.

The CLAWS++ detector consists of a three-by-three centimeter scintillator tile attached to a readout board. A Silicon Photo Multiplier (SiPM) on the readout board detects the scintillation light. We based the technology on the highly granular calorimeter technology the Max Planck Institute developed for the CALICE collaboration [60, 61].

CLAWS++ can detect ionizing particles and is primarily sensitive to charged particles, ideal for measuring beam background rates. Figure 2.3 illustrates the CLAWS++ hardware and location at Belle II. The left side shows a disassembled CLAWS++ module and how to assemble it. The right side indicates one of the eight positions of CLAWS++ module on one of the two QCS magnets.

The initial design purpose for CLAWS ++ was to monitor beam background rates caused by the continuous injection scheme. This monitoring is vital because parts of the Belle II-detector withstand only a specific background rate.

Many sources cause beam backgrounds, and the beam steering impacts these sources. For example, injecting new particle bunches into the collider ring increases the beam background while the new bunches

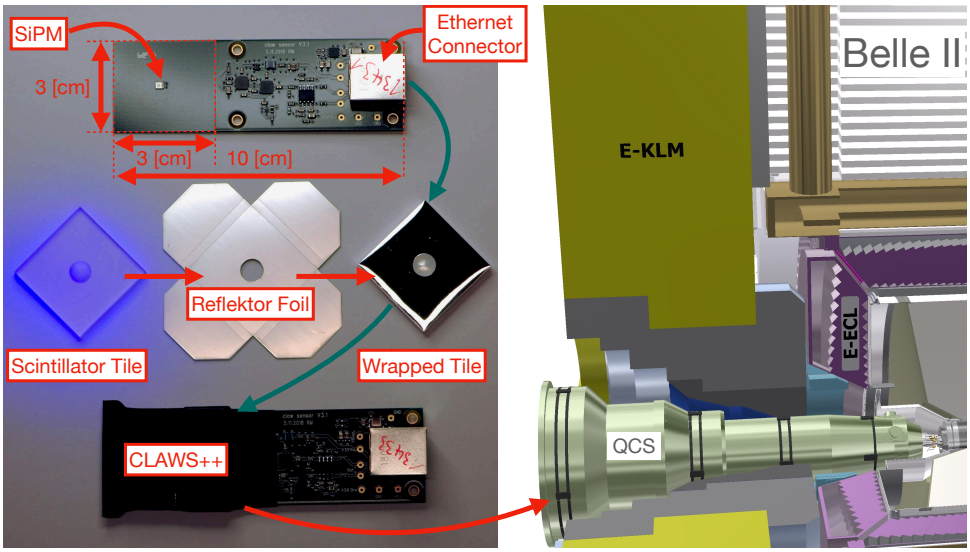


Figure 2.3: Left shows the main components of the CLAWS ++ detector. The top left part displayes the readout board with all electronic components, including the SiPM. The middle part is the scintillator tile and its light-reflective wrapping, and the bottom part is the final CLAWS detector. It consists of the readout board and wrapped scintillator tile packaged in light-tight tape. Right shows the CLAWS++ detectors attached to the QCS magnet; the arrow indicates a position [62].

cool down and adjust. SuperKEKB can improve the time needed for the cooldown and adjustment of new bunches. So, while SuperKEKB increases the instantaneous luminosity, it must simultaneously improve the beam steering to decrease the background rate. CLAWS++ helps to improve the continuous-injection induced beam-background rates.

Furthermore, CLAWS++ also serves as a beam-abort trigger. This task is essential because an unstable beam can go astray and deposit particle bunches in the Belle II-detector. A direct hit of a particle bunch causes significant damage to the detector and must be avoided. The CLAWS++ detectors system monitors the beam stability and triggers an abort if it detects an unstable beam.

There are other possible use-cases for CLAWS++ that are currently explored. Detectors installed at some beam collimators could help understand the accelerator performance by identifying background sources.

2.2 Particle Detection

Particle detection is the fundamental basis of experimental high-energy physics. The Belle II-detector needs to detect and identify particles to measure any particle interaction. There are four main processes of how particle interact with matter. Ionisation through charged particles, electromagnetic showers, hadronic showers and the emittance of Cherenkov light.

When passing through the detector, charged particles lose energy by ionization. Here, the charged particle can interact through the electromagnetic force with the charged components of the atoms, most probably the shell electrons. The energy loss relation is given by the well known Bethe Bloch formula. The particle that mainly interact through ionization is the $\mu^{\pm}$, which is long lived enough to pass the detector and heavy enough that Bremsstrahlung is a negligible effect.

Electromagnetic showers are caused by highly energetic $e^{\pm}$ or γ. The photon can undergo pair production, $\gamma \rightarrow e^{+}e^{-}$, if its energy is above 1.022 MeV. Highly energetic $e^{\pm}$ loose a considerable fraction of their energy through Bremsstrahlung, which produces highly energetic γ in the process. Electromagnetic showers are an iterative process in which highly energetic $e^{\pm}$ produce highly energetic γ, which in turn undergo pair production. This process prolongs until the daughter particles have an energy below the $e^{+}e^{-}$ production. In the end, one particle split into many ligher particles, causing many interaction with the detector material.

Hadronic showers follow a similar concept. Here hadrons interact with the atomic nucleus through the strong force which causes the production of many secondary particles. These particles interact with the detector material. Detectors specfically built for the detection of hadrinic showers use materials with high density to increase the probability of a hadron interacting with the detector.

Furthermore, charged particles can also emit Cherenkov light when passing through a medium. A charged particle causes the emittance of Cherenkov light when it has a velocity, v, that is faster than the speed of light in the medium,

$$c_n = \frac{c_0}{n}, \tag{2.3}$$

with the speed of light in a vacuum, c_0, and the refractive index, n. The light is emitted in a cone. The opening angle of the cone is determined by

$$cos(\theta) = \frac{1}{n\beta}, \tag{2.4}$$

with β given by

$$\beta = \frac{v}{c_0}. \tag{2.5}$$

Since the momentum p is conserved and it is related to v by

$$v = \frac{p}{m}, \tag{2.6}$$

each particle type has its specific velocity for a given momentum, translating into a specific Cherenkov angle θ. We can rewrite Equation (2.4) in terms of the mass

$$m = \frac{p\cos(\theta)}{c_n}. \tag{2.7}$$

Each particle type has its unique mass, so by identifying the particle mass through the Cherenkov cone, Cherenkov-based detectors identify the particle.

2.3 Belle II

The $\tau \to \ell\alpha$-search needs good particle identification capabilities, a good track reconstruction to reliably select leptons and reconstruct the momentum of the τ daughter, and a high amount of $\tau^+\tau^-$ events to test this hypothetical decay.

B-factories are built for high precision physics in a high luminosity environment. Belle II proposed an ambitious physics program [46] that requires equally good or better performances in a higher beam-background environment than its predecessor Belle.

The Belle II-detector offers an excellent vertex resolution providing a more precise measurement of charged particles's spatial location. It is capable of a high efficiency for charged-particle reconstruction over a wide range – down to 50 MeV of momentum, and it has an exceptional momentum resolution over the whole kinematic range – up to 8 GeV. These capabilities increase precision in various measurements, such as particle mass measurements. Furthermore, the Belle II-detector offers an efficient photon-detection above 30 MeV and precise measurements

of energy and direction up to 8 GeV, contributing to precision measurements.

The Belle II-detector has a highly efficient particle identification. It covers the full kinematic range of the experiment to separate pions, kaons, protons, electrons, and muons. This particle identification is crucial for the physics program of Belle II. In searches for physics beyond the Standard Model it enables probing new depths of parameter space.

Especially searches for physics beyond the Standard Model with a missing particle profit. The Belle II-detector has a broad coverage of the solid angle and a fast and efficient trigger system. Furthermore, it has a high-performance data storage system capable of storing large amounts of data.

The Belle II-detector resides at the interaction point of SuperKEKB. Belle II substantially upgraded or replaced all components of the Belle detector to withstand the higher luminosity environment and improve performance. Figure 2.4 outlines the Belle II-detector, indicating the detector parts described in the following. Belle II implemented a new vertex, tracking, and particle identification system and updated the electronics of the electromagnetic calorimeter. The listed components are within the 1.5 T magnetic field, enclosed by the K_L^0 and μ^- detector, guiding the magnetic field with its iron plate absorbers. The K_L^0 and μ^- detector were also updated with state-of-the-art scintillation read out by SiPMs.

The following section describes the different subsystems of the Belle II-detector starting from the vertex detector closest to the interaction point.

2.3.1 Vertex Detector

Excellent vertex reconstruction is key to improving many precision measurements in the B-, D-, or τ-sector. The vertex detector, VXD, performs the central part of particle vertex determination. Belle II reconstructs a

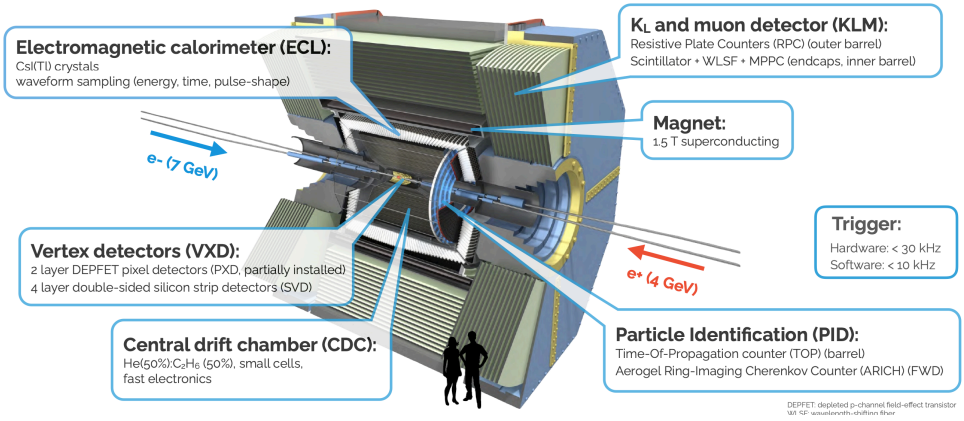


Figure 2.4: The Belle II-detector [63, 64].

vertex using the tracks from the central drift chamber and the VXD data to extrapolate to the interaction point.

The VXD, shown in Figure 2.5, consists of two subsystems. The pixel

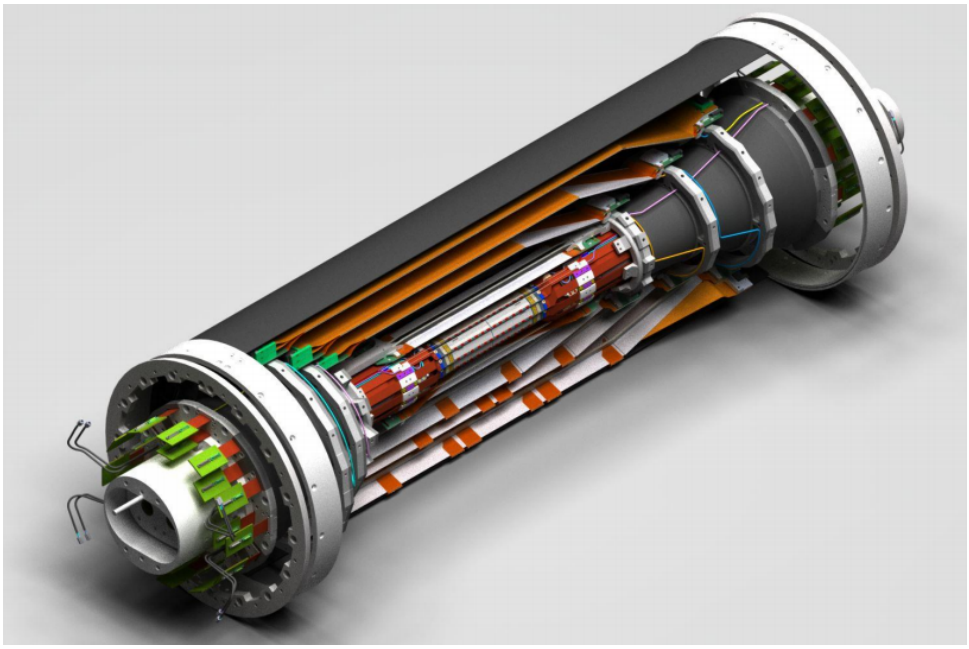


Figure 2.5: The Belle II vertex detector [65]

detector, PXD [66], is the inner part of the VXD. It consists of pixel silicon detectors made from depleted field-effect transistors, DEPFET [67]. The silicon vertex detector, SVD [68], is the outer part; it consists of double-sided silicon strips.

The VXD can reconstruct low transverse momentum, p_t, tracks down to 10 MeV, which is important for D^*-daughter reconstruction. The VXD enables to pinpoint the track of a particle precisely. Precise tracking is necessary to determine the particle decay vertex. The VXD increases the vertexing precision by almost a factor of two compared to Belle. This precision enables Belle II to measure world-leading results, for example, the $\overline{D}$-lifetime [69].

Pixel Detector

SuperKEKB's beam background necessitates a tracking detector that can differentiate between different particle tracks within a short amount of time. Furthermore, to improve the vertex resolution, Belle II placed the detector close to the interaction point, increasing occupancy. The resulting occupancy yields a tight time resolution requirement. A fast readout pixel detector has a high spatial resolution, relaxing the time resolution condition for the detector.

Belle II decided to build a two-layer pixel detector, with pixel sizes of $50 \times (55-85)\,\mu\text{m}^2$, to increase the vertex resolution and handle the beam

background. The second layer improves the pattern recognition needed for the vertex measurement in a high occupency environment.

The pixel detector modules consists of DEPFET detectors sensors. The DEPFET concept enables thin detector plates of 75 μm thickness, reducing particle scattering and, consequently, the detector material's impact on the track. A detector module consists of a fully depleted silicon substrate with a p-channel metal-oxide-semiconductor field-effect transistor, MOSFET. The left of Figure 2.6 shows the electrical circuit's technical drawing of a pixel. The right illustrates how the pixel looks. A field-effect transistor, FET, is a transistor using an electric field $\vec{E}$ to

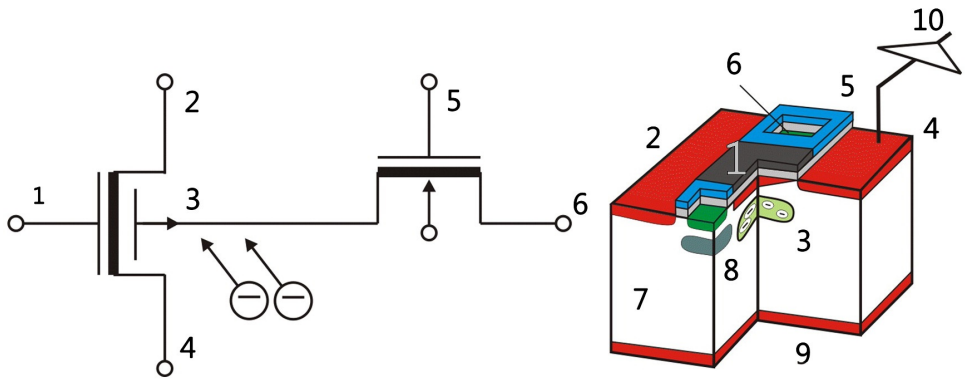


Figure 2.6: Principle of the DEPFET Belle II pixel detector. (1) external FET gate, (2) p+ source, (3) deep n-doped internal gate, (4) p+ drain with connection to external amplifier, (5) clear gate, (6) n+ clear, (7) depleted n-Si bulk, (8) deep p-well, (9) p+ backside contact, (10) amplifier. Taken from [67].

control the electric behavior.

The detection principle is that charged particles pass through the detector and free e^-, which gather at the internal gate. The MOSFET modulates current at the – e^- collecting – internal gate.

The entire pixel detector is a two-layer structure with the radius of the inner shell $r_1 = 14$ mm and the radius of the outer shell $r_2 = 22$ mm. The inner shell consists of eight planar sensors with a 15 mm width and 90 mm sensitive length. The outer shell consists of twelve planar sensors with 15 mm width and 123 mm sensitive length. Belle II mounted the sensor on integrated support and cooling structures. The acceptance region determines the final sensitive length, given by the Belle II-detector's polar angle of 17-150 degrees.

Silicon Vertex Detector

The SVD's primary purpose is to measure vertices together with the PXD. The detector has a similar basic detection principle as the pixel detector, but the SVD consits of double-sided silicon strip detector modules. The sensor are between 300-320 μm tick, have an active length of 123 mm, and an active width between 38-57 mm. To survive the hos-

tile environment of SuperKEKB, the SVD fulfills the following requirements:

- low dead time in high beam background rate environment;
- high trigger rate with about 30 kHz frequency;
- light mass.

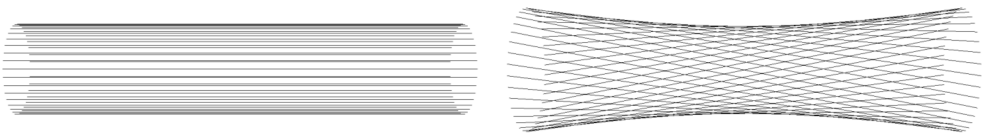
As for the PXD, the polar angle coverage is between 17 and 150 degrees for the SVD. The inner radius for the SVD is $r_{min} = 38\,\text{mm}$. It is determined by the pixel detector. The outer radius for the SVD is $r_{min} = 140\,\text{mm}$, determined by the inner radius of the central drift chamber.

The SVD has four layers, with the outer three layers tilted towards the beamline in the direction of the beam boost, as indicated in Figure 2.5. The tilt improves vertex precision in the boost direction.

2.3.2 Central Drift Chamber

The central drift chamber, CDC, is the primary tracking device of the Belle II detector. With an outer radius of 1130 mm, the large volume comprises small drift cells defined by the charged sense wires. The Belle II-detector has two orientations for its sense wires. Axial wires have the same alignment as the solenoid's magnetic field orientation, as shown in Figure 2.7a. Stereo wires are skewed to the axial wires as shown in Figure 2.7b.

The gaseous mixture in the chamber is composed of He-C2H6 in a 50:50 ratio. The average drift velocity of this gas is 3.3 cm/μs.



(a) Axial wire layer

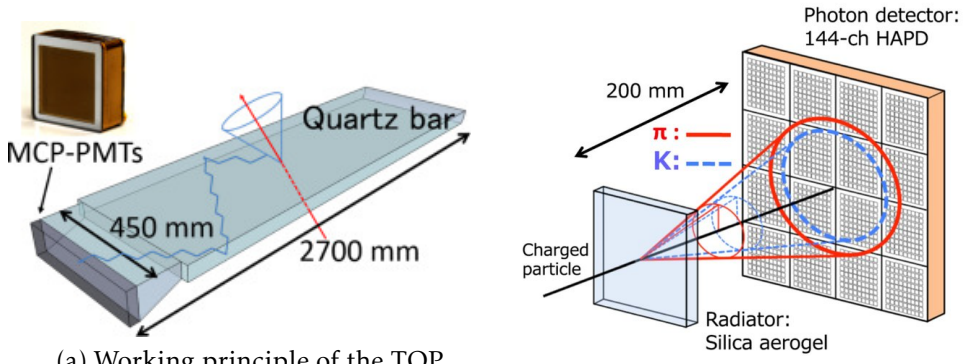
(b) Stereo wire layer

Figure 2.7: CDC wire configurations: The CDC has two configurations, the axial wire layers, and stereo wire layers. The axial sense wires are parallel to the beamline, and stereo wire layers are skewed to the beamline [70].

The CDC provides tracking information for the Belle II-detector and vital particle identification information. It measures the energy loss in the gas volume and the particle's path. By combining the energy loss with the spatial information, Belle II gets the $\mathrm{d}E/\mathrm{d}x$ information of the particle. This method can identify the particle type of low momentum tracks alone. Furthermore, Belle II uses the CDC information for the trigger system. The trigger system decides on recording or discarding an event.

2.3.3 Particle Identification System

The Belle II-detector has two dedicated detector systems for improved particle identification. The time of propagation counter, TOP, is located in the barrel, and the aerogel ring imaging Cherenkov detector, ARICH, at the front endcap is in the boost's direction. The fundamental principles of these two detectors are the same. Both detectors measure the Cherenkov angle, utilizing two different detection methods.



(a) Working principle of the TOP.

(b) Working principle of the ARICH.

Figure 2.8: Illustration of the detection principles for TOP [71] and ARICH [72]: Both detectors rely on the Cherenkov angle determination for particle identification. TOP covers the barrel region, and ARICH covers the endcap in the boost direction.

Time Of Propagation Counter

TOP covers the barrel region. The detector consists of 16 modules. Each module is a 2.6 cm x 45 cm x 2 cm quartz bar with a reflective mirror on one of its short ends and a linear array type photon detector on the other, as Figure 2.8a illustrates.

TOP determines the Cherenkov cone angle from a precise time of arrival measurement of the photons and the impact position of the particle, determined by the tracking system. These measurements require a time resolution of 100 ps and precise knowledge of the starting time with an accuracy of 50 ps. Furthermore, TOP can detect the light pattern on its two-dimensionally-resolved photon-detectors.

The ARICH Detector

The ARICH covers the forward endcap region. Requirements for the ARICH are a low momentum threshold for $\pi^{\pm}$ and a good $\pi^{\pm}$-K separation in the energy range of 0.4 to 4.0 GeV.

One detector module consists of two 2 cm thick aerogel layers with different refractive index $n_1 = 1.045$ and $n_2 = 1.055$, respectively. Belle II

uses two layers to increase the light yield at the hybrid avalanche photon detector. These detectors are made of bialkali photocathodes in a vacuum.

The ARICH detects the ring of the Cherenkov cone and measures the Cherenkov cone angle. This information, combined with the known refractive indices, results in the velocity of the particle.

2.3.4 Electromagnetic Calorimeter

The central purpose of the electromagnetic calorimeter, ECL, is to detect and measure the energy of γ and electrons. The detector is the primary discriminator to separate electrons from h, particularly $\pi^{\pm}$.

The electromagnetic calorimeter is an array of thallium doped cesium iodide (CsI(Ti)) crystals. It covers the polar angle of

$$12.4^\circ < \theta < 155.1^\circ,$$

except for two 1° small gaps between the barrel and endcaps.

The Belle II-detector inherited the CsI(Ti) crystals, preamplifiers, and support structure from Belle. Belle II updated the electronics and software to deal with overlapping signals from the increased beam background rates.

2.3.5 K_L^0 and Muon Detector

The long-lived neutral kaon, K_L^0, and μ detector, KLM, is the outermost detector. It is 4.7 m thick and consists of an iron plate and active detector element sandwich. The detector is placed outside the solenoid and serves two purposes.

The first one is to guide the magnetic flux of the 1.5 T magnet with the iron plates. The second one is to detect neutral hadrons with sufficient lifetime to reach the KLM and muons. The active detector components consist of scintillator strips with wavelength shifting fibers read out by SiPMs. The iron plates are an additional interaction length for hadronic showers, mainly from K_L^0. They increase the probability of a hadromnic-shower and therefore increase the volume we probe to detect the shower. In contrast to hadrons, the μ does not shower but minimally ionizes the material. The different behavior of hadron and muon particles help to separate e.g., $\pi^{\pm}$ from μ.

2.4 Detector Performance Studies

Detector performance studies ensure that the Belle II-detector operates adequately in tracking, energy measurement, and particle identification. Performance studies compare measured data to the expectations for simulated data using well-understood particle events called control modes.

The performance groups convert the differences in the measured to simulated data comparison to corrections. The analysts apply specific corrections to measured or simulated data, and the type of correction depends on the performance correction and recommendation of the experts.

This section introduces the particle identification method of Belle II and exemplifies a performance study with the electron identification study using radiative Bhabha events, which was part of this thesis work.

2.4.1 Charged Particle Identifaction

The following description is an overview of Belle II's standard particle identification method used in the $\tau \to \ell\alpha$ search. It is based on the Belle II physics book [46].

Belle II provides charged-particle identification, PID, as a probability distribution. There are six different charged-particle hypotheses for stable charged particles. The particle hypotheses are the e; μ; $\pi^{\pm}$; charged kaon, $K^{\pm}$; p; and deuteron, ${}^{2}D$.

Belle II calculates a global likelihood for each particle hypothesis to obtain the probability for a particle – the PID. It combines the particle likelihoods from all sub-detectors contributing to PID. The sub-detector likelihoods are based on various methods. A PID probability ultimately is the fraction of the global particle likelihood over all likelihoods.

Belle II uses two methods to determine the likelihood of a particle hypothesis. The first method is the energy loss of a particle through ionization, $\mathrm{d}E/\mathrm{d}x$. The main detector which provides this information is the CDC. There are also plans to include information from the VXD, particularly the SVD, but as of now, Belle II does not include this information in the global particle-hypothesis likelihood. Figure 2.9 shows the $\mathrm{d}E/\mathrm{d}x$ distributions for all particle hypothesis. $\mathrm{d}E/\mathrm{d}x$ has a good separation power below 1 GeV as Figure 2.9 indicates. The CDC likelihood for a particle hypothesis, $\mathcal{L}_i^{\mathrm{CDC}}$, is based on a look-up table generated from Monte Carlo simulations. The basic idea is to compare measured $\mathrm{d}E/\mathrm{d}x$ with the expectation from simulation for a given particle type. This comparison determines $\mathcal{L}_i^{\mathrm{CDC}}$.

The second Belle II method to determine the likelihood of a particle hypothesis relies on Cherenkov-cones. As described above, the TOP and ARICH provide a particle hypothesis based on the Cherenkov-cone's angle.

To determine the particle-hypothesis likelihood from ARICH, $\mathcal{L}_i^{\mathrm{ARICH}}$, Belle II extrapolates tracks from the CDC to the ARICH. Depending on the track parameters and a given particle hypothesis, different spatial distributions of the Cherenkov light pattern on the photon detector plane occur. $\mathcal{L}_i^{\mathrm{ARICH}}$, is based on comparing the observed light patterns in measured data to the expected one in the simulation.

The third method Belle II uses relies on measuring deposited energy

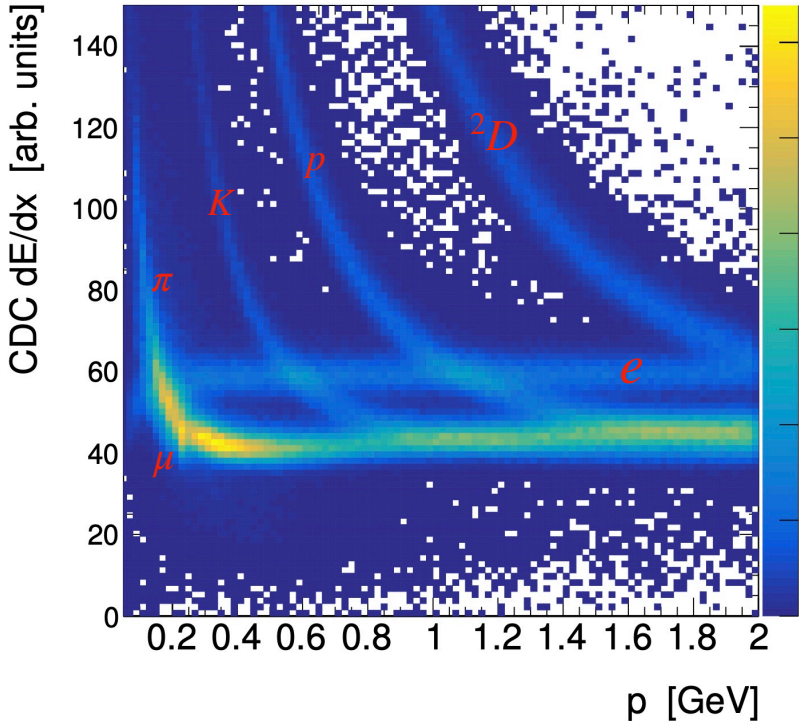


Figure 2.9: dE/dx-distributions for all 6 particle hypothesis provided by the CDC [46]. Each particle hypothesis is indicated in red to it's corresponding curve on the dE/dx-momentum-plane.

in the ECL. Here, the measured energy in the ECL over the particle momentum distribution, E/p-distribution, is different for each particle hypothesis for momenta greater than 1 GeV/c. The E/p-distribution is fitted and yields a fit quality that determines the log-likelihood of a particle hypothesis.

The fourth and last method relies on the difference in penetration range and scattering for $\boldsymbol{\mu}$ and charged $\boldsymbol{h}$ in the KLM. Here, a KLM cluster – an observed event in the KLM – is matched with a track. The likelihood of the particle hypothesis for the KLM, $\mathcal{L}_i^{\mathrm{KLM}}$, is determined based on the detector response compared to a precalculated probability density function.

In summary, Belle II obtains a likelihood for a particle hypothesis for each detector, $\mathcal{L}_i^d$. Simulation-based probability density functions, (high purity) measured data control samples, or analytical models determine $\mathcal{L}_i^d$. The global particle-hypothesis likelihood $\mathcal{L}_i$ is given by the product of all detector likelihoods, rewritten in terms of log-likelihoods as

$$\ln|\mathcal{L}_i| = \sum_d \ln|\mathcal{L}_i^d|;\ d \in [\mathrm{CDC}, \mathrm{ARICH}, \mathrm{TOP}, \mathrm{ECL}, \mathrm{KLM}]. \tag{2.8}$$

Bayes' theorem gives the global probability for a charged particle, with

$$P(i) = \frac{\mathcal{L}_i}{\sum_j \mathcal{L}_j};\ j \in [\boldsymbol{e}, \boldsymbol{\mu}, \boldsymbol{\pi}^{\pm}, \boldsymbol{K}^{\pm}, \boldsymbol{p}, {}^2\boldsymbol{D}]. \tag{2.9}$$

2.4.2 Electron Identification Study

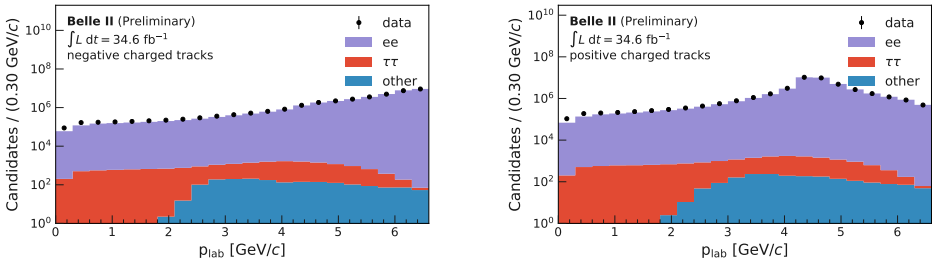
Electron identification, *electronID*, is the probability that the particle in question is an electron. This section discusses the determination of the *electronID* efficiency, which is a correction factor for the simulation data determined from measured data events.

The performance study uses a tag and probe method to determine the *electronID* efficiency with radiative Bhabha events. The efficiency is proportional to the ratio of the tag to the probe particle.

This method requires that one of the electrons – the tag – fulfills tight electron selection requirements of $electronID > 0.95$. The other electron – the probe – is selected with a predefined set of cuts, $electronID > [0.5, 0.9, 0.95]$, giving the electron identification efficiency for the corresponding probe cut.

Selection The selection requirements for the performance study are the following. Events must contain precisely two tracks that loosely originate from the interaction point within a few centimeters. A loos requirement for the squared mass of the recoiling system to be below 10 GeV^2 suppresses background events with missing particles and simultaniously ensures a high yield of the sample.

The tag particle must satisfy $electronID > 0.95$. Figure 2.10 shows the momentum distribution of positive and negative charged probe tracks after the selection. The number of candidates is on a log scale, which shows a highly suppressed remaining background of non-Bhabha events.



(a) Electron θ distribution. (b) Electron p distribution.

Figure 2.10: Momentum distribution of negative (a) and positive (b) charged probe tracks. Measured data is depicted as black dots. The distribution for simulated events is shown as a coloured stack plot. This simulation sample consists of e^+e^-, $e^+e^-e^+e^-$, $e^+e^-\mu^+\mu^-$, $\mu^+\mu^-$ and $\tau^+\tau^-$ events [73]

Belle II records Bhabha events with a low multiplicity trigger; it requires an ECL cluster with an energy above 2 GeV in the barrel region. Because of this trigger requirement, the performance study cannot provide corrections in the end caps. Once a suitable trigger is available,

corrections for the endcaps will be provided.

The performance study requires triggered events by the tag particle to minimize any bias in *electronID*.

Methodology The performance study computes the *electronID* efficiency, ε, as the number of events, N_{probe}, satisfying the *electronID* requirement placed on the probe, over the total number of selected events, N_{tag},

$$\varepsilon = \frac{p_{\text{probe}} \cdot N_{\text{probe}}}{p_{\text{tag}} \cdot N_{\text{tag}}}. \tag{2.10}$$

The purity p_{tag} is the probability that the tag particle is correctly identified as an electron. The purity p_{probe} is the probability that the probe is correct. We compute the purities using simulated e^+e^- , $e^+e^-e^+e^-$, $e^+e^-\mu^+\mu^-$, $\mu^+\mu^-$ and $\tau^+\tau^-$ samples as,

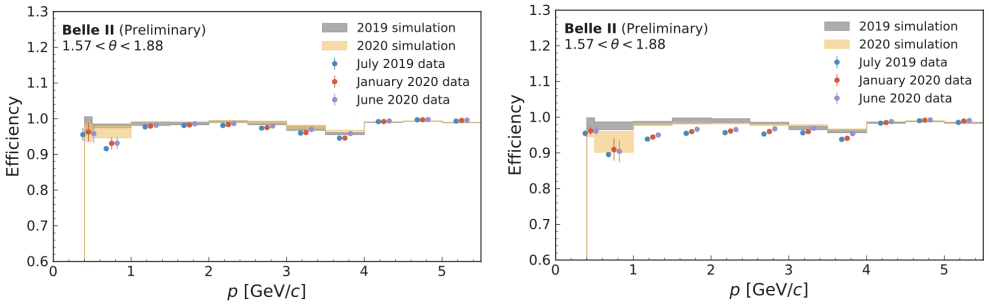
$$p_{\text{tag}} = \frac{N^{\text{Sig}}_{\text{tag}}}{N^{\text{BG}}_{\text{tag}} + N^{\text{Sig}}_{\text{tag}}},\ p_{\text{probe}} = \frac{N^{\text{Sig}}_{\text{probe}}}{N^{\text{BG}}_{\text{probe}} + N^{\text{Sig}}_{\text{probe}}}, \tag{2.11}$$

where $N^{\text{Sig}}_{\text{probe}}$ is the number of events with a correctly identified probe, $N^{\text{Sig}}_{\text{tag}}$ with a correct tag, $N^{\text{BG}}_{\text{tag}}$ the number of events with a mis-identified probe before applying the electron ID requirement on the probe track, and $N^{\text{BG}}_{\text{probe}}$ after the electron ID requirement.

Systematic uncertainties The performance study considers two sources of systematic uncertainties. First, it estimates the trigger's possible bias by calculating the *electronID* efficiency in the simulated Bhabha sample with the trigger switched off and on. The absolute difference between both efficiencies is a systematic uncertainty related to the trigger bias.

Second, the performance study calculates the *electronID* efficiency in measured data with and without applying the purity factors. The absolute difference is an additional source of systematic uncertainty related to the background contamination.

Results The performance study calculates the *electronID* efficiency separately for negative and positive charged probe candidate tracks to avoid systematic effects associated with the particle charge. Figure 2.11 shows the charge-dependent performance difference for the *electronID* efficiency, especially in the bins above 1 GeV/c. Error bands indicate uncertainties in the simulation data. The observed trigger bias dominates the uncertainty. The data samples's error bands show the systematic and statistical uncertainty added in quadrature. Systematic effects dominate the analysis.



(a) Positive Track Candidates (b) Negative Track Candidates

Figure 2.11: Example for the *electronID* efficiency in one θ bin. The bands indicate the electron ID efficiency for simulated data. The dots indicate measured data. The year indicates the different software releases of the simulation data production and the measured data processing. There are two raw measured data sets. The first raw measured data set corresponds to 5.2 fb^{-1}. The second raw measured data set corresponds to 34.6 fb^{-1}; it consists of the first measured data set with additionally recorded runs. The July 2019 data set is the first raw measured data set processed with the 2019 software release. The January 2020 data set is the first raw measured data set reprocessed with the 2020 software release. The June 2020 data set is the second raw measured data set processed with the 2020 software release [73].

We observe a general trend of increasing electron efficiency with momentum with a small dip at 4 GeV/c. At higher momenta, the E/p distribution has a higher probability of differentiating electrons from other particles because the electron distribution moves further apart from more massive charged particles. Furthermore, the electron efficiency is especially low for momenta around 0.7 GeV, because dE/dx gets less discriminatory as particle-hypothesis bands cross and the E/p-distribution is still limited as indicated in Figure 2.9. For momenta above 4 GeV/c, we observe good agreement between simulation and measured data, with a high efficiency close to one. Below 4 GeV/c, the agreement of simulation and measured data decreases for electrons towards the 0.7 GeV/c dip. The degeneration is not as severe in the case of positrons who display a recovering of simulation and measured data agreement towards 1 GeV/c until the 0.7 GeV/c peak.

Part III

The $\tau \to \ell\alpha$ Analysis at Belle II

In this part, we discuss the search for an invisible boson α emerging in a lepton decay. As described in Chapter 1, this kind of particle is well-motivated in prominent extensions of the Standard Model. If α is a spin $s = 0$ particle, we will refer to it as ALP, if α is a $s = 1$ particle we will call it a Z'.

This thesis searches for α in τ decays by looking for a $\tau \to \ell\alpha$ decay at the Belle II experiment. It is the central part of this thesis. Previous work on electron identification and a later discussion on the τ-mass are support and spinoff projects. The search for $\tau \to \ell\alpha$ decay is a blind analysis, meaning the whole analysis framework is developed using simulation data. To unblind – using the actual measured data from the Belle II-detector – needs the approval of the Belle II collaboration.

The $\tau \to \ell\alpha$ decay has no unique kinematic region, which would allow distinguishing it from the Standard Model $\tau \to \ell\nu_\tau\nu_\ell$ decay. Instead, the search strategy in this thesis is to evaluate the shape difference of the $\tau \to \ell\alpha$ decays' kinematic distributions to the $\tau \to \ell\nu_\tau\nu_\ell$ decay. To search for the $\tau \to \ell\alpha$ decay, we need an observable which is significantly different from the Standard Model three-body decay. The τ-rest frame exhibits the most significant shape difference. In the τ rest frame, the lepton momentum distribution of $\tau \to \ell\alpha$ is a monoenergetic peak. The peak position is at the α mass. This is in contrast to the $\tau \to \ell\nu_\tau\nu_\ell$ Michel spectrum shown in Chapter 1, Figure 1.3.

Belle II produces $\tau^+\tau^-$ events above threshold with an additional momentum of 3.5 GeV/c per τ. This boost leads to a clear separation of the two τ decays. It is common to divide $\tau^+\tau^-$ events into a tag and a signal τ-decay. We use the tag-τ to reconstruct the $\tau^+\tau^-$ event and the signal-τ for the study of interest.

The separated $\tau^+\tau^-$ at Belle II enables us to use the reconstructed tag-τ flight direction as a τ rest frame estimate. The rest frame boost necessitates the τ-decay on the tag side to consist of predominately visible particles. In the Standard Model, hadronic decays – all visible particles are hadrons – fulfill this condition. Here, only the unavoidable ν_τ is escaping detection.

A well-tested $\tau^+\tau^-$ event type with a hadronic τ decay is the so-called 3×1-prong topology. This topology is a $\tau^+\tau^-$ event with three charged tracks on one side and one on the other. The 3-prong τ is a hadronic decay, $\tau \to \nu_\tau hhh + \#\text{neutrals} \geq 0$, with the number of neutral particles, # neutrals, and a negligible amount of three charged lepton τ-decays. This topology also allows selecting a clean sample of $\tau^+\tau^-$ events. This thesis studied the selection using several approaches. Chapter 3 describes the considered approaches.

The missing ν_τ leads to a major problem in our $\tau \to \ell\alpha$ search: reconstructing the τ rest frame. Chapter 4 discusses several methods for reconstructing the τ-rest frame.

After the selection and τ rest frame reconstruction, Chapter 5 discusses the fitting procedure and study of systematic uncertainties to

search for the $\tau \to \ell\alpha$ decay. This analysis chooses a template fit, allowing for a complex fit shape.

At last, Chapter 6 presents the fit results for all studies in this thesis. First, we asse the preliminary fit result on $62.8\,\mathrm{fb}^{-1}$ of Belle II data and the Upper Limits, UL, for every mass hypothesis using the approved Belle II $\tau \to \ell\alpha$ search strategy. Afterward, we discuss the results expected from simulation data for the alternative selection and rest frame methods we discuss in Chapter 3 and Chapter 4.

This $\tau \to \ell\alpha$ search is a team effort, every member contributed to every step of the search. Appendix A details my contributions to the presented search.

Chapter 3

Event Selection

Event selection is the art of choosing detected and recorded events such that the event type of interest is selected. This selection process is crucial to avoid drowning the signal of interest in other processes.

The Belle II-selection strategy consists of two steps. The first step is referred to as preselection, which means that analysts apply cuts on the reconstruction level before reconstruction. The reconstruction level refers to the step in the analysis of reconstructing particle events from detector signals. The cuts on the preselection level rely on low-level selection criteria such as specific detector signals or track positions.

After the preselection and reconstruction of event and particle properties, the analyst also applies rudimentary event criteria such as PID, referred to as postselection, which are higher-level selection criteria available after reconstructing events. We collectively refer to the preselection and postselection as reconstruction level event selection. The second step is background suppression, and it typically utilizes higher-level variables that need reconstructed particles and events.

Belle II based the reconstruction on filling particle lists. The particle list concept associates detector signals with signatures for one particle, e.g., tracks and calorimeter information. The selected particle type's mass hypothesis yields the particle's properties derived for a specific hypothesis. For example, the momentum is calculated for all tracks, assuming it is an electron. Afterward, the analysts can select all reconstructed particle candidates within physically viable ranges, e.g., in the momentum. This concept evaluation allows for a accurate reconstruction of each particl's flavor.

The most abundant τ decays are 1-prong decays, followed by 3-prong decays. Higher-prong decays are relatively rare compared to the 1- and 3-prong decays. Depending on the study conducted, most $\tau^+\tau^-$ events are reconstructed with 1×1-prong, 3×1-prong, or 3×3-prong topologies. We select the 3×1-prong topology with a hadronic tag side because it enables a pure event selection without biasing the signal. The reason is that the tag-side 3-prong signature is well divisible from other physics processes, allowing for a selection and background suppression with the

tag side alone.

Also, a 3-prong τ decay allows reconstructing a τ rest frame. We calculate the pseudo-rest frame variable by assuming a certain tag-τ-decay, as Chapter 4 will discuss. The smaller the contamination of non-hadronic 3-prong and non-τ background is, the better the resulting pseudo-rest frame estimation. The tag-τ selection influences the quality of the τ pseudo-rest frame variable. Allowing for various τ-tag decays adds additional smearing in the τ pseudo-rest frame distribution. So, increased purity leads to a less smeared τ-rest frame estimate.

We chose a mass-independent optimization scheme by basing the selection on the irreducible $\tau \to \ell \nu_\tau \nu_\ell$ decay. Thus, for the selection, $\tau \to \ell \nu_\tau \nu_\ell$ is the signal decay. The following chapter bases all optimization work on the $\tau \to \ell \nu_\tau \nu_\ell$ decay and discusses the selection results for different $\tau \to \ell \alpha$ mass hypotheses afterward. We optimize for the following event type:

$$e^+e^- \to \tau^+(\to \nu_\tau \pi^+\pi^-\pi^+ + \#\text{neutrals} \geq 0)\tau^-(\to \ell^-\overline{\nu}_\ell \nu_\tau),$$

with the charged conjugate included and only a mild requirement for the hadrons to be $\pi^\mp$. We use the common Belle II software package BASF2 with the release-05-02-00 for the reconstruction.

Background-suppression studies are an essential part of this thesis. We must suppress all Standard Model processes faking the new physics signal[1]. Several strategies are available for the signal selection process. This chapter discusses studies using cut-based approaches and machine-learning methods using boosted-decision trees, BDTs. We discuss a total of four cases, which Figure 3.1 exemplifies by displaying the workflow for the event selection. Either we allow neutral particles candidates, neutrals, in the selection, or we reject them. In both cases, we will consider the performance of a cut-based and a BDT-based event selection.

We will learn that allowing for neutrals increases the available data and increases miss-identified events, degrading purity. Rejecting neutrals results in fewer data and fewer misidentified events, increasing purity. Furthermore, rejecting neutrals allows for defining a sideband. A sideband is a part of data that emulates properties of the signal itself or the signal region but is not part of the signal or signal region. The sideband allows for studying important properties of measured data, e.g., validating the simulated on measured data. A valide sideband definition is esstial because we performe a blind analysis, basing all developed tools and methods on simulated data. A blind analysis necessitates a careful verification of the selection and statistical treatment with measured data to avoide accidential unblinding. A natural choice for the sideband would be $\tau^\pm \to \pi^\mp \nu_\tau$ because it is also a τ decay and emulates the $\tau \to \ell \alpha$-decay signature. At the time of this analysis, Belle II did not provide a $\pi^\pm$ identification deemed reliable enough to utilize the

[1] Appendix H lists the physics processes we consider e.g. in Table H.1.

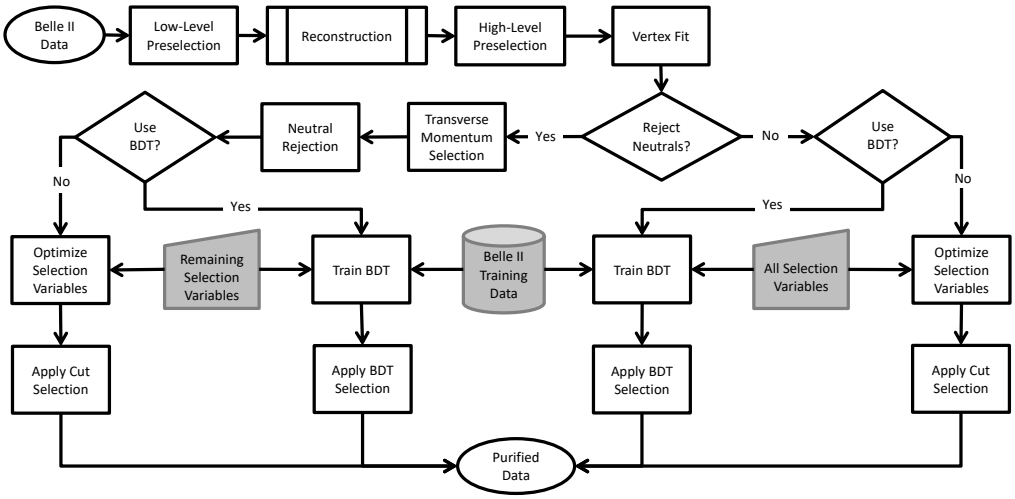


Figure 3.1: Flowchart displaying the different event selection strategies discussed: The start and end points are represented by the stadium shape, indicating the *state* of the data. A rectangle represents processes, a rectangle with double-struck vertical edges represents predefined processes, and a diamond represents decisions. Gray symbols represent external input. The manual input we provide is a quadrilateral, and the Belle II-training data is a cylinder.

$\tau^{\pm} \to \pi^{\mp}\nu_{\tau}$ decay as a sideband for the $\tau \to \ell\alpha$ analysis. We abstained from a custom $\pi^{\pm}$ selection as it would necessitate a dedicated performance study to understand simulated to measured data discrepancies. We made the final decision for the selection method after the selection studies.

We choose the variables empirically, and Figure 3.1 represents them as manual input (the gray quadrilateral). The BDT-based selection uses a training-data set (the gray cylinder) consisting of an additional Belle II-simulation-data batch from MC13a that other selection optimization processes do not use. We study two different BDTs with the variables of the cut-based selection. The first variant is to train a BDT after applying all cuts before the neutral rejection, and we train the BDT with the background suppression variables only. The second variant is a BDT trained after the reconstruction level event selection with all available variables used in the cut-based selection. We train both BDTs with variables regarded as safe to use, which are variables exhibiting no considerable deviation of the simulated to the measured data distributions.

3.1 Methodology

There are different schools on how to choose cuts to select signal events. The final goal is to select signal events and reduce background events. One school of thought strives for a maximized signal over background rate using some algorithm. We use this cut selection scheme but maximize the signal over background ratio only to the numerical precision of

two significant digits – above the detector resolution for all considered variables.

We apply a figure of merit, FOM, optimization Punzi suggestes in case of irreducable background to maximize the signal over background ratio [74]:

$$\text{FOM} = \frac{S}{\sqrt{S+B}}\,, \tag{3.1}$$

with the signal yield, S, and the background yield, B. The efficiency of selecting signal events is

$$\epsilon = \frac{N^{\text{reco}}}{N^{\text{gen}}}, \tag{3.2}$$

with the number of reconstructed and selected signal events, N^{reco}, and the number of generated signal events, N^{gen}.

The number of generated 3×1-prong $\tau \to \ell\nu_\tau\nu_\ell$ events is

$$N^{\text{gen}}_{\tau\to\ell\nu_\tau\nu_\ell} = N^{\text{gen}}_{\tau^+\tau^-} \cdot Br(\tau \to \ell\nu_\tau\nu_\ell) \cdot Br(\tau \to \nu_\tau hhh + \text{neutrals} \geq 0), \tag{3.3}$$

with the number of generated $\tau^+\tau^-$ events, $N^{\text{gen}}_{\tau^+\tau^-}$; the branching ratios for the $\tau \to \ell\nu_\tau\nu_\ell$-decay, $Br(\tau \to \ell\nu_\tau\nu_\ell)$; and the branching ratio for the hadronic 3-prong τ decay with any number of neutrals, $Br(\tau \to \nu_\tau hhh + \#\text{neutrals} \geq 0)$. The branching ratios are listed by the Particle Data Group [34].

The number of generated 3×1-prong $\tau \to \ell\alpha$ events is

$$N^{\text{gen}}_{\tau\to\ell\alpha} = N^{\text{gen}}_{\tau^+\tau^-} \cdot 1 \cdot Br(\tau \to \nu_\tau hhh + \#\text{neutrals} \geq 0), \tag{3.4}$$

with the branching ratio in the signal sample generation set to one.

The definition of purity is:

$$\mathcal{P} = \frac{N^{\text{signal}}}{N^{\text{reco}}}, \tag{3.5}$$

with the number of signal events , N^{signal}, and the number of all reconstructed and selected events, N^{reco}.

3.2 Reconstruction Level Event Selection

Reconstructing events from detector signals and tracks is computationally expensive; hence, selecting those events that fulfill the essential criteria of the signal event before reconstruction is more efficient than reconstructing all data. We refer to this process as a preselection. After reconstruction, it is common at Belle II to apply higher-level cuts. These cuts again fulfill the rudimentary criteria of the signal event, enhancing the ratio of signal events in data. We refer to this selection as postselection.

3.2.1 Preselection

At Belle II, reconstruction is the process of determining event properties from detector signals and tracks by applying particle hypotheses. Before reconstruction, we choose events fulfilling the requirements below to reduce the amount of reconstructed data. The preselection follows a three-step process discussed below.

1. We demand tracks from the interaction point;
2. select γ; and
3. reconstruct π^0.

First, we demand viable physics events requiring charged tracks to originate from the interaction point in order to suppress the beam background. The selection criteria are

- $|dz| < 3.0$ cm ; and
- $dr < 1.0$ cm ;

with the point-of-closest approach in the z-direction to the interaction point, dz, and the transverse distance between interaction point and vertex or track, dr. Tracks fulfilling the dz and dr requirements are referred to as "good tracks".

Second, we select all γ satisfying

- $\mathrm{E}(\gamma) > 100$ MeV ;
- $-0.8660 < \cos\theta < 0.9565$ ($\cong$ CDC acceptance) ; and
- clusterNHits > 1.5 ;

with the photon energy, $\mathrm{E}(\gamma)$; the momentum's cosine of the polar angle $\cos(\theta)$; and the sum of weights[2] of all crystals in an ECL cluster[3], clusterNHits[4].

Third, we reconstruct π^0 candidates and saves the remaining γ, which fulfill an energy threshold. It reconstructs π^0 candidates, from now on referred to as π^0, by combining two γ and requiring the invariant mass of two combined γ to be $115 < M_{\gamma\gamma} < 152\,\mathrm{MeV/c^2}$. Afterward, we save all photons not combined into a π^0 and passing the energy threshold of $\mathrm{E}(\gamma) > 200\,\mathrm{MeV}$.

After the event selection, we calculate the event shape and kinematic variables of the event from all tracks, π^0, and γ.

[2] the weights w_i fulfill $w_i \leq 1$

[3] An ECL cluster is a part of the ECL consisting of several crystals, next to each other, in which a particle deposited energy.

[4] For non-overlapping ECL clusters, this is equal to the number of crystals in the cluster. In the case of energy splitting among nearby clusters, this can be a non-integer value.

3.2.2 Postselection: Cut Based Event Selection After Reconstructions

We demand more fundamental event properties in the postselection to decrease non-signal events in the remaining data set. For the 1x3-prong topology, we require exactly four "good tracks" and discards all events that do not meet this requirement.

The thrust vector separates the event into a signal and tag hemisphere. We use the thrust vector to sort the tracks into the signal and tag hemispheres. This sorting ensures that one track is in one hemisphere and three tracks are in the other, rejecting events that do not fulfill the requirement. The thrust axis, $\widehat{n}_{thrust}$, is defined by maximizing the thrust value, *thrust*:

$$thrust = \sum \frac{| \overrightarrow{p}_i^{CMS} \cdot \widehat{n}_{thrust} |}{\sum \overrightarrow{p}_i^{CMS}}. \tag{3.6}$$

Here, $\overrightarrow{p}_i^{CMS}$ is the three-momentum of each particle of the event in the center-of-mass reference frame. We select events with tracks fulfilling the requirement

$$| \overrightarrow{p}_{signal}^{CMS} \cdot \widehat{n}_{thrust} | \cdot | \overrightarrow{p}_{tag,i}^{CMS} \cdot \widehat{n}_{thrust} | < 0 \,, \forall \; i \in \text{tag}. \tag{3.7}$$

The hemisphere with one track is the signal side. Our PID requirement for a lepton on the signal side is

- $electronID > 0.9$ for electrons and
- $muonID > 0.9$ for muons.

To select hadrons on the 3-prong tag side it requires

- $clusterE/p \leq 0.8\,c$ for each track (*clusterE* in GeV, p in GeV/c);

where *clusterE* is the energy deposited in the ECL and p is the momentum of the particle in the laboratory system.

This requirement selects mainly hadrons over electrons because the energy deposition in the ECL is lower for hadrons than that of the electrons. Correspondingly the ratio *clusterE*/p for electron is close to c and for hadrons below $0.8\,c$.

3.2.3 Trigger

High energy physics experiments commonly do not record every detected signal, and doing so would overwhelm the data acquisition systems and flood the recorded data with many well-known, studied, and abundant events such as Bhabha radiation. Instead, trigger systems select those event types deemed attractive for further studies. Belle II implemented several triggers, which enable recording $\tau^+\tau^-$ events. The

relevant triggers for this thesis are the so-called low multiplicity triggers, which are optimized to record events with only a handful of tracks. The trigger requirement is only present in measured data and not for the simulation data of MC13a used in this thesis.

We require events in data to fire the logical OR of several low-multiplicity ECL, lml, triggers:

- **lml0** or **lml1** or **lml2** or **lml4** or **lml6** or **lml7** or **lml8** or **lml9** or **lml10** or **lml12**

We will refer to this combination of riggers as *lmlX*.

3.3 Cut Based Background Suppression

Different physical processes, such as a $\tau^+\tau^-$ event or a $q\bar{q}$ event, exhibit different physical properties such as momentum or energy distributions. Physics analyses use this different behavior to differentiate between those processes. The idea of background suppression is to utilize the difference in behavior, which results in differently shaped distributions, to enhance signal events in the data sample by suppressing the background events.

To set a cut is the process of selecting events manually. The analyst considers different variable distributions and selects the regions with an enhanced signal yield compared to the background. Most commonly, analyses set cuts on one-dimensional distribution. We aim at selecting the most significant sample for the search of α. In this section, we will consider the difference between the two cut-based selections displayed in Figure 3.1, which either allow for neutral particles or rejects them.

We have selection procedure with multible steps. First, we apply a vertex fit to select viable hadronic-3-prong τ decays. Next, we reject mostly two lepton gamma events with a transverse momentum selection. Afterward, a selection of neutral particles rejects mainly $q\bar{q}$ background. The final background suppression uses event and tag side variables to achieve high purity of signal events in the data set.

3.3.1 Vertex Fit Selection

The location in the detector at which a particle decays into several other daughter particles is commonly known as a vertex. To enable accurate reconstruction of the vertex Belle II built the VXD, which measures the tracks of charged daughter particles as precisely as possible. A vertex reconstruction determines the most probable vertex point through a fit using the spatial information given by the VXD. This vertex fit yields *ChiProb* a variable in Belle II that expresses the probability the observed distance between the hypothesis – hadronic 3-prong τ– and the measurement exceeds the expectations for a corrected model by change [75].

We reject all events without a converged 3-prong vertex fit of the tag side. The failed events are either composed of decay chains with long-lived intermediate particles or physics events containing additional tracks from the underlying beam background. The failed events can also be true three-track vertices, which display poor reconstruction quality.

Figure 3.2 shows the expected *ChiProb* distribution for

$$e^+e^- \to \tau^+(\to \nu_\tau hhh + \#\text{neutrals} \geq 0)\tau^-(\to \ell^-\nu_\tau\nu)$$

events, $q\bar{q}$ events, radiative Bhabha events, and other background processes. The red line indicates the cut, selecting all samples above or equal to zero. The bin at -1 indicate all failed vertex fits. Most failed vertex fit events are background processes, but there are also failed signal events. In the case of long-lived intermediate particles, rejecting a failed vertex fit suppresses a portion of τ tags decaying into K_S^0. This rejection results in a slight reduction of the reconstruction efficiency. We choose this approach for simplicity, as the channel would otherwise require exclusive treatment.

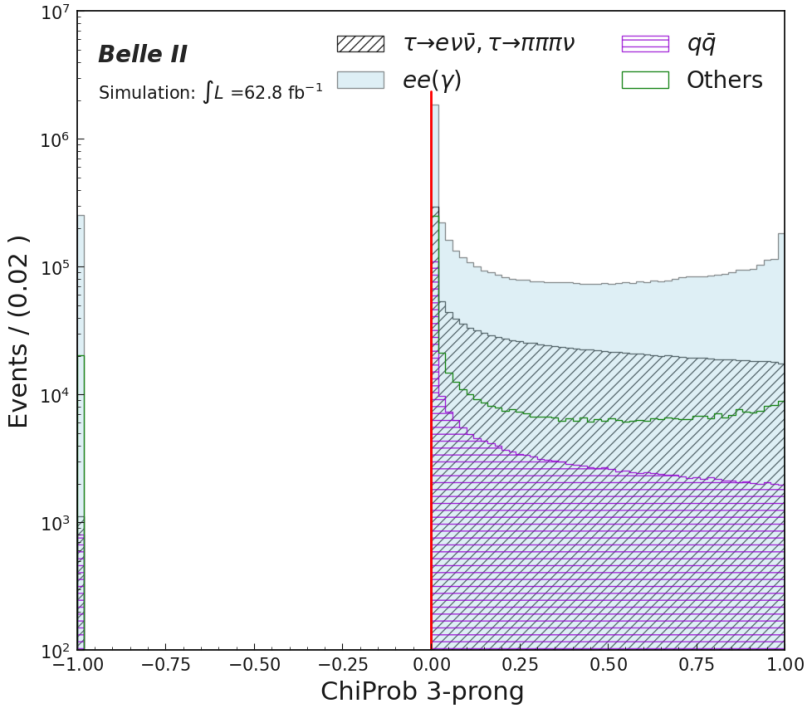


Figure 3.2: Common vertex probability *ChiProb* for 3-prong $\tau \to \pi\pi\pi\nu$ decay. The plot displayes the unstacked ChiProbe distributions for signal $\tau^+\tau^-$ events, $q\bar{q}$ events, $e^+e^-\gamma$ events, and other event types. A value of -1.00 corresponds to a failed vertex fit convergence. The red line indicates the cut, selecting all samples with a converged vertex fit [76].

3.3.2 Transverse Momentum Selection

When developing the selection strategy for the $\tau \to \ell\alpha$ analysis, Belle II did not provide hadron identification deemed reliable enough for this search. This lack of hadron identification necessitates an alternative selection, which enables differentiating hadronic 3-prong τ events from other physics processes.

We rank the three particles on the 3-prong tag side by their transverse momentum and uses the distributions for background suppression. These requirements help reduce significant contamination from the $e^+e^-(\gamma)$ and $\mu^+\mu^-(\gamma)$ final states. We obtain the following requirements by maximization of the FOM, Equation (3.1), for the electron channel of $\tau \to \ell\nu_\tau\nu_\ell$:

- $p_T^{\text{lead}} > 0.69$ GeV/c;
- p_T of the sub-leading track (p_T^{sub}) > 0.29 GeV/c; and
- p_T of the third track (p_T^{third}) > 0.08 GeV/c;

with the track having the highest p_T categorized as the leading track, p_T^{lead}; the track with the second highest p_T categorized as the sub-leading track, p_T^{sub}; and the track with the third highest p_T categorized as the third track, p_T^{third}. The muon channel results are

- $p_T^{\text{lead}} > 0.47$ GeV/c;
- $p_T^{\text{sub}} > 0.17$ GeV/c; and
- $p_T^{\text{third}} > 0.04$ GeV/c.

Figure 3.3 illustrates the selection and indicates the main background contribution for the electron channel. Figure 3.3a displays the background and signal distribution referred to as BG and $\tau \to \ell\nu_\tau\nu_\ell$, respectively. The two lepton gamma events, $\ell\ell\gamma$, heavily dominate the background. The $\ell\ell\gamma$ yield necessitates displaying the distributions in log scale. The red line indicates that the cut removes most background events. Figure 3.3b and Figure 3.3c also show an emerging divergence between the $\ell\ell\gamma$ and BG distributions. This divergence indicates a BG suppression power, but to a lesser extend than before. In Figure 3.3b, $\ell\ell\gamma$ events are still dominant, but we identify that the BG yield is also composed from other event types. Figure 3.3c at last shows the remaining background contributions, with background contributions other than $\ell\ell\gamma$, such as $q\bar{q}$, becomming more dominant. Figure G.1 in Appendix G shows the corresponding distributions for the muon channel. We observe a much lighter contamination of $\ell\ell\gamma$ events in the distributions on the muon channel, leading to softer cut requirements.

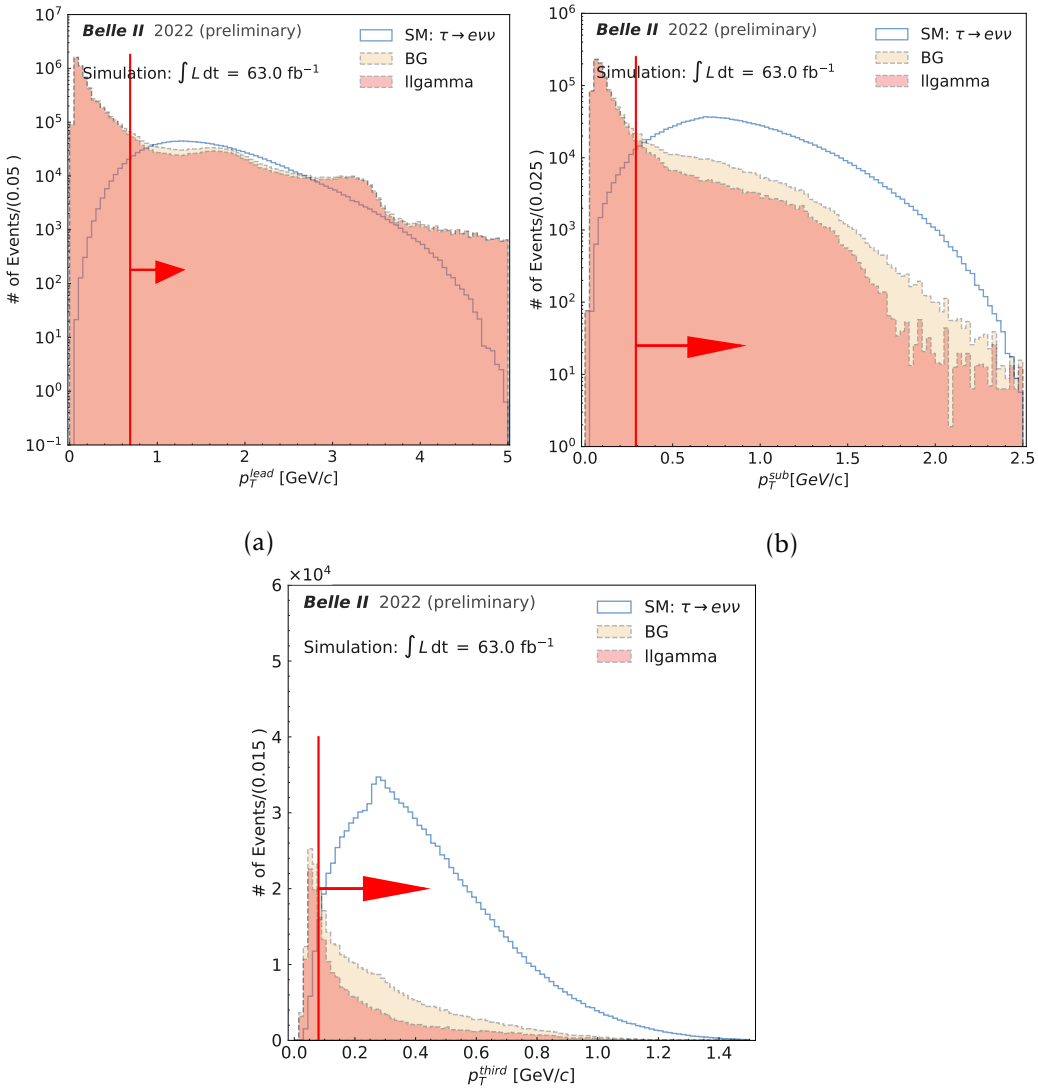


Figure 3.3: $\tau \to e\nu_e\nu_\tau$ channel distributions for the ranked p_T of the tag side (blue); the background, BG (bisque); and the main contributor to the background: two-lepton-gamma events, llgamma (light red). The red marker indicates the FOM optimized selection.

3.3.3 Neutral Selection

Neutral particles, such as γ or π^0, are produced in different processes with different yields, and this allows for differentiating between physics processes by considering the distribution of the number of produced neutral particles. As described in Section 3.2.1, we reconstruct two neutral particles, the γ, and π^0.

Two treatments for the neutrals are under consideration. The first treatment is to veto reconstructed neutral particles in the event. Typically vetoed decays are $\tau \to 3h\nu_\tau\pi^0$ on the 3-prong tag side and radiative decays such as $\tau \to \ell\nu_\ell\nu_\tau(\gamma)$ on the 1-prong side. We expect this veto to improve the resolution in the reconstructed τ rest frame and increase sensitivity.

The selection for the neutrals veto is:

- $\#\gamma_{1\text{prong}} = \#\gamma_{3\text{prong}} = 0$;
- $\#\pi^0{}_{1\text{prong}} = \#\pi^0{}_{3\text{prong}} = 0$.

The distribution of the number of photons, $\#\gamma$, and neutral pions candidates, $\#\pi^0$, for 1- and 3-prong sides show that these requirements also reduce the $q\bar{q}$ contamination. Figure 3.4 displays the neutral particle distributions for the electron channel and Appendix G in Figure G.1 displays the same distributions for the muon channel. We notice that especially $q\bar{q}$ events tend to have more neutral particles in the event than $\tau^+\tau^-$ events.

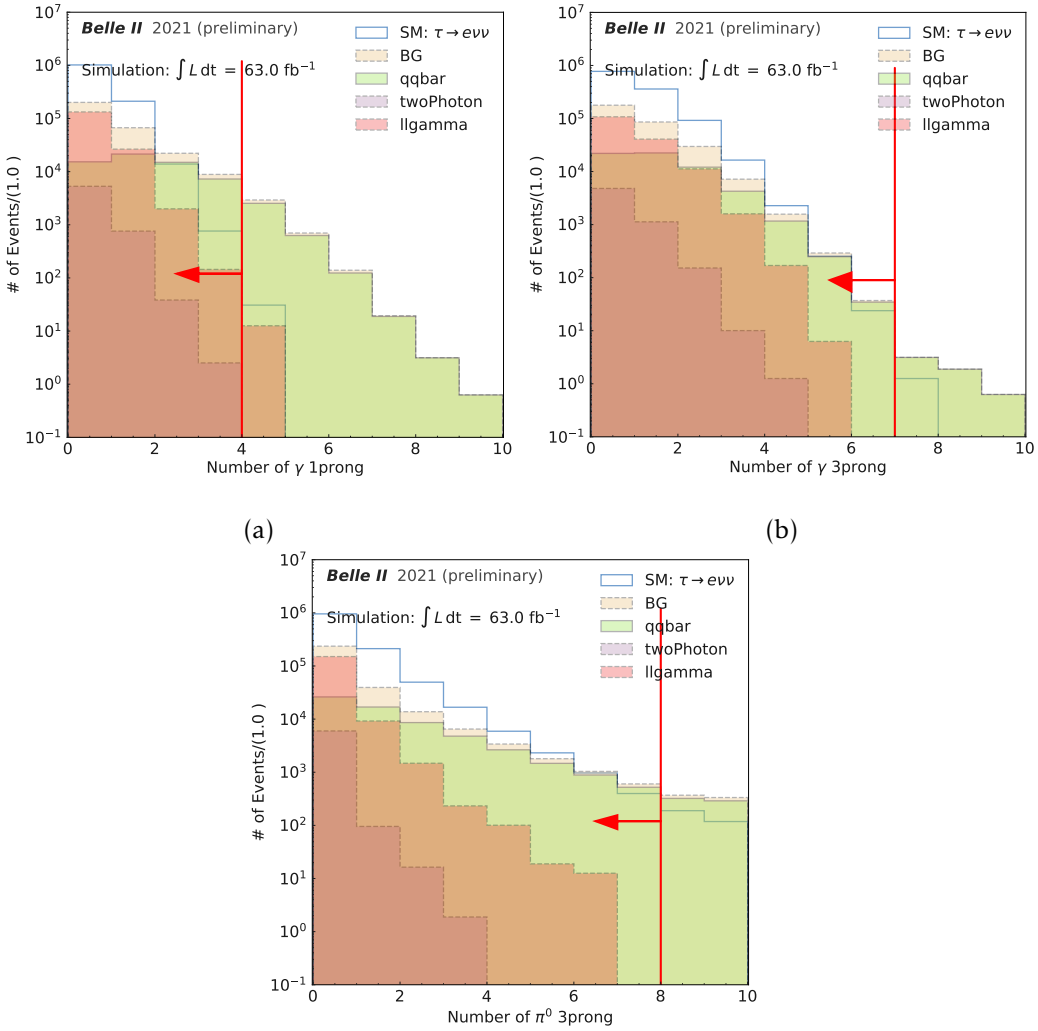
The second treatment is to place a cut on $\#\gamma$ and $\#\pi^0$ according to the maximized value of the FOM, Equation (3.1). Here, we reject any π^0 candidate on the signal side, because there is no such process in the model. The optimized selection for the electron channel is

- $\#\pi^0{}_{3\text{prong}} \leq 8$;
- $\#\gamma_{3\text{prong}} \leq 7$; and
- $\#\gamma_{1\text{prong}} \leq 4$.

The muon channel selection is

- $\#\pi^0{}_{3\text{prong}} \leq 6$;
- $\#\gamma_{3\text{prong}} \leq 6$; and
- $\#\gamma_{1\text{prong}} \leq 3$.

We observe a high number of potential γ and π^0 candidates in the event, which might come from beam background and double counting.



(a) (b)

(c)

Figure 3.4: $\tau \to e\nu_e\nu_\tau$ channel distributions for neutral particle candidates multiplicity of the event (blue). The background, BG (bisque), consists of several event types. They are $q\bar{q}$ events, qqbar (green); two-photon events, twoPhoton (violet); and two lepton gamma events, llgamma (light red). The red marker indicates the FOM optimized selection.

3.3.4 Tag-Side Selection

The tag-side selection is the final step of our event selection. It suppresses the background with any variable displaying a difference in distributions for the signal and background events. To not bias the distribution in the τ-rest frame of the signal, we decide to apply cuts only on tag-side variables and event variables with negligible dependence on the signal decay.

We investigated several variables and identified three effective ones for further suppressing background rates. The variables are *thrust*; the 3-prong energy in the centre-of-mass system (CMS), $E^{\text{CMS}}_{\text{3prong}}$; and the invariant mass of the 3 prong system, M_{3prong}.

We identify cut boundaries with the FOM optimization given by Equation (3.1). The order of listing corresponds to the order of optimization. We tested the order of optimization and found this to be the most significant one[5].

Table 3.1 lists the cut requirements for the electron and muon channels. The table shows both the neutrals rejection and neutral selection cases. The neutral selection case has diverging cut requirements from the neutral rejection case. The *thrust* selection is tighter, while the other selection requirements are more relaxed than the neutral rejection case.

Figure 3.5 allows a comparison of the two neutral treatments for the electron channel. The Appendix G in Figure G.2 shows the same for the muon channel. The left side displays the neutral veto case, and the right side the neutral selection case. We notice that the background distribution, BG in the figure, differs. In the neutrals rejection case, its distribution level close to the signal peak is on the order of one to ten percent compared to the signal for the *thrust* distribution of Figure 3.5a.

Furthermore, the tails end close to the lower cut bound. In contrast, the neutral selection case displays a distribution height about ten times higher, and the background distribution has a large tail extending to lower *thrust* values. Similar observations also hold for the other distributions. As expected, we observe that the background yield is higher for the neutral selection case.

Furthermore, we observe in Figure 3.5 that the *thrust* mainly reduces $\ell\ell\gamma$ processes, which have a *thrust* closest to one. M_{3prong} is an efficient cut for $q\bar{q}$ and two photon processes, such as $e^+e^- \to e^+e^-\ e^+e^-$. At last, $E^{\text{CMS}}_{\text{3prong}}$ again reduces the amount of $\ell\ell\gamma$ and two photon processes.

Table 3.2 lists the number of signal events and the purities and efficiencies of the cut-based selection. We treat the eelctron and muon channels individually and optimize the selection for the $\tau \to \ell\nu_\tau\nu_\ell$ decay. Differentiating the channels is necessary due to the different lepton identificantion, leptonID, performances at Belle II. We now test how well

[5] Simultaniouse optimization methods were also tested. The improvement was inferior to the results obtained with boosted decision trees – discussed further below – and hence dropped.

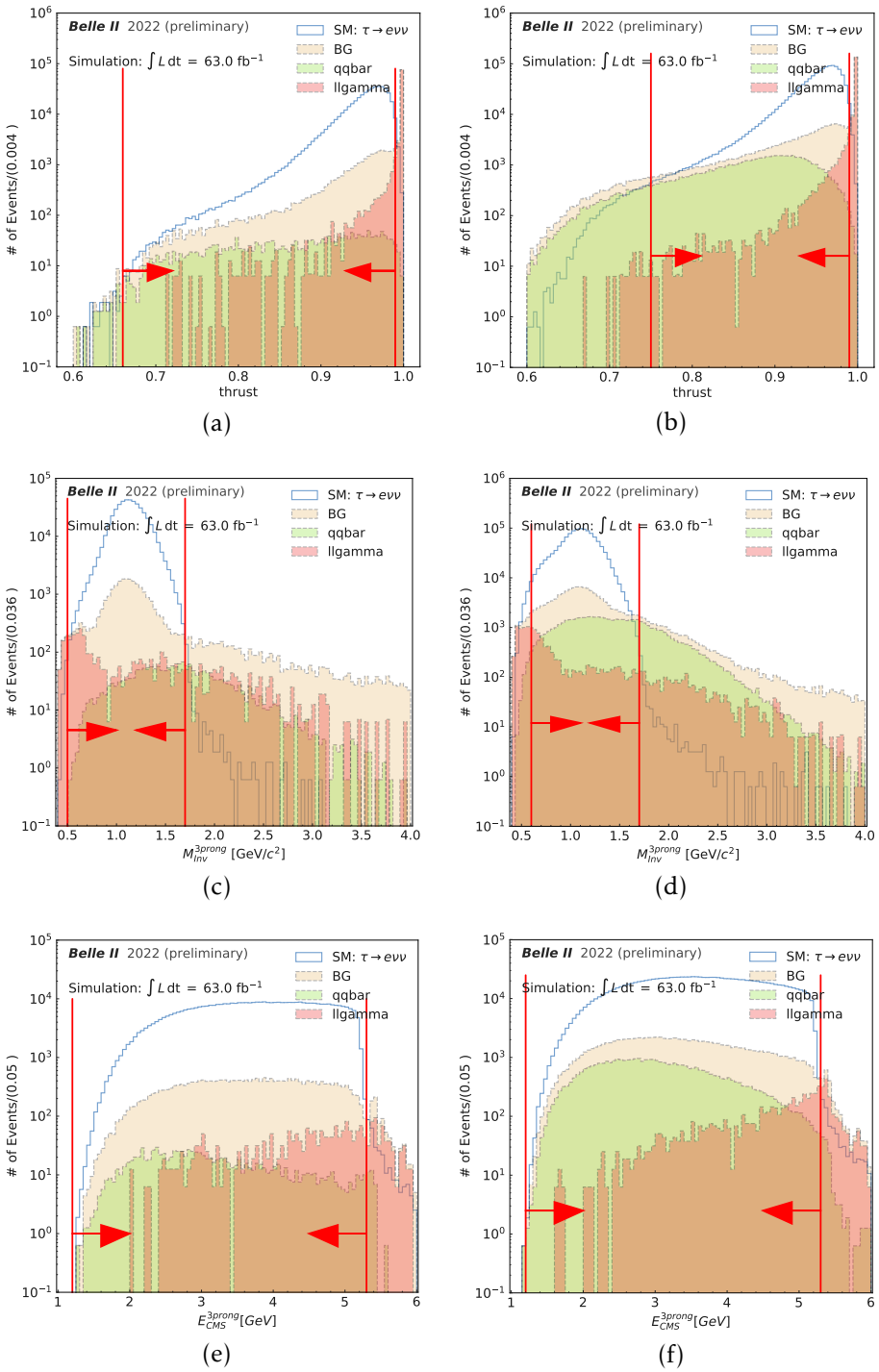


Figure 3.5: $\tau \to e\nu_e\nu_\tau$ channel distributions for the background suppression variables. The left side shows the case with a neutrals veto, the right side with FOM optimized neutrals. Red markers indicate the FOM optimized selection. The $\tau \to e\nu_e\nu_\tau$ channel distributions, SM: $\tau \to e\nu\nu$ is blue. The background, BG (bisque), consists of two main event types. They are $q\bar{q}$ events, qqbar (green) and two lepton gamma events, llgamma (light red). The red marker indicates the FOM optimized selection.

Table 3.1: Selection Ranges for the Tag-Side.

Channel:		e	μ	
Variant	Variable	Selection	Selection	Unit
no neutrals	*thrust*	(0.66, 0.99)	(0.69, 1.0)	
	M_{3prong}	(0.5, 1.7)	(0.4, 1.7)	GeV/c^2
	$E^{\text{CMS}}_{\text{3prong}}$	(1.2, 5.3)	(1.1, 5.3)	GeV
with neutrals	*thrust*	(0.75, 0.99)	(0.83, 1.0)	
	M_{3prong}	(0.6, 1.7)	(0.5, 1.7)	GeV/c^2
	$E^{\text{CMS}}_{\text{3prong}}$	(1.1, 5.4)	(1.1, 5.4)	GeV

Table 3.2: Results are listed for cut-based selection without neutrals, remaining event efficiencies are listed for all samples, and the purity is stated for the $\tau \to \ell \nu_\tau \nu_\ell$ sample. Section 3.1 defines efficiency and purity. The total remaining background events and selection efficiency are stated for comparison.

Channel		e			μ	
Sample $\tau \to l\alpha$	Events	Efficiency	Purity	Events	Efficiency	Purity
$M_\alpha = 0$	459589	15.12 %		591970	19.47 %	
$M_\alpha = 0.5$	460689	15.15 %		585720	19.27 %	
$M_\alpha = 0.7$	460396	15.14 %		581527	19.13 %	
$M_\alpha = 1.0$	459251	15.11 %		563060	18.52 %	
$M_\alpha = 1.2$	456160	15.01 %		537660	17.69 %	
$M_\alpha = 1.4$	423317	13.92 %		481741	15.85 %	
$M_\alpha = 1.6$	300784	9.89 %		292532	9.62 %	
$\tau \to \ell \nu_\tau \nu_\ell$	424057	17.04 %	95.69 %	525264	21.10 %	91.18 %
BG	19119	1.67e-07		50785	4.438e-07	

the Standard Model selection performs for the different $\tau \to \ell\alpha$ mass hypotheses. We observe that the results for the $\tau \to \ell\alpha$ mass hypothesis up to 1.2 GeV/c^2 show a similar cut efficiency as the $\tau \to \ell \nu_\tau \nu_\ell$ decay. The $\tau \to \ell\alpha$ mass hypotheses of 1.4 GeV/c^2 and 1.6 GeV/c^2 show a reduced efficiency. The reason is that their phase space, due to the high α mass, is notably different from the $\tau \to \ell \nu_\tau \nu_\ell$ decay, leading to a lower selection efficiency for the *thrust* and leptonID variables. The cut selection without neutrals achieves a high purity of above 90 % for the $\tau \to \ell \nu_\tau \nu_\ell$ sample, as was our selection goals.

Appendix H lists the same results for the cut-based selection with neutrals in Table H.2. We observe more than double the resulting efficiency for both the electron and muon channel but several percent degradations of the $\tau \to \ell \nu_\tau \nu_\ell$ purity. If the amount of data limits the sensitivity of our $\tau \to \ell\alpha$ search, then it is clear that the cut-based selection with neutrals is advantageous. If the number of misidentified events limits the sensitivity, then the cut-based selection without neutrals has the edge.

Appendix H also summarize the number of background events after the cut based selection with and without neutrals in the tables H.3 and H.1, respectively.

3.4 Background Suppresion Based on Boosted Decision Trees

Machine learning approaches can learn n-dimensional (nD) shapes. One method in the vast field of machine learning is gradient boosted decision trees [77]. The Extreme Gradient Boosting, XGBoost, package provides a library for learning boosted decision trees, BDTs; the documentation [78] gives a detailed description of the package and the principle of BDTs.

The BDT tool makes it possible to improve the background suppression performance further. The goal of this study is to improve purity and efficiency simultaneously. This study trained BDTs with two sets of variables in analogy to the cut-based background suppression study. To directly compare the BDT approach with the cut-based approach, the study based the variables on the cut-based background suppression. Both BDTs were hyper-parameter optimized for purity using a grid search based on [79]. The optimized parameters are the number of boosted decision trees, the depth of the boosted decision trees, and the learning rate.

The BDT without neutrals uses the three tag-side selection variables

- *thrust*;
- $E_{\text{3prong}}^{\text{CMS}}$; and
- M_{3prong}.

Prior to the application of the BDT, the data has to meet the event requirements of the neutral rejection case, using the reconstruction level event selection, the veto for failed vertex fits, ranked p_T cuts, γ-, and π^0-veto for the event, as is done for the cut selection without neutrals. Here, the ranked p_T cuts are included in the event requirement to ensure comparability with the cut selection without neutrals.

The BDT with neutrals uses all FOM optimized variables, similar to the cut-based selection with neutrals. The event requirement is more relaxed than before; they include the reconstruction level event requirements and the veto for failed vertex fits. In addition to the previous BDT, this BDT can optimize the ranked p_T and neutral cuts. The input variables are

- p_T^{lead};
- p_T^{sub};
- p_T^{third};
- $\#\pi^0{}_{\text{3prong}}$;
- $\#\gamma_{\text{3prong}}$;
- $\#\gamma_{\text{1prong}}$;
- *thrust*;

- $E^{\text{CMS}}_{\text{3prong}}$; and
- M_{3prong}.

Figure 3.6 shows the classification output for both BDT variants for the signal and background samples of the electron channel. Appendix G shows the classification for the muon channel in Figure G.3. We observe a good separation power for both BDT variants for most non τ-related background events. In the case of the BDT without neutrals, we observe some misclassification of $q\bar{q}$ events with an excess towards one, indicating that certain topologies of the $q\bar{q}$ trick the BDT. The BDT with neutrals seems less affected than the other, with the $q\bar{q}$ events more evenly distributed. The BDT with neutrals might suffer from the higher overall $q\bar{q}$ rate. Also, in figures G.4 and G.5 we clearly demonstrate that there

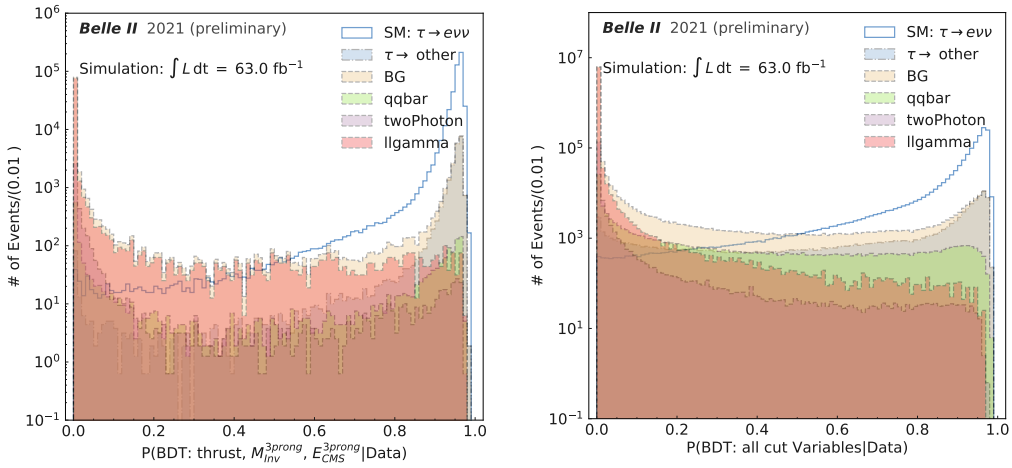


Figure 3.6: $\tau \to e\nu_e\nu_\tau$ channel distributions for the BDT without neutrals classification (a) and BDT with neutrals classification (b). The BDT classification results display the non τ background contributions.

is no over-training issue with the BDTs. Both variants are capable of discriminating τ-signal from other events.

The BDT-output selection requirement is such that it reproduces the best cut purity because we aim to achieve the purest possible sample, and the goal of this study is to compare the cut-based method directly to the BDT approach. In this section, we will compare only the cut-based selection without neutrals, which has the best purity. An overall comparison follows in the next section. The Table 3.3 and Appendix H in Table H.4 show the corresponding efficiencies when the selection requirement for the BDTs output reproduce the best purity, which is from the cut-based selection without neutrals. Table 3.3 displays the BDT with neutrals. We observe a higher efficiency, about twice as high when compared to the cut-based selection without neutrals. Furthermore, when comparing the relative efficiency differences, we see that the higher mass hypothesis for the $\tau \to \ell\alpha$ has a lower reduction rate than the cut-based

selection without neutrals. Appenmdix H in tables H.3 and H.1 shows in more detail the number of surviving background events. Furthermore,

Table 3.3: Results listed for BDT with neutrals, remaining events efficiencies, and purities are stated for signal samples. The total remaining BG events and selection efficiency is stated for comparison.

ok		e			μ	
Sample $\tau \to l\alpha$	Events	Efficiency	Purity	Events	Efficiency	Purity
$M_\alpha = 0$	1032530	33.96 %		1334719	43.91 %	
$M_\alpha = 0.5$	1032302	33.96 %		1326419	43.63 %	
$M_\alpha = 0.7$	1031086	33.92 %		1316952	43.32 %	
$M_\alpha = 1.0$	1023076	33.65 %		1282225	42.18 %	
$M_\alpha = 1.2$	1011809	33.28 %		1235427	40.64 %	
$M_\alpha = 1.4$	967694	31.83 %		1127877	37.1 %	
$M_\alpha = 1.6$	823620	27.09 %		784931	25.82 %	
$\tau \to \ell\nu_\tau\nu_\ell$	978463.56	39.31 %	95.69 %	1199573.57	48.19 %	91.18 %
BG	44111.35	3.85e-07		115978.79	1.0e-06	

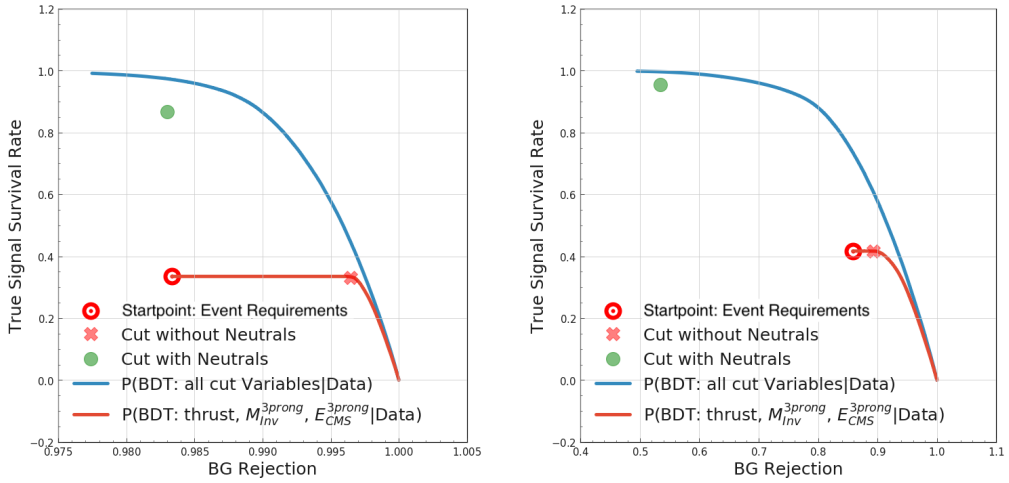
Appendix G exemplifies the capability of the BDT with neutrals to learn the nD-shape of the signal in the figures G.6 and G.7. They also show that the shapes are similar. The study must train dedicated BDTs for the electron and muon channels due to the different performances of the lepton identification, lepton ID. Appendix G also illustrates the shape differences between background and signal distributions. Figure G.8 show the ranked p_T variables in the muon channel case. Figure G.9 displays similar distributions for the electron channel.

3.5 BDT- vs. Cut-Based Background Suppresion

We compare the four background suppression studies to understand and verify the results. The BDT selections reproduce the best cut purity to compare the BDT and cut-based selection methods directly.

Figure 3.7 shows the receiver operating characteristic, ROC, curves for the electron channel (Plot 3.7a) and the muon channel (Plot 3.7b). The BDT without neutrals (red line) does not perform better than the same cut-based approach (red cross) and is worse than the background suppression methods allowing for neutrals. We reject the BDT without neutrals as a good background suppression tool. It introduces additional complexity that needs to be understood and has no advantage over the cut-based analysis.

The final decision for our cut approach depends on several considerations. First, the BDT approach introduces an additional level of complexity that must be understood and can introduce additional systematic uncertainty. A considerable weakness is a dependence on the Monte Carlo simulation in the early phase of an experiment. Belle II is still in the process of understanding its detector, and we are aware of considerable simulated to measured data discrepancies . This situation necessitates that every variable is studied and approved for the use in a BDT.



(a) $\tau \to e\,\nu_\tau\,\nu_e$ (b) $\tau \to \mu\,\nu_\tau\,\nu_\mu$.

Figure 3.7: The ROC curves for the two BDTs and the corresponding cut selection are displayed. The Startpoint from event requirements indicates the signal survival rate and BG rejection given by the ranked p_T cuts and the neutrals veto.

For these reasons, we choose to use the cut-based analysis in this first search iteration.

Second, the cut-based selections differ in their purity and efficiency. Although the cut-based selection with neutrals has higher efficiency, it suffers in purity. We choose to use the cut-based selection without neutrals because it results in higher purity and allows for a sideband definition by reversing the neutrals cuts. This sideband option is essential because other sidebands such as $\tau^\pm \to \pi^\pm \nu_\tau$ are unavailable due to insufficient $\pi^\pm$ identification. We were worried about unblinding the measured data during the simulated data validation phase with measured data.

A future iteration of the $\tau \to \ell\alpha$ search at Belle II should further study the background suppression with optimized neutrals and utilize BDTs. As Table 3.3 shows, the BDT with neutrals can increase the available data by more than a factor of two, displaying the potential for the next iteration of this search. This simplified study shows that the BDT-based approach is superior to the cut-based approach when allowing for neutral particles. Figures 3.7a and 3.7b indicate that the BDT can achieve better purity and efficiency than the cut-based approach.

For completeness, in Chapter 6 we will discuss the impact of the selection method when considering the resulting UL, which is an estimate for the sensitivity.

3.6 Validation of Simulation Using Measured Data

In high-energy physics, it is common to conduct searches for new physics events with blinded analysis, which means the whole process of setting up the analysis machinery uses simulated data. After presenting the analysis approach on simulated data to the collaboration and convincing them that the analysis is sound, the collaboration grants permission to unblind the measured data, allowing the use of the recorded experimental data set.

Typically, the unblinding process involves several steps to enhance confidence in the analysis setup and avoid burning the measured data due to discrepancies. The first step is to validate that after the selection, the simulated data set is a good representation of the measured data set. This step should still be blind to deviations indicating a signal because this blind state enables mitigating emerging deviations between simulated and measured data.

It is common to identify variable ranges, entire variable distributions, or introduce even entire physics processes with no discriminatory power for the search that enables validating the selection by comparing simulated to measured data distributions. Here, so-called data-driven cuts are introduced, which mitigate occurring differences.

Once analysts establish confidence in the data selection, the next step may include validating the signal distributions and checking whether the analyst's approach determined on simulated data is valid. A typical approach is to evaluate ten percent of the measured data. After establishing confidence in the method, the collaboration grants the search permission to consider the entire measured data set.

This section focuses on the selection validation, introduces our data-driven cuts, and compares the resulting distributions for selected variables, establishing confidence in the selection procedure.

3.6.1 Strategy

Establishing confidence in the Belle II simulation without unblinding the measured data is essential to our search. To this end, we developed a validation strategy for simulated data, and this section describes and justifies it.

The general idea is to compare simulated and measured data with two methods:

1. Considering 10 % of measured data ($6.3\,\mathrm{fb}^{-1}$) in the signal region and comparing only the variables where there is no discernible difference in shape between the background and signal+background hypothesis. We refer to such variables as "safe" (see below subsection). The selections listed in Table 3.1 define the signal region.

2. Considering the total measured data sample ($62.8\,\text{fb}^{-1}$) in a sideband region that is more background enriched. A looser, inverted version of the signal region selections defines the sideband. Figure 3.8 illustrates this idea, and the below section provides the set of selections.

These methods are complementary. The first allows us to check the agreement between measured and simulated data for the exact topology and kinematics of the signal region, although for a restricted set of "safe" variables. The second method allows for a more detailed study of the main backgrounds with no restrictions on the variables.

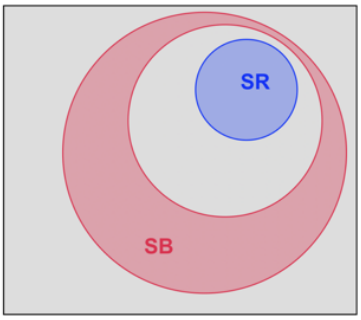


Figure 3.8: Simplified diagram showing the full phase space (gray box), signal region (SR, blue circle), and sideband (SB, red circle). Here, the signal region and sideband do not touch since the latter is defined by inverting a looser version of the signal region cuts.

Validation Approach in the Signal Region with "Safe" Variables

When comparing measured to simulated data in the signal region, we are cautious to study only those variables that will not reveal the presence or absence of the signal process. Typically, this unblinding will not happen when looking at tag-side variables agnostic to what happens on the signal side.

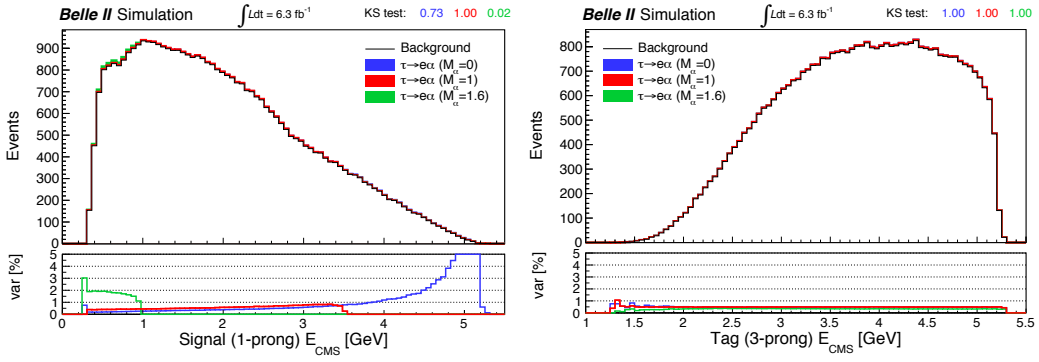
To decide whether or not a variable is "safe", we perform a statistical test of compatibility in shape between the background and signal+background distributions with simulated data. This test uses the so-called two-sample Kolmogorov-Smirnov test[6], KS test, [81, 82]. Here, the signal is normalized according to the upper limit from the ARGUS Collaboration [36]. If the KS test returns a number close to one, there is no discernible difference in shape, and we regard the variables as "safe".

Figure 3.9 provides two examples of the input distributions and KS test results. The left side shows an example distribution of an "unsafe"

[6]The test statistic is defined as $D_{n,m} = \sup_x |F_n(x) - F_m(m)|$. $F_i(x)$ is an empirical cumulative distribution function dependant of i and the identically distributed observations X_j. From $D_{n,m}$ we determine the level α. The test returns the $1-\alpha$ probability that $F_n(x)$ is from the same underlying cumulative distribution function as $F_m(x)$ [80].

variable. When considering the relative variation, we notice a shape difference between the Standard Model $\tau \to \ell \nu_\tau \nu_\ell$ and $\tau \to \ell \alpha$ decays. This difference is especially notable for α mass hypothesis at the edges of the possible mass spectrum given by the τ mass. We can confirm this observation by looking at the KS test result in the upper right corner of the plot. Here, the α mass hypothesis for 0.0 GeV/c^2 and 1.6 GeV/c^2 yield KS test results well below 1.0.

In contrast to the left side of Figure 3.9, there is no noticeable difference in the distribution on the right side. Furthermore, the KS test result



(a) (b)

Figure 3.9: Comparison of the expected background and signal+background distributions in the $\tau \to e\nu\bar{\nu}$ channel for (a) signal side E_{CMS} and (b) tag side E_{CMS}. We normalized the simulated data distributions to 10 % of the measured data's luminosity (6.3 fb^{-1}), and scaled the signal to the current upper limits from the ARGUS Collaboration. The results from the Kolmogorov-Smirnov test (KS test) are indicated on the top right. The lower panel shows the variation in percent of the signal+background concerning the background-only [76].

is 1.0 for all mass hypotheses. Appendix H summarizes the complete set of KS test results in Table H.7, from which we identify the following safe variables:

- **Tag side variables**: $E_{\text{CMS}}^{\text{3prong}}$, $M_{\text{Inv}}^{\text{3prong}}$, *ChiProb* , $\#\gamma$, $\#\pi^0$, p_T^{lead}, p_T^{sub} and p_T^{third}.
- **Event shape/kinematic variables**: *thrust* and the missing momentum in the CMS for the angle θ, *Missing momentum* θ_{CMS}.

Validation Approach in the Sideband

We took the following considerations into account when defining the sideband:

1. Choosing a region with the main backgrounds, mainly $q\bar{q}$ and $\tau^\pm \to \pi^\pm(\#\pi^0 \geq 0)\nu_\tau$. Here a charged pion fakes a lepton.

2. Ensuring separation from the signal region but still being as kinematically close to it as possible.
3. Having minimal signal leakages into the sideband.

Fulfilling these requirements necessitates events entering the sideband to satisfy most (but not all) of the signal region selections. We explicitly veto events that pass the complete signal region selections with looser neutrals cuts. Table 3.4 summarises the selections that define the sideband.

Table 3.4: Summary of the sideband selection criteria. The signal region (SR) refers to the selections defined earlier in Table 3.1.

Selection criteria	$\tau \to e\nu\bar{\nu}$	$\tau \to \mu\nu\bar{\nu}$
Event	$0.9 < thrust < 0.99$	$0.9 < thrust < 1.0$
Signal (1 prong)	$electronID > 0.9$	$muonID > 0.9$
	in the acceptance of lepton e-ID performance measurements	in the acceptance of lepton μ-ID performance measurements
Tag (3 prong) $\tau \to 3h(\nu)$	$clusterE/p \leq 0.8$	
	common vertex probability $\chi^2 > 0$	
	$p_T^{\text{lead}} > 0.69$	$p_T^{\text{lead}} > 0.47$
	$p_T^{\text{sub}} > 0.29$	$p_T^{\text{sub}} > 0.17$
	$p_T^{\text{third}} > 0.08$	$p_T^{\text{third}} > 0.04$
	$1.2 < E^{CMS}_{\text{3prong}} < 5.3$	$1.1 < E^{\text{CMS}}_{\text{3prong}} < 5.3$
SR orthogonality	veto events that fall into the SR with looser neutrals selections: $\#\boldsymbol{\pi^0}_{\text{1prong}} = 0$, $\#\boldsymbol{\pi^0}_{\text{3prong}} \leq 2$ $\#\boldsymbol{\gamma}_{\text{1prong}} =\leq 1$, $\#\boldsymbol{\gamma}_{\text{3prong}} \leq 1$	

3.6.2 Corrections

We apply all known corrections for simulated and measured data to estimate how well the Belle II simulation models the measured data. We apply several efficiency and track-related corrections. For these corrections, we follow the official recommendations of the Belle II Performance Group [83].

We apply the following corrections to simulated data:

- *lmlX* trigger efficiency determined in measured data. We bring this relative trigger efficiency to an absolute trigger efficiency by applying a small correction from the trigger simulation, TSIM. The Belle II note provides a detailed description of how we calculate this correction [76].
- Measured over simulated data ratio of the *electronID* and *muonID* efficiencies. We use the tables provided by the Belle II LeptonID Performance Group [84].

- Measured over simulated data ratios of the $\pi^{\pm} \to e^{\pm}$ and $\pi^{\pm} \to \mu^{\pm}$ fake rates. These are taken from the tables provided by the Belle II LeptonID Performance Group [84].
- Measured over simulated ratios of the π^0 reconstruction efficiency. The Belle II Neutrals Performance Group provides tables with these ratios [85].

We apply a global momentum scale correction to each track for measured data. This correction accounts for imperfections in the magnetic field map used to reconstruct measured data. When applying this correction, we follow the recommendations of the Belle II Tracking Group [86].

3.6.3 New Data-Driven Selections

When moving from simulated data studies to measured data, it is common to account for the miss-modeling of simulated data by rejecting the miss-modeled region. We refer to this rejection as a data-driven requirement.

We impose the following data-driven requirements known from previous studies:

- *thrust* > 0.9; and
- leptonID weight $w_i^{\ell ID} > -1$.

The first is a data-driven requirement that rejects the remaining two-photon exchange processes that are not simulated [87]. The second cut restricts the lepton kinematics to the $p - \theta$ region where lepton identification corrections are available [84]; for more details see Section 5.5.2.

After applying the known data-driven requirements, we find fewer events with a reconstructed γ than expected. This inefficiency leads to an excess in events without γ, as figures 3.10a and 3.10b show. Furthermore, we observe an excess of data in the *Missing momentum* θ_{CMS}, at a θ_{CMS} close to the beam pipe. We can see these deviation well by considering the lower ratio for measured and simulated data at $\theta_{\mathrm{CMS}} \sim 0°$ and $\theta_{\mathrm{CMS}} \sim 180°$. The following subsection explains how we treat these discrepancies.

Photon Isolation and Timing Selections

We investigated several detector variables which impact the reconstruction of γ to identify the miss-modeling of photons. An essential aspect in γ reconstruction is the number of ECL crystals in which the γ showered. The energy deposition above 10 MeV in one ECL crystal gets registered as a cluster. Neighboring-ECL crystals collectively form an ECL cluster by adding those with energies above 0.5 MeV to the cluster. Usually, particles depositing energy in the ECL produce an ECL cluster.

In measured data, we observe several kinds of backgrounds for ECL clusters, which are challenging to model precisely in simulated data. These background ECL clusters may occur due to hadronic split-off clusters, a single low momentum track resulting in several clusters, and beam background events.

We include two additional requirements to the photon reconstruction criteria – described in Section 3.2.1 – to suppress this kind of background cluster and account for this miss-modeling. We use the minimum distance between a cluster and the nearest track, *clusterC2TDist*, to ensure the cluster isolation of photon candidates:

- $[clusterC2TDist > 40\,\text{cm}]$ or $[E(\gamma) > 400\,\text{MeV}]$;

This requirement helps reject hadronic split-offs and multiple clusters from a single track that tend to mimic low-energy photons, as illustrated in Figure 3.11.

In addition, very out-of-time clusters from beam background events are rejected with the following loose timing requirement:

- $|clusterTiming| < 200$ ns;

which follows the official recommendation of the Belle II Neutrals Performance Group [88].

Cut on Missing Momentum Polar Angle

As mentioned above, we observe an excess in measured data when the missing momentum vector is close to the beam pipe, as illustrated by the measured vs. simulated data ratio plot in Figure 3.12 at the angles $\theta_{\text{CMS}} \sim 0°$ and $\theta_{\text{CMS}} \sim 180°$. This excess is likely due to two-photon processes not included or miss-modeled in Belle II simulations. We solve this issue by requiring events that satisfy:

- $20° < \textit{Missing momentum}\ \theta_{CMS} < 160°$.

Impact on Selection Optimization

The selection of our $\tau \to \ell\alpha$ analysis remains unchanged because after incorporating the new neutral selection, the optimization yielded only second decimal changes in some variables.

The new photon requirements and the *Missing momentum* θ_{CMS} cut improve the signal efficiency by about 3 %. We verify this by comparing Table 3.2 with Table 3.5. This counterintuitive result is due to the updated photon requirements leading to a less efficient selection of γ, and thus π^0 candidates. We enhance the number of selected events because a less efficient γ selection leads to fewer events affected by the neutrals veto, resulting in increased efficiency. The *Missing momentum* θ_{CMS} cut mitigates the effect.

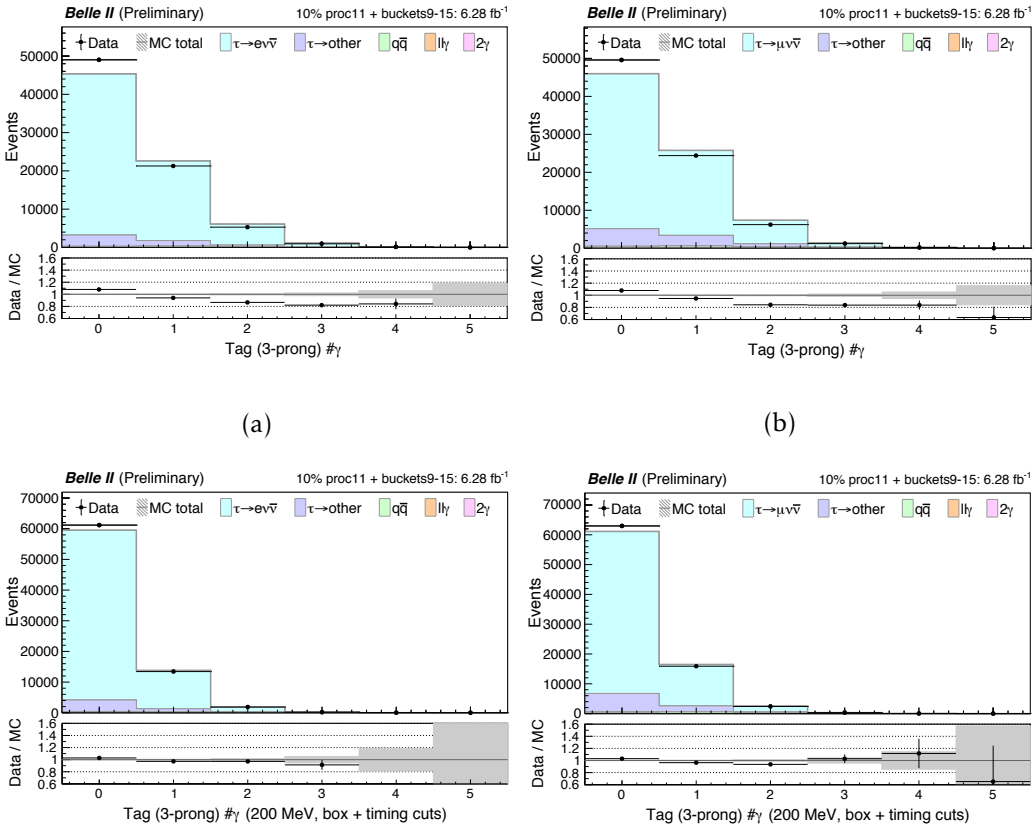


Figure 3.10: Measured vs. simulated data distributions of the tag-side #γ before, plots (a) and (b), and after, plots (c) and (d), including the photon-isolation and timing-criteria. The left shows the electron, and the right shows the muon channel signal regions. The error band includes the statistical and trigger efficiency-related uncertainties.

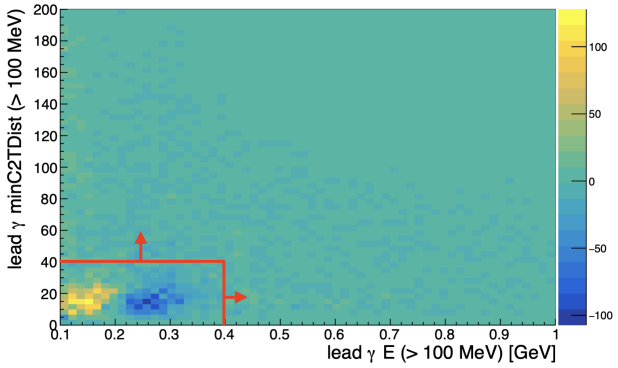
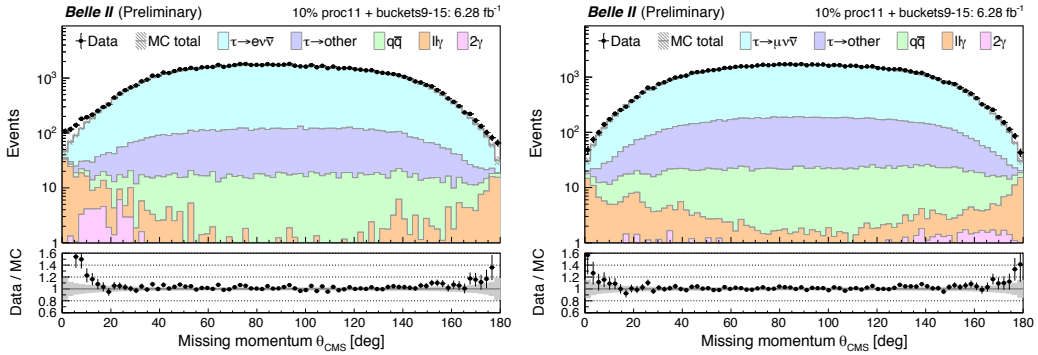


Figure 3.11: Measured vs. simulated data ratio distribution in the plane of highest energy and *clusterC2TDist* for the leading photon. The red box and arrows indicate the photon isolation requirement mentioned above.



(a) (b)

Figure 3.12: Measured (Data) and simulated (histograms) data distributions of the *Missing momentum* θ_{CMS} in log scale (top) for the electron channel (a) and muon channel (b) signal regions. The simulated data is a stacked histogram consisting of the $\tau \to \ell\nu_\tau\nu_\ell$ decay, the Standard Model decay we optimize for, in the electron channel ($\tau \to e\nu\nu$) and muon channel ($\tau \to \mu\nu\nu$); all misidentified τ decays ($\tau \to$ other); $q\bar{q}$ events, $q\bar{q}$; radiative lepton events, $ll\gamma$; and two-photon processes, 2γ. The simulation uncertainty is a gray band. The error band includes the statistical and trigger efficiency-related uncertainties. The ratio of the measured and simulated data (Data/MC) is in the bottom plot, and we expect it to be one. The top left corner specifies the Belle II data sample and indicates the fraction in percent and absolute numbers.

Table 3.5: New remaining event efficiencies and purities for signal samples.

ok		e			μ	
Sample $\tau \to l\alpha$	Events	Efficiency	Purity	Events	Efficiency	Purity
$M_\alpha = 0$	551115	18.13%	-	679090	22.34%	
$M_\alpha = 0.5$	556678	18.31%	-	675761	22.23%	
$M_\alpha = 0.7$	562005	18.49%	-	676363	22.25%	
$M_\alpha = 1.0$	574226	18.89%	-	668749	22.00%	
$M_\alpha = 1.2$	581010	19.11%	-	652625	21.47%	
$M_\alpha = 1.4$	546159	17.97%	-	598855	19.70%	
$M_\alpha = 1.6$	389773	12.82%	-	373125	12.27%	
$\tau \to l\nu\bar{\nu}$	547103	17.50%	95.93%	643186	21.08%	91.93%

3.6.4 Results of the Measured vs. Simulated Data Comparison

As discussed above, we aim to establish confidence in the simulation modeling. The first step is to check if the distributions of "safe" variables now agree with the measured data.

After verifying the agreement using ten percent of the measured data in the signal region with "save" variables – variables without discriminatory power for our $\tau \to \ell\alpha$ search –, we will consider the entire measured data set in the sideband. Later, we evaluate "unsafe" variables – with discriminatory power for our $\tau \to \ell\alpha$ search.

First, this section shows the measured to simulated data agreement in the signal regions. As described in the above sections, we consider only

ten percent of the total measured data set and include all corrections and the data-driven selection criteria.

Second, this section shows the measured to simulated data agreement in the sidebands. As described in Section 3.6.1, the entire data sample is considered. Again all corrections and the data-driven selection criteria described above are included.

In the plots presented below, the simulated data's error band includes the following uncertainties summed in quadrature:

- Statistical uncertainty of simulated data ;
- Trigger efficiency correction uncertainty;
- LeptonID efficiency correction uncertainties;
- $\pi^{\pm} \to \ell^{\pm}$ fake-rate correction uncertainties;
- π^0 efficiency correction uncertainty.

In Figure 3.12 we verified a good agreement between measured and simulated data within the data-driven requirements. The left side of Figure 3.13 further underlines this result. Both, Figure 3.13a for the electron channel and Figure 3.13c for the muon channel display good agreement between measured and simulated data, with the measured data mildly overshooting.

We also observe good agreement between measured and simulated data in the μ-channel sideband. However, as illustrated in Figure 3.13b, there is an overall deficit in the e-channel sideband. The reason for this observation is a miss-modeling of the $q\bar{q}$ background; the Belle II-note discusses a possible treatment in more detail [76].

Although $q\bar{q}$ has a significant contribution in the sideband, its contribution in the signal regions is minimal ($< 0.1\,\%$). This way, we do not expect a significant impact on the upper limit due to the $q\bar{q}$ normalization factor. We verified the negligible contribution in the Belle II-note [76].

Since we already verified an overall good agreement between measured and simulated data with Figure 3.13, the total collection of all measured to simulated data comparison plots for this section is presented in Appendix G. Figure G.16 and Figure G.17 display the signal region distributions for the electron and muon channel. Figure G.18 and Figure G.19 show the sideband region distributions for the electron and muon channel, respectively. In total, we demonstrated an excellent agreement between measured and simulated data and are allowed to unblind the data.

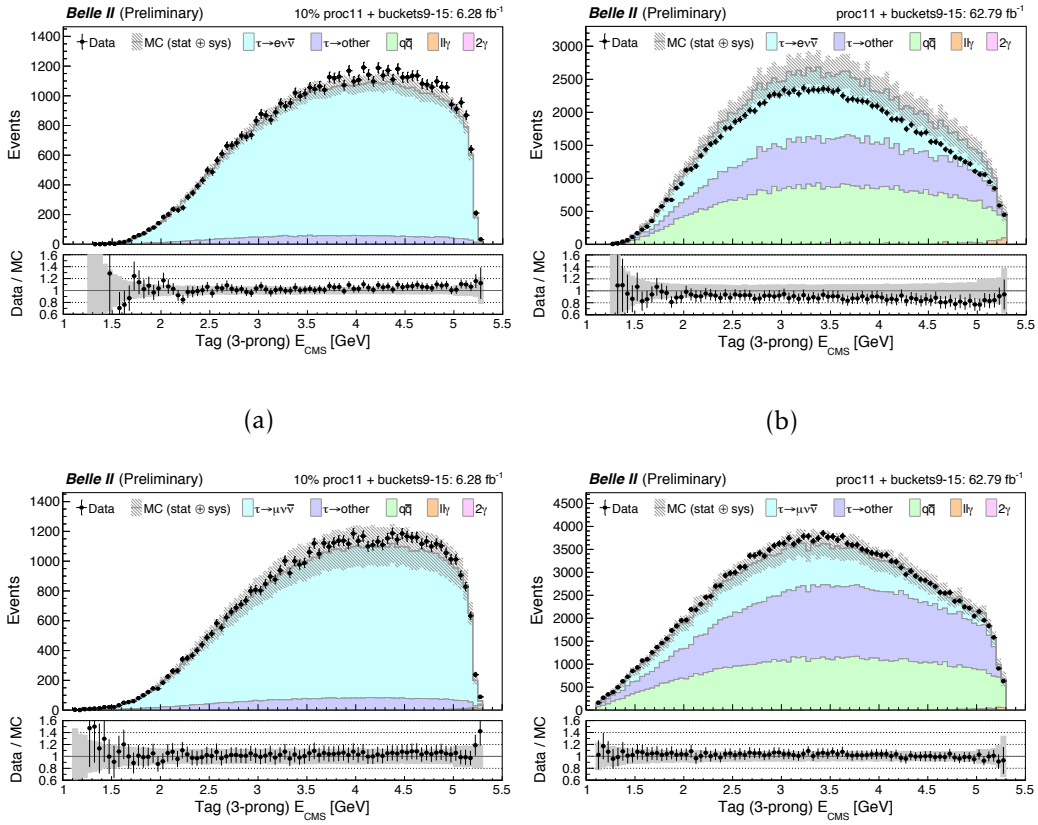


Figure 3.13: Measured (Data) and simulated (stacked histograms) data distributions of the tag E_{CMS} in the (a) e-channel signal region, (b) e-channel sideband, (c) μ-channel signal region and (d) μ-channel sideband. For each figure, the top plot is the resulting histogram, and the bottom is the measured vs. simulated data ratio. The simulated data is a stacked histogram consisting of the $\boldsymbol{\tau \to \ell \nu_\tau \nu_\ell}$ decay, the Standard Model decay we optimize for, in the electron channel ($\tau \to e\nu\nu$) and muon channel ($\tau \to \mu\nu\nu$); all misidentified $\boldsymbol{\tau}$ decays ($\tau \to$ other); $\boldsymbol{q\bar{q}}$ events, $q\bar{q}$; radiative lepton events, $ll\gamma$; and two-photon processes, 2γ. The total simulation uncertainty (MC(stat $\bigoplus$ sys)) is a gray band. The error band includes the statistical and all systematic uncertainties listed in the text. The ratio of the measured and simulated data (Data/MC) is in the bottom plot, and we expect it to be one. The top left corner specifies the Belle II data sample and indicates the fraction in percent and absolute numbers.

Chapter 4

Determining the τ-Rest Frame

We discussed in the previous chapters that boosting[1] into the τ rest frame provides the most significant shape difference of the lepton momentum between Standard Model $\tau \to \ell\nu_\tau\nu_\ell$-decay and the $\tau \to \ell\alpha$-decay. The reason is that the two-body decay signature is mono-energetic with one expected value for the lepton momentum; in contrast, the Standard Model $\tau \to \ell\nu_\tau\nu_\ell$ decay has a three-body decay structure with a spectrum of expected lepton momenta. For this boost, we need the flight direction of the signal-τ lepton.

We can use the reconstructed tag-particle for determining the flight direction in particle pair events. In the center-of-mass system of the event, we obtain the flight direction of the signal particle by reversing the tag particle's flight direction.

In $\tau^+\tau^-$ events, we face the difficulty that the tag-τ is not completely reconstructable due to the missing ν_τ. It is, however, possible to approximate the momentum of the τ and boost into a τ-pseudo-rest frame.

In the following, we will consider three methods to determine the τ rest frame. The first method is the ARGUS method developed by the ARGUS collaboration, and it uses the visible daughter particles of the tag-τ to estimate the signal-τ flight direction. The ARGUS collaboration used this method for their $\tau \to \ell\alpha$ search. The second method is a variant of the ARGUS method called the Thrust method, and it uses the thrust direction as a flight direction estimate. The third method is the Generalised Known Kinematics (GKK) method, a numeric sampling approach using all available kinematic information to obtain a probability distribution per event.

4.1 The ARGUS Method

In 1995, the ARGUS collaboration used a 3x1-prong signature to search for $\tau \to \ell\alpha$-decay [36]. They pioneered the approach to boost into a τ

[1] A Lorentz transformation into another restframe.

restframe to study the $\tau \to \ell\alpha$-decay, and their approach is the inspiration for the other methods discussed afterward.

The ARGUS collaboration approximates the τ-lepton direction with the direction of the 3-prong system's momentum vector

$$\widehat{p_\tau} \approx \frac{\vec{p}_{3h}}{|\vec{p}_{3h}|} \,. \tag{4.1}$$

In addition, they use the fact that half of the center-of-mass-beam energy, $\sqrt{s}$, is a reasonable estimate of the τ energy, E_τ – up to radiative corrections of the initial state:

$$E_\tau \approx \frac{\sqrt{s}}{2} \,. \tag{4.2}$$

and calculate the momentum vector of the τ lepton with this approximation. Combining these two approximations enables transforming to the so-called τ pseudo-rest frame.

Figure 4.1a shows the lepton momentum distribution of the $\tau \to \ell\alpha$ channel in the pseudo-rest frame for three mass hypotheses. We can observe a discernable shape difference for the lepton momentum in the τ-pseudo rest frame, $p^\star_\tau$. However, we also identify a considerable smearing effect, transforming the expected mono-energetic $\tau \to \ell\alpha$-decay distribution into a broad distribution. We attribute the smearing mainly to the missing directional information of the τ due to the ν_τ. Furthermore, we observe that the smearing effect affects the lower mass hypotheses more than the higher mass hypotheses. We can understand this as an impact due to the magnitude of the lepton momentum. The greater the lepton momentum, the more is $p^\star_\tau$ impacted by the directional approximation of the τ rest frame.

Since the ARGUS method shows some drawbacks, we aim to improve it in the following. The most crucial factor is the directional uncertainty due to the missing ν_τ, and this thesis found some ideas to mitigate the effect.

4.2 The Thrust Method

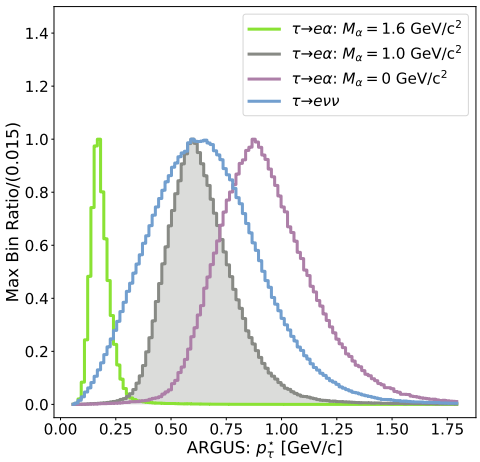
The idea of the Thrust method is to improve the flight direction's approximation by considering the decay products on the signal side.

The Thrust method uses the reconstructed thrust vector to approximate the τ-flight direction. Equation (3.6) gives the thrust vector, and it incorporates the momentum information of the event. For the Thrust method, we assume

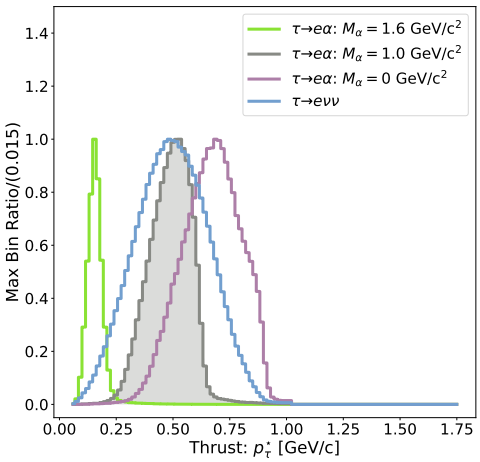
$$\widehat{p_\tau} \approx \widehat{n}_{trust} \,, \tag{4.3}$$

and derive the magnitude of the τ-momentum and the E_τ from $\sqrt{s}/2$.

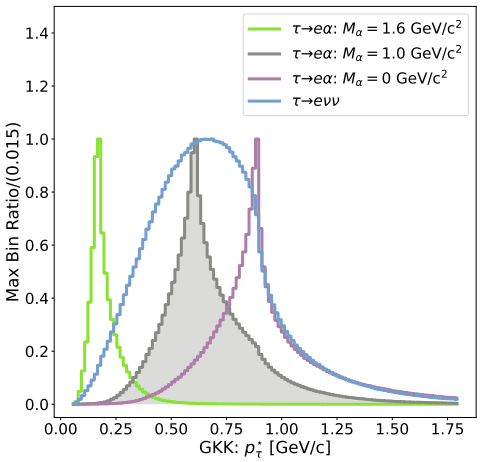
Figure 4.1b shows the emerging shape of the lepton momentum distribution of the $\tau \to \ell\alpha$ channel in the pseudo-rest frame for three mass



(a) ARGUS



(b) Thrust



(c) GKK

Figure 4.1: Pseudo-rest frame distributions for the (a) ARGUS , (b) Thrust , and (c) GKK method displaying the $\tau \to \ell\alpha$ ($\tau \to e\alpha : M_\alpha = \ldots [\mathrm{GeV}/c^2]$) and $\tau \to \ell\nu_\tau\nu_\ell$ ($\tau \to e\nu\nu$) channel. The distributions are normalized to the respective maximum value allowing for a shape comparison.

hypotheses. As expected, we observe a reduction of the smearing effect; however, we also observe a systematic shift of the distributions towards lower lepton momenta and that the Standard Model and $\tau \to \ell\alpha$-decay are affected similarly. The following section tries to tackle both problems simultaneously and reconstruct a pseudo-rest frame with improved smearing and no systematic shift.

4.3 The Generalised Known Kinematics (GKK) Method

The smearing of the ARGUS method is an excellent motivation to improve on the τ-pseudo-rest frame estimate. The first attempt to improve the flight direction's estimate by incorporating the signal side information with the Thrust method in Section 4.2 yields an improvement smearing-wise but also leads to a shift of all distributions towards lower values of the lepton momentum and shows no noticeable improvement of the shape difference between the $\tau \to \ell\alpha$ and the $\tau \to \ell\nu_\tau\nu_\ell$ $p_\tau^\star$ distribution. One issue with both methods is that they neglect the ν_τ completely. The following section includes the available information of the ν_τ in the τ-rest-frame estimate. Several advantages emerge from this ansatz.

First, the new method does not rely on the 3-prong hadron system of the tag-τ as the ARGUS method and, to a lesser extent, the Thrust method. We can apply the approach also to 1-prong hadron systems. This addition would significantly increase the available data in searches for $\tau \to \ell\alpha$ because hadronic 1-prong events are more abundant than 3-prong ones.

Second, by including the ν_τ information, the method has a general use case. Including the ν_τ information yields a full sufficient statistic, which by definition allows for measuring any decay property of the particle. We expect that other measurements benefit from a τ-rest-frame estimate without missing information.

This section closely resembles the article [89]. Since we expect a wide field of possible use cases, the description is generalized to underline the applicability to various scenarios, including other particle pair events.

4.3.1 Concept

For measuring any quantity, the optimal observable is a uniformly minimum-variance unbiased estimator of a statistic. The idea of GKK, as developed in the following, is inspired by the concept of such an estimator.

[1] The linked Exact Definition's summary is: An estimator with an expectation value equal to the actual random distribution's parameter of interests value and minimal variance.

A statistic T – used to determine the estimator of interest – must contain all available information to make this concept work [80]. However, particle pair events with invisible particles lose kinematic information. Due to physical constraints, we can recover some of the lost information and determine a statistic T' which is complete concerning the kinematic information.

Let us consider a sample of τ-pair events, with two hemispheres τ_1 and τ_2. We want to measure an observable – the τ-rest frame in this case – dependent on the momentum-spectrum of the visible daughters in the rest frame of τ_2.

We can identify particle-pair events by reconstructing one of the two particles undergoing a well-known decay. In this case, we demand the τ_1 hemisphere be a type τ_h decay, a hadronic τ-decay, which we call tag-side. Reconstruction of the tag-side allows studying the properties of the second particle τ_2, the so-called signal-side, without introducing a reconstruction bias.

We simplify the τ_1 decay into a two body decay, with the invisible particle I_1 and the visible-daughters-system V^1_{eff} by combining all $n_1 V_1$ daughters into an effective particle $V^1_{\text{eff}} = \bigoplus_{i=1}^{n_1} V_1^i$, with it's four-momentum $p^\mu_{V^1_{\text{eff}}}$ given by the set of four-momentum vectors $[p^\mu_{V_1^1}, ..., p^\mu_{V_1^{n_1}}]$:

$$p^\mu_{V^1_{\text{eff}}} = \sum_{i=1}^{n_1} p^\mu_{V_1^i}. \tag{4.4}$$

The missing information of I_1 translates into a probability distribution function for the momentum vector of the tag-particle τ_1 using energy-momentum conservation and the isotropic distribution of the decay of I_1 in the rest frame of τ_1. In the center-of-mass system, I_1 is in a cone-shaped momentum distribution around the τ_1 momentum. Now, we can turn the argument around and constrain the τ_1 momentum direction on a cone around the V^1_{eff} direction, as shown in Figure 4.2.

As a first step, the τ_1-energy, E_{τ_1}, is determined. In $\tau^+\tau^-$-events, it can be approximated as half of the centre-of-mass beam-energy $\sqrt{s}$:

$$E_{\tau_1} \approx \frac{\sqrt{s}}{2}. \tag{4.5}$$

This approximation also allows determining the magnitude of the τ_1-momentum in the centre-of-mass system via the known τ-mass, m_τ. Doing so, we can derive the angle θ between the V^1_{eff}-momentum $\vec{p}_{V^1_{eff}}$ and the τ_1-momentum $\vec{p}_{\tau_1}$ by using the law of cosines[2]:

$$cos(\theta) = \frac{\vec{p}_{\tau_1} \cdot \vec{p}_{V^1_{eff}}}{|\vec{p}_{\tau_1}||\vec{p}_{V^1_{eff}}|}. \tag{4.6}$$

[2] θ is given by $\cos(\theta)$ because in the polar coordinates it is confined in $[0, 1]$.

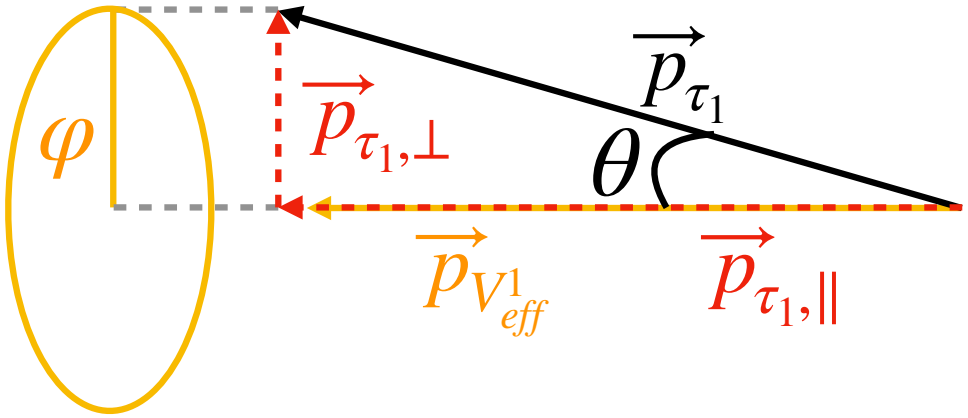


Figure 4.2: Visualization of the GKK principle. The true τ_1 momentum-vector $\vec{p}_{\tau_1}$ is shown in black. θ indicates the angle between the $\vec{p}_{\tau_1}$ and $\vec{p}_{V^1_{eff}}$, the resulting momentum vector sum for all visible daughters V^1 is yellow. The invisible particle I_1 involved in the decay leads to an angular offset, which we parameterise by θ and φ. The red dashed lines indicate the components of the τ-momentum. $\vec{p}_{\tau,\parallel}$ is parallel to $\vec{p}_{V^1_{eff}}$, $\vec{p}_{\tau,\perp}$ is orthogonal. The unknown component is then translated into the direction of $\vec{p}_{\tau,\perp}$, parameterised by the azimuth angle φ.

As Figure 4.2 shows, the three-momentum of τ_1 can be deconstructed into a parallel, $\vec{p}_{\tau_1,\parallel}$, and perpendicular, $\vec{p}_{\tau_1,\perp}$, component with respect to the V^1_{eff} momentum direction. Constraining τ_1 on a cone around V^1_{eff} allows us to parameterize $\vec{p}_{\tau_1}$ such that the unknown direction is expressed in terms of the azimuth angle φ in a cylindrical coordinate system parallel to V^1_{eff}.

We know that each possible $\vec{p}_{\tau_1}$ on the cone is equally probable, so we can sample the $\vec{p}_{\tau_1}$-distribution by stepping through the equally distributed φ, providing a set of equally probable candidates for the τ_1 momentum. This numeric probability distribution function determination is the concept of GKK.

The finite detector resolution can cause this approach to yield $\vec{p}_{\tau_1}$-momentum candidates which deviate considerably from the actual $\vec{p}_{\tau_1}$ of the event. This deviation makes it worthwhile to add further physical constraints. We can utilize the signal side for these constraints in the case of particle pair events.

First, we consider a special case and generalize afterward. In the case of $e^+e^- \to \tau^+\tau^-$ events, we can have events in which both τ_1 and τ_2 decay hadronically. In this case, we can reconstruct the τ_2 momentum in the same way as done for τ_1 with the corresponding angles θ' and φ'. This approach results in two momentum cones, as depicted in Figure 4.3. If the momenta of V^1_{eff} and V^2_{eff} were perfectly known, we could reconstruct the τ-momentum by inverting the momentum on one of the two sides and looking for the momentum-vectors that fulfill the

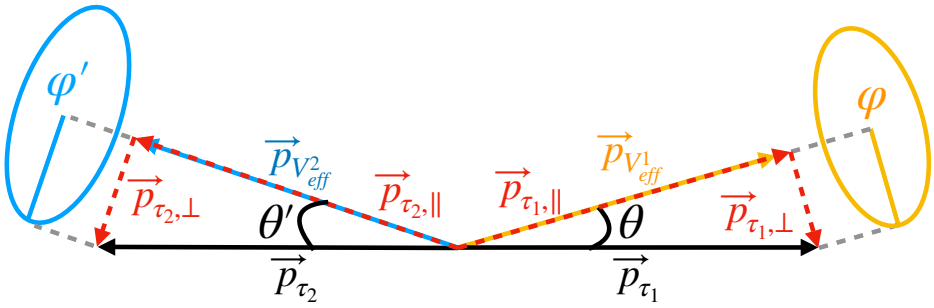


Figure 4.3: Visualization of the GKK method in the special case, in which both τ are of type τ_h. Here, we depict a $\tau^+\tau^-$ event in the centre-of-mass system (CMS). The true τ-momentum-vectors are shown in black. The tau-pair events are divided into a τ_1- and a τ_2-side. θ and θ' indicate the angle between the true τ momentum vector and the sum of all reconstructed daughters on the tag and signal side, depicted in yellow and blue respectively. The τ_i-momentum can be deconstructed into two components indicated in red. A parallel component ($\vec{p}_{\tau_i,\parallel}$) to the daughter momentum ($\vec{p}_{V^i_{eff},\parallel}$), and a perpendicular ($\vec{p}_{\tau_i,\perp}$) one. The unknown direction of $\vec{p}_{\tau_i}$ is translated into the direction of $\vec{p}_{\tau_i,\perp}$. The direction of $\vec{p}_{\tau_i,\perp}$ can be parametrized by the azimuth angles φ and φ', for τ_1 and τ_2, respectively.

requirements imposed by energy and momentum conservation. Both τ momenta must be back-to-back in the center-of-mass frame and lie on their respective cones. In general, this leads to two solutions. In extreme cases, we obtain either one solution (the cones touch each other) or infinite (the cones are on top of each other).

Finite detector resolution smears the reconstructed values. This smearing implies that we do not have a perfect knowledge of $\vec{p}_{V^i_{eff}}$, with $i \in \{1,2\}$. The lost information and detector smearing cause an overlap, leading to an overestimation of the angles θ and θ' and a slight misalignment of the cones, as indicated in Figure 4.4. Instead of looking for a single $\vec{p}_\tau$-candidate which is in both statistics of τ_1 and τ_2, we are thus looking for those $\vec{p}_{\tau_1}$-candidates which are on or within the $\vec{p}_{\tau_2}$-cone, as indicated in Figure 4.4.

This approach can be generalized to the case in which only the τ_1 decay is of type τ_h, and τ_2 has an unspecified number of invisible particles $m_2 I_2$. In this case, we cannot determine a cone of τ_2-momentum-candidates. Instead, we can give a constraint to the momentum-candidate cone of τ_1 by maximizing the θ' to a θ'_{max} and defining a maximized τ_2-momentum cone, as shown on the right side of Figure 4.5. How θ' is maximised is discussed in Section 4.3.2. We constrain the $\vec{p}_{\tau_1}$-cone by demanding that it has to be within or on the cone of $\vec{p}_{\tau_2}$-candidates given by θ'_{max}, indicated on the left of Figure 4.5. This method rejects all $\vec{p}_{\tau_1}$-candidates outside the momentum constraints of the event. We say physical candidates if events pass the momentum constraint; if not, we call them non-physical candidates. As a further refinement step, we

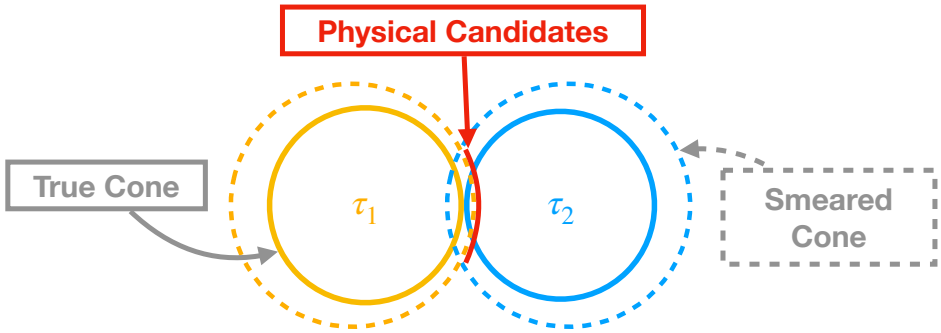


Figure 4.4: Sketch of detector smearing effects on the reconstructed τ-momentum candidates distributed on a cone. Here, we consider the special case of a $\tau^+\tau^-$ event, where both τ_1 and τ_2 are of type τ_h. In an ideal case, indicated by the solid-lined circles denoted as *True Cones*, we expect the τ-cones to touch each other just at one point. The idealized case enables the determination of the true τ-momentum from the common point of the τ-cones. However, the cones can be wider and misaligned due to detector smearing. This case is indicated as *Smeared Cones* with dashed-lined circles. The smeared cones do not touch but overlap, giving rise to a physically sound range, indicated in red.

can restrict the sampling range of φ to the range given by the physical candidates φ_{new} and redo the sampling, giving more weight to physical φ-ranges.

4.3.2 Mathematical Description

After we present the concept of the GKK method, we will discuss the implementation in more detail. We condensed the τ_1-decay's visible and invisible decay partners into an effective two-particle decay problem if τ_1 is of type τ_h. Here, we express the invisible daughter's I_1 four-momentum $p^\mu_{I_1}$ as

$$p^\mu_{I_1} = p^\mu_{\tau_1} - p^\mu_{V^1_{\text{eff}}}. \tag{4.7}$$

We can determine the angle θ, displayed in Figure 4.2, by calculating the mass of I_1, m_{I_1}, with the sum $p^\mu_{I_1} p_{\mu,I_1}$, and the scalar product $p^\mu_{\tau_1} \cdot p^\mu_{V^1_{\text{eff}}}$. First, we derive

$$m^2_{I_1} = m^2_{\tau_1} + m^2_{V^1_{\text{eff}}} - 2(p^\mu_{\tau_1} \cdot p^\mu_{V^1_{\text{eff}}}) \tag{4.8}$$

to replace $p^\mu_{\tau_1} \cdot p^\mu_{V^1_{\text{eff}}}$ in

$$p^\mu_{\tau_1} \cdot p^\mu_{V^1_{\text{eff}}} = E_{\tau_1} E_{V^1_{\text{eff}}} - \cos(\theta)|\vec{p}_{\tau_1}||\vec{p}_{V^1_{\text{eff}}}|, \tag{4.9}$$

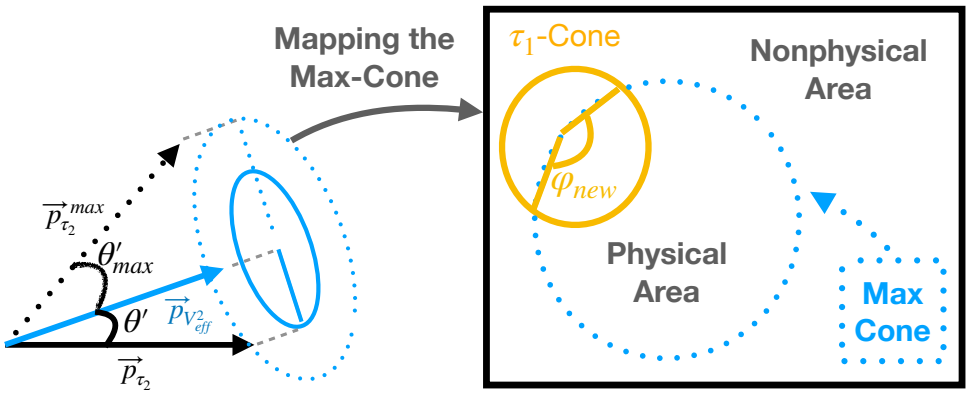


Figure 4.5: On the left, the construction of the maximized cone, using θ_{max} is displayed. We do this in the case of two or more invisible particles in the τ_2-decay because the missing information prevents us from constructing the τ_2-candidate cone. Instead, we can construct a cone with the maximum possible deviation of the τ_2-momentum and the momentum of all visible τ_2-daughters. We map both cones – the hadronic τ_1 and the non-hadronic τ_2 – on a 2D-plain. As defined in the text, the *Max-Cone* gives us the physical area, constraining the τ_1-cone to a certain range φ_{new}. We can sample this physical φ-range again, resulting in a narrower probability distribution of the $\vec{p}_{\tau_1}$.

which we solve for $\cos(\theta)$, resulting in

$$\cos(\theta) = \frac{2\,E_{\tau_1}E_{V^1_{\text{eff}}} + m^2_{I_1} - (m^2_{\tau_1} + m^2_{V^1_{\text{eff}}})}{2\,|\vec{p}_{\tau_1}||\vec{p}_{V^1_{\text{eff}}}|}. \tag{4.10}$$

We can now study how to get θ_{max}. To maximize θ, we have to minimize $\cos(\theta)$. All components of Equation (4.10) are given by the detected event, except for m_{I_1}. So we minimize $\cos(\theta)$ by setting $m_{I_1} = 0$.

The expression derived from hereon is dependent on the reference frame; thus, the following considerations are only valid in the center-of-mass frame, and we consider all appearing quantities to be in the center-of-mass frame. The τ_1 momentum vector $\vec{p}_{\tau_1}$ is decomposed as

$$\vec{p}_{\tau_1} = \vec{p}_{\tau_1,\parallel} + \vec{p}_{\tau_1,\perp}\,. \tag{4.11}$$

With the parallel component of $\vec{p}_{\tau_1}$ to $\vec{p}_{V^1_{\text{eff}}}$

$$\vec{p}_{\tau_1,\parallel} = |\vec{p}_{\tau_1}|\cos(\theta)\,\vec{n}_{\parallel}^{\,V^1_{\text{eff}}}, \tag{4.12}$$

and the orthogonal component of $\vec{p}_{\tau_1}$ to $\vec{p}_{V^1_{\text{eff}}}$

$$\vec{p}_{\tau_1,\perp} = |\vec{p}_{\tau_1}|\sin(\theta)\,\vec{n}_{\perp}^{\,V^1_{\text{eff}}}, \tag{4.13}$$

we get

$$\vec{p}_{\tau_1} = |\vec{p}_{\tau_1}|\cos(\theta)\,\vec{n}_{\parallel}^{\,V^1_{eff}} + |\vec{p}_{\tau_1}|\sin(\theta)\,\vec{n}_{\perp}^{\,V^1_{\text{eff}}}. \tag{4.14}$$

The angle θ is given by Equation (4.10), $\vec{n}_{\parallel}^{V^1_{\text{eff}}}$ is a unit vector in the direction of $\vec{p}_{V^1_{\text{eff}}}$, and $\vec{n}_{\perp}^{V^1_{eff}}$ perpendicular to $\vec{p}_{V^1_{\text{eff}}}$. We estimate the absolute value of the τ_1 momentum $|\vec{p}_{\tau_1}|$ as

$$|\vec{p}_{\tau_1}| = \sqrt{\left(E_{\tau_1}\right)^2 - m^2_{\tau_1}}, \tag{4.15}$$

using Approximation (4.5) and the decay-topology. This leaves only one unknown, the unit vector $\vec{n}_{\perp}^{V^1_{eff}}$. We can define a basis for $\vec{p}_{V^1_{eff}}$, with a parallel basis vector

$$\vec{e}_{\parallel}^{V^1_{\text{eff}}} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{4.16}$$

and the orthogonal basis vector

$$\vec{e}_{\perp}^{V^1_{\text{eff}}} = \begin{pmatrix} cos(\varphi) \\ sin(\varphi) \\ 0 \end{pmatrix}. \tag{4.17}$$

In this basis, the basis vectors and the unit vectors of the combined tag particles are identical. So, by determining the basis transformation from the detector basis to $\vec{e}_{\parallel}^{V^1_{\text{eff}}}$, presented in Equation (4.18), we get an expression of $\vec{n}_{\perp}^{V^1_{\text{eff}}}$, shown in Equation (4.19), which incorporates the ambiguity of $\vec{p}_{\tau_1}$.

$$D_y(\rho) \cdot D_z(\xi) \cdot \vec{n}_{\parallel}^{V^1_{\text{eff}}} = \vec{e}_{\parallel}^{V^1_{\text{eff}}} \tag{4.18}$$

$$\vec{n}_{\perp}^{V^1_{\text{eff}}} = D_z(\rho)^{\mathrm{T}} \cdot D_y(\xi)^{\mathrm{T}} \cdot \vec{e}_{\perp}^{V^1_{\text{eff}}} \tag{4.19}$$

Here, $D_y(\rho)$ and $D_z(\xi)$ are the rotation matrices around the y- and z-axis of the detector respectively. They are

$$D_y(\rho) = \begin{pmatrix} \cos(\rho) & 0 & -\sin(\rho) \\ 0 & 1 & 0 \\ \sin(\rho) & 0 & \cos(\rho) \end{pmatrix} \tag{4.20}$$

and

$$D_z(\rho) = \begin{pmatrix} \cos(\xi) & \sin(\xi) & 0 \\ -\sin(\xi) & \cos(\xi) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{4.21}$$

The angles ρ and ξ are the two polar angles of the detector which rotate $\vec{n}_{\parallel}^{V^1_{eff}}$ into the basis $\vec{e}_{\parallel}^{V^1_{\text{eff}}}$. By combing the above results, we obtain the expression

$$\vec{n}_{\perp}^{V^1_{\text{eff}}} = \begin{pmatrix} \cos(\xi)\cos(\rho)\cos(\varphi) - \sin(\xi)\sin(\varphi) \\ \sin(\xi)\cos(\rho)\cos(\varphi) - \cos(\xi)\sin(\varphi) \\ -\sin(\rho)\cos(\varphi) \end{pmatrix}. \tag{4.22}$$

With the general expression of $\vec{p}_{\tau_1}$ and the knowledge of the distribution function $f(\varphi) = \text{const.}$ we can sample $\vec{p}_{\tau_1}$. We obtain a statistic $[{}^1p^{\mu}_{\tau_1}, ..., {}^np^{\mu}_{\tau_1}]$ of size N for each event, with the number of samples $\vec{p}_{\tau_1}$, N. By sampling the tag momentum with $f(\varphi)$ we obtain a statistic independent of φ, so the statistic is only dependent on the momentum of V^1_{eff}.

As discussed in our example, we can use the calculated τ_1-momentum statistic to determine the τ_2-daughters' momentum statistic in the rest frame of their mother particle. The resulting distribution function of the rest frame momentum $p^{\star}(\varphi)$ is non-invertible as we have no knowledge of a closed-form for this problem. In principle, we expect that the limiting distribution has an analytic description, which could be part of future studies of the GKK method. Figure 4.1c shows the lepton momentum distribution of the $\tau \to \ell\alpha$ channel in the pseudo-rest frame for three mass hypotheses.

4.4 Method Comparison

We discussed the three different τ-rest-frame methods and will compare the resulting distributions. We compare the resulting τ-daughter lepton momenta distributions for 10000 events in the $\tau \to e\,\alpha$ channel for a zero mass hypothesis in Figure 4.6. We observe that the ARGUS-

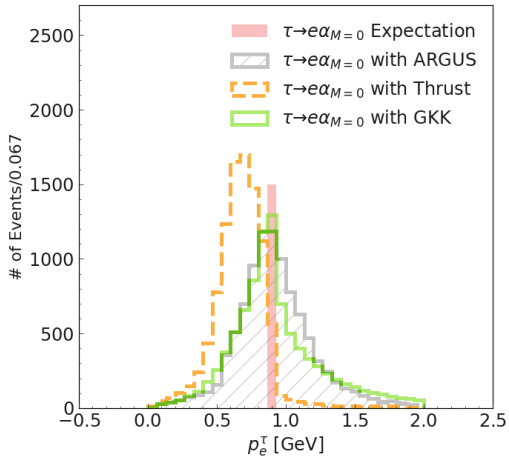


Figure 4.6: The plot shows the reconstruction level distributions of the ARGUS, Thrust, and GKK method. The $\tau \to \ell\alpha$ channel for the e has a zero mass hypothesis for the α.

and GKK-method's first mode reproduces the expectation value of the $\tau \to \ell\alpha$ channel. The first mode of the Thrust method introduces a systematic shift towards lower lepton momenta in the reconstructed τ-reft frame. The distributions vary in width. The ARGUS method has the widest distribution and lowest peak, the Thrust method has a smaller width and the highest peak, and the GKK method has a higher peak than the ARGUS method and the smallest width of the three distributions.

For the $\tau \to \ell\alpha$ analysis, we expect the method with the most significant shape difference of the $\tau \to \ell\alpha$ and $\tau \to \ell\nu_\tau\nu_\ell$ distributions to be the most sensitive method for the search. We can compare the $\tau \to \ell\alpha$ and $\tau \to \ell\nu_\tau\nu_\ell$ distributions for the three methods in Figure 4.1. By eye, the method with the most significant shape difference is the GKK method. In chapter 6, we will compare the performance of all three methods in the $\tau \to \ell\alpha$ analysis. The preliminary conclusion we can draw is as follows.

The ARGUS method provides an easy way for a τ rest frame estimate. It conserves essential properties such as the first mode of the distributions, but severe smearing leads to the broadest distribution of the three methods affecting parameter estimation. The Thrust method uses more information about the event and is less smeared than the ARGUS method. The resulting distributions are sharper but shifted towards lower lepton momenta in the pseudo-rest frame. This shift might not concern the template-based search, but it is problematic to use this distribution to measure the properties of the momentum distribution in a later iteration. The GKK method can retain features such as the first mode and is less affected by smearing effects than the ARGUS method, and it seems to have the most significant shape difference. One disadvantage of the GKK method is that it is computationally intensive, expanding the data by the factor of samples (e.g., $N = 100$ or $N = 1000$).

Chapter 5

Model Fitting and Upper Limit Setting

When we want to measure a property, we typically have direct access to the parameter of interest for this property in the measurement. For example, when we measure the length of a table with a scale, we directly measure the length. In high-energy physics, we often face the challenge that we do not have this direct access to the parameter of interest. For example, in our $\tau \to \ell\alpha$ search, we are interested in the number of $\tau \to \ell\alpha$ decays, but we cannot select a pure sample that would allow us to count the $\tau \to \ell\alpha$ decays directly.

An alternative way of extracting the parameter of interest is to build a model that best describes our data. Our variable of interest is $p_\tau^\star$ – the lepton momentum in the pseudo-rest frame of the τ, reconstructed with the ARGUS method as described in Chapter 4. The parameter of interest is part of the model's distribution function. The distribution function describes the resulting data distribution in our measurement.

We extract the parameter of interest by adjusting all model parameters until the model is the best possible description of the observed data distribution. We call this procedure fitting. A common way to evaluate the best possible set of model parameters is by comparing the likelihoods for different model parameter settings and choosing the one with the maximum likelihood. We determine the likelihood for each model parameter setting as follows. Each model parameter setting provides us with a probability density function. We can determine a probability for every observed point in $p_\tau^\star$. The product of probabilities for all observed points of data in $p_\tau^\star$ provides us with the likelihood of the model given the data [80].

The most probable parameter setting describing our data best will most likely result in the parameter of interest equaling zero. That is why we want to determine a value of the parameter of interest, which tells us the maximal decay strength of the $\tau \to \ell\alpha$ decay with 95% confidence. We call the value the 95% confidence-level upper limit. This upper limit constrains the possible parameter space for models beyond the Standard

Model predicting a lepton flavor violating decay $\tau \to \ell\alpha$.

This chapter describes how this thesis tests the existence of the $\tau \to \ell\alpha$ decay. We first define the parameter of interest we want to measure to achieve this goal. Afterward, we describe methods for evaluating an upper limit set for the $\tau \to \ell\alpha$ decay.

A standard method to search for the $\tau \to \ell\alpha$ decay is counting the number of identified events. We can express the number of $\tau \to \ell\alpha$ decays as a fraction of all τ decays, which we refer to as the branching ratio for $\tau \to \ell\alpha$, $Br(\tau \to \ell\alpha)$. In our case, the parameter of interest is the ratio of branching ratios,

$$R = \frac{Br(\tau \to \ell\alpha)}{Br(\tau \to \ell\nu_\tau\nu_\ell)}, \tag{5.1}$$

with the branching ratio of the Standard Model decay $\tau \to \ell\nu_\tau\nu_\ell$ being $Br(\tau \to \ell\nu_\tau\nu_\ell)$. Using R as the parameter of interest has some advantages. We have seen in Chapter 3 that the $\tau \to \ell\alpha$ and $\tau \to \ell\nu_\tau\nu_\ell$ decays are very similar, often indistinguishable. This property renders the $\tau \to \ell\nu_\tau\nu_\ell$ decay irreducible in the selected data set. Using R, we can profit from this observation because we expect both branching ratios to be affected by systematic uncertainties similarly. The systematic uncertainties cancel out by taking the ratio of both branching ratios. Furthermore, the PDG states values for $Br(\tau \to \ell\nu_\tau\nu_\ell)$ which were not determined specifically as a $\tau \to \ell\nu_\tau\nu_\ell$ decay. Instead, the PDG states the branching ratio for $\tau \to \ell\nu_\tau\nu_\ell$ as the branching ratio for a τ decay into one ℓ, leading to uncertainty about the number of actual $\tau \to \ell\nu_\tau\nu_\ell$ decays. In principle, $Br(\tau \to \ell\nu_\tau\nu_\ell)$ in the PDG could already include the $\tau \to \ell\alpha$ decay. We can account for this by measuring $Br(\tau \to \ell\alpha)$ and $Br(\tau \to \ell\nu_\tau\nu_\ell)$ simultaneously and allowing for a down fluctuation of $Br(\tau \to \ell\nu_\tau\nu_\ell)$. R in this scenario is a natural way to express the fraction of $\tau \to \ell\alpha$ decays.

To measure R, we need a variable distinguishing between $\tau \to \ell\alpha$ and $\tau \to \ell\nu_\tau\nu_\ell$. In Chapter 4 we learned that in the τ-pseudo rest frame $\tau \to \ell\alpha$ and $\tau \to \ell\nu_\tau\nu_\ell$ display different shapes for the lepton momentum, which allows discriminating them. We consider the $p_\tau^\star$ distribution, which is the lepton momentum of the τ daughter in the pseudo-rest frame of the signal τ, to discriminate between the $\tau \to \ell\alpha$ and $\tau \to \ell\nu_\tau\nu_\ell$ decays. We choose the ARGUS method for determining the pseudo rest frame as our default method.

The function $F(p_\tau^\star)$ models the expected distribution in $p_\tau^\star$ for a given amount of data – $62.8\,\mathrm{fb}^{-1}$ in our case. We differentiate between three physics processes to describe the $p_\tau^\star$ distribution in data. They are the $\tau \to \ell\alpha$ decay, the $\tau \to \ell\nu_\tau\nu_\ell$ decay, and all other misidentified physics processes. We extract the expected shape and the number of events from simulations for the Standard Model process. The extracted, binned distribution functions are referred to as templates. We fit these templates with a maximum likelihood fit to extract R. The fit allows the number of events for each of the three physics processes to float, enabling ad-

justments as described above. The resulting log-likelihood distributions also yield the uncertainty for every fitted parameter. In this thesis, we determine R from the best fit values for $Br(\tau \to \ell\alpha)$ and $Br(\tau \to \ell\nu_\tau\nu_\ell)$.

This thesis follows the common high-energy physics discovery rules. The Standard Model expects $R = 0$. If the measurements yield a positive mean value of R deviating more than three standard deviations, we report the mean value of R and claim to have found evidence for new physics within our data. If the measurement yields a deviation of R by five standard deviations, we report the discovery of the $\tau \to \ell\alpha$ decay. If we do not find evidence, we report the mean value and determine a 95% confidence-level upper limit, UL.

We pursue a robust statistical approach for the upper limit setting, applying a Frequentist CL_S and a Bayesian interference approach. We expect that both statistically independent approaches yield the same result for an infinite amount of data and some deviations for a finite amount of data. The advantage of the two independent approaches is that we can identify problems in the statistical method if the two approaches are not converging. Furthermore, Frequentist and Bayesian results yield a robust and complementary interpretation. While a frequentist confidence level identifies the range in which the experiment can identify the signal in 95 of 100 cases, the Bayesian confidence level allows assessing the model's probability.

We study the capability of extracting parameters through a fit and set a 95% confidence-level upper limit of R to validate the statistical methods in a controlled environment. We sample toy data sets for two cases, the null-hypothesis, H_0, with $R = 0$ and the signal hypothesis, H_S, with $R > 0$. The toy data are randomly sampled from $F(p_\tau^\star)$ which provides the expected number of events and the shape of the distribution. If for many toy-data sets, the mean of every parameter corresponds to the value of $F(p_\tau^\star)$, we have validated the capability of the method. Furthermore, we also evaluate the expected 95% confidence-level upper limit through the Frequentist and the Bayesian approach.

Finally, we must also consider systematic uncertainties that emerge from many sources. The most significant systematic uncertainties are the lepton identification and the trigger uncertainties.

This chapter first presents the statistical model in Section 5.1. It introduces the function $F(p_\tau^\star)$ which describes the expected distribution of data for $62.8\,\text{fb}^{-1}$. In sections 5.2 and 5.3 we consider the implementation of the Frequentist and Bayesian approach, respectively. They explain the CL_S and Bayesian interference methods. Section 5.4 presents the fit performance evaluation and displays the expected 95% confidence-level upper limit in case of $R = 0$. We will verify that fitting the model yields the expected parameter values, and the confidence-level upper limit estimation of both the Frequentist and the Bayesian method yield similar results. At last, Section 5.5 discusses systematic uncertainties, focusing on the relevant systematic uncertainties coming

from leptonID and trigger.

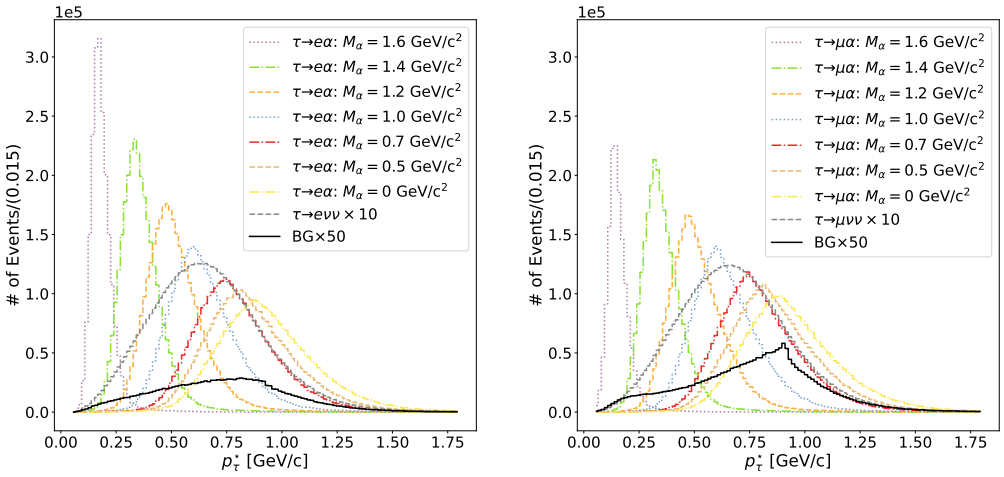
5.1 Statistical Model

To measure R in data, we model the data by considering all possible physics processes, which Belle II identifies as a 1-prong τ decay containing one ℓ. We split the expected physics processes into three categories (1.) the $\tau \to \ell\alpha$ process, (2.) the $\tau \to \ell\nu_\tau\nu_\ell$ process, and (3.) all misidentified events we refer to as background, BG. These physics processes have distributions in $p_\tau^\star$ described by three probability density functions, pdf:

1. $f_\alpha(p_\tau^\star)$: $\tau \to \ell\alpha$;
2. $f_{\ell\nu\nu}(p_\tau^\star)$: $\tau \to \ell\nu_\tau\nu_\ell$; and
3. $f_{\mathrm{BG}}(p_\tau^\star)$: $\tau^\pm_{\mathrm{BG}}$, $q\bar{q}$, $B\bar{B}$, $B^0\bar{B}^0$, $\ell^+\ell^-\gamma$, $e^+e^-\ell^+\ell^-$, e^+e^-hh

Here, $f_i(p_\tau^\star)$ denotes the corresponding probability density function in $p_\tau^\star$ for the above-defined categories.

We model the measured data through the resulting histogram distributions of $p_\tau^\star$ from simulated data. We call these templates. Figure 5.1 displays the resulting distribution functions $f_i(p_\tau^\star)$ for the electron (a) and muon (b) channel. We display the $\tau \to \ell\alpha$ distributions with their



(a) $\tau \to e\alpha$ (b) $\tau \to \mu\alpha$

Figure 5.1: Templates of the distribution functions $F_i(p_\tau^\star)$ for the electron (a) and muon (b) channel in the pseudo-rest-frame given by the ARGUS method. We display all considered mass hypotheses for the $\tau \to e\alpha$ and $\tau \to \mu\alpha$ distributions with their original size of 10^7 events. We indicate the respective $\tau \to \ell\alpha$ mass hypothesis for the α mass with M_α. The Standard Model distributions ($\tau \to e\nu\nu$, $\tau \to \mu\nu\nu$, and BG) are normalized to 62.8 fb^{-1}. We scale them up by a factor of 10 for the $\tau \to \ell\nu_\tau\nu_\ell$ distributions ($\tau \to e\nu\nu \times 10$, $\tau \to \mu\nu\nu \times 10$) and a factor of 50 for the background distribution (BG×50) to compare directly to the originial sized $\tau \to \ell\alpha$ distributions.

original size of 10^7 events. The Standard Model $\tau \to \ell \nu_\tau \nu_\ell$ distribution template, normalized to $62.8\,\mathrm{fb}^{-1}$, is scaled up to compare the shape to the unscaled $\tau \to \ell \alpha$ distributions. We observe a transitioning of the signal distributions towards lower $p_\tau^\star$ for higher α mass hypothesis. We also observe that the background template distribution (BG) covers most parameter space. Its peak is close to $p_\tau^\star = 1.0\,\mathrm{GeV}/c$. The electron channel's distribution is without notable shape features, which could lead to difficulties differentiating between the background distribution and a $\tau \to \ell \alpha$ signal. The muon channel has a distinctive peak and left tail shape. We attribute the differences in the background distributions to the different selections for the respective $\tau \to \ell \alpha$ channel. For example, the muon channel has a relaxed *thrust* requirement. The $\tau \to \ell \nu_\tau \nu_\ell$ distributions coincide with the mass hypothesis of α for $1.0\,\mathrm{GeV}/c^2$. We expect a reduced sensitivity for this mass hypothesis.

For the $\tau \to \ell \alpha$ and $\tau \to \ell \nu_\tau \nu_\ell$ decays with the 1-prong τ decay i and the 3-prong (tag) τ decay j the number of observed events is

$$N_i = 2 \int \mathbf{L} dt \cdot \sigma(e^+ e^- \to \tau^+ \tau^-) Br(\tau_i \to \ell \,\text{invisible}) Br(\tau_j \to hhh) \epsilon_{ij}. \quad (5.2)$$

We encode efficiencies and detector acceptances with ϵ_{ij}, and collectively refer to the α or the ν_τ and ν_ℓ as invisible.

For the variable $p_\tau^\star$, we model the data as

$$F(p_\tau^\star) = N_\alpha \times f_\alpha(p_\tau^\star) + N_{\ell\nu\nu} \times f_{\ell\nu\nu}(p_\tau^\star) + N_{BG} \times f_{BG}(p_\tau^\star), \quad (5.3)$$

where N_i is the number of events corresponding to the i-th sample, and it holds that

$$\int F(p_\tau^\star) dp_\tau^\star = N_T = N_\alpha + N_{\ell\nu\nu} + N_{BG}.$$

Here N_T is the total number of events in data.

Given the signal channel $\tau \to \ell \alpha$, and the normalization channel $\tau \to \ell \nu_\tau \nu_\ell$, we get from Equation (5.2):

$$\frac{N_\alpha}{N_{\ell\nu\nu}} = \frac{Br(\tau \to \ell \alpha)}{Br(\tau \to \ell \nu_\tau \nu_\ell)} \frac{\epsilon_\alpha}{\epsilon_{\ell\nu\nu}}. \quad (5.4)$$

By solving for the branching fraction ratio, we define the relative branching ratio as

$$R \equiv \frac{Br(\tau \to \ell \alpha)}{Br(\tau \to \ell \nu_\tau \nu_\ell)} = \frac{\epsilon_{\ell\nu\nu}}{\epsilon_\alpha} \frac{N_\alpha}{N_{\ell\nu\nu}}, \quad (5.5)$$

which is the parameter of interest.

At this point, we identify two options to measure R. We could either measure R directly from data as a fitting variable or measure the yields of the signal and normalization channels and compute R afterward. We explore both options, but this thesis will focus on the second option because it directly yields $Br(\tau \to \ell \alpha)$ if we find a signal and is unbiased in determining the relative efficiency ratio from simulation data.

5.2 Frequentist Approach Using pyhf

We use the HistFactory software package to construct our fit model described in the last section. HistFactory provides a framework for template-based fitting [90]. We use two implementations, the python based pyhf [91, 92] and the C++ (Root) based RooStats [93, 90]. This thesis uses the pyhf implementation, which provides the fit model, a fitting tool called MINUIT [94, 95], and a Frequentist-leaning approach for determining the upper limit of R. For the Frequentist approach, we use the Frequentist leaning asymptotic CL_S-technique. We introduce it in the following, relying on [96, 97].

The CL_S method tries to solve the problem of setting a signal decay's upper limit with a certain confidence level when a considerable background is present. In our case, we set a 95% confidence level for the signal decay. The idea of the CL_S method is to estimate the signal confidence level by normalizing the signal + background, S+B, with the background, B:

$$CL_S = \frac{CL_{S+B}}{CL_B}, \tag{5.6}$$

where CL_{S+B} and CL_B are the p-values [97] of the data for the signal + background hypothesis, H_{S+B}, and the background hypothesis, H_B, respectively. We evaluate the values at fixed points for our parameter of interest. For the evaluation, we use a test statistic $q(\mu)$ according to [97], which is dependent on the value of the parameter of interest, μ. For a fixed parameter of interest, μ', we determine the test statistic distributions for H_{S+B} and H_B: $q(\mu')_{S+B}$ and $q(\mu')_B$. There are two methods to determine the respective test statistic distributions. The first method is to generate many toy-data sets for the respective hypothesis and values of μ' and calculate the values of q. We refer to q as toy-based when calculated with this method. The second is an approximation based on Wilk's theorem, which assumes that we can describe the distribution of q with a χ^2-distribution. This method determines the specific χ^2-distribution for the specific $q(\mu')$ of H_{S+B} or H_B and uses this analytic description for $q(\mu')_{S+B}$ or $q(\mu')_B$.

With the test statistic value of the data, $q(\mu')_{Data}$, we calculate the p-value for H_{S+B} and H_B given by $q(\mu')_{S+B}$ and $q(\mu')_B$, respectively. The p-values evaluated at many values of μ yield the CL_{S+B} and CL_B distributions. We exclude the signal decay strength at the confidence level, CL, when

$$1 - CL_S \leq CL. \tag{5.7}$$

In our $\tau \to \ell\alpha$ search we set a 95% confidence level with $CL_S \leq 0.05$.

In the case of a likelihood function without nuisance parameters, the CL_S method reproduces the well-accepted Frequentist Feldman Cousins [98] interval results [97], which showcases the CL_S method's validity. Furthermore, by definition, the CL_S method shows the physically desirable property that the probability of our parameter of interest equals

100% for $R = 0$ [99]. Although widely used within the high-energy physics community, we should be aware of some limitations the CL_S method has.

The method does not follow a precise theoretical framework, sometimes referred to as *ad hoc* [99]. We can understand the CL_S method as an approximate confidence level of the signal-only hypothesis that provides a conservative upper limit [96]. HistFactory uses the CL_S-technique for the upper limit setting. It assumes Gaussian constraints for our model's relevant nuisance parameters. Due to these constraints, HistFactory is not purely Frequentist but rather Frequentist leaning. We do not assume a parameter's probable distribution in a purist Frequentist approach.

As we will learn in this and the next chapter, the Gaussian constraint assumption in HistFactory is justifiable because the nuisance parameters emerge from measurements that show Gaussian uncertainties. However, it also impacts the search, as it demands symmetric uncertainties, which do not necessarily emerge when combining several measurements of a nuisance parameter. Acknowledging these method properties is crucial to enable a meaningful interpretation of our final result.

This thesis uses the pyhf software to construct the likelihood of the fit model from templates, perform the fit, and determine the 95% confidence level upper limit. Based on the likelihood given by the pyhf package, we fit the model to data using the minute optimizer. By default, pyhf is based on the concept of asymptotic formulas for upper limit calculations from "Asymptotic formulae for likelihood-based tests of new physics" [97], which assumes that the test statistics distribution in the hypothesis tests is according to a χ^2 distribution based on Wilk's theorem [97]. If the asymptotic assumption fails, it can also run on toy-based test statistics. The toy-based approach produces many toy data sets by randomly sampling the model, and fitting these toy data allows for the respective hypothesis to produce test statistic distributions $q(\mu')_{S+B}$ or $q(\mu')_B$.

This thesis tested both concepts and validated the asymptotic assumption. Appendix E covers a detailed discussion on the pyhf fitting procedure and complementary studies.

The pyhf software prefers numbers between -10 and 10 for the fit parameters. Hence, this work rewrites the model in Equation (5.3) in terms of constant N_i and a fit factor μ_i:

$$F(p_\tau^\star) = \mu_\alpha N_\alpha \times f_\alpha(p_\tau^\star) + \mu_{\ell\nu\nu} \times N_{\ell\nu\nu} \times f_{\ell\nu\nu}(p_\tau^\star) + \mu_{BG} \times N_{BG} \times f_{BG}(p_\tau^\star). \quad (5.8)$$

With the fixed number of N_i and the factors, μ_i, giving the relation between the input and for each expected branching fraction, $Br(i)$:

$$Br(i) = \mu_i \times Br'(i). \quad (5.9)$$

Here, the branching ratios Br' correspond to the simulation input. We assume that the Standard Model $Br'(\tau \to \ell\nu_\tau\nu_\ell)$ corresponds to the measured branching ratio listed in the PDG [34]. Due to this definition, our

expectation values for the Standard Model samples are $\mu_{\ell\nu\bar{\nu}} = \mu_{BG} = 1$. We allow for fluctuations in the range of [0,2], implying no contribution or double the amount. We chose this range because it ensures that the MINUITE minimization cannot reach a bound known to cause problems for the minimizer.

In this thesis's implementations for the measurement of R, it holds that $\mu_\alpha = Br(\boldsymbol{\tau \to \ell\alpha})$. This equivalence holds because the new physics simulation input in Equation (5.9) is $Br'(\boldsymbol{\tau \to \ell\alpha}) = 1$. We calculate R of Equation (5.5) now as

$$R = \frac{\mu_\alpha}{\mu_{\ell\nu\nu} Br'(\boldsymbol{\tau \to \ell \nu_\tau \nu_\ell})}. \tag{5.10}$$

We determine the fit factors μ_i in the fit simultaneously.

In this thesis, if not stated otherwise, we obtain all examined upper limit results with pyhf based on the asymptotic approach. We can expect the approximation of Wilk's theorem to hold since our fit results, presented in Section 5.4 and Chapter 6, are well-behaved likelihoods following a Gaussian distribution.

5.3 Bayesian Approach

The Bayesian approach uses Bayesian interference to determine the 95% confidence-level upper limit. This method needs the analyst to assume a prior probability distribution, referred to as prior, for the model parameters. The measurement updates the prior and yields a new, updated probability distribution we call posterior. The assumed prior impacts the posterior distributions. Although we expect the posterior for an infinite amount of data to converge to the actual probability distribution, we observe differences for a finite amount of data. We mitigate this problem by considering two uninformative priors. The first one is a flat prior, considered uninformative regarding the fit value, and the second is a log uniform prior, considered uninformative regarding the fit parameter's resulting order of magnitude.

In our $\boldsymbol{\tau \to \ell\alpha}$ search, we use the Bayesian Analysis Toolkit's, BAT, C++ version [100] to verify the CL_S method results. The C++ BAT implementation uses the Multi-Template Fitting functionality of the BAT package [100]. Utilizing the Metropolis-Hastings algorithm [101], our C++ BAT implementation estimates the parameter of interest via marginalization with a Markov Chain Monte Carlo. Additionally, we perform a MINUIT fit to maximize the full posterior likelihood; this fit only provides a means of comparison with other fitters such as pyhf. The C++ BAT implementation verifies all results we obtain in the fit and the upper limit estimation with pyhf.

This thesis implemented a template fit with BAT in Julia [102], similar to the C++ BAT version, for comparison studies with pyhf. We perform the Bayesian-based fit using the same templates as the Frequentist

approach, and the Julia implementation uses the LiteHF software package [103]. The Julia BAT templates and likelihood function are the same as the pyhf implementation ones. LiteHF enables using the same pyhf workspace with Julia BAT. The Julia BAT implementation uses the same priors as the C++ BAT implementation.

The three components for the fit are as described in Section 5.1. For every parameter, we assign a prior representing the knowledge of the respective contribution.

- We assign two different priors to assess the influence of the prior choice. We set the flat prior probability of $Br(\tau \to \ell\alpha)$ in a range of $[0, 1]$. We use these huge bounds to ensure that the prior is as uninformative as possible, consciously ignoring any known upper bounds provided by experimental or physical constraints. We set the log uniform prior with a bound well above any expected value, 5× the upper limit of the ARGUS collaboration, and well below, 10^{-9}.
- We assign a Gaussian prior probability for $Br(\tau \to \ell\nu_\tau\nu_\ell)$; the mean is the model expectation described below. The standard deviation is the model uncertainty added in quadrature with the luminosity uncertainty of 1.5%.
- We assign a Gaussian prior probability for the expected number of background events, N_{BG}. The Julia BAT implementation's mean is one. The standard deviation is the luminosity uncertainty of 1.5% and the normalized uncertainty of the expected number, $\sqrt{N_{BG}}/N_{BG}$. We add both in quadrature.

In the $Br(\tau \to \ell\nu_\tau\nu_\ell)$ prior determination for the Bayesian fitting procedure, we impose a Gaussian constraint on the parameter $N_{\ell\nu\nu}$. This constraint guarantees the stability of the fitting procedure and avoids biases in the fitting parameters due to potentially low sample size. We use Equation (5.2) and the following values of the involved parameters to determine the mean and width of the Gaussian constraint:

- for the integrated luminosity: $\int \mathbf{L}dt = 62.8 \pm 0.942\ fb^{-1}$;
- for the cross section: $\sigma(ee \to \tau\tau) = 0.919 \pm 0.003\ nb^{-1}$;
- for the tag side branching ratio: $Br(\tau \to 3-prong) = 0.152 \pm 0.0006$
- for the efficiency of the electron: $\varepsilon_{e\nu\nu} = 0.1265 \pm 0.00015$;
- or the muon: $\varepsilon_{\mu\nu\nu} = 0.1618 \pm 0.00017$;
- and for the branching ratio of the electron channel: $Br(\tau \to e\nu\nu) = 0.1782 \pm 0.0004$;
- or muon channel: $Br(\tau \to \mu\nu\nu) = 0.1739 \pm 0.0004$.

By using error propagation, and assuming no correlation, the expected number of events and uncertainty in the electron channel, $E[N_{e\nu\bar{\nu}}]$, and

muon channel, $E[N_{\mu\nu\bar{\nu}}]$, are:

$$E[N_{e\nu\bar{\nu}}] = 395,500 \pm 6,328 \tag{5.11}$$

and

$$E[N_{\mu\nu\bar{\nu}}] = 493,659 \pm 7,898. \tag{5.12}$$

5.4 Fit Performance Evaluation and Upper Limit Sensitivity

After introducing the fit concepts, we are now interested in evaluating the framework's performance. We will consider two cases. First, we evaluate the fit performance for the background-only case (null hypothesis, H_0), and afterward, the fit performance when injecting a signal (signal-hypothesis, H_S). Here, we evaluate the capability of extracting the correct signal yield. At last, we determine the expected upper limit's mean and standard deviation from many toy experiments and compare the implemented methods.

Using the templates given by simulation, we produce for each hypothesis a set of 1,000 pseudo-data simulations[1], which are sampled randomly from F($p_\tau^\star$). We use the pyhf generator, which utilizes the templates as pdf-inputs and generates a random distribution of pseudo-data according to the total number of events in $F(p_\tau^\star)$. We refer to this new, statistically independent distribution $F'(p_\tau^\star)$ as toy data. We evaluate the parameter estimation performance with a set of 1,000 toy data.

In the following, we give an example of the fit results for one randomly chosen toy data distribution. Furthermore, we evaluate the mean performance to identify systematic deviations due to the fitting method.

5.4.1 Null Hypothesis Test

In this subsection, we verify that the Frequentist and Bayesian implementations can extract the null hypothesis, $R = 0$, evaluate the fit parameter estimation performance, and assess potential biases in the parameter extraction methods. Furthermore, we evaluate the 95% confidence-level upper limit extracted with both methods.

Null Hypothesis Test with pyhf

First, we evaluate the pyhf package's fit performance for the simulation-based template model to the 1,000 toy-data distributions. This step is crucial since the templates and model are the basis for our BAT implementations.

[1] We observed that a set of 100 pseudo-data simulations would also be sufficient to determine the mean upper limit and use this number in later studies.

The fit returns approximately Gaussian distributed fit parameters centered around the expected values of $\mu_\alpha = 0$, $\mu_{SM} = 1$, and $\mu_{BG} = 1$. Figure 5.2 shows example profile likelihood function distributions for each parameter. We can verify that within expected statistical fluctua-

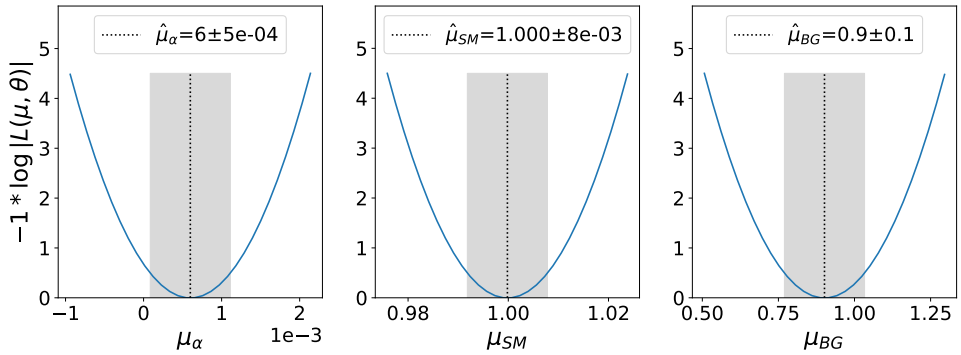


Figure 5.2: Example profile likelihood distributions for all three fit parameters from one of the 1,000 toy-data sets for H_0. The expected fit parameter values are $\mu_\alpha = 0$, $\mu_{SM} = 1$, and $\mu_{BG} = 1$. The dotted line indicates the best fit value, the gray band is the uncertainty, and the blue curve is the likelihood function. The range of μ_α is at the order of $\mathcal{O}(10^{-3})$ indicated as 1e-3.

tions, the extracted fit parameters are consistent with the expected values in the case of H_0. Table 5.1 displays the correlation matrix for this example. We observe a strong correlation between μ_α and μ_{SM} and a

Table 5.1: Example correlation matrix for one fit of the 1,000 toy-data sets for H_0, using the pyhf package.

	μ_α	μ_{SM}	μ_{BG}
μ_α	1	-0.82	0.62
μ_{SM}	-0.82	1	-0.94
μ_{BG}	0.62	-0.94	1

mild correlation between μ_α and μ_{BG}. To some degree, we expect this result due to the similar physical properties of the samples.

Figure 5.3 displays the mean fit value for each sample of μ_α (black dot), determined from the 1,000 toy-data simulations. We determine the displayed uncertainties of μ_α (the error bar) as the mean of the toy-data set for the upper and lower uncertainties, respectively. We compare the mean fit value to the input value of μ_α (red line) in the toy-data generation. We present the mean fit results for the Standard Model distributions in Appendix G in Figure G.10. We observe that the fit parameters are consistent with the initial input for the toy data set. These results show that the fit strategy has no bias. Figure 5.4 is an example of one of the fitted toy-data samples. We observe a consistency between the fitted templates and the toy data in the top plot. We confirm this by considering the ratio of the fitted templates with the toy data, N^{Data}/N^{MC}, in the

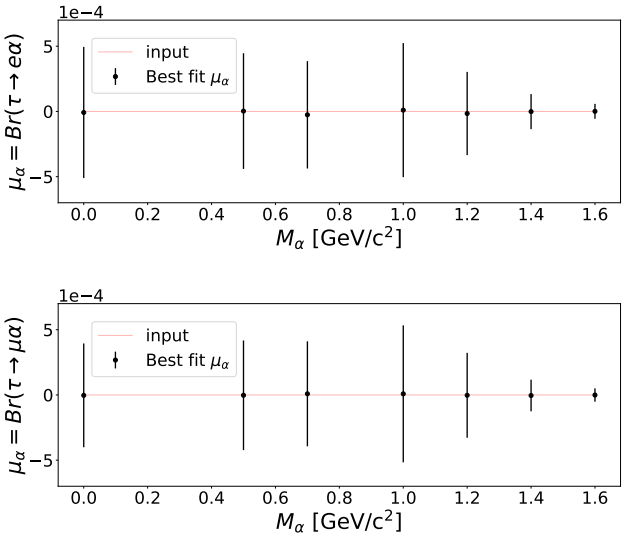


Figure 5.3: Mean (black dots) and mean uncertainty (error bar) of pyhf's μ_α fit-parameter results for 1,000 H_0 toy-data distributions, compared to the toy-data simulation input (red line).

middle plot. Except for some fluctuations on the right tail, we observe excellent consistency between the template and the toy data, verifying the capabilities of the fit. The bottom plot further justifies this conclusion, which shows the toy data subtracted with the fitted Standard Model template distributions. We observe no excess, as expected for H_0.

We confirm that our model can reproduce the background model. Appendix E explains in more detail the fit studies in this thesis. We also demonstrate that the asymptotic upper limit determination based on Wilk's theorem is equivalent to the results obtained with a toy-based approach.

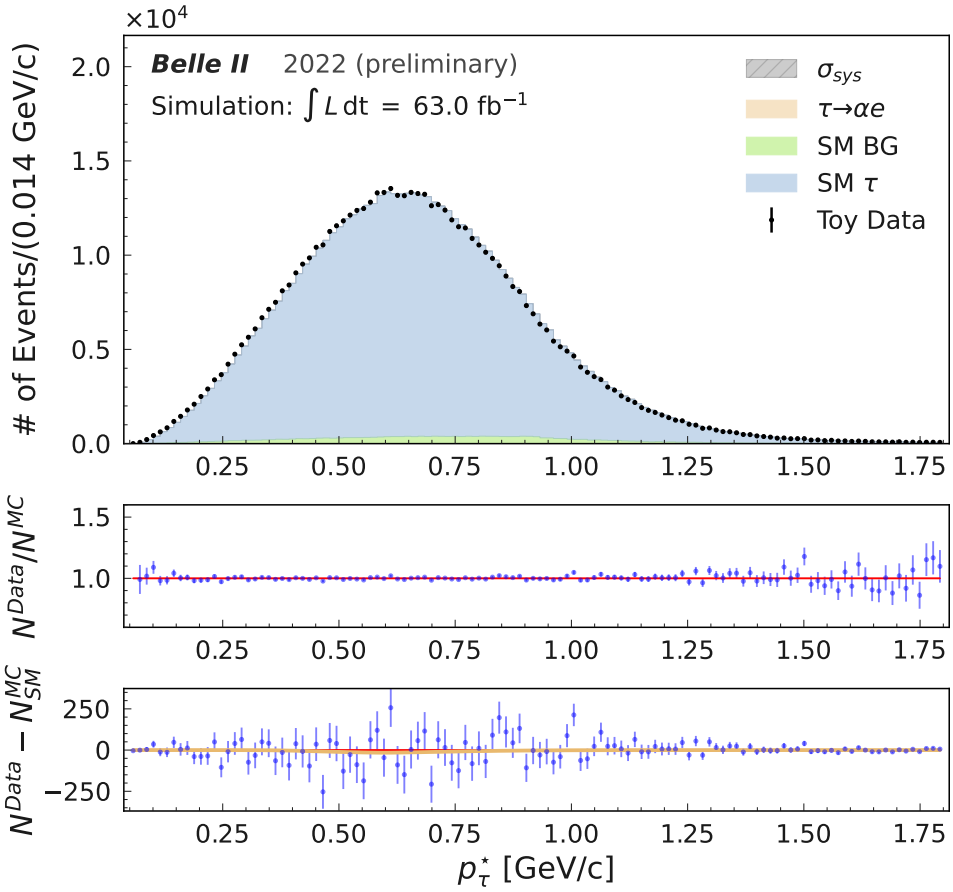


Figure 5.4: Fit to a background only (H_0) toy-data sample using the $m_\alpha = 1.0$ GeV/c^2 mass hypothesis for the new physics. The upper plot of the figure shows the systematic uncertainty (σ_{sys}), the $\tau \to \ell\alpha$ hypothesis ($\tau \to \alpha e$), the fit model's background distribution (SM BG), the distribution for the $\tau \to e\,\nu_\tau\,\nu_e$ (SM τ), and the sampled toy data set (TotData). σ_{sys} is zero in the evaluated case. The two lower plots allow evaluating toy data distribution (blue dots) and uncertainty (blue error bar) with the expectation for H_0 (red line). They show the ratio of toy data to fit model in the central plot, and the toy data subtracted with the SM distributions in the bottom plot. The bottom plot also shows the best fit $\tau \to \ell\alpha$ distributions (orange line), which allows comparing it to the subtracted data.

Null Hypothesis Test with Julia BAT

Next, we evaluate the results of the posterior distributions using the Julia BAT implementation. Figure 5.5 presents an example result for the flat prior case. In the subplots displaying $p(\mu_i)$, the posterior distribution for the fit parameter μ_i, we can observe that all posteriors are approximately Gaussian distributed. The ranges for the 68%, 95%, and 99% confidence levels are green, yellow, and red, respectively.

The 2D distribution of μ_{SM} and μ_{BG} shows a notable deviation from a symmetric two-dimensional Gaussian with an identifiable slope, indicating a mild correlation between μ_{SM} and μ_{BG}. The 2D distribution of μ_α and μ_{SM} is oval and with a slope close to 45°, indicating a strong correlation. Finally, the 2D distribution of μ_α and μ_{BG} deviates from a symmetric two-dimensional Gaussian with a barely noticeable slope, indicating a low correlation. In general, the 2D distributions show a defined spread with well-defined confidence intervals.

Table 5.2 lists the mean and standard deviation of the posteriors. We can observe a good consistency – within two standard deviations – of the

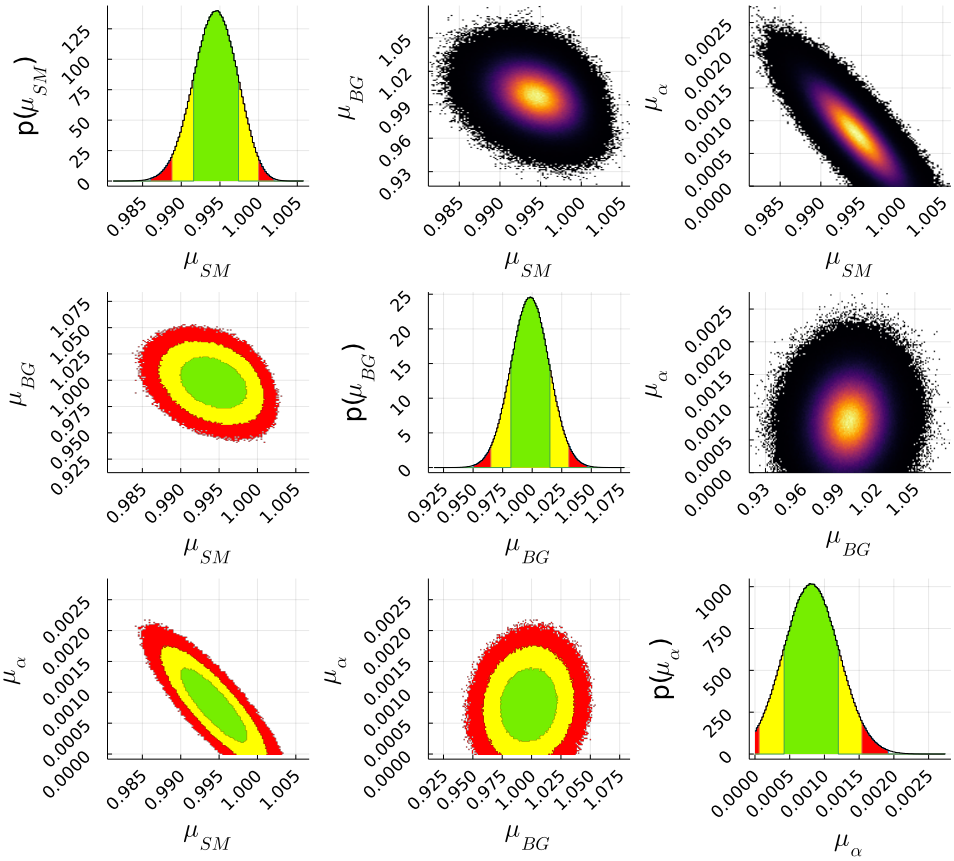


Figure 5.5: BAT Markov Chain Monte Carlo sampling result for the background only toy-data sample using the $m_\alpha = 1.0$ GeV/c^2 mass hypothesis for the new physics. The 68%, 95%, and 99% confidence intervals are green, yellow, and red, respectively.

Table 5.2: Julia BAT mean posterior values with standard deviation in brackets. Example of one of the 1,000 toy data H_0 distributions.

	flat	log uniform
μ_α	$8(4)\times10^{-4}$	$3(3)\times10^{-4}$
μ_{SM}	0.994(0.003)	0.998(0.003)
μ_{BG}	1.00(0.02)	1.00(0.02)

mean parameters with their respective input value of $\mu_\alpha = 0$, $\mu_{SM} = 1$, and $\mu_{BG} = 1$. The two τ to ℓ processes show a mild deviation of two standard deviations. The fit result leads to a deviation of the fitted to the expected $Br(\tau \to e\nu_\tau\nu_e)$. For the flat prior, we get

$$\Delta Br(\tau \to e\nu_\tau\nu_e) = (1 - \mu_{SM}) \cdot 17.82\% = 0.11 \pm 0.05\%,$$

which is consistent within uncertainties with the measured branching fraction of $Br(\tau \to e\alpha) = 0.08 \pm 0.04\%$. We can attribute this to the visible correlation of μ_α and μ_{SM}. Due to the flat prior of μ_α, bounded at zero, the parameter tends to get overestimated. This overestimation causes the underestimation of the τ Standard Model parameter μ_{SM}. Overall, the findings for the linear prior case align with what we observed with MINUITE using the pyhf package.

Figure 5.6 presents an example result for the log uniform prior case. We observe a more complex behavior of the posteriors for this prior. $p(\mu_i)$ is only approximately Gaussian distributed for μ_{BG}. The μ_{SM} parameter displays an enhanced tail for lower values. As expected for the log uniform prior, μ_α is exponentially distributed, with the most probable values close to zero. The general behavior of the 2D distributions is similar to the flat prior case. In addition, the 2D distribution of μ_{SM} and μ_{BG} shows a wider spread for lower values of μ_{SM}. The 2D distribution for μ_α shows a notable deformation towards lower values of μ_α.

Although the posterior distributions show a more complex behavior, we find the log uniform prior case also in line with what we observed with MINUITE using the pyhf package. The mean of the τ samples is closer to what we expect from the simulation input. Nevertheless, the interpretation of the mean is less straightforwards. In case of the μ_{SM} parameter, there is a mild deviation from the Gaussian distribution, leading to a mild deviation of the mean and first mode of the distribution. In the case of μ_α, the posterior distribution is exponential, leading to a substantial deviation of the first mode (close to zero) and the mean. We observe that the mean we have calculated for the posterior corresponds to the expectation value for all fit parameters. Furthermore, we can still identify the impact of the correlation between μ_α and μ_{SM}, as described in the flat-prior case. Although both posterior distributions do not allow for a naive interpretation of the mean and the standard deviation, the main takeaway is that BAT can reproduce the simulation input for the mean and first mode for the presented example.

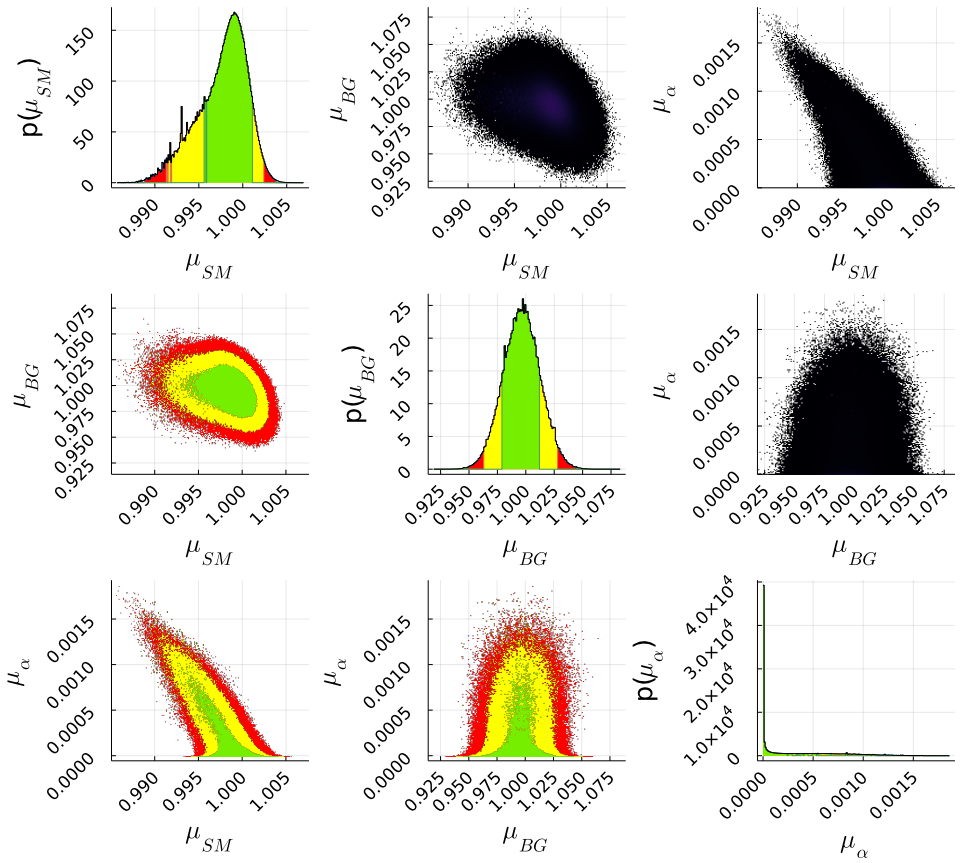
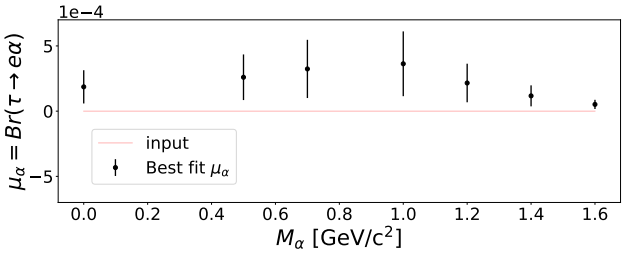
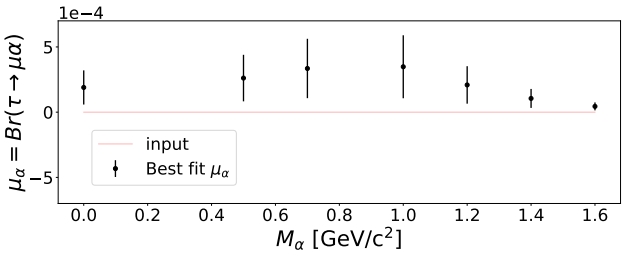


Figure 5.6: BAT Markov Chain Monte Carlo sampling result for the background only toy-data sample using the $m_\alpha = 1.0$ GeV/c^2 mass hypothesis for the new physics and a log uniform prior for μ_α. The 68%, 95%, and 99% confidence intervals are green, yellow, and red, respectively.

Figures 5.7 and 5.8 display the mean value for μ_α (black dot) for the 1,000 toy data simulations. We determine the uncertainties of μ_α (the error bar) as the mean of the toy-data set for the upper and lower uncertainties. The expectation value is the simulation input (red line). Appendix G shows the mean of the sampling result's Standard Model distributions fit result in the figures G.11 and G.12. The results show a bias



(a) $\mu_\alpha = Br(\tau \to e\alpha)$

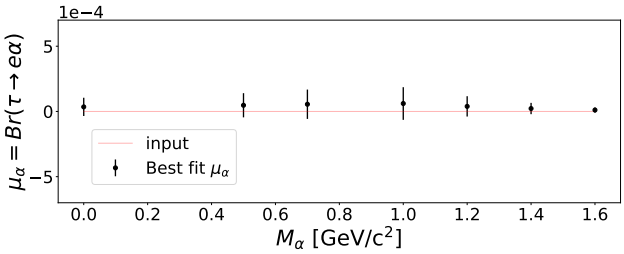


(b) $\mu_\alpha = Br(\tau \to \mu\alpha)$

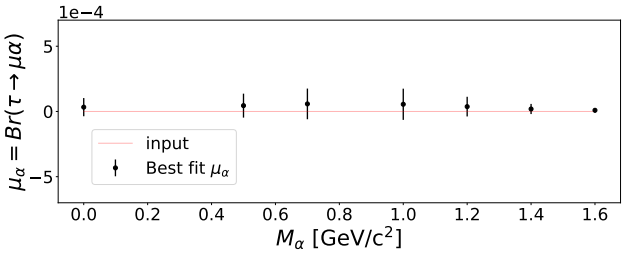
Figure 5.7: Mean (black dots) and mean uncertainty (error bar) of Julia BAT's sampling result's mean and standard deviation with flat prior for 1,000 H_0 toy-data distributions, compared to toy-data simulation input (red line).

of this parameter estimation strategy. We observe an agreement of the sampling result's mean value with the initial input for the toy data set in case of the log uniform prior, although with a mild upward trend. In the case of the flat prior, the situation displays a more pronounced bias towards higher values of μ_α. The results are consistent with the initial toy data set input within two standard deviations.

Overall we observe that the Julia BAT implementation can reproduce the input values of H_0, but μ_α is upwards biased. We expect this result from our choices of positive definite priors. Furthermore, the choice of priors also explains the degree of observed bias. The flat prior accentuates higher values of μ_α and we observe a notable bias. Compared to the flat prior, the log uniform prior accentuates lower values, and we observe a milder bias of the mean μ_α value.



(a) $\mu_\alpha = Br(\tau \to e\alpha)$



(b) $\mu_\alpha = Br(\tau \to \mu\alpha)$

Figure 5.8: Mean (black dots) and mean uncertainty (error bar) of Julia BAT's sampling result's mean and standard deviation with log uniform prior for 1,000 H_0 toy-data distributions, compared to the toy-data simulation input (red line).

5.4.2 Signal Hypothesis Test

The second validation step is to validate that the fit methods can extract an injected signal. In this case, we again sample 1,000 toy data distributions. We use the ARGUS collaboration upper limit to calculate the number of signal events for H_S:

$$Br(\tau \to \ell\alpha) = ARGUS\ UL \times Br(\tau \to \ell\nu\nu), \tag{5.13}$$

where the *ARGUS UL* corresponds to the upper limit result of the ARGUS collaboration for the considered α mass hypothesis, M_α. For the example we choose the electron channel's $M_\alpha = 1.0$ and expect $\mu_\alpha = 0.0064$.

Signal Hypothesis Test pyhf

We verified that the pyhf fit can return the injected H_0 fit parameters with $\mu_\alpha = 0$. In this section, we check if it can recover an injected signal.

The fit returns approximately Gaussian distributed fit parameters centered around the expected values. Figure 5.9 shows an example of the profile likelihood distributions for each parameter. We can verify that within statistical fluctuations, the extracted fit parameters are consistent with the expected values of H_S. As expected, the correlation in this fit is the same as in the H_0 case, comparing the significant numbers.

Figure 5.10 presents an example of the fitted data challenge sample. We observe consistency between the templates and the example toy data set. The simulation distribution, N_{MC}, is consistent with the data distribution, N^{Data}, as the ratio of one in the second plot indicates. Further-

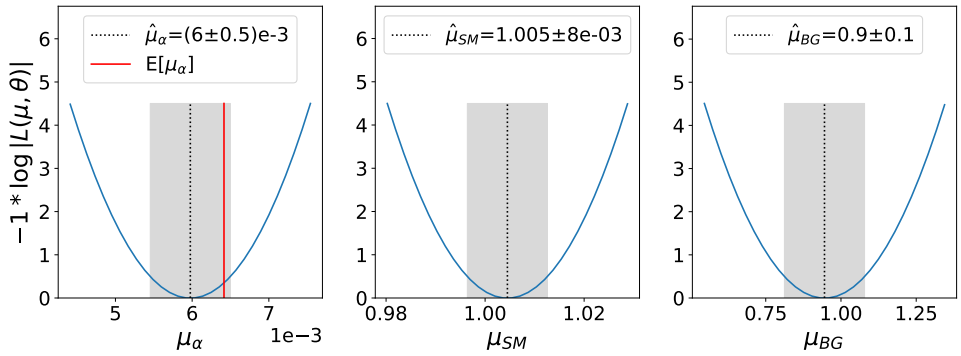


Figure 5.9: Example profile liklihoods distributions for all fit parameters. The dotted line indicates the best fit value, the gray band is the uncertainty, the blue curve is the likelihood function, and the red line indicates the injected signal.

more, the third plot shows that the fit can correctly identify the $\tau \to \ell\alpha$ excess. We observe a clear excess in the $N^{Data} - N^{MC}_{SM}$ distribution, with the Standard model distributions, N^{MC}_{SM}.

We present the resulting mean μ_α (plack dot) with the mean upper and lower uncertainty (error bar) from all of the 1,000 toy data distributions fits in Figure 5.11. We observe a high consistency between the observed μ_α and the input value (red line).

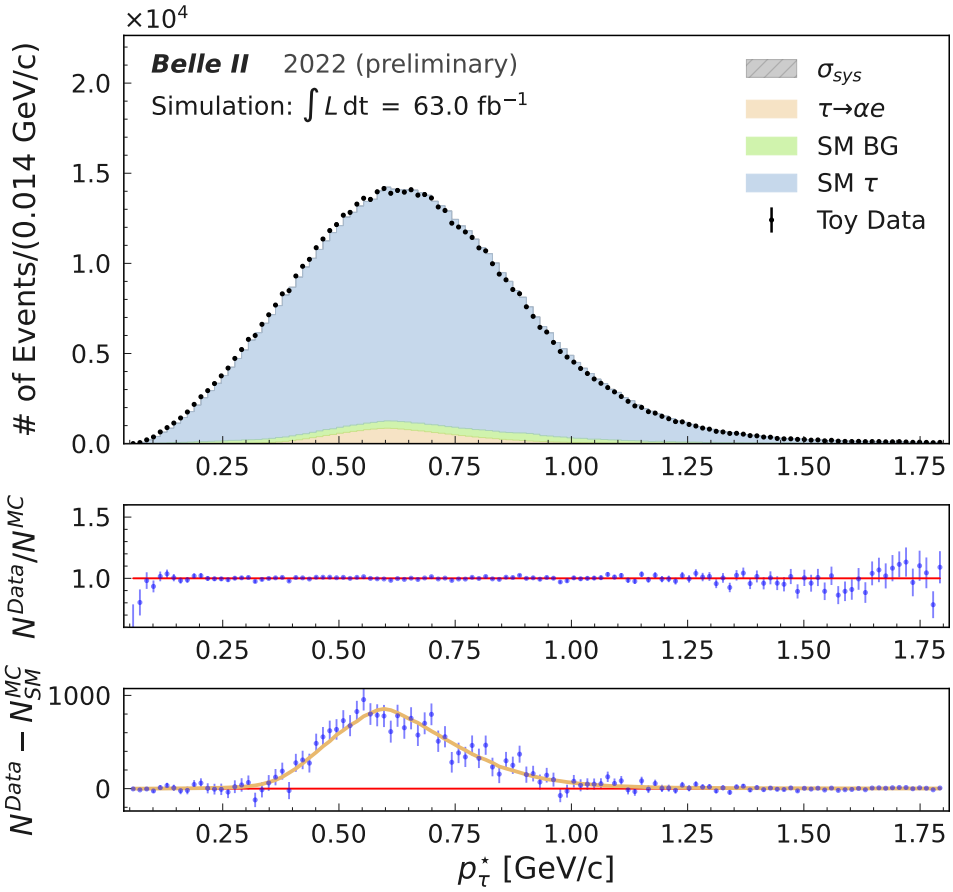
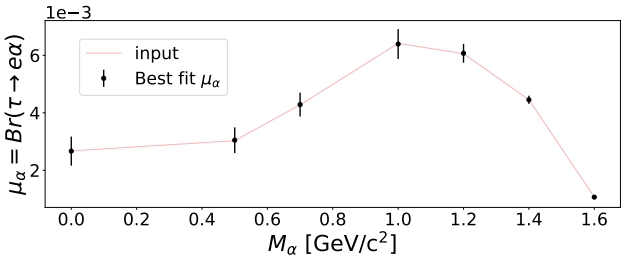
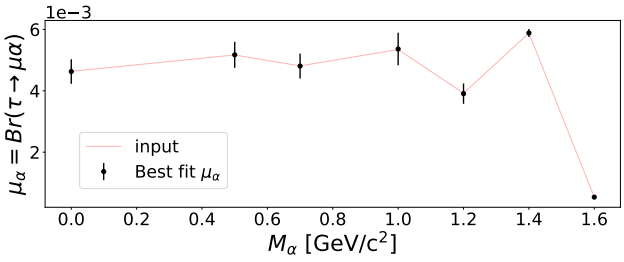


Figure 5.10: Fit to one of the 1,000 toy data distributions for H_S with M_α = 1.0 GeV/c^2 mass-hypothesis for the new physics. The upper plot of the figure shows the systematic uncertainty (σ_{sys}), the $\boldsymbol{\tau \to \ell \alpha}$ hypothesis ($\tau \to \alpha e$), the fit model's background distribution (SM BG), the distribution for the $\boldsymbol{\tau \to e\, \nu_\tau\, \nu_e}$ (SM τ), and the sampled toy data set (TotData). σ_{sys} is zero in the evaluated case. The two lower plots allow evaluating toy data distribution (blue dots) and uncertainty (blue error bar) with the expectation for H_S (red line). They show the ratio of toy data to fit model in the central plot, and the toy data subtracted with the SM distributions in the bottom plot. The bottom plot also shows the best fit $\boldsymbol{\tau \to \ell \alpha}$ distributions (orange line), which allows comparing it to the subtracted data.



(a) $\tau \to e\alpha$



(b) $\tau \to \mu\alpha$

Figure 5.11: Mean pyhf fit-parameter results (black dots) and uncertainty (error bar) for 1,000 toy data, compared to H_S toy-data simulation input (red line).

Signal Hypothesis Test BAT

After evaluating the pyhf performance for the H_S, we now consider the BAT results. For the H_0 case, we observe a mild bias of the estimated parameters in the Julia BAT implementation. This bias is not present in the H_S test. Again, we choose the electron channel's $M_\alpha = 1.0$ for the example plots and expect that $\mu_\alpha = 0.0064$.

By comparing the results of Table 5.3 with the pyhf fit results in Figure 5.9, we observe that both priors can estimate the parameter well. The extracted μ_α mean agrees within uncertainties with the expectation

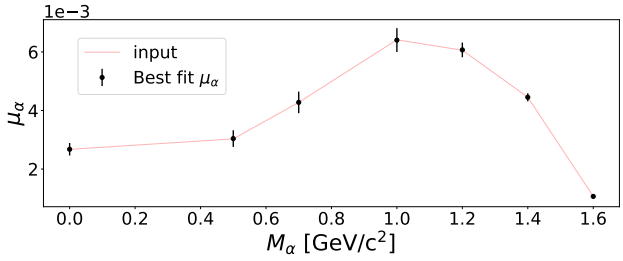
Table 5.3: Example of H_S parameter results with Julia BAT. The table displays mean posterior values with standard deviation in brackets for the electron channel's $M_\alpha = 1.0$.

	flat	log uniform
μ_α	$6.1(0.4)\times10^{-3}$	$6.1(0.4)\times10^{-3}$
μ_{SM}	1.001(0.003)	1.002(0.003)
μ_{BG}	1.00(0.02)	1.00(0.02)

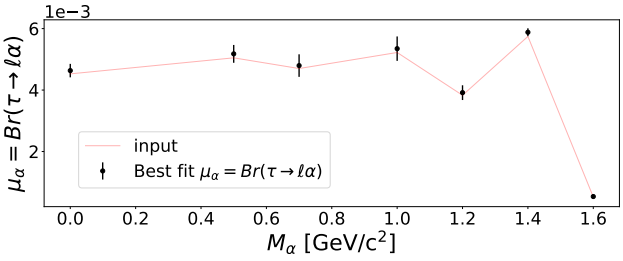
from the simulation input. We observe approximately Gaussian posterior distributions for all fit parameters in both prior cases. We can verify the distributions in Appendix G by considering the figures G.13 and G.14. As expected, the overall trend of the 2D distributions is the same as for the H_0 test.

Figure 5.12 displays the mean (black dots) of the sampling result's mean value for μ_α in each sample, determined from the 1,000 toy-data

simulations for the flat prior. The uncertainties (error bar) are the mean of the sampling result's standard deviation. As discussed in the exam-



(a) $\mu_\alpha = Br(\tau \to e\alpha)$



(b) $\mu_\alpha = Br(\tau \to \mu\alpha)$

Figure 5.12: Mean (black dots) and mean uncertainty (error bar) of Julia BAT's prior sampling result's mean and uncertainty with flat for 1,000 toy-data distributions, compared to the H_S toy-data simulation input (red line).

ple, we observe a high consistency between the injected signal and the extracted parameter values. Appendix G shows the corresponding plot for the log uniform prior in Figure G.15. The distributions are very similar to the flat prior case.

5.4.3 Conclusion Fit Results

We tested the capabilities of the pyhf fit and Julia BAT sampling to extract the input values for H_0 and H_S from a set of 1,000 toy data, samples for each hypothesis, respectively. In the H_0 test, we observe that the MINUITE fit with pyhf can reproduce the input values of the sampled model. We do not observe any bias for mean fit parameters for the 1,000 toy data samples. In contrast, we observe a systematic bias in the Julia BAT sampling results. In the case of the flat prior, the mean μ_α values with uncertainties are not consistent with zero, we observe a systematic bias toward higher values. In the case of the log uniform prior, we observe a similar but reduced bias. The posterior's mean agrees within its uncertainty with zero; due to μ_α's exponential nature of the posterior, we examine the first model of the distribution. It is close to zero, reproducing the simulation input value. The different capabilities of the Julia BAT and pyhf parameter extraction in the H_0 case are due to the difference in available parameter space. The pyhf μ_α parameter space in the fit spans

into negative values, allowing downward fluctuations. The negative parameter space is necessary to avoid problems in the MINUITE minimization. In Julia BAT, the parameter space is bound at zero, leading to a bias towards positive μ_{α} values, influencing the prior choice. The prior accentuates higher μ_{α} values, leading to a more significant bias than the prior log uniform, which accentuates lower μ_{α} values.

In the H_S test, we observe only minor differences between pyhf and Julia BAT. Both methods can extract the injected signal well.

5.4.4 Upper Limit Sensitivity

If we find no signal in the data, we determine an upper limit on the relative branching ratio with 95% confidence given our measured data. This approach is common practice to provide the maximum strength of a possible $\tau \to \ell\alpha$ decay. We use the 1,000 toy data distributions from the H_0 fit study, compute the upper limit of R for each distribution, and calculate the mean and standard deviation of the 1,000 upper limit results to study the sensitivity of our search for each mass hypotheses. We perform this study using the Frequentist and Bayesian approaches. In case of the Bayesian approach, the upper limit is the highest value of the smallest interval (range) for the μ_{α} posterior distribution. In case of the Frequentist approach, we perform a CL_S scan to estimate a value for the μ_{α} distribution. Figure 5.13 is an example of the expected CL_S scan distributions. It shows the observed CL_S (black line) and the corresponding CL_{S+B} (blue line) and CL_B (orange line) curves in one of the toy data distributions using pyhf. As described in Section 5.2, we determine the p-value at many fixed μ_{α} for the signal + background, $S + B$, and background hypothesis, B. We denote the resulting p-value curves for the $S + B$ hypothesis CL_{S+B} and the B hypothesis CL_B. We approximate the μ_{α} dependent confidence level for the signal, CL_S, as the ratio of CL_{S+B} and CL_B.

For $\mu_{\alpha} = 0$, we expect CL_{S+B} and CL_B to have the same value because both hypotheses are identical. In Figure 5.13, we observe for low values of μ_{α} that the $S + B$ hypothesis is still viable, leading to a constant CL_{S+B}. That means we cannot reject the $S + B$ hypothesis. In turn, the CL_B curve rises, indicating that rejecting the B hypothesis becomes less likely. The CL_{S+B} and CL_B behavior constantly decrease the CL_S curve. Once the CL_{S+B} decreases, the CL_B curve keeps constant. We expect the CL_B curve to be constant for high μ_{α} and the CL_{S+B} to converge towards zero, leading to a decreasing CL_S. We set the 95% confidence-level upper limit of μ_{α} (green line) at $CL_S = 0.05$. The dotted line is the expected CL_S, with the respective uncertainty as one and two σ deviation in green and yellow (Brazilian band), respectively.

We estimate the expected mean upper limit on 1,000 toy data simulations. Then, we show the average and standard deviation of the obtained limits as an uncertainty band. As described before, we use the

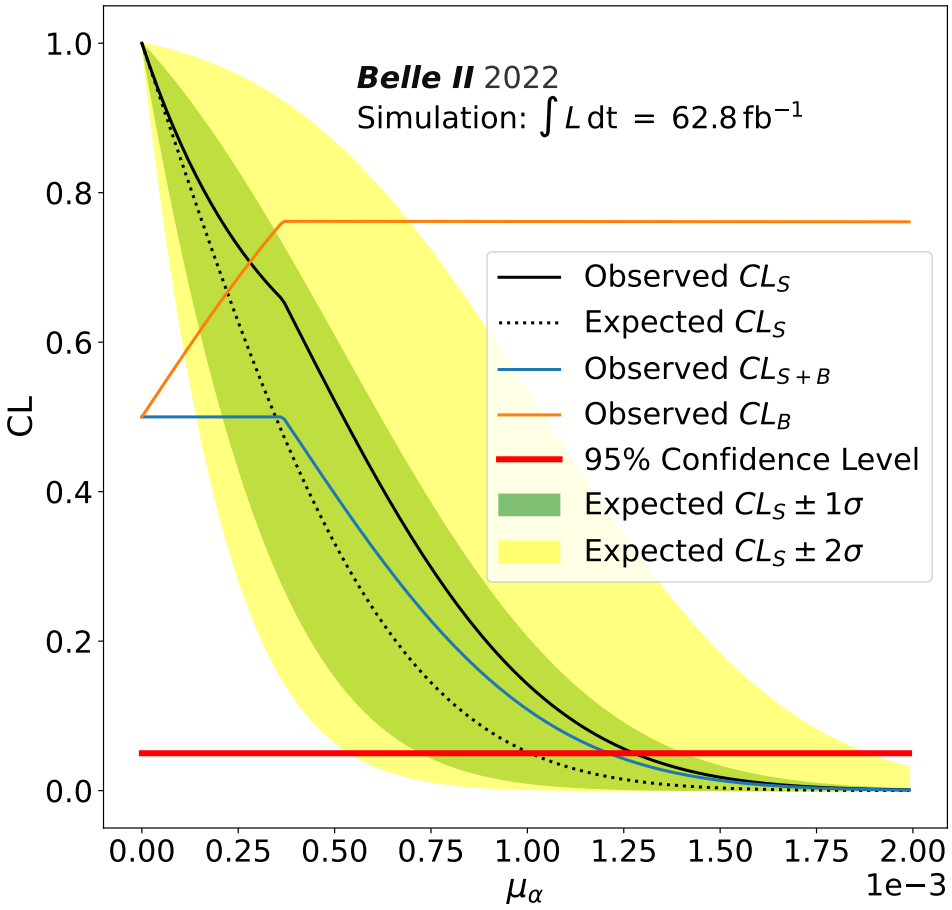
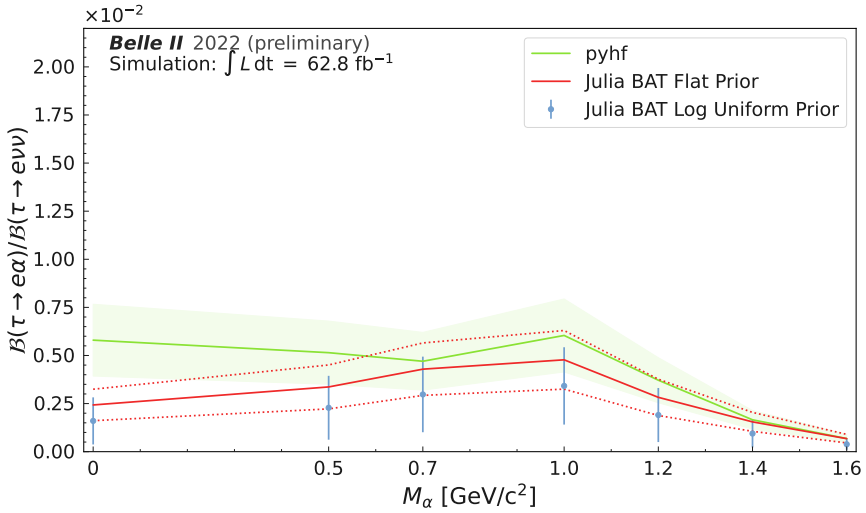
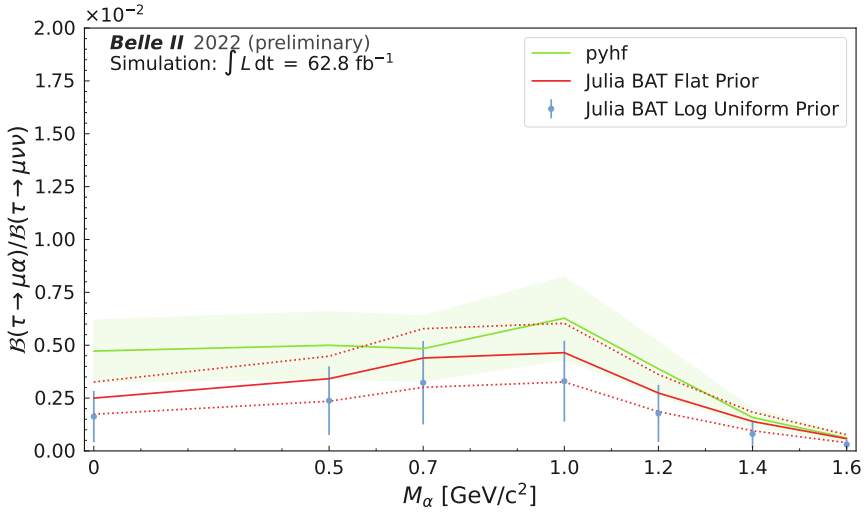


Figure 5.13: Example of a CL_S scan distribution. The *Observed* curves result from the evaluated toy data set. CL_S is the ratio of p-values for the H_0, denoted as CL_B, and the H_S, denoted as CL_{S+B}. The test statistic distributions give the *Expected* CL_S curve. The green and yellow bands denote the 1σ and 2σ variation of the Expected CL_S curve, respectively.

CL_S method to determine the Frequentist upper limit. In the case of the Bayesian Julia BAT implementation, we choose the posterior's highest value of the smallest interval covering the 95% confidence level. Figure 5.14 compares the mean upper limit for pyhf and this thesis's Julia BAT implementation. This thesis's Julia BAT implementation predicts



(a) Results for $R = Br(\tau \to e\alpha)/Br(\tau \to e\nu_e\nu_\tau)$



(b) Results for $R = Br(\tau \to \mu\alpha)/Br(\tau \to \mu\nu_\mu\nu_\tau)$

Figure 5.14: Comparison of the mean upper limit (solid lines or dots) and standard deviation (band, dashed line, or error bar) for the three upper limit estimation methods. We use the Frequentist CL_S method based on the pyhf HistFactory implementation (pyhf) and the Bayesian interference methods based on the Julia BAT implementation for the flat prior (Julia BAT Flat Prior) and the log uniform prior (Julia BAT log Uniform Prior). The lines show an interpolation of the upper limit between the evaluated mass hypothesis.

a slightly lower upper limit than the pyhf implementation. An essential factor for this result is the constraint on the $\tau \to \ell\nu_\tau\nu_\ell$ distributions and background distributions given by the prior choice. Another reason

for this effect might be the known tendency of the CL_S method towards more conservative upper limit estimates.

5.5 Systematic Uncertainties

After setting up the procedure to measure R or, depending on what we observe in data, determine an upper limit, the next step is to consider the systematic uncertainties and how to include them in the analysis.

Systematic uncertainties are all uncertainties introduced for other reasons than the data sample size. These uncertainties may stem from measured quantities such as lepton identification, leptonID, or uncertainties due to mismodeling in simulation. Due to their inherent diverse nature, we must take care of how to treat these uncertainties.

The following section describes how we treat our systematic uncertainties from various sources. We focus on systematic uncertainties coming from the leptonID and trigger, which are the dominant systematic uncertainties. We determine the impact of a systematic uncertainty by evaluating the resulting upper limit with and without systematic uncertainties. We use the upper limit because it estimates the sensitivity for our fit model.

5.5.1 Approach

In sections 5.1 and 5.3, we discussed how we set up our fit model with HistFactory and determine the upper limit. The HistFactory package includes nuisance parameters for every systematic uncertainty in the likelihood determined from the fit model.

Hence, we can reinterpret the systematic uncertainties as extra unknown parameters of the model and include them in the likelihood function as additional nuisance parameters of the fit. These new parameters can be determined or constrained from extra inputs or separated auxiliary measurements. Hence, in the combined likelihood function, the nuisance parameters of the systematic uncertainties are treated similarly to $\mu_{\ell\nu\nu}$ and μ_{BG}.We therefore include the systematic uncertainties in a unified approach that works for measuring or setting an upper limit. At this point of the search, we identify the following possible sources of systematic uncertainties.

- **Lepton identification and fake-rate corrections**:
  The leptonID is not modeled perfectly in simulated data. Belle II observes a difference between data and simulation samples. We apply the correction factors and include one systematic uncertainty for each leptonID correction by combining the statistical and systematic uncertainties. Depending on the leptonID efficiency or fake-rate correction, the normalized uncertainty is on the same order of magnitude as the correction.

- **Trigger efficiency**:
  We consider that the trigger logic is not 100% efficient, and we only have the events in measured data that pass any of the available triggers requirements. As Section 5.5.3 discusses in detail, we obtain correction factors for our simulated data and apply them to the templates. We consider the trigger correction's statistical uncertainty as a systematic uncertainty on the order of several percent.
- **Momentum scale** (evaluated on data):
  We consider the tracking momentum bias derived from the wrong magnetic field map, B-field map, used in the data reconstruction process. Belle II measured the magnetic field and generated a B-field map for the detector. This measured B-field map has inherent uncertainties from the measurement. In the reconstruction, this uncertainty leads to a deviation between the B-field map's magnetic field strength and the actual magnetic field strength by which the reconstructed particle was affected. This deviation leads to a bias, which is not present in the simulation data where the B-field map is perfectly known. Belle II provides a global correction factor, which we apply to the measured data. The normalized uncertainty of the correction factor is on the order of about 0.5 per mill.
- π^0 **efficiency**:
  We first reconstruct neutral candidates in our selection and reject them afterward, as Chapter 3 describes. Due to this approach, we must consider the overall π^0 reconstruction efficiency correction provided by the Belle II Neutrals Performance Group and the associated total uncertainty [104]. The overall correction factor has a normalized uncertainty of about two percent. This correction affects us in the following way: The lower the efficiency, the more π^0 candidates are not reconstructed, and we count more events without π^0 candidates, leading to an increased data set. Potentially, this correction results in a shape variation for the background sample, which might affect the sensitivity of our search. We do not expect any impact on pure samples because the floating yield parameter in the fit allows us to account for this global correction. We observe a negligible impact in simulation studies when including the variations as a shape correlated modifier.
- **Tracking efficiency**:
  We consider the Belle II track finding reconstruction efficiency. It is the probability of reconstructing a track. Belle II measures the efficiency and uncertainty. It provides a tool that accounts for the efficiency during the event reconstruction. The normalized uncertainty is on the order of about one percent, and we find the impact of this systematic uncertainty to be negligible.
- **Beam energy**:
  We consider the effect of a possible beam energy spread. The actual

center-of-mass energy of SuperKEKB, $\sqrt{s} = 10.58\,\text{GeV}$, can fluctuate by 0.1 per mill. Since we use $\sqrt{s}$ as input in our analysis for the τ pseudo-rest frame, we evaluate the impact of a variation on our result. We find that the impact of this systematic uncertainty is negligible.

- **Relative efficiency**:
  We consider that the sample size of simulated data we use to generate our templates is finite. Due to that, we must evaluate the impact of the finite data size on our result. We find that this systematic uncertainty is negligible for a dataset of $200\,\text{fb}^{-1}$ of simulation data.

In the studies on simulated data, the first two items on above list are the dominant sources of systematic uncertainties, and the others are negligible in their effect on the upper limit estimation, which we use to estimate sensitivity. The following sections discuss leptonID and trigger corrections. The momentum scale is evaluated on data. We plan to treat the momentum scale correction on data similarly to all other systematic uncertainties. The Belle II performance group provides the correction factors on data. Chapter 6 presents the results. Appendix F discusses the sources with negligible effects.

We use the upper limit estimation to quantify the impact of the systematic uncertainty on the sensitivity. Our approach of evaluating the systematic uncertainties in the upper limit estimation is similar to Section 5.4.4. We determine the upper limit for a set of 1,000 H_0 toy-data distributions sampled with the respective fit model $F(p_\tau^\star)$. We compare the resulting mean upper limits and standard deviations of the fit function without systematic uncertainties, nominal, with the fit functions including the systematic uncertainty, SM fit σ_{sys}.

5.5.2 Lepton Identification

Excellent leptonID is one of the critical requirements for our $\tau \to \ell\alpha$ search. Due to various reasons, there are discrepancies between measured and simulated data. Belle II accounts through them in performance measurements, providing two-dimensional binned weights, w^{ij}, in the space of the particle's polar angle θ and momentum – both in the laboratory frame – applied on simulated data samples. These weights correct the simulation-leptonID efficiency to the actual efficiency determined in measured data. Furthermore, we get fake-rate correction weights that account for misidentifying a particle as a lepton. We apply the corrections to the simulation data as follows:

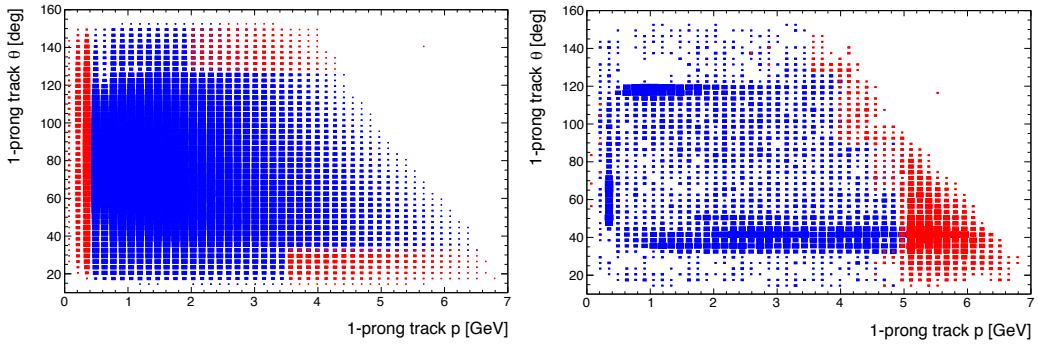
- If the track's generator level particles ID matches that of an electron (muon), we apply the leptonID correction.
- If the track's generator level particles ID does not match an electron (muon), we apply the lepton fake-rate correction.

The leptonID and fake-rate corrections each have two uncertainties. One is the statistical uncertainty, σ^{stat}, of the measurement. The other is a systematic uncertainty, σ^{sys}, mainly related to discrepancies between the different methods used to obtain these measurements. In our statistical model, we combine the two uncertainties as

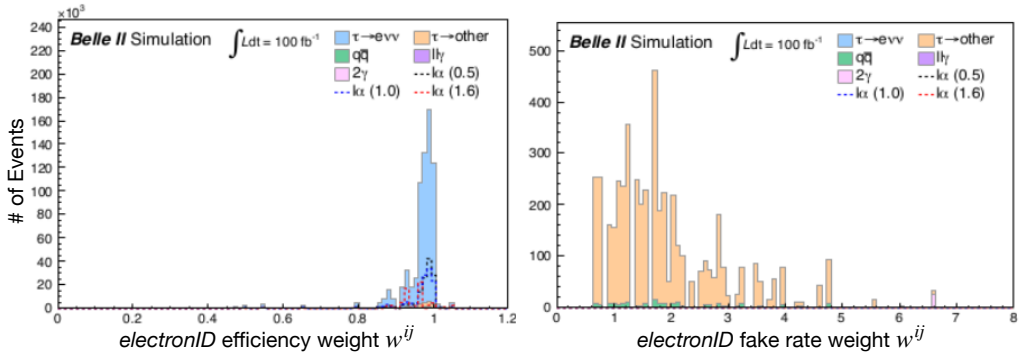
$$\sigma_{\text{lid}} = \sqrt{(\sigma_{\text{lid}}^{\text{stat}})^2 + (\sigma_{\text{lid}}^{\text{sys}})^2}, \tag{5.14}$$

with *lid* indicating the leptonID or fake-rate correction, respectively.

Figure 5.15 shows the correction coverage and weight distributions for the electron. Plot 5.15a shows the covered (blue) and uncovered



(a) *electronID* Efficiency Coverage. (b) *electronID* Fake Rate Coverage.



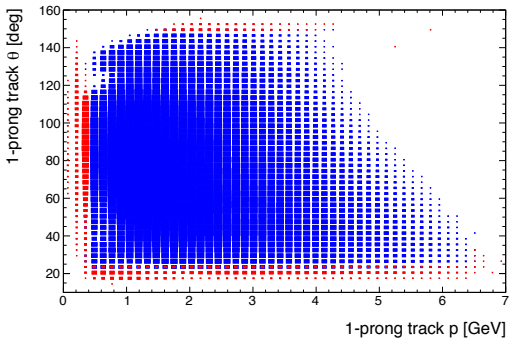
(c) *electronID* Efficiency and Fake Rate Available Weights.

Figure 5.15: *electronID* efficiency and fake-rate corrections. The plots (a) and (b) show the available coverage of the *electronID* efficiency and fake-rate correction, with blue bins indicating a covered and red bin indicating an uncovered region. The plots show the two-dimensional space of the event's 1-prong momentum (*1-prong track p*) and 1-prong θ (*1-prong track θ*) in the laboratory rest frame. The lower two plots show the distributions of the available weights w^{ij} for various samples. The left plot shows the distribution of the *electronID* efficiency correction, and the right plot shows the *electronID* fake-rate correction. The distributions in Plot (c) are the $\tau \to e\nu_\tau\nu_e$ distribution ($\tau \to e\nu\nu$), the background distribution split into misidentified τ 1-prong events ($\tau \to$ other), quark pair events ($q\bar{q}$), radiative two lepton events (llγ), two-photon events (2γ), and three mass hypothesis for the $\tau \to \ell\alpha$ decay (lα), with the mass hypothesis $M_\alpha = \{0.5, 1.0, 1.6\}[\text{GeV}/c^2]$ indicated by the number in brackets.

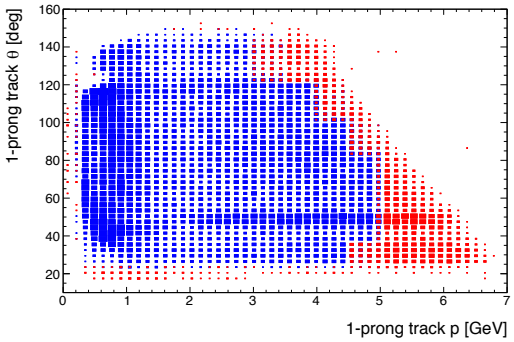
(red) regions of the electron efficiency corrections. Plot 5.15b shows the same for the electron fake-rate correction. The two-dimensional parameter space of the two plots is the same as the leptonID group's binned weights w^{ij} parameter space. By comparing plots (a) and (b), we generally observe a better coverage for the electron efficiency than for the fake-rate correction. The reason is that the Belle II fake rate study relies on hadronic 3-prong τ decays, which are limited in the momentum range. Furthermore, the box sizes in the plot indicate the number of events in the bin, allowing us to evaluate the cause of the misidentified particle. We notice that most of the misidentified events are close to *1-prong track* θ values of 40° and 120°. These values correspond to the gaps between the detector's barrel and endcap ECL, which is the most important detector for calculating the *electronID*. Plot 5.15c shows the distributions of the electron efficiency correction weights (left) and the electron fake rate weights (right). In the case of the efficiency corrections, we observe a relatively narrow distribution close to one, with a small tail towards lower values, indicating a reasonable description of the lepton identification in simulation data. In the case of the fake-rate corrections, we observe a broad distribution from below one up to seven. This wide range indicates a severely underestimated fake rate in simulation data. The y-axis ranges are different because we evaluate the simulation data set after selection, which has a highly suppressed background sample contribution.

Figure 5.16 shows the correction coverage and weight distributions for the muon channels. In general, we observe similar trends as in the electron channel. Plot (b) in both figures shows the same coverage plot for the fake-rate corrections. The *muonID* efficiency covers more parts of the endcap region than the *electronID* because it is not affected by the unavailable trigger as is the case for the *electronID*. The coverage of the *muonID* fake-rate correction is similar to the *electronID* correction. We observe a reduced dependence of misidentified particles on *1-prong track* θ and an increased dependence on *1-prong track p*. We noticeable an increased misidentification for *1-prong track p* below 1.0 GeV/c. This region corresponds to the momentum range in which $\pi^{\mp}$ and $\mu^{\pm}$ have a very similar detector signature. As $\pi^{\mp}$ and $\mu^{\pm}$ do not reach the KLM detector, we can barely distinguish them. The w^{ij} distributions for the correction factors in the muon channel are generally closer to a Gaussian distribution than in the electron channel. This observation can be mainly attributed to the *muonID* fake-rate correction, which leads to a considerably narrower distribution, between 0.2 and 1.5, and a defined peak close to one. The *muonID* efficiency correction shows a peak close to one and a mildly larger spread than the *electronID*.

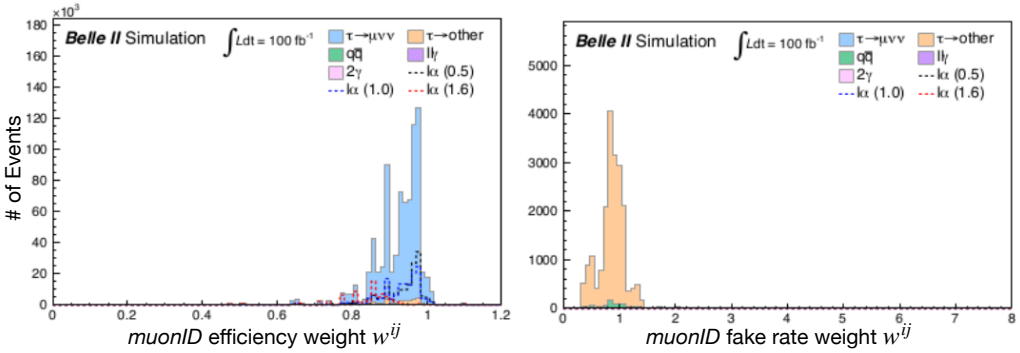
In general, we observe a more consistent behaviour of the *muonID* corrections than for the *electronID* correction. The reason is the observed fake-rate-correction distribution that indicates a better understanding of the particles faking muons than faking electrons in simu-



(a) *muID* and Efficiency Coverage.



(b) *muID* and Fake Rate Coverage.



(c) *muID* and Fake Rate Available Weights.

Figure 5.16: *muonID* efficiency and fake-rate corrections. The plots (a) and (b) show the available coverage of the *muonID* efficiency and fake-rate correction, with blue bins indicating a covered and red bin indicating an uncovered region. The plots show the two-dimensional space of the event's 1-prong momentum (*1-prong track p*) and 1-prong θ (*1-prong track* θ) in the laboratory rest frame. The lower two plots show the distributions of the available weights w^{ij} for various samples. The left plot shows the distribution of the *muonID* efficiency correction, and the right plot shows the *muonID* fake-rate correction. The distributions in Plot (c) are the $\tau \to \mu\nu_\tau\nu_\mu$ distribution ($\tau \to \mu\nu\nu$), the background distribution split into misidentified τ 1-prong events ($\tau \to$ other), quark pair events ($q\bar{q}$), radiative two lepton events (llγ), two-photon events (2γ), and three mass hypothesis for the $\tau \to \ell\alpha$ decay (lα), with the mass hypothesis $M_\alpha = \{0.5, 1.0, 1.6\}[\,\mathrm{GeV}/c^2]$ indicated by the number in brackets.

lation data. The *electronID* efficiency correction shows a narrower distribution, close to one, indicating a better understanding of the *electronID* in the covered detector. This observation can be an artifact stemming from the rejection of the high momentum-endcap regions – shallow and high *1-prong track* θ values – in which the precision of the *electronID* can suffer in measured data.

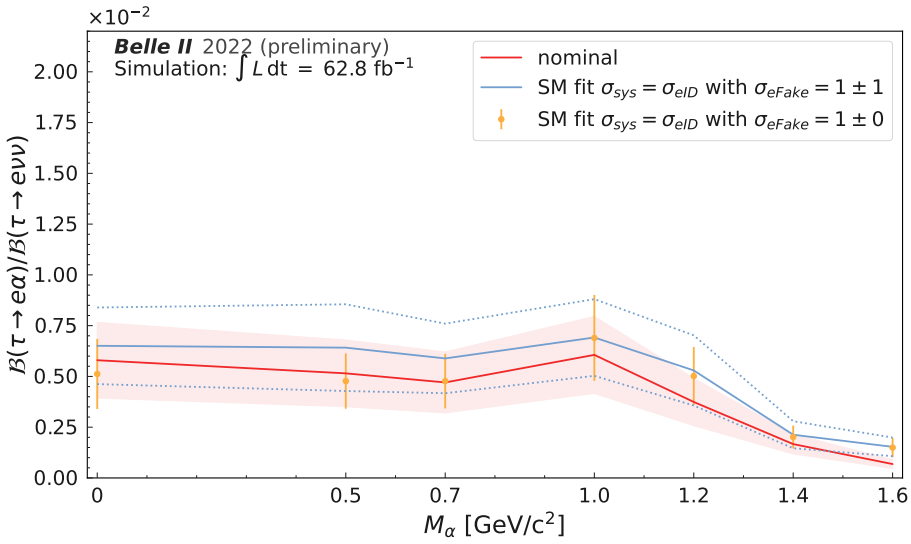
We have limited coverage of weights (red region) for both leptonID fake-rate corrections. In the case of the lepton fake-rate corrections, the uncovered regions pose a problem, as rejecting events from these regions would cost a decent part of our data set. We address this problem in the following discussion.

We study two extreme cases to account for the poor coverage of the fake-rate correction. For the fake-rate corrections in the uncovered phase space, we assign:

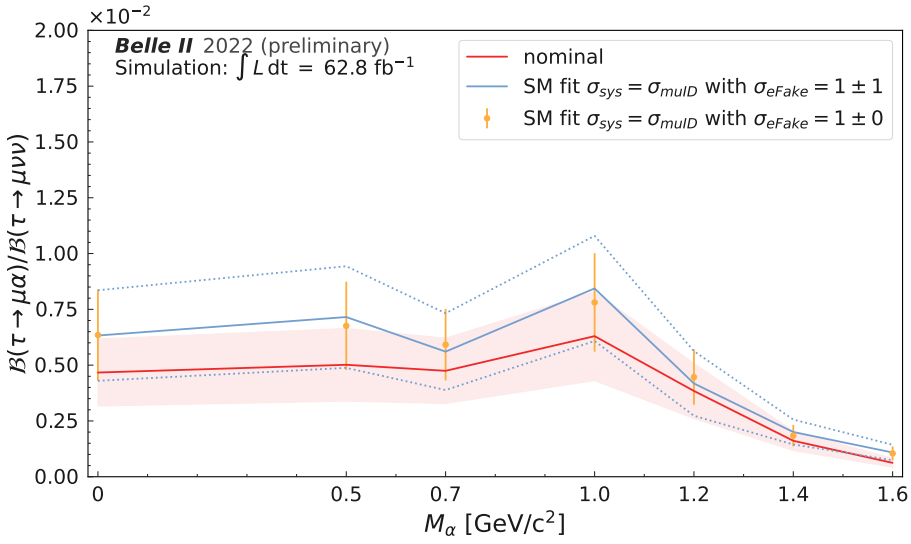
1. a weight of $w^{ij}_{\text{fake}} = 1$, and an uncertainty of $\sigma^{ij} = 0$.
2. a weight of $w^{ij}_{\text{fake}} = 1$, and an uncertainty of $\sigma^{ij} = \pm 1$.

Judging from the distributions of the available w^{ij}_{fake} in the left plots of figures 5.15c and 5.16c, we assume that the uncertainty of 100% for the electron channel is reasonable; however, a normalized uncertainty of 100% could result in an overestimation of the effect of the muon channel. Thanks to the relatively comprehensive coverage of the fake-rate correction, we consider the potential impact to be minor in our estimation.

Following the above procedure and based on the pyhf package, we report the results for the electron and muon channels. Results denoted as *nomina* are without systematic uncertainties taken into account, and results denoted as *SM fit* σ_{sys} take the systematic uncertainties into account. We include the leptonID efficiency uncertainty and the fake-rate correction each as a systematic uncertainty. We select the correlated shape modifier for each systematic, which yields one nuisance parameter per systematic uncertainty in the fit model. In total, we have two nuisance parameters for the leptonID. One nuisance parameter for the leptonID efficiency (σ_{eID} or σ_{muID}) and one for nuisance parameter for the leptonID-fake rate (σ_{eFake} or σ_{muFake}). We observe no significant reduction in estimation power for R for H_0 or H_S in toy-data sets for both leptonID studies. We compare the two cases for the estimated leptonID fake-rate-correction uncertainty when calculating the upper limit for H_0 in Figure 5.17. Plot 5.17a shows the results for the electron channel. We observe that an *electronID* fake-rate-correction uncertainty of zero results in an upper-limit estimation similar to the nominal case. In contrast, in the case of an *electronID* fake-rate-correction uncertainty of one, we observe a systematic shift towards higher upper limits, indicating a significant effect of this fake-rate-correction uncertainty treatment for the *electronID*. This observation shows that the electron channel is significantly affected by misidentified electrons and the uncovered *electronID*



(a) $\tau \to e\alpha$



(b) $\tau \to \mu\alpha$

Figure 5.17: Comparison of the mean upper limit (solid line or dots) and standard deviation (band, dashed line, or error bar) for the fit model without systematic uncertainties (nominal) and the fit models including the leptonID systematic uncertainties (SM fit $\sigma_{\text{sys}} = \sigma_{\text{muID}}$ with σ_{eFake} ...). The lines show an interpolation of the upper limit between the evaluated mass hypotheses.

fake-rate correction. In Plot 5.17b we observe a systematic shift towards higher upper limit estimations for both *muonID* fake-rate-correction uncertainty treatments. We observe only marginal differences between the two treatments. This observation shows that the provided *muonID* correction impacts the muon channel the most. The effect of the estimated *muonID* fake-rate-correction uncertainty is negligible.

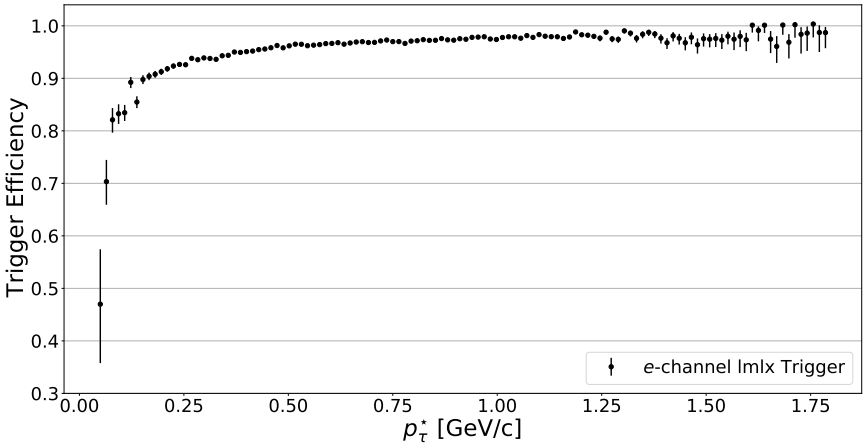
We choose to include the leptonID systematic uncertainties with the fake-rate-correction uncertainty treatment using the 100% normalized uncertainty ($\sigma_{\text{eFake}} = 1 \pm 1$). This way, we obtain a treatment for both channels, which has a similar impact on our sensitivity, and it is the more conservative treatment, allowing for more fluctuations in measured data.

5.5.3 Trigger Correction

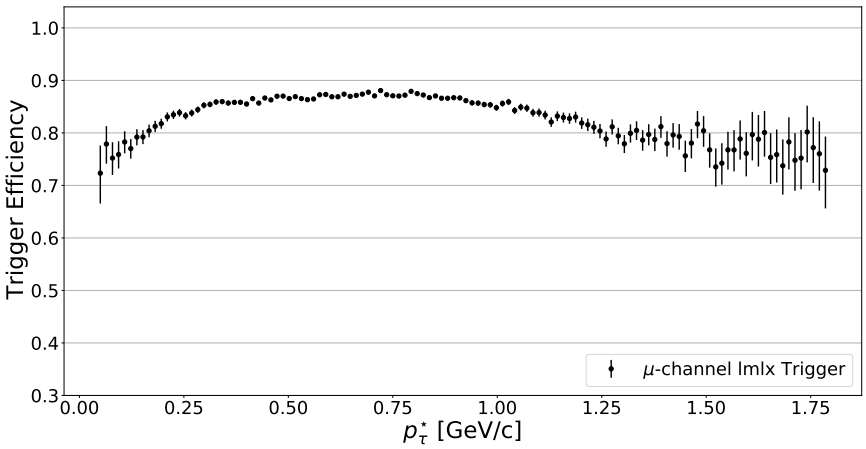
Our simulation does not include trigger effects, meaning that the simulation assumes 100% trigger efficiency for our event type. Before comparing simulated with measured data, we must evaluate the trigger inefficiencies in our measured data and account for differences in the simulation data. The Belle II Note [76] describes in detail the methodology followed to obtain the efficiency curves. We determine the trigger correction and apply it directly to the templates, altering the shape and introducing systematic uncertainties due to the trigger correction uncertainty.

The trigger correction curves in Figure 5.18 result from comparing the events that fired the logical OR of all triggers we use in simulated and measured data. We use them to correct our fit model templates. Similar to the leptonID and fake-rate corrections, we introduce a new systematic uncertainty into our fit model for the trigger correction's uncertainty. We use the shape correlated modifier yielding one nuisance parameter to propagate the efficiency curves' uncertainty to estimate the upper limits.

Figure 5.19 shows the resulting mean upper limit (solid band) when accounting for trigger inefficiencies and the resulting systematic uncertainty. When accounting for the trigger correction, we observe a systematic shift upwards for the upper limits. Hence, we observe a significant effect of the trigger on the upper limit.

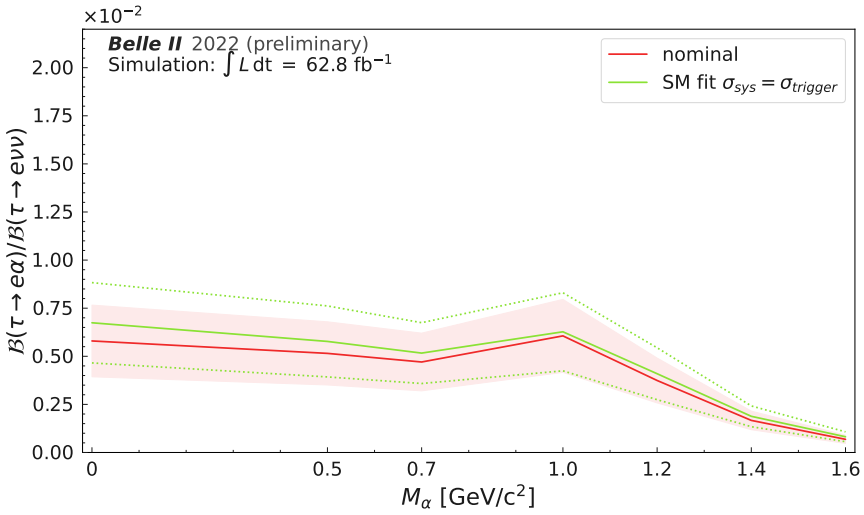


(a) e channel

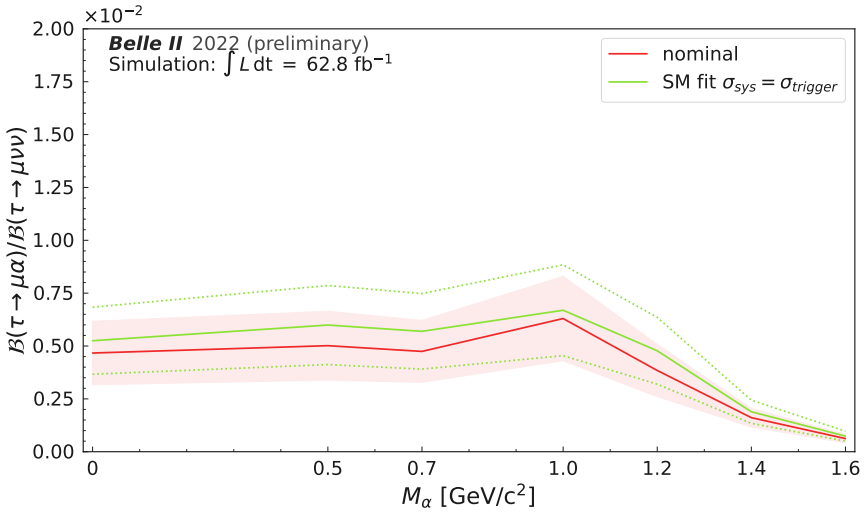


(b) μ channel

Figure 5.18: Trigger efficiency as a function of the $p_\tau^\star$ variable. The trigger requirement is the logical OR of the Belle II-low-multiplicity triggers. The uncertainty indicates the statistical error.



(a) $\tau \to e\alpha$

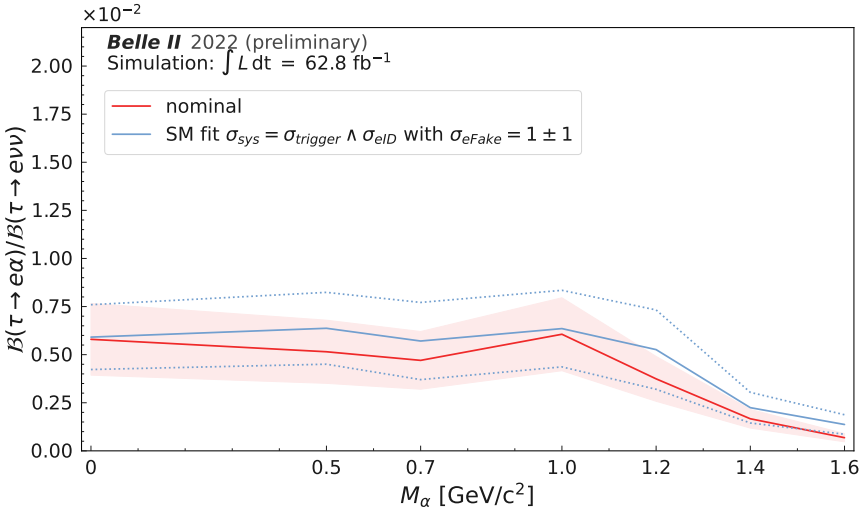


(b) $\tau \to \mu\alpha$

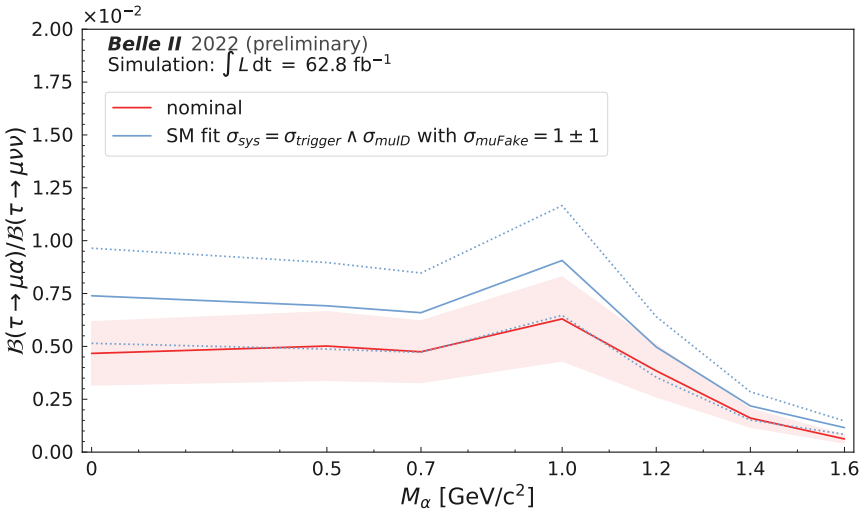
Figure 5.19: Comparison of the mean upper limit (solid lines) and standard deviation (band or dashed line) for the fit model without systematic uncertainties (nominal) and the fit model including the trigger correction uncertainty (SM fit $\sigma_{\text{sys}} = \sigma_{\text{trigger}}$). The lines show an interpolation of the upper limit between the evaluated mass hypotheses.

5.5.4 Combination of Trigger and LeptonID Corrections

After evaluating the individual effect of the dominant systematic uncertainties, we estimate the total effect on the upper limit estimation. Here, we include the systematic uncertainties stemming from the leptonID efficiency, leptonID fake-rate, and trigger correction uncertainty into the fit model. We show the resulting mean upper limit (solid band) results in Figure 5.20. We observe a systematic shift upwards for the



(a) $\tau \to e\alpha$



(b) $\tau \to \mu\alpha$

Figure 5.20: Comparison of the mean upper limit (solid lines) and standard deviation (band or dashed line) for the fit model without systematic uncertainties (nominal) and the fit model including the dominant systematic uncertainties being the trigger and leptonID correction uncertainties, with the leptonID estimating uncovered fake-rate correction with a weight of one and uncertainty of one (SM fit $\sigma_{\text{sys}} = \sigma_{\text{trigger}} \wedge \sigma_{\text{leptonID}}$ with $\sigma_{\text{eFake}} = 1 \pm 1$). The lines show an interpolation of the upper limit between the evaluated mass hypotheses.

fit model when including the dominant systematic uncertainties. We do not identify a simple rule for increasing the upper limit when combing the systematic effects. The systematic shift upwards seems mainly attributed to the leptonID correction and its corresponding uncertainties, but we also observe an impact from the trigger correction. In the electron channel, we observe countering effects when applying all systematic uncertainties. For $M_{\alpha} = 0.0\,\mathrm{GeV}/c^2$, we observe a similar upper limit result for the nominal fit model and the one including systematic uncertainties, which means the upper limit decreased compared to the single treatment of the leptonID and trigger correction uncertainties. We identify also for for $M_{\alpha} = 1.0\,\mathrm{GeV}/c^2$ and $M_{\alpha} = 1.6\,\mathrm{GeV}/c^2$ a decreasing upper limit when comparing Figure 5.17a with Figure 5.20a. That means we observe an improved sensitivity for the fit model when including all systematic uncertainties compared to the results with the fit model only using the *electronID* systematic uncertainties. In the muon channel, this countering effect is not visible. In contrast, the combined systematic uncertainties lead to the highest upper limit result. In Figure 5.20b we observe that the fit model with systematic uncertainties leads to an almost constant upwards shift such that the lower blue dashed line indicating the standard deviation is approximately equal to the nominal-fit model upper-limit results. One possible reason might be the differences we discussed for the upward shifting when applying the leptonID systematic uncertainty. The electron channel was mainly affected by the estimated fake-rate-correction uncertainties. We attribute the effect in the muon channel to other sources when applying the *muonID* correction. Hence, it could be that the combination of trigger correction and *electronID* correction uncertainties mitigates the effects of the estimated fake-rate-correction uncertainties, leading to a less pronounced effect of the systematic uncertainties on the upper limit. In the muon channel, this effect is negligible; hence, we observe no improvement.

Chapter 6

Results

In the last chapters, we discussed the selection, the τ-rest-frame reconstruction, and the fit procedure. This chapter pesents the final result for our $\tau \to \ell\alpha$ search with measured data in Section 6.1. Section 6.2 shows sensitivity estimations using simulated data.

In Section 6.1 we find no evidence for the $\tau \to \ell\alpha$ decay and determine a 95% confidence-level upper limit for the parameter of interest R. We obtain this result using the cut-based selection without neutrals, described in Chapter 3, and the ARGUS τ-rest-frame method, described in Chapter 4. This corresponds to the approved Belle II $\tau \to \ell\alpha$ search strategy.

Section 6.1 consists of three parts. First, Section 6.1.1 discusses necessary changes to the treatment of systematic uncertainty in the fit model. These were necessary because we observed fit instabilities we had to overcome. Next, Section 6.1.2 evaluates the impact of the momentum-scale-correction uncertainty on the fit and the 95% confidence-level upper limit results. We find a negligible impact of the below-percent-level uncertainty on the fit results. Finally, Section 6.1.3 presents the final result of the measurement with the updated fit, including all relevant systematic uncertainties, i.e., leptonID and trigger corrections.

In Section 6.2 we compare the approved Belle II $\tau \to \ell\alpha$ search strategy to alterntives we developed in this thesis. In Chapter 3 we found two alternative selection strategies, and in Chapter 4 we developed two alternative τ-rest-frame estimates. The alternative selection strategies we found allow for neutral particle candiates. We discussed a cut-based and a BDT-based selection variant. Section 6.2.1 discusses the results for the 95% confidence-level upper limit with toy data for each selection variant, giving an estimate of the respective sensitivity. We find a focus on purity is not favourable. A selection optimizing all selection variables with our figure of merit yields an improved sensitivity. In Chapter 4 we discussed the reconstruction of the τ-rest frame and developed the Thrust and the GKK method. In Section 6.2.2 we evaluate their respective sensitivity using toy data by considering the 95% confidence-level upper limit results for the three variables schemes. We find no signif-

icant improvement of the ARGUS method. The GKK method showed even worse performance for the $\tau \to \ell\alpha$ search than the ARGUS method. This suggests, that the GKK method is disadvantageous for a search. Chapter 7 will show that GKK performs well in precisely measuring a parameter.

6.1 Results with 62.8 fb^{-1} of Belle II Data

This section describes the steps toward the final result of the $\tau \to \ell\alpha$ search. Section 6.1.1 describes the results of and reasoning behind necessary adjustments in our fit model after unblinding the data. Section 6.1.2 includes the momentum-scale-correction uncertainty into our fit model as described in Chapter 5.

As our analysis was a blind analysis, we needed the approval of the review committee to unblind Belle II's measured data. The review committee approved our unblinding request after we included the π^0 uncertainty in the fit model described in the previous chapter. The resulting likelihood function for the measured data showed discontinuities and double minima, leading to an unstable fit. This instability necessitates modification of the fit function such that the likelihood function is well-behaved. Section 6.1.1 describes our treatment of systematic uncertainties for the $\tau \to \ell\alpha$ search, which got accepted by the Belle II collaboration.

Section 6.1.2 describes the momentum-scale correction-uncertainty study. The momentum scale correction is applied to measured data, so the uncertainty of the correction cannot be evaluated on simulated data. Therefore, we can only incorporate the uncertainty into the fit model after unblinding. We want to estimate the impact of this systematic uncertainty on our parameter of interest to examine if we have to incorporate it into our fit model. We treat this uncertainty as we treated the leptonID or trigger in Chapter 5. We determine the variation for the upward and downward templates from data and include this uncertainty into the fit model. We are concerned that the log-likelihood function of the fit model is not well-behaved. To mitigate statistical fluctuations, we want to overcome this problem in advance and rebin the uncertainties' up and down variation.

Section 6.1.3 presents the results for the approved $\tau \to \ell\alpha$ search at Belle II. First, we check for fit instabilities in our model parameters' log-likelihood curves. For this step, we use phyf. Afterward, we estimate our parameter of interest. We find no evidence for a signal. Therefore, we determine a 95% confidence-level upper limit with the CL_S method. We compare the final upper limit results obtained with pyhf with our approved Belle II results obtained with RooStats. We observe no statistically significant deviation comparing both methods.

6.1.1 Post-Unblinding Changes

After unblinding, the likelihoods of the nuisance parameters showed double minima or discontinuities in the likelihood function when profiled over each nuisance parameter. These likelihoods destabilize the minimization during the fit, leading to an unstable fit. Furthermore, this causes problems in the CL_S scan leading to the erratic behavior of the curves and hence the upper limit estimation. Shape variation with large statistical fluctuations, asymmetric upward and downward variations[1], or both features cause not well-behaved likelihoods. Therefore, we must update our treatment and take the following measures:

- we increase our background data set by a factor of five by considering the full 1 ab^{-1} of Belle II simulation data for the background simulation samples listed in Table C.3 of Appendix H;
- we increase the bin sizes, re-bin, for the trigger efficiency correction to reduce the statistical fluctuations in low and high regions of $p_\tau^\star$;
- we smooth [105] the relative up and down systematic variations of the leptonID efficiency, the fake rate, and the π^0 efficiency corrections;
- and we symmetrize the templates of the upward and the downward variations for all nuisance parameters. That means we force the systematic uncertainty to be symmetric for each bin.

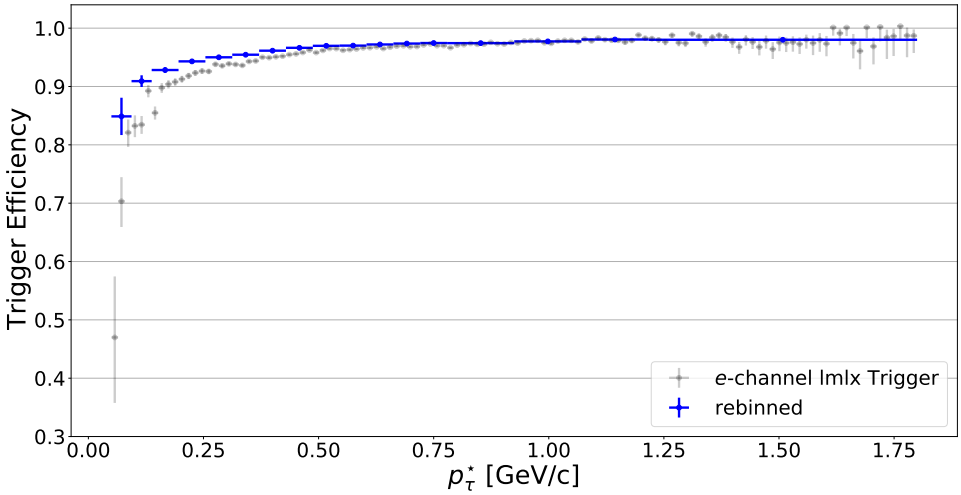
We update our systematic uncertainties by either re-binning as is done with the trigger efficiency or smoothing the systematic uncertainty. We apply re-binning or smoothing depending on how the Belle II performance groups decided to provide the correction. We rebin those systematic corrections applied directly on the template and smoothen those corrections applied to the event weight. We determine the trigger efficiency correction ourselves as described in Chapter 5 and apply it to the template. We apply all other systematic uncertainties discussed up until now on the event weight. Additionally, if the resulting uncertainty is not symmetric, we symmetrize it.

Rebinning the trigger efficiency correction

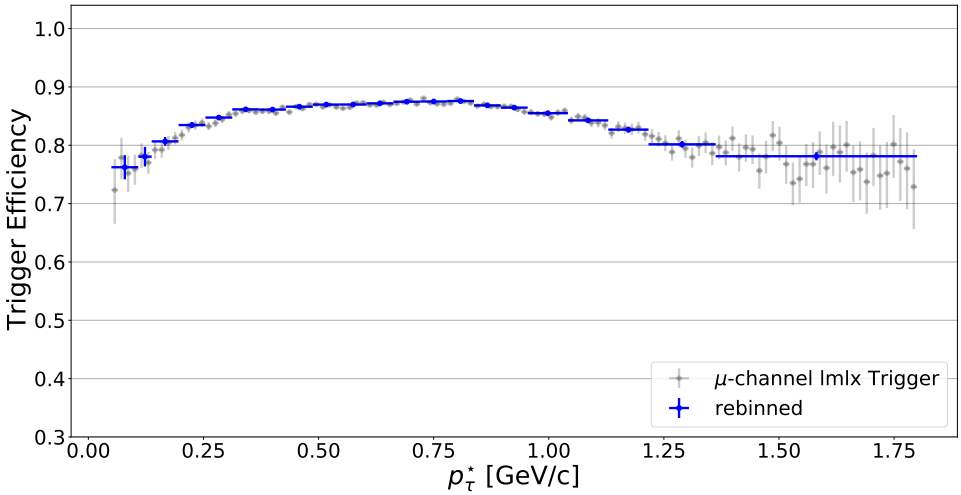
We want to obtain a reduction in statistical fluctuations of the trigger efficiency. In order to preserve the features of the efficiency curves, we decided to manually rebin the trigger efficiency correction. One of the efficiency curve features we want to preserve includes the shape of the trigger "turn-on" region at low $p_\tau^\star$ and the plateau at high $p_\tau^\star$. We can preserve the observed shape of the trigger efficiency with bins of variable width. Bins at low $p_\tau^\star$ are typically narrower while those at high

[1] An asymmetry that breaks the Gaussian approximation of the nuisance parameter constraint.

$p_\tau^\star$ are wider. Figure 6.1 shows the distributions of the trigger efficiency before (gray dots) and after the rebinning (blue dots). In Figure 6.1a we



(a) Electron Channel



(b) Muon Channel

Figure 6.1: Trigger efficiency before (gray dots) and after (blue dots) the rebinning. Plot (a) is the electron channel and Plot (b) shows the muon channel.

observe that the rebinned trigger efficiency for the electron channel has a similar shape, but a higher efficiency in $p_\tau^\star$ below 0.5 GeV/c. From that point onward, the rebinned trigger efficiency coincides with the original trigger efficiency. In Figure 6.1b we observe that the rebinned trigger efficiency for the muon channel follows the shape of the original trigger efficiency and that it reduces statistical fluctuations substantially. In both channels, we observe a substantial reduction of the trigger-efficiency uncertainty when comparing our rebinned to our original trigger-efficiency uncertainty. The resulting trigger efficiency is symmetric. Therefore,

there is no need to symmetrize it.

Smoothing and Symmetrization

To ensure well-behaved log-likelihood functions of the fit model, we must eliminate erratic variations and asymmetries in the systematic uncertainties. We eliminate these erratic variations and asymmetries by smoothening and symmetrization. We use the standard Root smoothening algorithm [105]. We smooth the variation to preserve the systematic uncertainties' variation shape and reduce statistical fluctuations. Afterward, we symmetrize the uncertainty.

When symmetrizing the shape variations for systematic uncertainties, the conservative approach compares the upward and the downward variations in each template bin to the nominal distribution and takes the largest of the two variations. This procedure gives the symmetric uncertainty for the bin in question. This procedure removes double minima or discontinuities in the profile likelihood function of the systematic uncertainties' nuisance parameter.

Figure 6.2 exemplifies the procedure by illustrating the *electronID* correction before and after the smoothing and symmetrization. It shows the variations of the *electronID* for the $\tau \rightarrow e\nu_\tau\nu_e$ sample. The two top plots compare the smoothed (green line) to the orginial (black dotted line) *electronID* up-variation (left) and down-variation (right). We observe a mediation of the statistical fluctuations of the up-variation for high $p_\tau^\star$. The shape of the *electronID* is preserved. The bottom plot shows the smoothed and symmetrized *electronID* variations. We find that the new *electronID* variation covers the orginial one, mitigates the statistical fluctuations in the higher $p_\tau^\star$, but also expands the uncertainty substantially. This leads to an overestimation of the *electronID* uncertainty. Figure G.35 in Appendix G is a RooStats result that exemplifies the impact of the shape variation systematics on our upper limit.

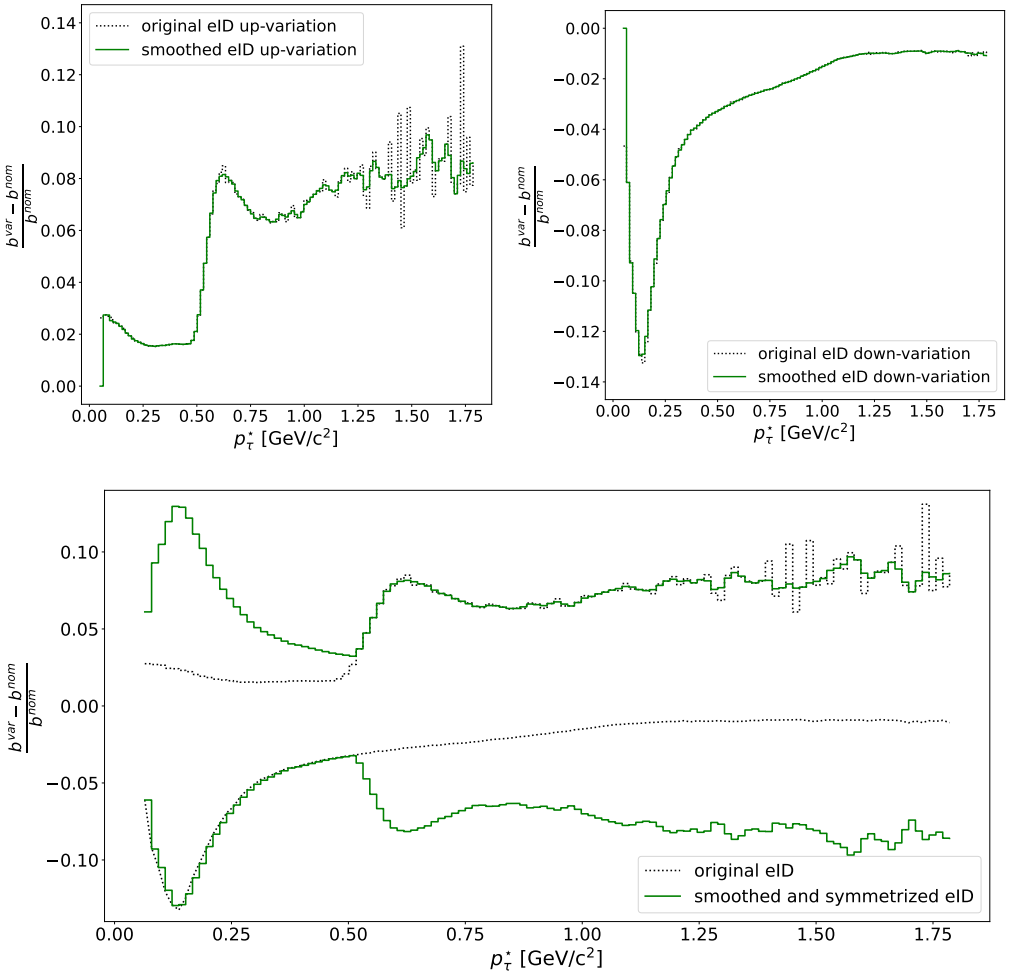


Figure 6.2: Relative variation due to *electronID* correction uncertainty; relative systematic variation before (original eID) and after the smoothing procedure (smoothed eID) for the up and down variation (top left and right). The bottom plot shows the smoothed and symmetrization variation (smoothed an symmetrized eID). We chose this particular systematic uncertainty as an example. We display how the bin entry of the *electronID* correction uncertainty (b^{var}) and the nominal template (b^{nom}) varry, normalized to the nominal templat's bin ($[b^{\text{var}} - b^{\text{nom}}]/b^{\text{nom}}$).

6.1.2 Systematic Uncertainty Due to Momentum-Scale Correction

We discussed most of the systematic uncertainties in the previous chapter, in Section 5.5, because we could evaluate them on simulated data. The only missing systematic uncertainty is the momentum-scale correction, pScale. The overall momentum scale correction applied to tracks in data (see Section 5.5.1) has an associated total uncertainty [106]. As with the previous corrections, we include this systematic uncertainty as a shape-correlated systematic uncertainty in the likelihood function. Figure 6.3 shows the resulting variations for the upward and downward variation of the momentum-scale correction as the normalized uncertainty given by:

$$\sigma_i^{\text{pScale } j} = \frac{b_i^{\text{pScale } j} - b_i^{\text{nom}}}{b_i^{\text{nom}}}, \tag{6.1}$$

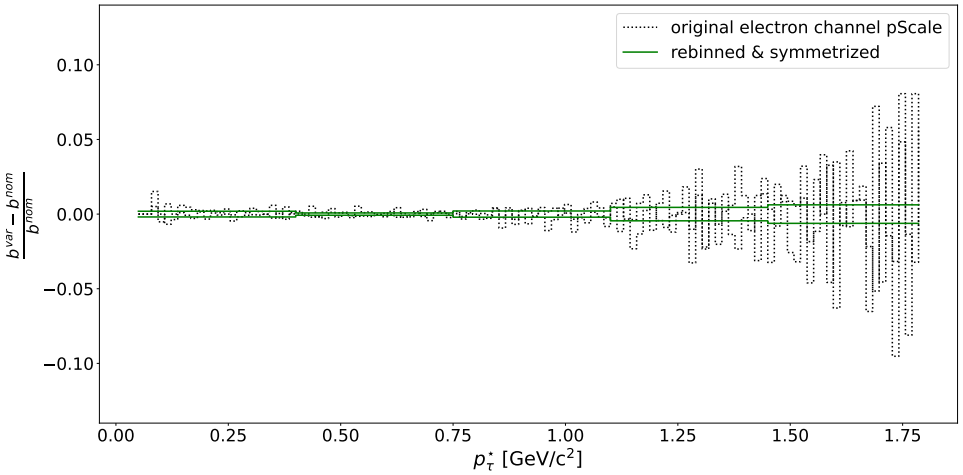
with the numer of events in the bin i, b_i; for the nominal correction, b_i^{nom}; and the variation due to the momentum scale correction, $b_i^{\text{pScale j}}$, with j indicating the upward or downward variation. We re-bin the original momentum-scale-correction uncertainty (black dotted lines) to five equally sized bins (green lines) to avoid problems in the fit. We include this normalized uncertainty in the fit model such that the fit allows for an additional momentum-scale correction-uncertainty parameter. We use the shape correlated modifier [92] to account for this systematic uncertainty in the fit model. The resulting uncertainty is below one percent, and we expect a negligible impact on the result.

We verify this expectation by comparing the fit result for μ_α either including the momentum scale uncertainty (pScale) or excluding the momentum scale uncertainty (nominal). Table 6.1 presents the fit results. We observe almost identical results except for minor variations in the

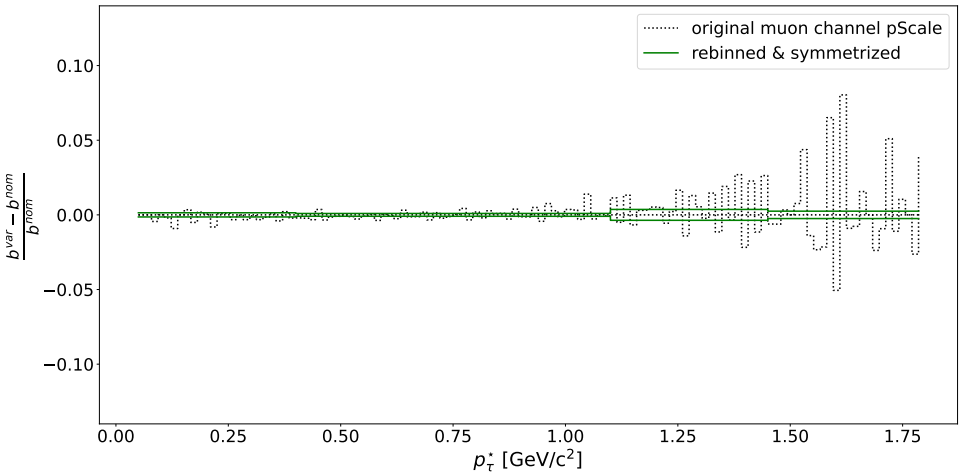
Table 6.1: μ_α values for the likelihood including the momentum scale (pScale) and excluding the momentum scale (nominal) uncertainty, obtained with pyhf.

Variant	$M_\alpha = 0$	$M_\alpha = 0.5$	$M_\alpha = 0.7$	$M_\alpha = 1.0$	$M_\alpha = 1.2$	$M_\alpha = 1.4$	$M_\alpha = 1.6$
$\tau \to e\alpha$							
$\mu_\alpha \times 10^{-4}$							
nominal	-10 ± 8	0.0 ± 9	4 ± 7	3 ± 7	-3 ± 5	-0.6 ± 2	0.4 ± 0.9
pScale	-10 ± 8	0.2 ± 9	4 ± 7	3 ± 7	-3 ± 5	-0.6 ± 2	0.3 ± 0.9
$\tau \to \mu\alpha$							
$\mu_\alpha \times 10^{-4}$							
nominal	-20 ± 7	-7 ± 7	5 ± 6	10 ± 10	-3 ± 4	1 ± 2	-0.4 ± 1
pScale	-20 ± 7	-7 ± 7	4 ± 6	10 ± 10	-3 ± 4	1 ± 2	-0.5 ± 1

last digits of the fit parameter μ_α. The results match well within their uncertainty. Furthermore, we calculate the upper limits of R for nominal and pScale and compare the results. They are almost identical, as Table 6.2 demonstrates. There are only minor, insignificant variations in the last digits. We observe a slight improvement in the upper limit, e.g., for $M_\alpha = 1.0\,\text{GeV}/c^2$. We attribute this to a similar effect as we observed



(a) Electron Channel



(b) Muon Channel

Figure 6.3: Normalized correction-uncertainty variation taken from measured data. Plot (a) shows the electron, and Plot (b) the muon channel. We display how the bin entry of the momentum-scale's correction-uncertainty (b^{var}) and the nominal template (b^{nom}) varry, normalized to the nominal templat's bin ($[b^{\text{var}} - b^{\text{nom}}]/b^{\text{nom}}$). The original variation (black dotted line) shows high statistical fluctuations because we extract the variation from measured data, which is limited to a measured data set of 62.8 fb^{-1}. We rebin and symmetrize the variation into five equally sized bins (green line). We observe that the rebinning and symmetrizing leads to a variation below one percent.

Table 6.2: Upper limit for R determined with the likelihood including the momentum scale (pScale) and excluding the momentum scale (nominal) uncertainty, obtained with pyhf.

Variant	$M_\alpha = 0$	$M_\alpha = 0.5$	$M_\alpha = 0.7$	$M_\alpha = 1.0$	$M_\alpha = 1.2$	$M_\alpha = 1.4$	$M_\alpha = 1.6$
$\tau \to e\alpha$							
UL: $R \times 10^{-3}$							
nominal	6.17	9.12	9.70	9.40	4.31	1.82	1.17
pScale	6.35	9.24	9.72	9.37	4.29	1.84	1.15
$\tau \to \mu\alpha$							
UL: $R \times 10^{-4}$							
nominal	3.10	5.66	8.70	1.65	3.83	2.37	1.06
pScale	3.29	5.70	8.54	1.63	3.96	2.40	1.05

in the previous chapter, in Section 5.5.4, when combining the two dominant systematic uncertainty sources. Here we observed an improvement in the sensitivity for the combined leptonID and trigger systematic uncertainty fit model in the electron channel.

Concluding, we observed in this subsection that the inclusion of the momentum-scale uncertainty as a nuisance parameter yields an almost identical outcome compared to excluding it. The impact is negligible because our dominant uncertainty – the leptonID – overshadows the impact of this systematic uncertainty. We, therefore, can conclude that this systematic uncertainty has a negligible impact. Hence, we drop the momentum scale uncertainty for our final evaluation.

6.1.3 Final Result

After considering the necessary changes to the systematic treatment in Section 6.1.1 and evaluating the momentum-scale-correction uncertainty in Section 6.1.2, we evaluate the final results for measured data. In subsection Fit Results we first evaluate the likelihood function distributions and the result for μ_α. Then, we compare the resulting model with the measured data and verify that no significant signal is observed. Afterward, in sub-section Upper Limit Estimation, we evaluate the 95% confidence-level upper limit. At last, in sub-section Conclusion, we summarize the results and conclude.

Fit Results

Chapter 5 stated how we approach the determination of

$$R = \frac{Br(\tau \to \ell\alpha)}{Br(\tau \to \ell\nu_\tau\nu_\ell)}$$

and the upper limit. The first step is to evaluate the resulting log-likelihood distributions. The second step is to check if a significant signal – above three standard deviations – is present.

The Frequentist pyhf determination yields well-behaved log-likelihood functions for every mass hypothesis in the maximum likelihood fit

for the treated systematic uncertainties. Figure 6.4 exemplifies the profile likelihood functions for every fit parameter with the electron channels' zero mass hypothesis, $M_{\alpha} = 0.0\,\mathrm{GeV}/c^2$. We observe a mild negative excess of the μ_{α} and mild positive excess of the μ_{SM} fit parameter. The deviation is barely above one standard deviation. Within their uncertainty, the other fit parameters correspond to their expected values.

Figure 6.5 displays the resulting model distribution for the pyhf fit. We observe a generally good capability of the template model to describe the data. When considering the ratio of the fit model and measured data, $N^{\mathrm{Data}}/N^{\mathrm{MC}}$, we observe a ratio close to one and the data points within the expected total uncertainty. These observations indicate that the model can describe the data. When subtracting the fitted standard model yield from data, $N^{\mathrm{Data}} - N^{\mathrm{MC}}_{\mathrm{SM}}$, we observe a slight excess of the data points within the total uncertainty. This excess behavior might lead to overestimating Standard Model distributions, countered by the $\tau \to \ell\alpha$ distribution, as we have observed in the resulting likelihood distributions.

We do not observe any significant signal. Table 6.3 lists the results for the fit parameter μ_{α} and the parameter of interest R, which is the ratio of the $\tau \to \ell\alpha$ branching ratio and the $\tau \to \ell\nu_{\tau}\nu_{\ell}$ branching ratio calculated according to Equation (5.5), for the pyhf fit results. We calculate the uncertainty of R with error propagation and the correlation matrix resulting from the fit. All considered mass hypotheses' branching ratios are equivalent to zero, within two standard deviations of their respective uncertainty. We observe no evidence of the $\tau \to \ell\alpha$ decay; hence we will determine a 95% confidence upper limit.

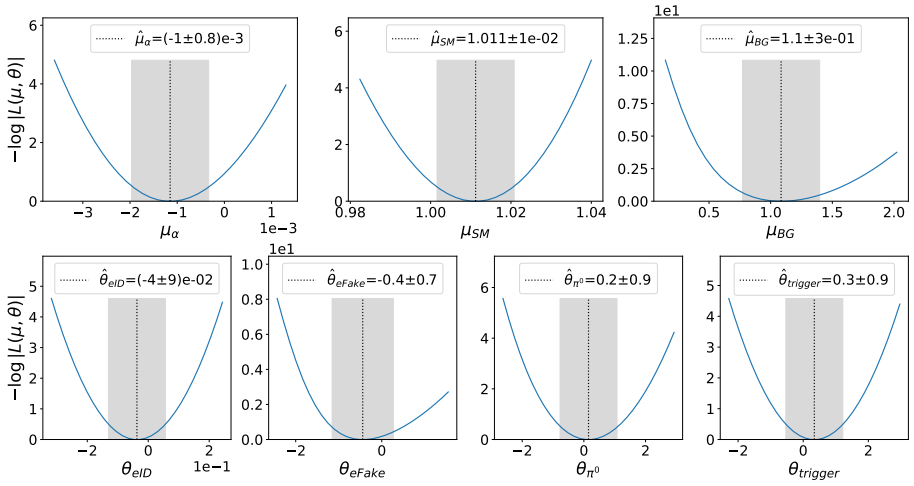


Figure 6.4: Profile liklihood (blue) for all model parameters with the corresponding best fit value (black) and error (gray) for the mass hypothesis $M_{\alpha} = 0.0\,\mathrm{GeV}/c^2$ of the electron channel determined on measured data with pyhf.

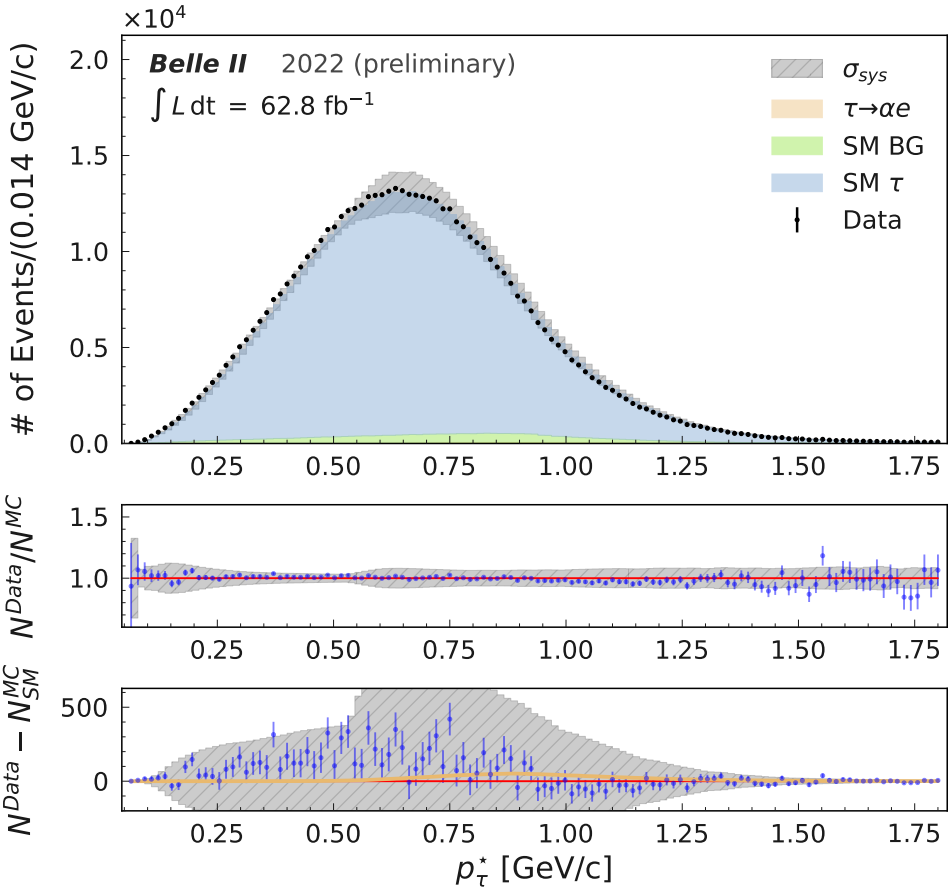


Figure 6.5: Best-fit distribution for the electron channel's $M_{\alpha} = 0$ hypothesis determined on measured data. The top plot shows the actual distribution with the systematic uncertainty (σ_{sys}), the distribution for the $\tau \to \ell\alpha$ hypothesis ($\tau \to \alpha e$), the fit model's background distribution (SM BG), and the distribution for the $\tau \to e\,\nu_\tau\,\nu_e$ (SM τ). The two lower plots allow evaluating the measured data (Data) distribution (black dots) with the templates. They show the ratio of measured data to fit model ratio (blue dots) with the statistical uncertainty (blue error bar) in the central plot, and the measured data subtracted with the Standard Model distributions (blue dots) with the statistical uncertainty (blue error bar) in the bottom plot. The red line in the middle plot is the expected ratio for successfully modeling the data. The red line in the bottom plot is the H_0 expectation. The bottom plot also shows the best fit $\tau \to \ell\alpha$ distributions, which allows comparing this to the subtracted measured data.

Table 6.3: Final μ_α and R values obtained with pyhf on measured data. The parameter of interest R is the ratio of branching ratios according to Equation (5.5). The uncertainty is determined with error propagation and the resulting correlation matrix from the fit.

Variant	$M_\alpha = 0$	$M_\alpha = 0.5$	$M_\alpha = 0.7$	$M_\alpha = 1.0$	$M_\alpha = 1.2$	$M_\alpha = 1.4$	$M_\alpha = 1.6$
$\tau \to e\alpha$							
$\mu_\alpha \times 10^{-4}$	-10 ± 8	0.0 ± 9	4 ± 7	3 ± 7	-3 ± 5	-0.6 ± 2	0.4 ± 0.9
$R \times 10^{-3}$	-6 ± 5	$0. \pm 5$	2 ± 4	2 ± 4	-2 ± 3	-0.3 ± 1	$0.2 \pm 0..5$
$\tau \to \mu\alpha$							
$\mu_\alpha \times 10^{-4}$	-20 ± 7	-7 ± 7	5 ± 6	10 ± 10	-3 ± 4	1 ± 2	-0.4 ± 1
$R \times 10^{-3}$	-10 ± 4	-4 ± 4	3 ± 3	6 ± 6	-2 ± 2	0.6 ± 1	-0.2 ± 0.6

Upper Limit Estimation

Since we found no signal, we determine a 95% confidence-level upper limit. By determining this upper limit, we want to find the maximum strength of a possible $\tau \to \ell\alpha$ decay given our measured data. The upper limit is determined using the CL_S method. Figure 6.6 shows an example of the observed CL_S (black line) and the corresponding CL_{S+B} (blue line) and CL_B (orange line) curves for the electron channel's $M_\alpha = 0$ hypothesis using pyhf. As described in Chapter 5, Section 5.2, we determine the p-value at many fixed μ_α for the signal + background, $S + B$, and background hypothesis, B. We denote respective p-value curves for the $S + B$ hypothesis, CL_{S+B}, and for the B hypothesis, CL_B. We approximate the μ_α dependent confidence level for the signal, CL_S, as the ratio of CL_{S+B} and CL_B. We expect that CL_{S+B} and CL_B have the same value for $\mu_\alpha = 0$ because both hypotheses are the same. In Figure 6.6, we observe a decreasing CL_{S+B} for increasing μ_α, which means the $S + B$ hypothesis becomes less likely. If no signal is present, we expect the CL_{S+B} to converge towards zero. We expect the CL_B curve to be constant for no signal in data. The observed behavior of CL_{S+B} and CL_B leads to a decreasing CL_S, and we set the 95% confidence-level upper limit of μ_α (red line) at $CL_S = 0.05$. The dotted line is the expected CL_S, with the respective uncertainty as one and two σ deviations in green and yellow (Brazilian band), respectively.

We determine this result with the asymptotic method using the asymptotic CL_S calculation in pyhf. It is based on Wilks's theorem, which assumes that all test statistics in this evaluation are distributed according to a χ^2-distribution. We double-check with toy-based calculations, which generate the test statistics with many toy experiments, and they verify the result. We obtained almost identical results for the CL_S curves and upper limits with the toy-based calculations; hence they are not included in this thesis. Appendix G.4 lists the summary of all log-likelihood distributions, the fit distributions, and the CL_S curves calculated with the asymptotic CL_S calculation.

After evaluating the upper limit setting procedure with the above-given example, we now discuss the results for all $\tau \to \ell\alpha$ mass hypotheses tested with the measured Belle II data. First, we present the approved

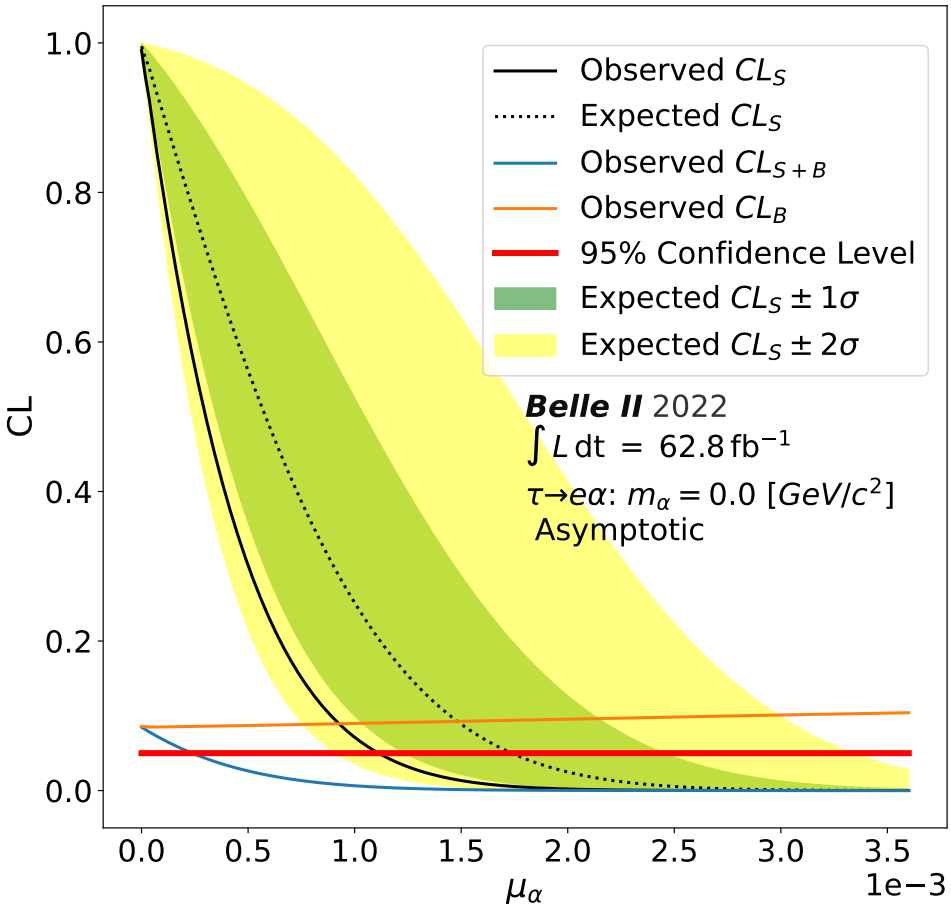


Figure 6.6: The CL_S curve scan distribution for the electron channel's $M_\alpha = 0$ hypothesis determined on measured data. The *Observed* curves result from the evaluated measured data set. CL_S is the ratio of p-values for the H_0, denoted as CL_B, and the H_S, denoted as CL_{S+B}. The test statistic distributions give the *Expected* CL_S curve. The green and yellow bands denote the 1σ and 2σ variation of the Expected CL_S curve, respectively.

Belle II result for the upper limit of R, calculated with the RooStats implementation using the accepted Belle II search strategy[2]. Second, we compare these results to the previous most stringent results provided by the ARGUS collaboration. Third, we compare the RooStats results to the pyhf results determined in this thesis to check the robustness of the results.

Table 6.4 lists the numerical values of our approved Belle II result for the upper limit intervals of R for all mass hypotheses calculated with RooStats.

Table 6.4: Offical Belle II upper limit values of R for the final result obtained with RooStats on measured data.

Channel	$M_\alpha = 0$	$M_\alpha = 0.5$	$M_\alpha = 0.7$	$M_\alpha = 1.0$	$M_\alpha = 1.2$	$M_\alpha = 1.4$	$M_\alpha = 1.6$
$\tau \to e\alpha$	0.0053	0.0078	0.0090	0.0097	0.0045	0.0018	0.0011
$\tau \to \mu\alpha$	0.0034	0.0062	0.009	0.0122	0.0036	0.0025	0.0007

Figure 6.7, compares our approved Belle II result to the previous most stringent upper limit provided by the ARGUS collaboration [36]. In the electron channel (top), we observe an improvement of about one-third for most mass hypotheses by Belle II. In the muon channel (bottom), we observe an improvement of about one-fifth for most hypotheses by Belle II.

Table 6.5 lists the numerical values of the upper limit intervals of R for all mass hypotheses obtained in this thesis with pyhf. We can compare the observed upper limit results for RooStat and pyhf. In the case of the electron channel, the upper limit distribution varies up to 15 %. In the case of the muon distribution, we observe a variation of up to 36%. Figure 6.8 allows us to put the observed deviations of the measured 95% confidence-level upper limits calculated via RooStats and pyhf into perspective. The figure compares the observed upper limits from RooStats and pyhf with the pyhf-calculated expected upper limit and uncertainty from the fit model. The pyhf and RooStats results are similar within the upper limit's expected uncertainty. We can attribute the observed deviations to the differences in the modeling. The RooStat implementation extracts R directly from the fit and CL_S calculation whereas the pyhf implementation determines $\mu_{\boldsymbol{\alpha}} = Br(\boldsymbol{\tau \to \ell\alpha})$ and calculates R according to Equation (5.10) using the best fit value of μ_{SM}.

Table 6.5: Upper limit values of R for the final result obtained with pyhf on measured data.

Variatione	$M_\alpha = 0$	$M_\alpha = 0.5$	$M_\alpha = 0.7$	$M_\alpha = 1.0$	$M_\alpha = 1.2$	$M_\alpha = 1.4$	$M_\alpha = 1.6$
$\tau \to e\alpha$	0.0062	0.0091	0.0097	0.0094	0.0043	0.0018	0.0012
$\tau \to \mu\alpha$	0.0031	0.0057	0.0087	0.0165	0.0038	0.0024	0.0011

[2]The Belle II collaboration chose this implementation as the to be published result. The publication is currently in preparation.

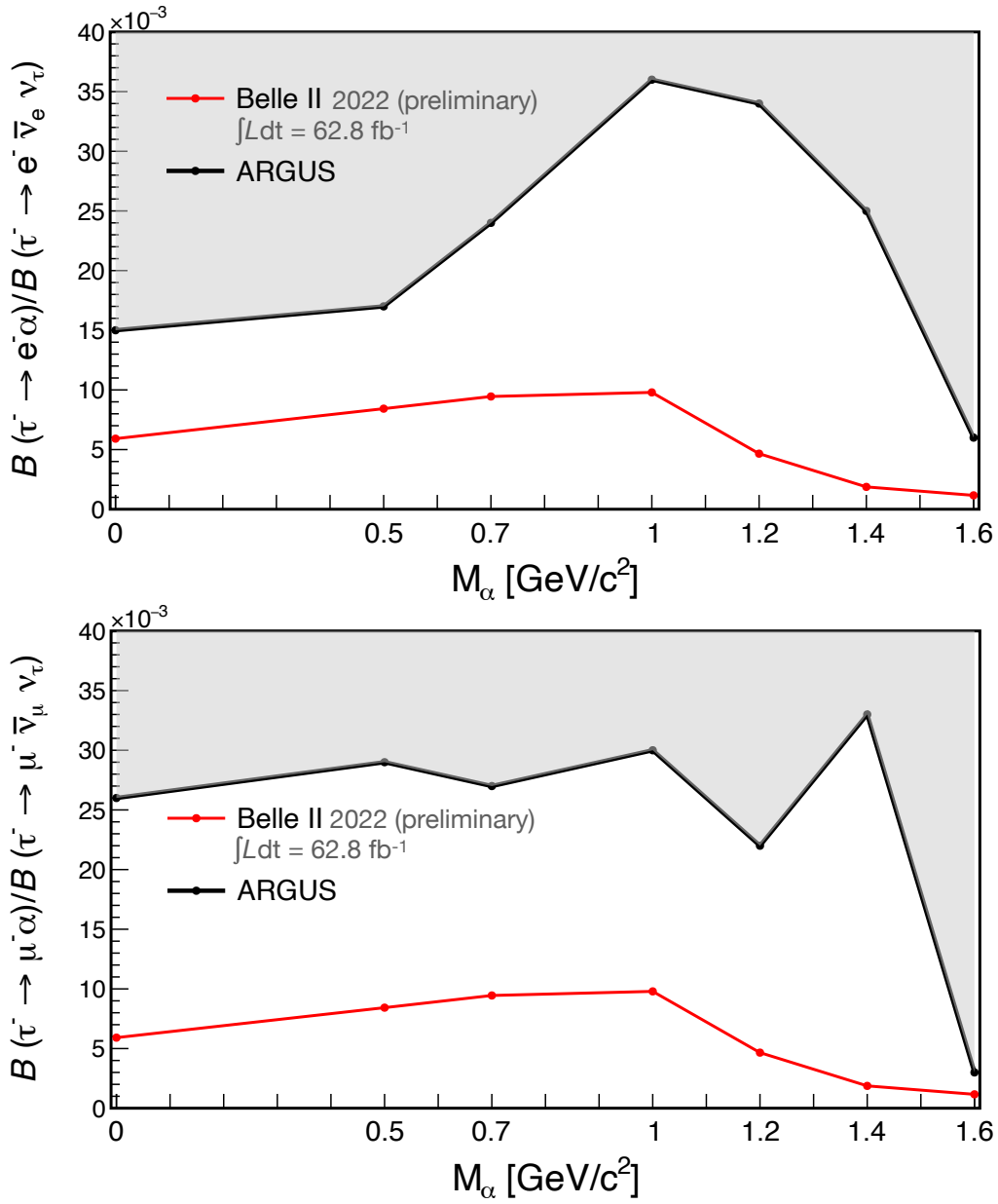


Figure 6.7: Offical Belle II 95% upper limit result of the ratio for branching ratios R for the electron (top) and the muon (bottom) channels, with the previous result from the ARGUS collaboration as a comparison.

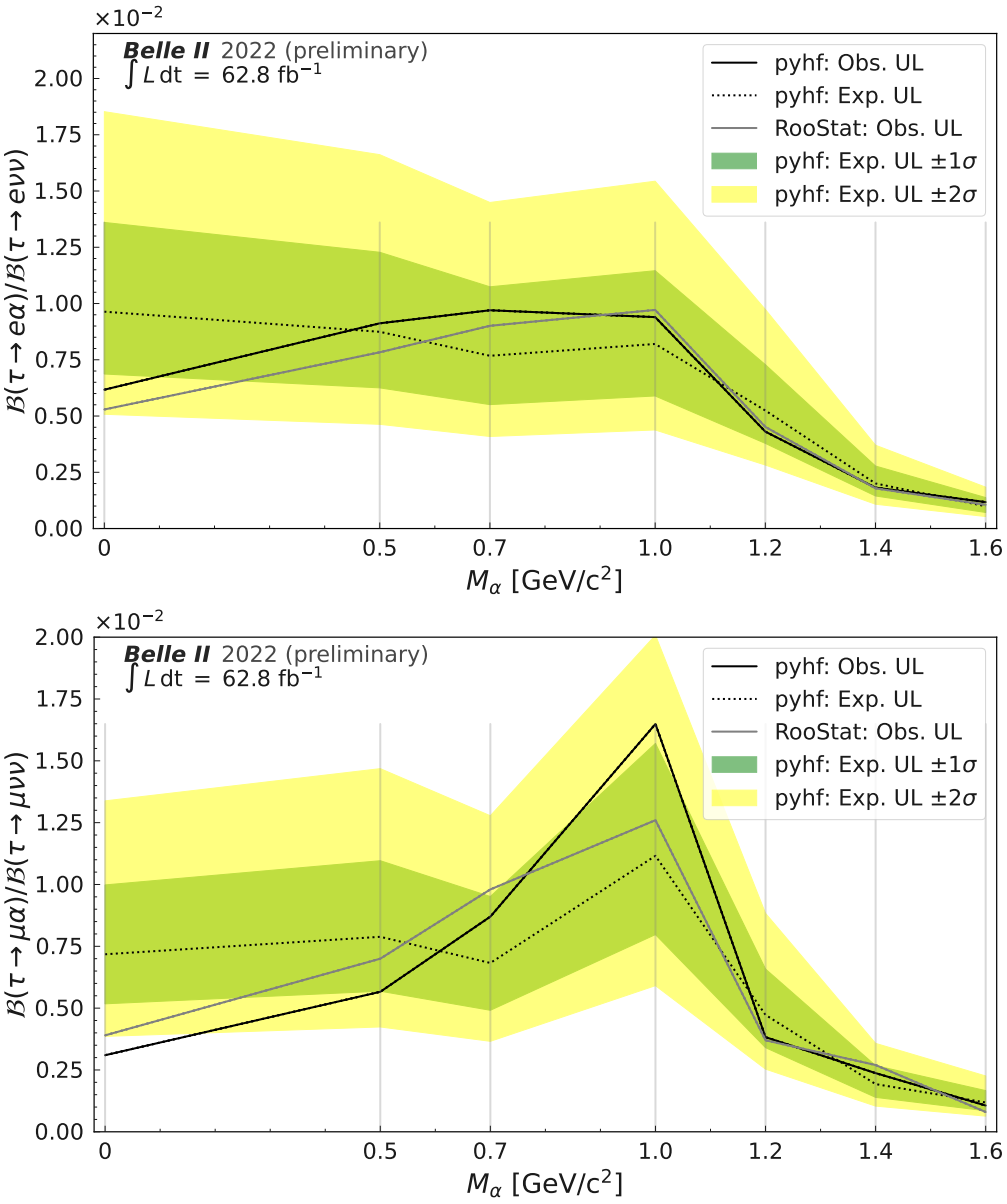


Figure 6.8: Comparison of the pyhf and RooStat upper limit result for measured data (Obs. UL) of the ratio for branching ratios R for the electron (top) and the muon (bottom) channels, with the pyhf expectation (Exp. UL) and variance for one (Exp. UL $\pm 1\sigma$) and two (Exp. UL $\pm 2\sigma$) standard deviations. For each channel we evaluate $M_\alpha = \{0.0, 0.5, 0.7, 1.0, 1.2, 1.4, 1.6\}\,[\mathrm{GeV}/c^2]$.

Conclusion

We determined R for seven mass hypotheses in the electron and muon channel. As we neither find a 5σ nor 3σ deviation, we report no evidence for the $\tau \to \ell\alpha$ decay process and determined a 95% confidence-level upper limit using a Frequentist CL_S method. The results of all implementations of our $\tau \to \ell\alpha$ search at the Belle II-detector are similar, setting the most stringent 95% confidence-level upper limit to date. We update the ARGUS collaboration's previous result by two to fourteen. The size of the data set mainly limits our search. In addition, the systematic uncertainties introduced by the leptonID also have a considerable impact. A future iteration of the $\tau \to \ell\alpha$ search at Belle II will profit from the increased available data set.

6.2 Event Selection and τ-Rest Frame Study Results

This section evaluates the alternative selection variants and τ-rest frame estimation methods we introduced in this thesis. The approved Belle II $\tau \to \ell\alpha$ search strategy is the reference point for the following results. We calculate the upper limit of the parameter of interest R to estimate the sensitivity of our $\tau \to \ell\alpha$ search. We either change the selection variant or the τ-rest-frame method, to calculate this upper limit. Then, we compare the results to determine the most sensitive search strategy.

Section 6.2.1 focuses on the comparison of the selection variants introduced in Chapter 3. The approved Belle II event selection [107] for this $\tau \to \ell\alpha$ search is the cut-based event selection rejecting all reconstructed neutral particle candidates and results in a high purity above 90%. In contrast to the approved Belle II selection, our alternative selection variants allow for neutral particle candidates. We find that the cut-based selection with neutrals yields a substantial increase in available data (it more than doubles) but suffers in purity. The BDT-based selection with neutrals shows a similar increase in data but can achieve the same purity as the approved event selection. We test the respective sensitivity for each selection variant and evaluate if focus on purity leads to the highest sensitivity.

Section 6.2.2 compares the τ-rest frame methods described in Chapter 4. As described in Chapter 5, we use the CL_S approach implemented in pyhf and the Bayesian interference approach implemented in Julia BAT, to determine the upper limit of R. These two statistical approaches calculate the upper limit differently. We use them to provide a full picture of the τ-rest frame methods' sensitivity.

In order to reconstruct the τ-rest frame, we utilize the hadronic tag-side 3-prong-τ decay, which allows us to approximate the τ-rest frame with a pseudo-rest frame. The ARGUS method is the approved pseudo-rest frame reconstruction used in the approved Belle II $\tau \to \ell\alpha$ search

strategy. It approximates the τ flight direction via the flight direction of the three-hadron system. We developed two more rest-frame estimation methods. They include more event information, such as the momentum direction of the signal-τ side. The first developed method is the Thrust method, which approximates the τ flight direction with the *thrust* direction. The second is the GKK method, which incorporates the available neutrino information and provides an estimate through the limiting distribution function. We find no significant improvement to the ARGUS method for the $\tau \to \ell\alpha$ search. We observe that the GKK method systematically performs worse than the ARGUS method, which indicates that the GKK method is not the best choice for a search.

6.2.1 Event Selection Study Results

In Chapter 3 we introduced three selection variants, which yield different results. They are all based on the same variables and only differ in the background suppression method. Two variants are cut-based selections based on one-dimensional cuts determined by figure-of-merit optimizations. The first variant is the cut-based selection which rejects reconstructed neutral particle candidates, and the second allows for these neutrals and optimizes the number of neutral candidates on the tag-side τ decay.

The other variant is a BDT trained with all figure-of-merit-optimized variables. Here, we first apply all cuts determined without a figure of merit optimization. Afterward, we train the BDT. We tried to ensure good comparability of the selection variants; hence we sacrificed performance to conformity.

This section compares the three variants' upper limit of R to each other to estimate their respective sensitivity. The BDT-based selection with neutrals is such that it emulates the purity of the cut-based selection without neutrals. The cut-based selection without neutrals is the approved selection strategy for the Belle II $\tau \to \ell\alpha$ search. The selection study determines the upper limit results from 100 toy data sets using the ARGUS τ-rest-frame method and compares the mean and standard deviation of the upper limit results. This study uses 100 toy data sets because this number gives us sufficient statistical precision.

Figure 6.9 presents the results of the upper limit of R for the electron (top) and muon channel (bottom) and their respective seven mass hypotheses for α, M_α. We observe a lower upper limit for the cut-based selection variant allowing for neutrals (Cut with Neutrals) and the BDT (BDT with Neutrals) compared to our approved $\tau \to \ell\alpha$ search selection (Cut no Neutrals). The upper limit improvement is most significant for lower α mass hypothesis, M_α between $0.0\,\mathrm{GeV}/c^2$ and $0.7\,\mathrm{GeV}/c^2$. In this range, we observe that the gain in sensitivity of the BDT with neutrals (BDT with Neutrals) corresponds to the increased data set of $\sim \sqrt{2}$. The cut-based selection with neutrals (Cut with Neutrals) displays the low-

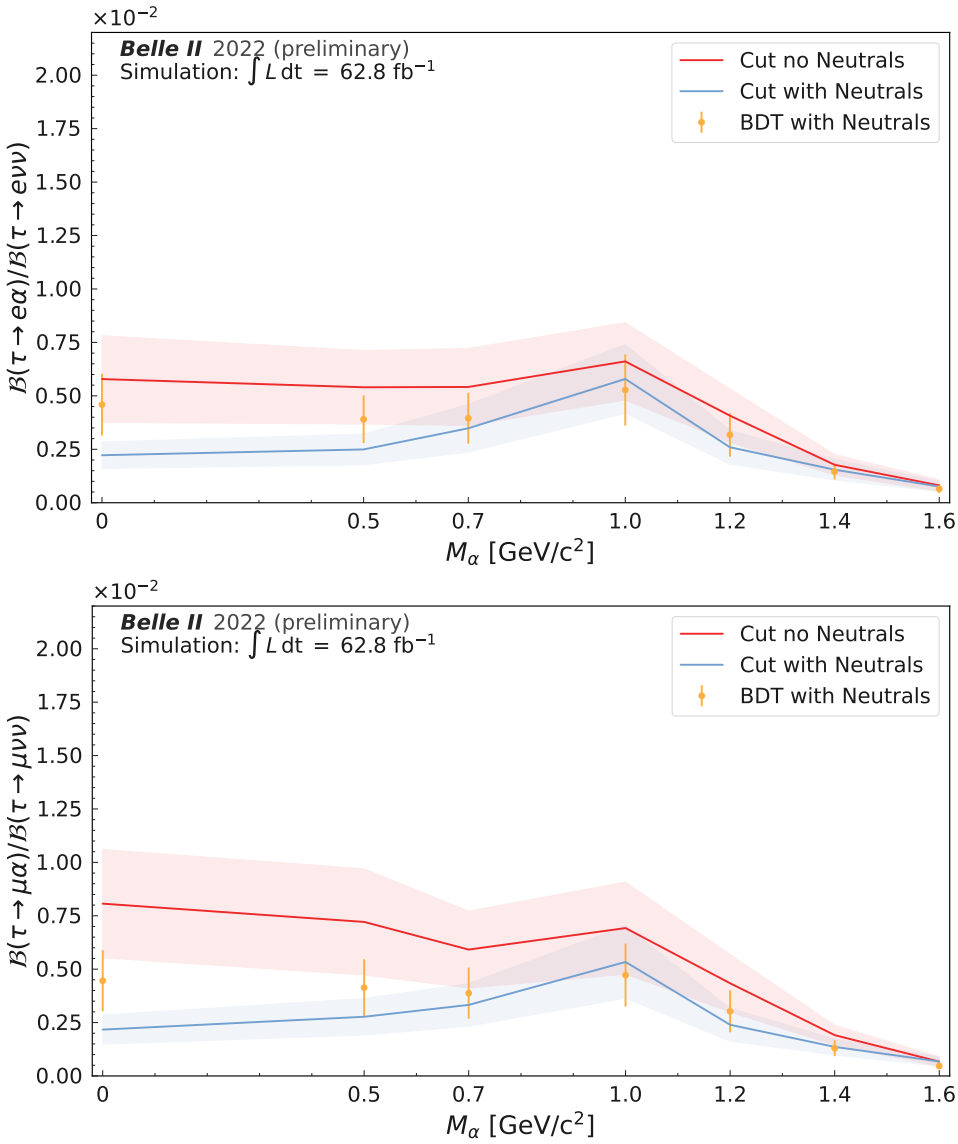


Figure 6.9: Comparison of the 95% confidence-level upper limit of $R = Br(\tau \to \ell\alpha)/Br(\tau \to \ell\nu_\tau\nu_\ell)$ results for the selection variants described in the text. We consider the cut-based selection without neutrals (Cut no Neutrals), the cut-based selection which allows for neutral candidates (Cut with Neutrals), and a BDT trained with the figure-of-merit optimized variables of the cut-based selection which allows for neutral candidates (BDT with Neutrals). The electron channel results are in the top plot, and the muon channel results are in the bottom. For each channel we evaluate $M_\alpha = \{0.0, 0.5, 0.7, 1.0, 1.2, 1.4, 1.6\}[\mathrm{GeV}/c^2]$. A line or dot in the plot indicates the mean upper limit. The lines show an interpolation of the upper limit between the evaluated mass hypothesis. A band, dotted lines, or an error bar indicate the standard deviation of 100 upper limits determined from toy data.

est upper limit in this range. For all selection variants, we observe that the sensitivity gain is antiproportinal to M_{α} in the range of $0.0\,\mathrm{GeV}/c^2$ to $1.0\,\mathrm{GeV}/c^2$. At $M_{\alpha} = 1.0\,\mathrm{GeV}/c^2$, we observe a decreased sensitivity gain for the selection methods allowing for neutrals. We observe a gain of sensitivity at $M_{\alpha} = 1.2\,\mathrm{GeV}/c^2$. The selection variant's upper limit results converge for the higher mass hypotheses, indicating a decreasing dependence on the selection method for increasing M_{α} above $1.0\,\mathrm{GeV}/c^2$.

We can attribute the sensitivity loss at $M_{\alpha} = 1.0\,\mathrm{GeV}/c$ to the coinciding of the background and $M_{\alpha} = 1.0\,\mathrm{GeV}/c$ distribution in $p_{\tau}^{\star}$. Here we observe that all selection variants converge towards the approved Belle II selection.

The results of the selection study show that we gain sensitivity when allowing for neutral particle candidates. The increased data set can explain the increased sensitivity. The results of the BDT with neutrals compared to the cut-based selection with neutrals for M_{α} below $1.0\,\mathrm{GeV}/c^2$ show that optimizing for high purity is not ideal. The BDT optimization sacrificed data for purity; we choose a value of the BDT classification to reflect the purity of the approved Belle II selection for the $\tau \to \ell\alpha$ search. This optimization results in a lower sensitivity than the more straightforward, less pure cut-based selection. If we optimize the selected BDT classification output value with a figure-of-merit optimization, we will probably gain sensitivity, as the ROC-curve in Chapter 3 indicates. The BDT allows for a better trade between purity and available data.

Furthermore, we observe that the high M_{α} are less sensitive to the optimization choice. The higher M_{α}, the less the $\tau \to \ell\alpha$ distributions in $p_{\tau}^{\star}$ coincide with the $\tau \to \ell\nu_{\tau}\nu_{\ell}$ and background distributions, effectively decoupling the sensitivity from the background dependence observed in lower mass hypothesis. The observed upper limit reflects this observation in the $\tau \to \ell\alpha$ distribution.

In this section, we observe that the selection variants allowing for neutrals show a lower upper limit and hence a higher sensitivity than our approved Belle II strategy. This result indicates that in the future, we should not focus on purity but instead follow the FOM optimization given by Equation (3.1). By combining machine learning methods, such as the BDTs considered in this thesis, with a FOM-based optimization for the BDT output, we can expect a significant increase in sensitivity due to an optimized selection. A crucial requirement for this sensitivity improvement by the BDT is that Belle II improves its understanding of the detector and decreases systematic uncertainties. Aside from the side-band argument, we also reject neutral particles because it simplifies the treatment of systematic uncertainties related to these particles. As we have seen in the previous section, we are already limited to a considerable degree by the systematic uncertainties. If the sensitivity gained by increasing the data set is countered by introducing a new systematic effect, we do not gain an advantage.

6.2.2 τ-Rest Frame Study Results

We need to boost into the τ-rest frame to search for the $\tau \to \ell\alpha$ decay. For $\tau^+\tau^-$ events, we can use the tag-τ side to reconstruct the rest frame of the signal τ. Due to the ν_τ, we cannot reconstruct the actual τ-rest frame, but we can approximate it. We call this approximated-τ-rest frame pseudo-rest frame. In Chapter 4 we discussed three methods to determine the pseudo-rest frame. All methods rely on the approximation $E_\tau = \sqrt{s}/2$. The crucial difference between the methods is their respective direction estimate of the τ's flight direction.

The ARGUS collaboration introduced the ARGUS method. It is the approved method to determine the pseudo-rest frame in Belle II's $\tau \to \ell\alpha$ search. The ARGUS method approximation demands a 3-prong decay with hadrons on the tag-τ side. The ARGUS method approximates the hadronic system's flight direction as the direction of the τ.

The Thrust method uses the thrust direction to approximate the τ's flight direction. This approach also utilizes signal-side information. In Chapter 4 we observed a systematic shift impacting precision measurements, but we also observed that the Thrust method's resulting distributions are narrower, which might lead to an increased sensitivity for our $\tau \to \ell\alpha$ search.

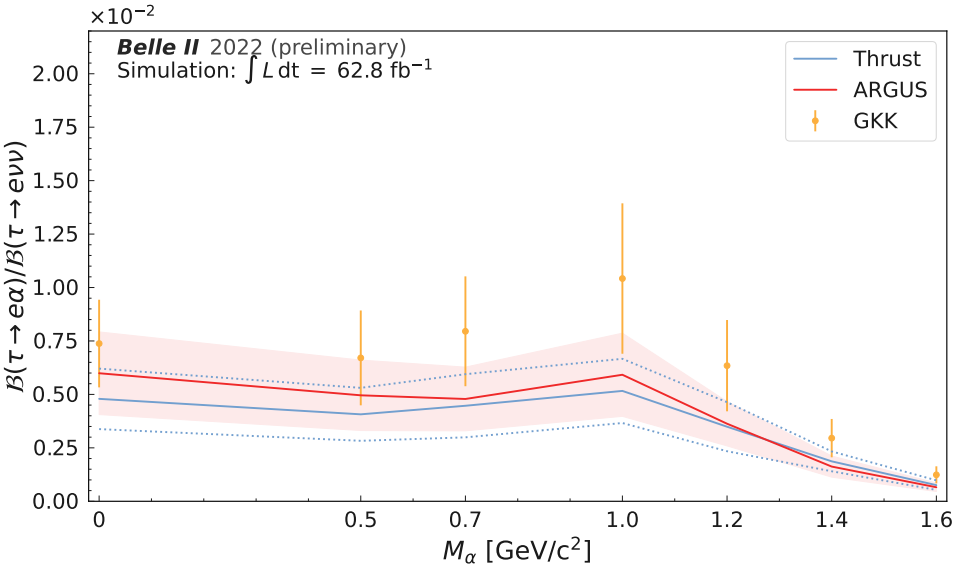
The GKK method includes the available information of the ν_τ and generates a probability density function per event. By stacking the probability density functions of many events on top of each other, we obtain a limiting function that mitigates the effect of the ν_τ in the observed effect in the ARGUS method. We find a significant shape difference between the resulting $\tau \to \ell\alpha$ and $\tau \to \ell\nu_\tau\nu_\ell$ decay distributions, and we observe no signs of a bias.

In the following, we compare the sensitivity of the $\tau \to \ell\alpha$ search using the three different τ-rest-frame methods. This consideration focuses on the ability to detect a new physics signal but neglects the possible capabilities of precision measurements.

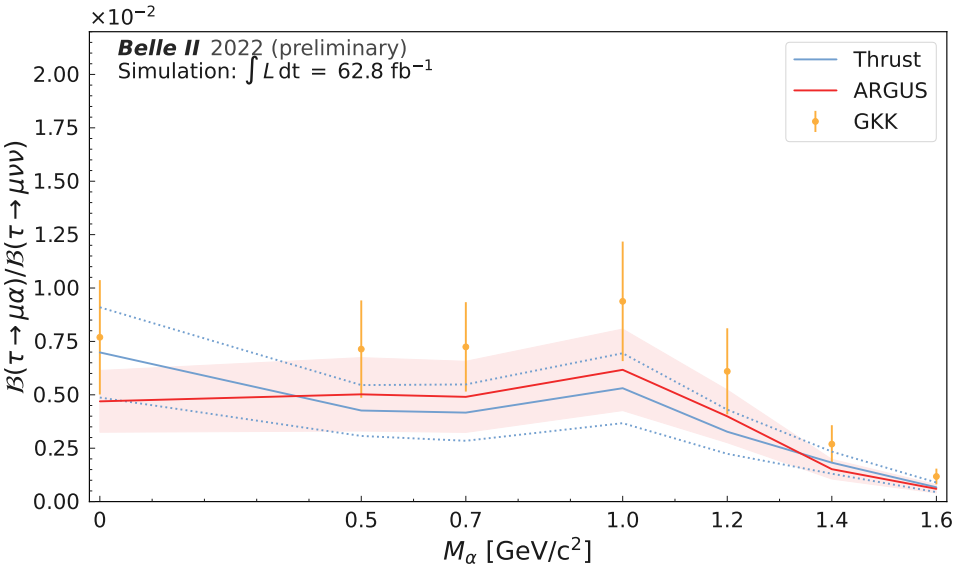
This section compares the results with the CL_S implementation in pyhf and the Bayesian interference implementation in Julia BAT to double-check the results with a second, statistically independent method[3].

Figure 6.10 presents the pyhf results of the upper limit for R in the electron (6.10a) and muon (6.10b) channel for seven mass hypothesis of α, M_α. We observe no immediate advantage for the sensitivity with any of the new methods. In the electron channel, we observe that the τ rest-frame estimate with the consistently highest upper limit of R is the GKK method (GKK). The ARGUS method (ARGUS) shows an advantage compared to the GKK method. The Thrust method (Thrust) results in the mildly lower upper limit result for M_α below $1.4\,\mathrm{GeV}/c^2$. For equal or

[3]We did not consider the Julia BAT implementation in Section 6.2.1 because the statistically independent samples of the electron and muon channel show the same behavior for the different selection variants, rendering a robustness check with Julia BAT implementation unnecessary.



(a) Electron channel results.



(b) Muon channel results.

Figure 6.10: Upper limit comparison for different τ rest-frame approaches of Chapter 4 obtained with pyhf. The electron channel results are in Plot (6.10a), the muon channel results are in Plot (6.10b). For each channel we evaluate $M_\alpha = \{0.0, 0.5, 0.7, 1.0, 1.2, 1.4, 1.6\}\,[\mathrm{GeV}/c^2]$. A line or dot in the plot indicates the mean upper limit, and the lines show an interpolation of the upper limit between the evaluated mass hypothesis. A band, dotted lines, or an error bar indicate the standard deviation of 100 upper limits determined from toy data.

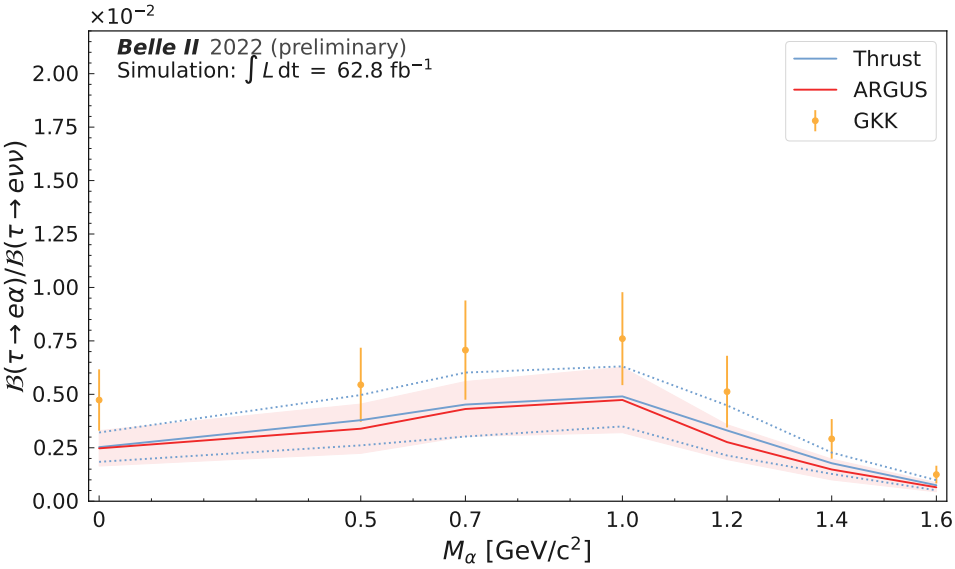
higher M_{α} we observe that the ARGUS method results in a slightly lower upper limit than the Thrust method. When using the GKK method, we observe the same behavior in the muon as in the electron channel. The results are consistently above the ARGUS method's mean upper limit results. When we compare the Thrust method with the ARGUS method, we also observe a similar behavior of the upper limit results as we did in the electron channel. The exception is $M_{\alpha} = 0.0\,\mathrm{GeV}/c^2$. Here, the Thrust method results in a slightly lower upper limit than the GKK method and a considerably higher upper limit than the ARGUS method.

The overall trend of the pyhf results agrees with the Julia BAT results for both the flat and the log uniform prior. Figure 6.11 presents the Julia BAT result for the flat prior, with the upper limit for R in the electron (6.11a) and muon (6.11b) channel. In both channels, the GKK method shows a consistently higher upper limit of R than the ARGUS and Thrust method. Although we observe differences in the upper limit result for the ARGUS and Thrust method, both show similar performance and are consistent within their respective uncertainty. We observe no advantage for one of the new pseudo-rest frame methods for the $\tau \to \ell\alpha$ search. The Julia BAT result for the log uniform prior agrees with the flat prior results, as Figure G.20 in Appendix G shows.

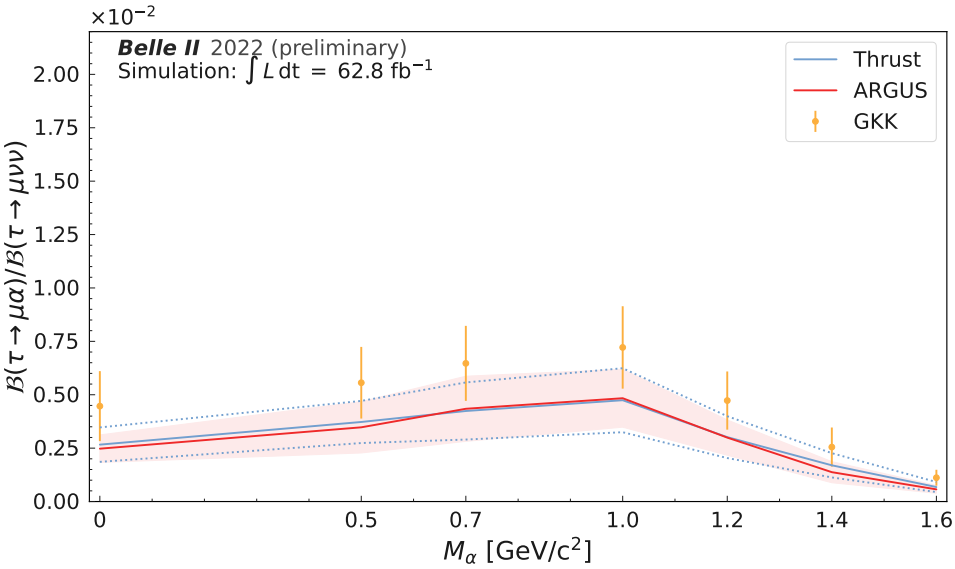
The ARGUS method displays a surprisingly good performance – compared to the other methods – considering its discussed shortcomings in Chapter 4, especially the broad smearing for the low mass hypothesis. In contrast, the GKK method does not improve the sensitivity, and its performance is worse in all upper limit estimations. The Thrust method shows a mild but consistent lower upper limit for α mass hypothesis between $0.5\,\mathrm{GeV}/c^2$ and $1.2\,\mathrm{GeV}/c^2$ in the upper limit results obtained with the pyhf implementation. We observe a less clear picture for the Thrust method with the Julia BAT implementation than the pyhf implementation. When comparing the ARGUS and Thrust method, we observe the Thrust method's mean fluctuating around the ARGUS method's mean. The mean values are close, and the methods are consistent when considering the respective upper limit uncertainty.

We attribute the differences between the two implementations in the upper limit results to the different statistical methods. In pyhf, we use the CL_S scan method to determine the upper limit. Based on Wilks's theorem, it estimates the test statistic distributions as χ^2-distributions, allowing for an analytical estimate. In contrast, Julia BAT samples the posterior likelihood distribution.

The standard deviations of the upper limits for all methods overlap, leading to no significant advantage for any of the methods. The GKK method's upper limits are higher than the ARGUS method's upper limits, which means the GKK method's sensitivity is consistently worse in the respective result compared to the ARGUS method. The Thrust method's upper limit is slightly lower in the pyhf results, which might indicate a slightly higher sensitivity, but we observe an equivalent up-



(a) Electron channel results.



(b) Muon channel results.

Figure 6.11: Upper limit comparison for different τ Rest-Frame Approaches of Chapter 4 for Julia BAT using the flat prior. The electron channel results are in Plot (6.11a), the muon channel results are in Plot (6.11b). For each channel we evaluate $M_\alpha = \{0.0, 0.5, 0.7, 1.0, 1.2, 1.4, 1.6\}\,[\mathrm{GeV}/c^2]$. A line or dot in the plot indicates the mean upper limit, and the lines show an interpolation of the upper limit between the evaluated mass hypothesis. A band, dotted lines, or an error bar indicate the standard deviation of 100 upper limits determined from toy data.

per limit with an upwards tendency in the Julia BAT results, which does not confirm this indication. On top of the unclear sensitivity gain by the Thrust method, we discussed in Chapter 4 its shortcomings due to a systematic shift.

The above considerations lead to a clear conclusion. In our $\tau \to \ell\alpha$ search, the ARGUS method is the best-evaluated strategy to estimate our upper limit of R. We observe that the ARGUS method shows good sensitivity without biasing the result in a one-dimensional fit. The Thrust method does not show any clear advantage and biases the signal; we find it is not a good τ-rest frame estimate. The GKK method shows lower sensitivity than the ARGUS method; it is not ideal for a search. However, in Chapter 7 we show that it is a suitable choice for precision measurements.

In a future iteration, new ideas for sensitive variables that show shape differences may improve the sensitivity of the $\tau \to \ell\alpha$ search. Chapter 7 presents some ideas for new variables and methods. If we observe a signal, the analysts should reevaluate the GKK method to precisely determine M_α.

Part IV

Outlook and Conclusion

Chapter 7

Outllok for Potential Future Improvements and Applications of Developed Techniques

This chapter intends to provide some prospects for $\tau \to \ell\alpha$ and τ mass measurements exploiting techniques described in this thesis, and parallel developments this thesis could not include in its discussion. First, we discuss the prospects of the $\tau \to \ell\alpha$ search. Afterward, we discuss the possible application of the GKK method in a maeasurement of the τ mass.

7.1 Improvements for Future $\tau \to \ell\alpha$ Searches

This section briefly discusses improvements a future iteration of a $\tau \to \ell\alpha$ search can implement. A better understood Belle II detector will lead to an improved $\tau \to \ell\alpha$ sensitivity. Starting with the selection, the search profits from a better understanding by enabling the usage of BDT-based selections. As demonstrated, this selection method increases the remaining data set after selection by approximately a factor of two. Also, a BDT optimization would allow for a reasonable 1×1-prong topology identification, adding a new channel to the search. Another aspect is considering mass hypothesis-dependent selections, which could improve the available data per tested hypothesis and improve sensitivity further.

Furthermore, a better-understood detector leads to improved particle ID variables; for example, BDT-based particle ID shows a better identification of lepton fakes, which decreases the background yield. The treatment of the leptonID at Belle II is one of our main constraints. Significantly reduced systematic uncertainties will improve the total leptonID uncertainty and benefit the $\tau \to \ell\alpha$ sensitivity. Furthermore, a future $\tau \to \ell\alpha$ search should reevaluate the systematic uncertainty treatment. For example, by treating the systematic and statistical uncertainties of the leptonID seperately.

The 1×1-prong topology also necessitates a better rest-frame esti-

mate. The GKK method is a possible candidate, but other exciting ideas are also worth evaluating. Guadagnoli, Park, and Tenchini propose the M_2 variable for the search [108]. Furthermore, using all variables that display a shape difference, machine learning might be able to differentiate a $\tau \to \ell\nu_\tau\nu_\ell$ from a $\tau \to \ell\alpha$ decay. Recently a Punzi-loss bassed neural network, Punzi-net, was proposed [109]. The Punzi-net or similar approaches may further improve sensitivity.

Depending on the approach used for the search, a future $\tau \to \ell\alpha$ analysis might also investigate an analytic fit function for the search, enabling an unbind fit with a floating m_α.

A future iteration of the $\tau \to \ell\alpha$ search at Belle II will also provit from the increased amount of available measured data. At the time of writing this thesis, SuperKEKB delviered a total integrated luminosity of $427{,}79\,\text{fb}^{-1}$, which increases the available Belle II data set by a factor of about seven.

7.2 GKK: A New Method to Measure the τ-Mass

When developing the GKK method for the $\tau \to \ell\alpha$ search, we designed the method such that the peak of the two-body $\tau \to \ell\alpha$ decay is precisely at the expected lepton momentum value in the τ-rest frame. While working on the $\tau \to \ell\alpha$ search, we had the idea to use the GKK-method to measure the τ mass in $\tau \to \pi\nu_\tau$ decays. The following describes how we would conduct a possible τ mass measurement. In closely follows and expands this authors pre-print [89].

We consider the momentum spectrum of a $\tau^- \to \pi\nu_\tau$ decay – the signal-side – in the τ-rest-frame. As described in Section 4.3.1, we first sample a set of p^μ_τ candidates of the tag-side τ, considering the decay mode $\tau^- \to \pi\pi\pi\nu_\tau$. We use the set of p^μ_τ to boost the signal-side π-momentum, p^μ_π, into the rest frame of the signal τ. We denote the momentum of the signal side in the τ-rest frame as p^τ_π. In the process $\tau^- \to \pi\nu_\tau$, we expect a peak at $p^\tau_\pi = \frac{m_\tau - m_\pi}{2}$. Figure 7.1 illustrates how the resulting GKK limiting distribution, referred to as GKK-distribution, forms. We do this by stacking each event's p^τ_π distributions. As the number of events increases, a limiting distribution emerges, which should only depend on the parameter of interest – in this particular case, p^τ_π. With 25 events a clear peak emerges around the expected $p^\tau_\pi \approx 0.82\,\text{GeV}$. If there are enough events and the input mass for m_τ is the actual mass, as is the case with the blue line in Figure 7.2a, a sharp peak emerges at the expected momentum.

In order to use the GKK method for τ-mass measurements, we need to understand the behavior of the GKK. We studied the influence of the τ-mass input, which is needed to calculate the GKK distribution. This is done by considering the input masses $m_{\tau_{\text{input}}}$, which deviate by Δm from the mass value $m_{\tau_{\text{PDG}}} = 1776.86 \pm 0.12$ MeV [110] used in the event

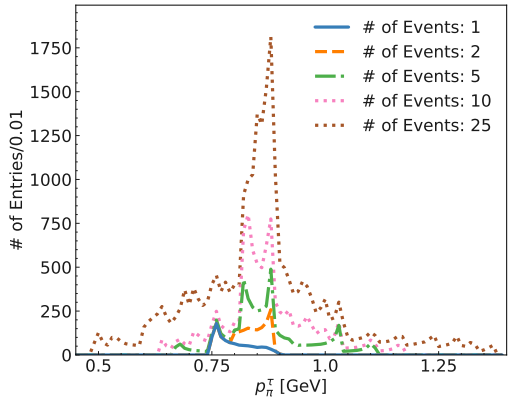
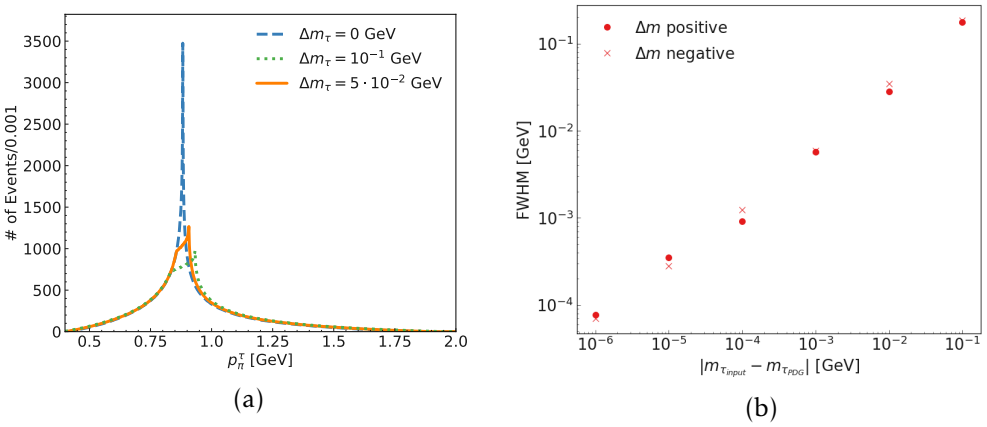


Figure 7.1: Step by step formation of the GKK limiting distribution for the $\tau \to \pi\nu_\tau$ decay. Here, for every $\tau^+\tau^-$ event, we calculate a set of 1000 possible π-momentum candidates. The blue curve gives an example distribution for one event. By stacking one event on top of the next, the limiting distribution emerges for a large enough number of events, as indicated by the distributions, step by step, in the plot.

generation:

$$\Delta m = m_{\tau_{\text{input}}} - m_{\tau_{\text{PDG}}}. \tag{7.1}$$

Figure 7.2a illustrates the influence of Δm. Here, we compare the result-



(a) (b)

Figure 7.2: The smearing effect of the GKK distribution for different deviations of the mass hypothesis is used to calculate the GKK distribution. (a) GKK distribution for the π-momentum of the $\tau \to \pi\nu$ decay for 3×10^5 events in the τ-rest-frame. The effect of Δm is demonstrated for the GKK distribution. (b) Full-Width Half-Maximum, FWHM, for different Δm. Here, Toy MC denotes simulation data.

ing limiting distribution for 3 different $\Delta m \in \{0, 0.05, 0.1\}$ [GeV]. We also calculate the Full-Width Half-Maximum, FWHM, of the resulting limiting distributions to quantify the influence of Δm. We can understand this quantity as a measure of the spread or smearing of the distribution. Figure 7.2b shows the relation of FWHM versus Δm.

Figure 7.2a shows two essential aspects of the GKK distribution. First, the true/expected momentum $p^{\tau}_{\pi,exp}$ is within the GKK-distributions peak region. We define the peak region between the two exponentially decreasing flanks of the distribution. Second, a simple peak position search for the determination of the τ-mass will not yield reliable results since a miss-match of the assumed and actual value of the mass, Δm, leads to a distortion of the peak shape. For non-zero Δm, the maximum's position of the distribution does not represent the actual $p^{\tau}_{\pi,exp}$. Instead, we observe that the width of the distribution increases with increasing Δm.

Figure 7.2b further quantifies this increase in width, which shows the Full-Width Half Maximum, FWHM, of the GKK distribution for different values of Δm. We calculate the FWHM of the distribution numerically. Here, the FWHM depends linearly on Δm for negative and positive Δm values. We do not observe any significant differences in the behavior of a positive or negative Δm.

We interpret the GKK-distribution's behavior as follows: the boost calculation incorporates the τ-mass to compute the boost in the τ-rest-frame and determine the candidate-cone of the τ-momenta. A mismatch between the true and the input value leads to a smearing effect in both cases. We use the distribution's width to quantify the smearing.

This result makes it possible to determine the τ-mass by scanning through $m_{\tau_{\text{input}}}$ hypothesis. For example, we can determine the τ-mass by minimizing the FWHM in a numerical approach with different $m_{\tau_{\text{input}}}$. Figure 7.3c illustrates the FWHM distribution close to the actual simulation value. The minimum of the distribution corresponds to $m_{\tau_{\text{PDG}}}$. Furthermore, we have found that the maximum peak value of the GKK-distributions is similar to the FWHM. Again, Figure 7.3d clearly shows that the extremum is at $m_{\tau_{\text{PDG}}}$. Both figures show two example distributions, the mean (black dots) and the 68%, 90% and 95% confidence intervals (gray bands) produced from 100 independent toy simulations without detector resolution[1]. We note that the GKK method clearly distinguishes the $m_{\tau_{\text{PDG}}}$ in the plots, indicating a precision of at least the step size of 10 keV for this method. All 100 toy simulations showed an extremum in the same peak, indicating the result's high robustness. The red dot with the error bar indicates the PDG average and corresponding (statistical) uncertainty for the m_{τ} mass. With the generated data set, the GKK method is already much less limited by the statistical uncertainty than methods proposed in the past. We expect that systematic uncertainties will dominate future measurements.

The proposed new method uses the information of all available events of the chosen decay topology. This property contrasts the ARGUS- and CLEO methods for measuring the τ-mass [112, 113]. The ARGUS method uses only the subsample of all events close to the endpoint of the distri-

[1] Please note that the $m_{\tau_{\text{input}}}$ scan yields a distribution as indicated by the black dots. The confidence interval shows the variance in the slope of the FWHM or maximum peak value distribution.

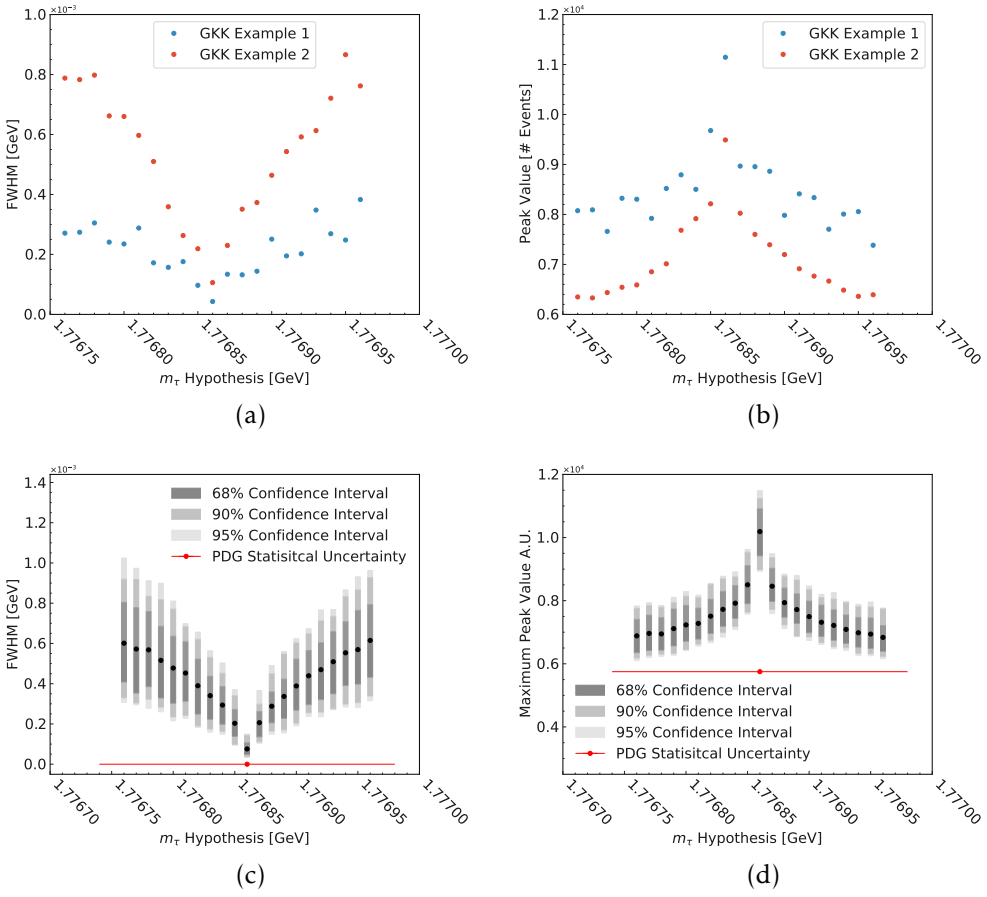


Figure 7.3: GKK-distribution's Full-Width Half Maximum (FWHM), plots (a) and (c), and Maximum Peak Value, plots (b) and (d), for a τ-mass hypothesis ($m_{\tau_{\text{input}}}$) scanned around the initial simulation value of 1.77686 GeV/c^2 in 10 KeV/c^2 steps. Plots (a) and (b) show the GKK-distribution's FWHM and maximum peak value for two data sets with a shallow FWHM slope (Example 1) and a steep FWHM slope (Example 2), exemplifying the possible variations. The extrema for the FWHM and Maximum Peak Value are clearly at $m_{\tau_{\text{PDG}}}$ for both data sets. Plots (c) and (d) show the confidence intervals indicating the expected slope variation for this scan's FWHM or Maximum Peak value. The PDG Statistical Uncertainty indicates the world average (dot) and the corresponding lowest statistical uncertainty measured by BaBar [111] (red line). The displayed uncertainty also coincides with the total uncertainty of the PDG's average mass value.

bution. The CLEO method considers only a fraction of events contributing to its peak position.

Today, the ARGUS method is used at e^+e^--colliders well above the $\tau^+\tau^-$-production threshold [114, 111]. It has worse precision than the one used by the BES III collaboration exploring the $\tau^+\tau^-$-production threshold [115, 116]. The GKK method has the potential to significantly improve the precision of the τ-mass measurement compared to the ARGUS method.

Chapter 8

Conclusion

This thesis searches for the lepton-flavor-violating beyond the Standard Model process $\tau \to \ell\alpha$, with α being an invisible particle, most probably a boson. We search with $62.8\,\text{fb}^{-1}$ of data at and around the $\Upsilon(4S)$ resonance at the SuperKEKB and Belle II experiment. We neither find a five nor three sigma deviation. Hence we report no evidence for the $\tau \to \ell\alpha$ decay and provide a 95% confidence-level upper limit in the range of one per mill to one percent for

$$R = \frac{Br(\tau \to \ell\alpha)}{Br(\tau \to \ell\nu_\tau\nu_\ell)},$$

depending on the tested mass hypothesis of the α. This result is the most stringent upper limit for R, increasing the constraints on lepton flavor violating models. This thesis describes the substantial contributions to the Belle II $\tau \to \ell\alpha$ search in evaluating the selection strategy, the τ-rest frame estimate, and validating the fit and upper limit estimation. Appendix A outlines this thesis's contributions in more detail. A crucial part of this search is the excellent detector performance at Belle II. This thesis contributes to the beam background monitoring system and the electron identification performance measurement. To improve the $\tau \to \ell\alpha$ search sensitivity, we develop and test new τ-rest frame estimation methods. One of the developed methods, the GKK method, also shows a perspective for substantially improving the statistical precision of the τ mass measurement.

Relevant theories predicting the $\tau \to \ell\alpha$ decay are Z' and ALP theories, depending on the spin of α. In contrast to the irreducible Standard Model three-body $\tau \to \ell\nu_\tau\nu_\ell$ decay, we assume a two-body τ-decay into a lepton, ℓ, and an invisible particle, α. In the search for the $\tau \to \ell\alpha$ decay, we can utilize these differences to differentiate between the $\tau \to \ell\nu_\tau\nu_\ell$ and $\tau \to \ell\alpha$ decay. The most significant difference between a two and three-body decay emerges in the rest frame of the τ. Because of the undetectable ν_τ, this thesis must define a pseudo rest frame of the τ. We demand $\tau^+\tau^-$ events with one τ being the signal decay $\tau \to \ell\alpha$ decay and the other τ decaying into three hadrons, h, most probably $\pi^\mp$. By demanding a three h decay, we can approximate the τ-flight direction with

the direction of a three h system. Furthermore, we use the well-known initial state of $e^+e^- \to \tau^+\tau^-$ to approximate the τ energy with half of the center-of-mass energy, ignoring initial state radiation. The ARGUS collaboration proposed this method, which we refer to as the ARGUS method. The ARGUS method's momentum spectrum of the lepton in the τ restframe can differentiate between two and three-body decays.

Our selection optimization utilizes the irreducible $\tau \to \ell\nu_\tau\nu_\ell$ decay to optimize for $\tau \to \ell$ + invisible decays, leading to an unbiased selection regarding the mass hypothesis of α. The detector signal is a 1-prong τ decay, with the charged particle being a lepton. We want to reduce biases in our signal distribution due to the event selection. Thus, we only use particle ID to identify the lepton for the 1-prong decay. We choose the 3×1-prong $\tau^+\tau^-$ decay topology because it is a well-tested topology and allows selecting a pure sample of $\tau^+\tau^-$ events. Furthermore, it enables reconstructing a τ-pseudo-rest frame with the 3-prong τ decay. Our selection mainly focuses on the 3-prong τ decay to identify the $\tau^+\tau^-$ event, and this focus reduces biases in the selection for the signal decay.

We reconstruct γ and π^0 candidates for every event – collectively referred to as neutrals. We discuss four different selection variants in this thesis. A cut-based selection that rejects the reconstructed neutrals, a cut-based selection that optimizes the number of neutrals, and two boosted decision tree variants. To enable comparability, we train the boosted decision trees with the same variables used in the cut-based selections. One BDT emulates the neutral rejection case. We first apply all cut variables consecutively until the neutral rejection, and then we train the BDT with the remaining variables; the other BDT uses all cut variables in the cut-based selection allowing for neutrals to be optimized. The thesis evaluates the sensitivity of each variant with a simulation-based study. We found that a figure of merit-based selection of the neutrals yields the most sensitive result and that BDT-based selection allowing for neutrals can provide both high purity and a high amount of reconstructed and selected data. The approved Belle II selection for the $\tau \to \ell\alpha$ search uses a cut-based approach rejecting neutrals, and it achieves a purity above 90% and a good simulated to measured data agreement. We chose this selection for its robustness and good simulation to measured data agreement.

We identify the reconstruction of the τ restframes as a challenge, and this thesis develops two new methods for the estimation. The Thrust method approximates the τ direction with the thrust direction of the event. The GKK method infuses the known ν_τ information and generates a probability density function for each event. Many event density functions stacked on each other yield a limiting distribution which allows estimating the actual parameter of interest. The sensitivity test for the ARGUS, the Thrust, and the GKK methods showed no significant advantage for the new estimation methods. Hence, we use the ARGUS method for the $\tau \to \ell\alpha$ search with data.

We use templates to describe the shape of the $\tau \to \ell\alpha$, $\tau \to \ell\nu_\tau\nu_\ell$, and all other physics processes collectively treated and referred to as background. We use a maximum likelihood approach for the fit, and we study the determination of a 95% confidence-level upper limit for the branching ratio of the $\tau \to \ell\alpha$ decay with a frequentist leaning CL_S and a Bayesian interference method. The fit is unbiased in choosing the yield for the three fit distributions in the frequentist leaning approach and uses Gaussian priors for the Standard Model distributions in the Bayesian approach. This thesis tests the mass hypothesis of $0.0\,\mathrm{GeV}/c^2$, $0.5\,\mathrm{GeV}/c^2$, $0.7\,\mathrm{GeV}/c^2$, $1.0\,\mathrm{GeV}/c^2$, $1.2\,\mathrm{GeV}/c^2$, $1.4\,\mathrm{GeV}/c^2$, and $1.6\,\mathrm{GeV}/c^2$ of the $\tau \to \ell\alpha$ decay, and sets a 95% confidence level upper limit if we find no evidence for the $\tau \to \ell\alpha$ decay. We show that the fit is unbiased, and both hypothesis test methods agree.

After unblinding, we needed to update the systematic uncertainty modeling. The uncertainties are smoothed and symmetrized or rebinned according to the specific systematic. The updated model shows no problems in the fit, and the fit on measured data does not show any sign of the $\tau \to \ell\alpha$ decay. We determined a 95% confidence level upper limit with the CL_S method. We set upper limits ranging from one per mill to one percent in both channels, depending on the mass hypothesis of α.

To explore the potential of the GKK method for other applications, we conduct a feasibility study and show that the GKK method will reduce the τ mass uncertainty.

Bibliography

[1] D. L. Couprie, "Anaximander," (2021).

[2] D. W. Graham, "Anaximenes," (2021).

[3] P. O'Grady, "Thales of miletus," (2021).

[4] D. W. Graham, "Heraclit," (2021).

[5] S. Berryman, "Leucippus," (2016).

[6] S. Berryman, "Democitus," (2016).

[7] F. Grabowsk, "Plato: The timaeus," (2021).

[8] J. Humphreys, "Aristotl," (2021).

[9] R. Boyle, *The sceptical chymist* (London : Dent; New York : Dutton, 1964).

[10] D. Bernoulli, *Hydrodynamics* (Dover Publications, 1968).

[11] A. Laurent, *Traité élémentaire de chimie* (Chez Cuchet, 1789).

[12] R. Brown, The Philosophical Magazine **4**, 161 (1828).

[13] C. Wiener, Poggendorffs Annalen. **118**, 79 (1863).

[14] A. Einstein, *Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen*, Annalen der Physik (Johann Ambrosius Barth, 1905).

[15] A. Einstein, Ann. Phys. **324**, 289 (1906).

[16] J. Perrin, English edition: atoms, tr. D. Hammick. London: Constable (1923).

[17] D. I. Mendelejew, Zeitschrift für Chemie **12**, 405 (1869).

[18] H. Becquerel, Compt. Rend. Hebd. Seances Acad. Sci. **122**, 420 (1896).

[19] J. J. Thomson, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **44**, 293 (1897).

[20] P. Curie and M. Curie, Comptes rendus hebdomadaires des séances de l'Académie des sciences. **127**, 1215 (1898).

[21] E. Rutherford, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **47**, 109 (1899).

[22] E. Rutherford and T. Royds, Phil. Mag. **17**, 281 (1909).

[23] E. Rutherford, Philosophical Magazine **21**, 669 (1911).

[24] E. Rutherford, Phil. Mag. Ser. 6 **37**, 581 (1919).

[25] J. Chadwick, Nature **129**, 312 (1932).

[26] M. Gell-Mann, Physics Letters **8**, 214 (1964).

[27] G. Zweig, *An SU_3 model for strong interaction symmetry and its breaking; Version 1*, Tech. Rep. (CERN, Geneva, 1964).

[28] B. Abi *et al.* (Muon $g-2$ Collaboration), Phys. Rev. Lett. **126**, 141801 (2021).

[29] R. Aaij *et al.* (LHCb), arXiv (2021), arXiv:2103.11769 [hep-ex] .

[30] L. Calibbi *et al.*, Journal of High Energy Physics **09**, 173 (2021), arXiv:2006.04795 [hep-ph] .

[31] W. Altmannshofer *et al.*, Physics Letters B **762**, 389 (2016).

[32] G. Arcadi *et al.*, arXiv (2021), arXiv:2104.04456 [hep-ph] .

[33] K. Asai *et al.*, Physical Review D **99**, 055029 (2019).

[34] P. A. Zyla *et al.*, Progress of Theoretical and Experimental Physics **2020** (2020), 10.1093/ptep/ptaa104, 083C01, https://academic.oup.com/ptep/article-pdf/2020/8/083C01/34673722/ptaa104.pdf .

[35] J. Heeck, Physics Letters B **758**, 101 (2016).

[36] H. Albrecht *et al.*, Zeitschrift für Physik C - Particles and Fields **68**, 25 (1995).

[37] I. J. R. Aitchison and A. J. G. Hey, *Gauge theories in particle physics: a practical introduction; 4th ed.* (CRC Press, Boca Raton, FL, 2013).

[38] J. E. Dodd and B. Gripaios, *The Ideas of Particle Physics*, 4th ed. (Cambridge University Press, 2020).

[39] D. Goldberg, *The standard model in a nutshell* (Princeton University Press, Princeton, NJ, 2017).

[40] M. Thomson, *Modern Particle Physics* (Cambridge University Press, 2013).

[41] G. Aad *et al.* (ATLAS), Phys. Lett. B **716**, 1 (2012), arXiv:1207.7214 [hep-ex] .

[42] S. Chatrchyan *et al.* (CMS), Phys. Lett. B **716**, 30 (2012), arXiv:1207.7235 [hep-ex] .

[43] Y. Nir, CERN Courier **60**, 41 (2020).

[44] P. Salucci, The Astronomy and Astrophysics Review **27**, 2 (2019).

[45] B. L. Roberts and W. J. Marciano, eds., *Lepton dipole moments*, Vol. 20 (2009).

[46] W. Altmannshofer *et al.* (Belle-II), PTEP **2019**, 123C01 (2019), [Erratum: PTEP 2020, 029201 (2020)], arXiv:1808.10567 [hep-ex] .

[47] M. Aartsen *et al.*, Physical Review Letters **111** (2013), 10.1103/PhysRevLett.111.021103.

[48] M. Aartsen *et al.*, Science **342**, 1242856 (2013), https://www.science.org/doi/pdf/10.1126/science.1242856 .

[49] A. Kamada, Physical Review D **92** (2015), 10.1103/PhysRevD.92.113004.

[50] A. DiFranzo, Physical Review D **92** (2015), 10.1103/PhysRevD.92.095007.

[51] T. Araki, Physical Review D **91** (2015), 10.1103/PhysRevD.91.037301.

[52] T. Araki, Physical Review D **93** (2016), 10.1103/PhysRevD.93.013014.

[53] J. Heeck and W. Rodejohann, Physics Letters B **776**, 385 (2018).

[54] Z. G. Berezhiani and M. Y. Khlopov, Zeitschrift für Physik C Particles and Fields **49**, 73 (1991).

[55] O. Davidi *et al.*, Journal of High Energy Physics **2018**, 153 (2018).

[56] B. Grinstein, J. Preskill, and M. B. Wise, Physics Letters B **159**, 57 (1985).

[57] N. Braun *et al.*, Belle II Public Notes (2018).

[58] K. Akai *et al.*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **907**, 188 (2018), advances in Instrumentation and Experimental Methods (Special Issue in Honour of Kai Siegbahn).

[59] T. Abe *et al.* (Belle-II), arXiv (2010), arXiv:1011.0352 [physics.ins-det] .

[60] M. Gabriel *et al.*, The European Physical Journal C **81**, 972 (2021), arXiv:2012.10948 [physics.ins-det] .

[61] L. M. S. de Silva and F. Simon, Journal of Instrumentation **15**, P06030 (2020), arXiv:2004.05066 [physics.ins-det] .

[62] H. U. Windel, *CLAWS - An injection background monitoring system for the second and third phase of the SuperKEKB commissioning*, Dissertation, Technische Universität München, München (2021).

[63] D. Matvienko, EPJ Web of Conferences **191**, 02010 (2018).

[64] T. Ferber, "First results from belle ii," Conference Talk (2020).

[65] H. Ye and C. Niebuhr, "Thermal mockup studies of belle ii vertex detector," Conference Talk TIPP17 (2017).

[66] B. Wang *et al.*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **1032**, 166631 (2022).

[67] L. Andricek *et al.*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **638**, 24 (2011).

[68] L. Zani *et al.*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **1038**, 166952 (2022).

[69] F. Abudinén *et al.*, Physical Review Letters **127** (2021), 10.1103/PhysRevLett.127.211801.

[70] H. Svidras, "The central drift chamber of belle 2," Talk at Belle II Starterkit Workshop (2020).

[71] S. Hirose, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **766**, 163 (2014), rICH2013 Proceedings of the Eighth International Workshop on Ring Imaging Cherenkov Detectors Shonan, Kanagawa, Japan, December 2-6, 2013.

[72] S. Iwata *et al.*, Progress of Theoretical and Experimental Physics **2016** (2016), 10.1093/ptep/ptw005, 033H01, https://academic.oup.com/ptep/article-pdf/2016/3/033H01/19301325/ptw005.pdf .

[73] J. Skorupa, *Characterization of the Particle Identification of the Belle II Detector with Bhabha Radiation and First Measurement of* $\sin(2\phi_1)$ *and* Δm_d *Using Belle II Data*, Master thesis, Technische Universität Münchent (2021).

[74] G. Punzi, eConf **C030908**, MODT002 (2003), arXiv:physics/0308063 .

[75] The Belle II Colloboration, "Belle ii software documentation," (2022).

[76] E. De La Cruz Burelo *et al.*, "Search for lepton flavor violating decay $\tau \to l\alpha$," https://docs.belle2.org/record/1326?ln=en (2022).

[77] J. H. Friedman, The Annals of Statistics **29** (2001), 10.1214/aos/1013203451.

[78] xgboost developers, "Introduction to boosted trees," (2021).

[79] scikit-learn developers, "Tuning the hyper-parameters of an estimator," (2021).

[80] A. Mood, F. Graybill, and D. Boes, *Introduction to the Theory of Statistics*, International Student edition (McGraw-Hill, 1974).

[81] A. Kolmogorov, G. Ist. Ital. Attuari. **4**, 83 (1933).

[82] N. Smirnov, The Annals of Mathematical Statistics **19**, 279 (1948).

[83] T. Ferber *et al.*, "Conference readiness," (2021).

[84] M. Milesi *et al.*, "Recommendations for leptonid - moriond 2021," (2021).

[85] Z. Gruberova *et al.*, *Study of the tau lepton decays at the Belle II experiment*, Master's thesis, Faculty of Mathematics and Physics Charles University (2021).

[86] P. Rados *et al.*, "Tracking and vertexing performance," (2021).

[87] F. Abudinén *et al.* (Belle-II), arXiv (2020), arXiv:2008.04665 [hep-ex] .

[88] Z. Gruberova, P. Rados, and A. Rostomyan, "Photon timing selection recommendations," (2021).

[89] T. Kraetzschmar *et al.*, arXiv (2021), arXiv:2109.14455 [hep-ex] .

[90] K. Cranmer *et al.* (ROOT Collaboration), *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, Tech. Rep. (New York U., New York, 2012).

[91] L. Heinrich, M. Feickert, and G. Stark, "pyhf: v0.6.3," (2021).

[92] L. Heinrich *et al.*, Journal of Open Source Software **6**, 2823 (2021).

[93] L. Moneta *et al.*, in *Proceedings of 13th International Workshop on Advanced Computing and Analysis Techniques in Physics Research — PoS(ACAT2010)* (Sissa Medialab, 2011).

[94] F. James and M. Roos, Comput. Phys. Commun. **10**, 343 (1975).

[95] H. Dembinski and P. O. et al., (2020), 10.5281/zenodo.3949207.

[96] A. L. Read, Journal of Physics G: Nuclear and Particle Physics **28**, 2693 (2002).

[97] G. Cowan *et al.*, The European Physical Journal C **71**, 1554 (2011).

[98] G. J. Feldman, Physical Review D **57**, 3873 (1998).

[99] Wikipedia user Nurg *et al.*, "Cls method (particle physics)," (2020).

[100] A. Caldwell, D. Kollár, and K. Kröninger, Computer Physics Communications **180**, 2197 (2009).

[101] W. K. Hastings, Biometrika **57**, 97 (1970), https://academic.oup.com/biomet/article-pdf/57/1/97/23940249/57-1-97.pdf .

[102] O. Schulz *et al.*, SN Computer Science **2**, 210 (2021).

[103] J. Ling *et al.*, "Litehf.jl," github (2022).

[104] S. Longo, "Neutrals performance confluence page." .

[105] J. H. Friedman, "Data analysis techniques for high energy particle physics," (1974).

[106] P. Rados, "Tracking performance confluence page." Webpage (2021).

[107] F. Tenchini, "New limit on $\tau \to \ell$ + invisible - recent tau lepton results at belle ii," Slides for a talk at ICHEP 2022 (2022).

[108] D. Guadagnoli, C. B. Park, and F. Tenchini, Physics Letters B **822**, 136701 (2021), arXiv:2106.16236 [hep-ph] .

[109] F. Abudinén *et al.*, (2021), arXiv:2110.00810 [hep-ex] .

[110] P. A. Zyla *et al.* (Particle Data Group), PTEP **2020**, 083C01 (2020).

[111] B. Aubert *et al.* (BaBar), Phys. Rev. D **80**, 092005 (2009), arXiv:0909.3562 [hep-ex] .

[112] H. Albrecht *et al.* (ARGUS), Physics Letters B **292**, 221 (1992).

[113] R. Ballest *et al.* (CLEO), Phys. Rev. D **47**, R3671 (1993).

[114] K. Abe *et al.* (Belle), Phys. Rev. Lett. **99**, 011801 (2007), arXiv:hep-ex/0608046 .

[115] V. V. Anashin *et al.*, JETP Lett. **85**, 347 (2007).

[116] M. Ablikim *et al.* (BESIII), Phys. Rev. D **90**, 012001 (2014), arXiv:1405.1076 [hep-ex] .

Appendix A

Thesis Work in Context of Collaborative Projects

Almost every task in a high-energy physics experiment is collaborative, meaning several people contribute, and it is not always obvious who contributed what. In this chapter, I describe in more detail what I contributed to the work described in this thesis.

At the beginning of this Ph.D. project, I helped Hendrik Windel to build and install the permanent CLAWS++ detectors and contributed to the software update.

Furthermore, I also started a performance study to improve Belle II's understanding of electron identification. A well-understood electron identification is a crucial ingredient for our $\tau \to \ell\alpha$ search. In the later discussion of systematic uncertainties for $\tau \to \ell\alpha$, lepton identification is a significant systematic effect.

The performance study for electron identification uses radiative-Bhabha events. I chose a selection independent of specifically selecting radiative events and developed the methodology for the study with the help of other colleges at DESY. Justin Skorupa took over the performance study, developed it further, and included it in his masters-thesis project. Together with Justin Skorupa and the valuable help of Thibaud Humair, we contributed to the lepton performance paper, which is still in preparation.

In the $\tau \to \ell\alpha$ analysis, I contributed to most of the analysis steps. I studied selection methods together with Marcela García. I developed the GKK methods and wrote, with the valuable help of my co-authors, a publication intended for submission to JHEP. Furthermore, I set up the software to create templates for the template fit and implanted a fit with pyhf and Julia BAT. Francesco Tenchini also uses the templates for his C++ BAT-fit. My pyhf implementation was a crucial verification for the approved Belle II results obtained with RooStats.

We decided deliberately in the $\tau \to \ell\alpha$ analysis that those who work on the analysis do not contribute to the validation of measured and simulated data. This work was predominately done by Marcela García and

Petar Rados. The corresponding section is a reproduction of their work.

Appendix B

Theory Complementary

B.1 Standard Model Tau Lepton Decay-Width

The Standard Model distribution of $\tau \to \ell \nu_\tau \nu_\ell$ is given by the weak interaction and reads as

$$\frac{\Gamma_{\tau\to l\nu\nu}}{d\Omega dx} = \frac{G_F^2 m_\tau^5}{192\pi^4} x^2 \{3(1-x) + \frac{1}{2}(4x-3) + \pm P_\tau \cos(\theta)[1-x+\frac{1}{2}(4x-3)]\}, \quad \text{(B.1)}$$

with the the Michel parameters given to be

$$\rho = \delta = \frac{3}{4},\ \eta = 0,\ \xi = 1.$$

G_F is the Fermi constant, m_τ the τ mass, θ is the angle of the spin to the lepton momentum, P_τ the initial state polarisation and x is given by

$$x = \frac{E_l^\star}{E_l^{max}}, \quad \text{(B.2)}$$

with $E_l^\star$ denoting the lepton energy in the tau rest frame and E_l^{max} the maximum energy the lepton can have. In the case of the τ it is given by

$$E_l^{max} = \frac{m_\tau}{2}. \quad \text{(B.3)}$$

Reformulating x in terms of the lepton momentum yields

$$x = \frac{2\sqrt{\left(p_l^\star\right)^2 + m_l^2}}{m_\tau} \quad \text{(B.4)}$$

Furthermore Equation B.1 can be simplified if there is no initial polarisation $P_\tau = 0$ and integrating over $d\Omega$ gives

$$\frac{d\Gamma_{\tau\to l\nu\nu}}{dx} = \frac{G_F^2 m_\tau^5}{48\pi^3} x^2 \{\frac{3}{2} - x\} \quad \text{(B.5)}$$

Rewriting Formula B.5 in terms of $p_l^\star$ with

$$\frac{dx}{dp_l^\star} = \frac{2p_l^\star}{m_\tau} \frac{1}{\sqrt{(p_l^\star)^2 + m_l^2}} \tag{B.6}$$

yields

$$\frac{d\Gamma_{\tau\to l\nu\nu}}{dp_l^\star} = \frac{G_F^2 m_\tau^5}{48\pi^3} \frac{4}{m_\tau^2} ((p_l^\star)^2 + m_l^2) \left\{ \frac{3}{2} - \frac{2\sqrt{\left(p_l^\star\right)^2 + m_l^2}}{m_\tau} \right\} \frac{2}{m_\tau} \frac{p_l^\star}{\sqrt{(p_l^\star)^2 + m_l^2}} \tag{B.7}$$

and simplifies to

$$\frac{d\Gamma_{\tau\to l\nu\nu}}{dp_l^\star} = \frac{G_F^2 m_\tau^2}{6\pi^3} ((p_l^\star)^2 + m_l^2) \left\{ \frac{3}{2} \frac{p_l^\star}{\sqrt{(p_l^\star)^2 + m_l^2}} - 2\frac{p_l^\star}{m_\tau} \right\}. \tag{B.8}$$

Appendix C

Data Samples and Monte Carlo Simulations

C.1 Data

The analysis is based on the Moriond2021 data set, corresponding to a total integrated luminosity of 62.8 fb^{-1} (see Table C.1). For this study, we only use runs without any problems in the data taking or callibration – categorised as "good" runs. We perform the measured to simulated data comparison studies as explained in Section 3.6 following a blinded procedure.

Table C.1: Data sets used in the analysis with the corresponding integrated luminosity and uncertainty.

Data sample	$\int Ldt$
exp7	$425.5 \pm 0.3\,pb^{-1}$
exp8	$4597.4 \pm 0.9\ pb^{-1}$
exp10	$3741.3 \pm 1.1\ pb^{-1}$
Buckets 9-11, 13-15	$54030.5 \pm 4.8\ pb^{-1}$

C.2 Monte Carlo Simulations

C.2.1 Signal

In this analysis we use the official MC13a production for the $\tau \to \ell\alpha$ decay. For each of the seven different masses hypotheses

$$M_\alpha = \{0, 0.5, 0.7, 1.0, 1.2, 1.4, 1.6\}[\,\mathrm{GeV}/c^2]$$

, we generate 10^7 events with nominal, run-independent beam background overlays (`BGx1'). Event type codes and production identification codes of those samples are tabulated in Table C.2; the corresponding

grid location of these samples is given by
/belle/MC/release-04-02-01/DB00000757/MC13a/prod000<ProdID>/s00/e1003/4S/r00000/<EvType>/mdst/sub00.

Table C.2: Event type codes and production identifiers of Early Phase III MC13a $\tau \to \ell\alpha$ samples. Each sample comprises 10^7 events.

Signal	Event Type			
	e^+	μ^+	e^-	μ^-
	3410910000	3410910010	3410910020	3410910030
$M_\alpha = 0.5$	3410910001	3410910011	3410910021	3410910031
$M_\alpha = 0.7$	3410910002	3410910012	3410910022	3410910032
$M_\alpha = 1.0$	3410910003	3410910013	3410910023	3410910033
$M_\alpha = 1.2$	3410910004	3410910014	3410910024	3410910034
$M_\alpha = 1.4$	3410910005	3410910015	3410910025	3410910035
$M_\alpha = 1.6$	3410910006	3410910016	3410910026	3410910036
	Production ID			
	e^+	μ^+	e^-	μ^-
$M_\alpha = 0$	13303	13304	16060	16067
$M_\alpha = 0.5$	13305	13306	16061	16068
$M_\alpha = 0.7$	13307	13308	16062	16069
$M_\alpha = 1.0$	13309	13310	16063	16070
$M_\alpha = 1.2$	13311	13312	16064	16071
$M_\alpha = 1.4$	13313	13314	16065	16072
$M_\alpha = 1.6$	13315	13316	16066	16073

In the official samples the $\tau \to \ell\alpha$ decay occurs through a phase space model, while the tag-side τ decays according to the general KKMC tau decay table. Due to the peculiarities of the Monte Carlo generation for simulated data, the charge conjugate samples are produced independently of each other.

C.2.2 Background Processes

In order to study event selection and background suppression, we use generic and low multiplicity samples from the official MC13a production with nominal beam backgrounds (`BGx1') (see Table C.3). These samples are produced with BASF2 *release-04-00-03*. We reconstruct two batches, each corresponding to 100 fb^{-1} after appropriate rescaling.

Processes	Production Name	Integrated luminosity (per batch)	Production IDs batch 1	Production IDs batch 2
$e^+e^- \to \tau^+\tau^-$	taupair	100 fb^{-1}	9441	9494
$e^+e^- \to u\bar{u}$	uubar	100 fb^{-1}	9436	9486
$e^+e^- \to d\bar{d}$	ddar	100 fb^{-1}	9438	9488
$e^+e^- \to s\bar{s}$	ssbar	100 fb^{-1}	9439	9490
$e^+e^- \to c\bar{c}$	ccbar	100 fb^{-1}	9437	9492
$e^+e^- \to B\bar{B}$	mixed,charged	100 fb^{-1}	9434, 9435	9482, 9484
$e^+e^- \to e^+e^-\gamma$	ee	10 fb^{-1}	12385	12831
$e^+e^- \to \mu^+\mu^-\gamma$	mumu	100 fb^{-1}	10385, 10386	10387, 10388
$e^+e^- \to e^+e^-e^+e^-$	eeee	100 fb^{-1}	12386	12801
$e^+e^- \to e^+e^-\mu^+\mu^-$	eemumu	100 fb^{-1}	12387	12812
$e^+e^- \to e^+e^-\pi^+\pi^-$	eepipi	1 ab^{-1}	13376	
$e^+e^- \to e^+e^-K^+K^-$	eeKK	1 ab^{-1}	13374	
$e^+e^- \to e^+e^-p\bar{p}$	eepp	1 ab^{-1}	13378	

Table C.3: MC13a samples with their equivalent luminosities. Bhabha(ee) samples are scaled up 10x to match the rest. For eehh samples, only one large 1 ab^{-1} batch has been produced.

Appendix D

Simulation Truth Definition for Signal

In order to estimate the reconstruction efficiency and purity of

$$e^+e^- \to \tau^+(\to \nu_\tau \pi^+\pi^-\pi^+ + \#\text{neutrals} \geq 0)\tau^-(\to \ell^-)$$

, where the $\tau \to \ell$ includes both, the Standard Model $\tau \to \ell\nu_\tau\nu_\ell$ and beyond the STandard Model $\tau \to \ell\alpha$ processes, the 3- and 1-prongs in simulated data are defined as:

- τ 3-prong:
  - tau(Plus/Minus)MCProng = 3 ;
  - all the tracks have a τ ancestor: hasAncestor(15) > 0;
- $\tau \to e\nu\nu$ (SM):
  - mcPDG = ± 11 ;
  - the track has a τ parent: hasAncestor(15) = 1 ;
  - tau(Plus/Minus)MCMode = 1;
- $\tau \to e\alpha$ (NP):
  - mcPDG = ± 11 ;
  - genMotherPDG = ± 15;

Then the "signal" event should satisfy the combined requirements on the 3-prong and either $\tau \to \ell\nu_\tau\nu_\ell$ or $\tau \to \ell\alpha$ 1-prong sides.

Appendix E

pyhf Fitting

pyhf is a recently developed software[92, 91], which allows to us the HistFactory template fitting tool [90]. The following chapter is aimed as a complementary which demonstrated the capabilities of pyhf.

pyhf allows to provide templates which are translated into a likelihood funktion. The likelihood funktion allows to include the systematic uncertainties of the analysis. The simplified function reads:

$$\mathcal{L}(\mu,\theta|X) = \textit{Likelihood of Model}(\mu) \times \prod \textit{Sys. Err.}(\theta), \tag{E.1}$$

whith the model parameters, μ, the nuisance parameters, θ, the product of the poission probabilities for every bin, $\textit{Likelihood of Model}(\mu)$, and the probability given the systematic uncertainty, $\textit{Sys. Err.}(\theta)$. The systematic uncertainty's probability is given by the constrain term we use. In our case we predominately use the shape correlated uncertainty, which translates two templates we provided – an up and down variation due to the systematic uncertainty – into a Gaussian shaped nuisance parameter.

E.1 Toy Sampling

When sampling a toy data set with pyhf, we want to ensure that the bin content and the total number of events in the sample vary. We test that the number of events in the toy data set is not fixed but fluctuates. Figure E.1 shows that the fluctuations are present and distributed normally.

E.2 Toy Vs. Asimov

In the main body, from Chapter 5 onward, we use the asymptotic CL_S method to determine our 95% confidence upper limit. We verified the validity of the toy-based implementation. In the following section, we present some selected examples that showcase the validity of this assumption.

We compare a set of 100 toy data and evaluate the resulting upper limit distributions for the asymptotic and toy-based methods in Figure

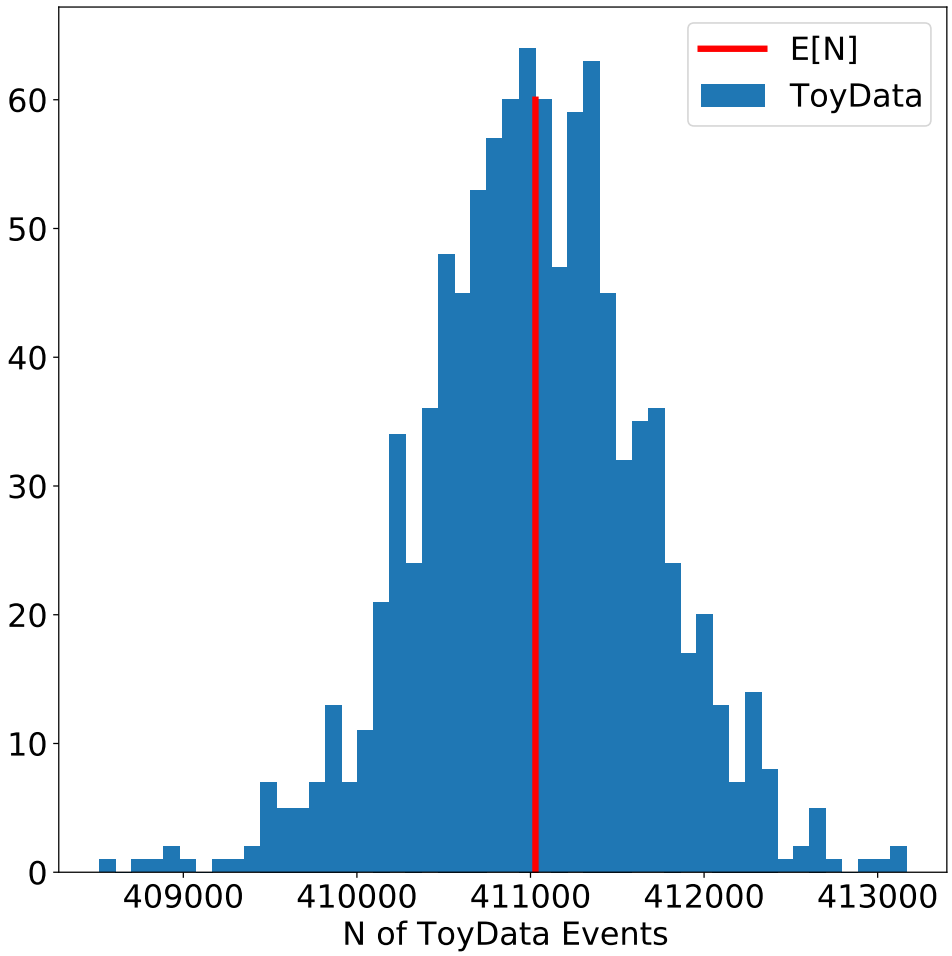


Figure E.1: Number of Events in each Toy Data set (blue) and the expectation value E[N] (red)

E.2. Figure E.3 shows the mean UL for each mass candidate with the corresponding uncertainty.

E.3 Comparison of SciPy vs. Minuit Optimisation

pyhf offers several optimization packages for usage. The default is the scipy optimizer, but in High energy physics, minuit is the default optimizer. We studied both for this analysis, finding that scipy has problems reliably finding a minimum. The following showcases two examples of what we encountered.

First, we tested the H_0 hypothesis and compared the results of the MINUITE with the scipy optimizer. The resulting μ_α of the fitting approach is shown in Figure E.4. We find that MINUITE has no bias, whereas scipy shows a considerable bias towards higher values of μ_α

When we tested to inject a signal,H_S, into the toy data, we found that scipy could not find it. By varying the initial start values, we could sometimes recover this problem. Figure E.4 shows the value of the likelihood function $-2\ln L(\Theta^{obt}|X)$ for a scan of different μ_α start values. The minima are the only converged fit results.

These examples show that scipy yields unreliable results.

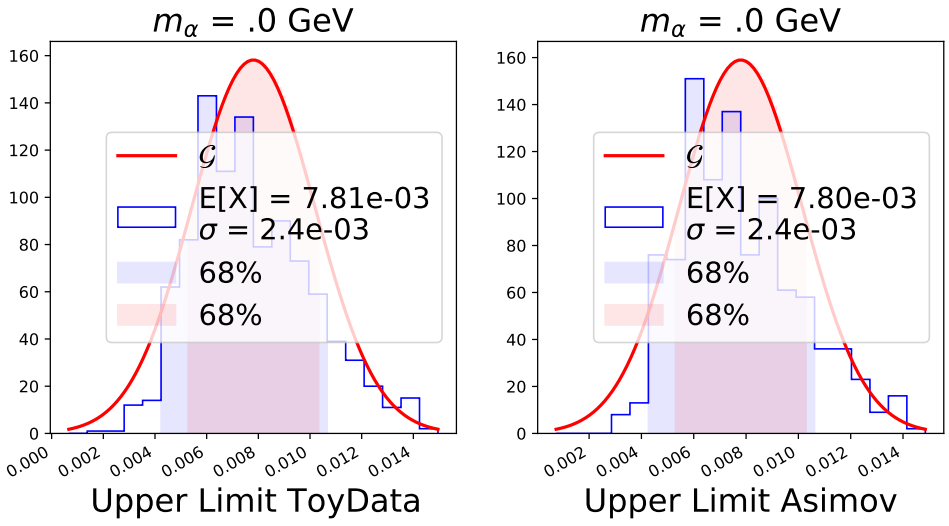


Figure E.2: Comparison of the upper limit distributions for the toy-based and asymptotic methods in the nominal sampling case in a selected example.

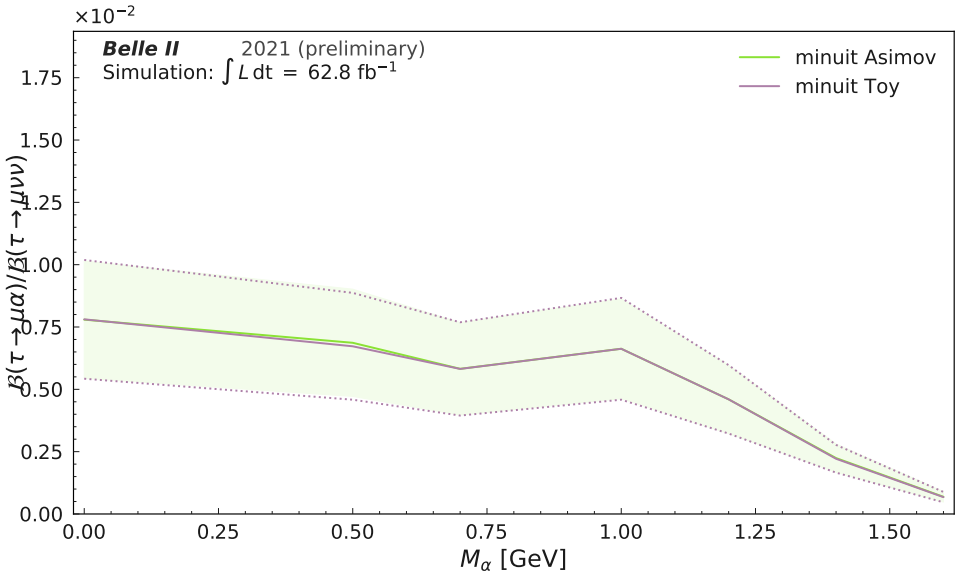


Figure E.3: Mean upper limit for $\tau \to \mu\alpha$ for toy- and asymptotic-based determination.

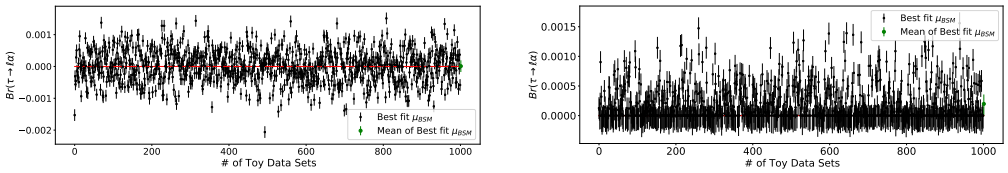


Figure E.4: μ_{α} distributions for both the scipy optimizer (right) and the minuit optimizer (left) for all mass-hypothesis.

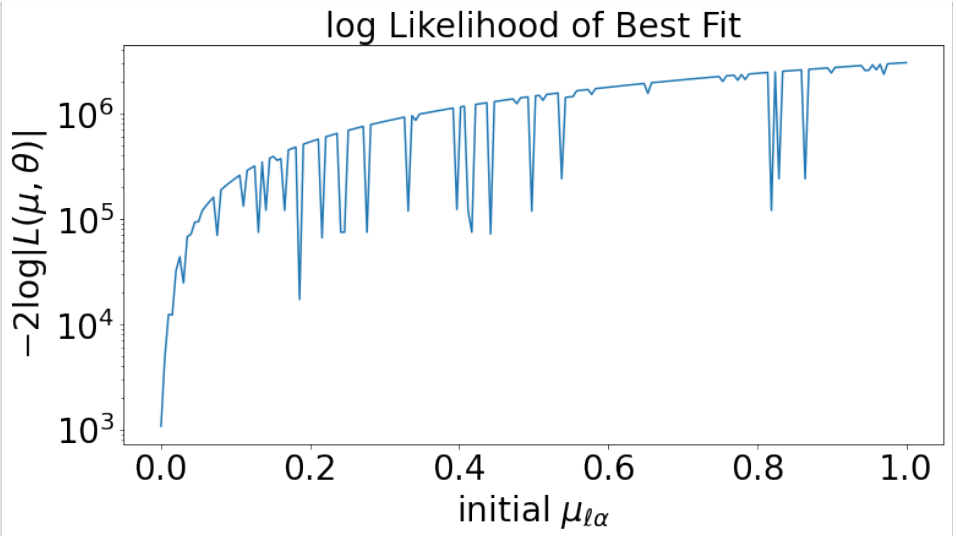


Figure E.5: Distribution of $-2\ln L(\Theta^{obt}|X)$ local minima on the μ_{α}-plane found by the SciPy optimiser.

Appendix F

Negligible Systematic Uncertainties

This chapter briefly summarizes how we tested the impact of the negligible systematic uncertainty. If not stated otherwise, we treat all systematic uncertainties in the same way as we do for the leptonID or trigger uncertainties as described in Chapter 5, Section 5.5.

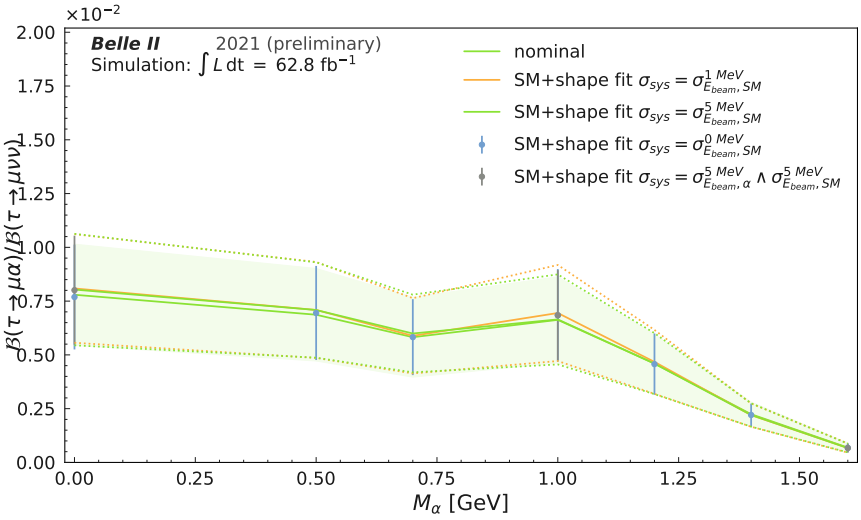
F.1 Beam Energy Shift

Although the beam energy, out of necessity, is a relatively stable quantity of the collider, it slowly shifts over time. Because we use the beam energy as an estimate for the tau energy in our search, it is vital to study the potential effect of the uncertainty on the upper limit. We performed two independent studies on the effect.

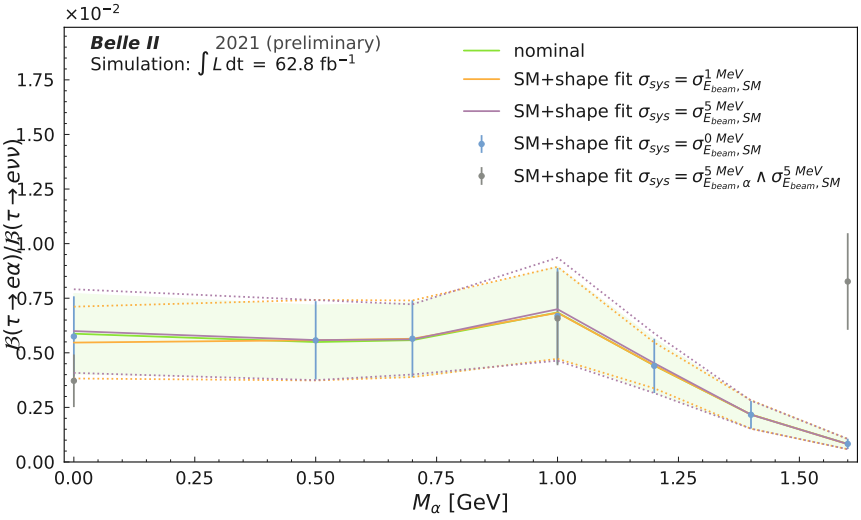
We investigated the effect of E_{beam} uncertainty in two configurations. The first configuration focuses on the $\tau \to \ell \nu_\tau \nu_\ell$ sample only. It assigns an E_{beam} uncertainty of 0.0 MeV, 1.0 MeV, and 5.0 MeV to the $\tau \to \ell \nu_\tau \nu_\ell$ decay. Figure F.1 presents the corresponding upper limit expectations. The results of the mean upper limit of R agree within statistical uncertainty for 100 toy data samples. In the second configuration, we consider an uncertainty for both the $\tau \to \ell \nu_\tau \nu_\ell$ and $\tau \to \ell \nu_\tau \nu_\ell$ sample at 5.0 MeV. The result of the second configuration agrees with the first study, which means that the E_{beam} uncertainty has almost no impact on the upper limit result.

F.2 Tracking Efficiency

Track efficiency is related to the probability of identifying a track. As there is a certain probability of missing a trick, we lose signal events on the one hand and get more background events introduced on the other. The consequence could be a lower sensitivity resulting in a worse upper limit in case of no observation due to the tracking inefficiency. We



(a) E_{beam} shift affecting tau SM and tau to lepton alpha.



(b) E_{beam} shift affecting only tau SM

Figure F.1: Comparison of the mean upper limit (solid lines and dots) and standard deviation (band, dashed line, or error bar) for the fit model without systematic uncertainties (nominal) and the fit model including the E_{beam} uncertainty configuration we study. We evaluate the three cases when the E_{beam} uncertainty is applied only to the $\tau \to \ell\nu_\tau\nu_\ell$ sample (SM+shape fit $\sigma_{\text{sys}} = \sigma^{i,\text{MeV}}_{E_{beam},SM}$ with $i \in \{0,1,5\}$) and one when we apply it to the $\tau \to \ell\nu_\tau\nu_\ell$ and $\tau \to \ell\alpha$ sample ($\sigma_{\text{sys}} = \sigma^{5\,\text{MeV}}_{E_{beam},\alpha} \wedge \sigma^{5\,\text{MeV}}_{E_{beam},SM}$). The lines show an interpolation of the upper limit between the evaluated mass hypotheses.

studied this by running a reconstruction with the tracking efficiency tool switched on. The sensitivity in the data challenge is not reduced by a recognizable margin. In the case of the H_0, the upper limit of the tracking efficiency uncertainty study is comparable to the nominal result. We compare the mean upper limits and respective uncertainties from a set of 100 toy data in Figure F.2.

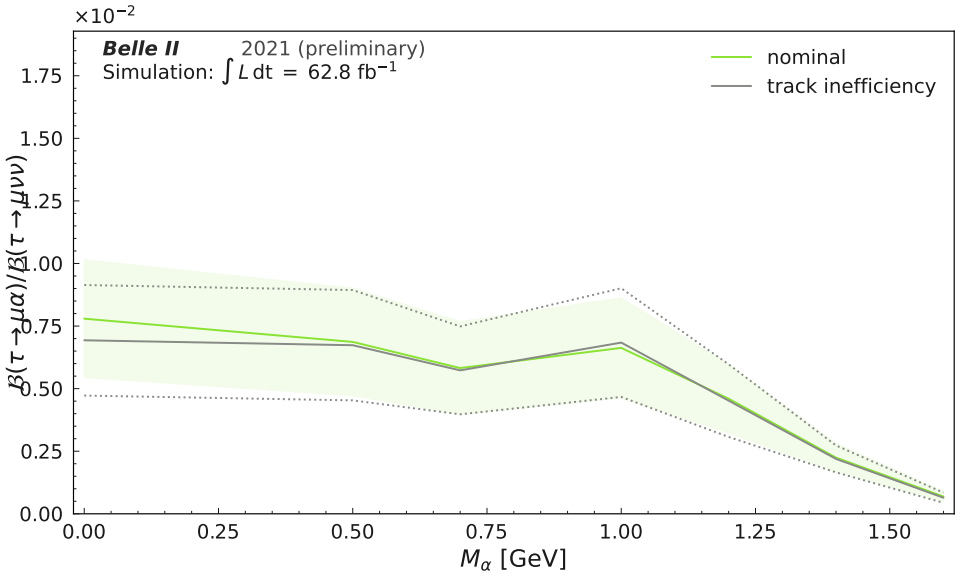


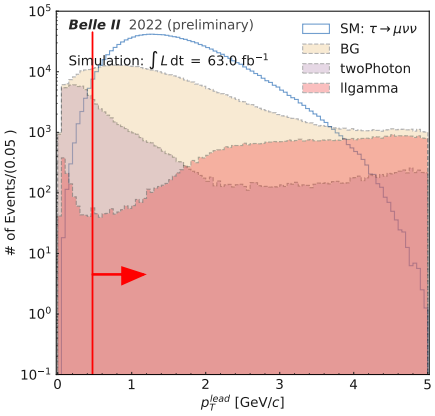
Figure F.2: Comparison of the mean upper limit (solid lines) and standard deviation (band or dashed line) for the fit model without systematic uncertainties (nominal) and the fit model including tracking inefficiency (track inefficiency). The lines show an interpolation of the upper limit between the evaluated mass hypotheses.

Appendix G

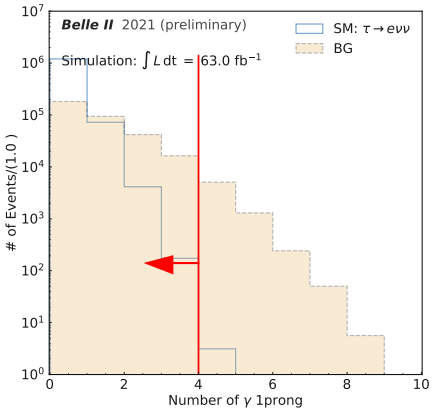
Figures

G.1 Event Selection

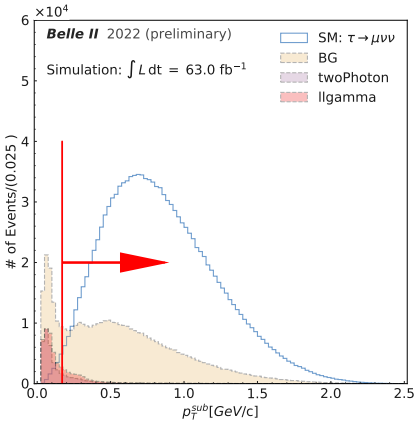
G.1.1 Cut Based Event Selection



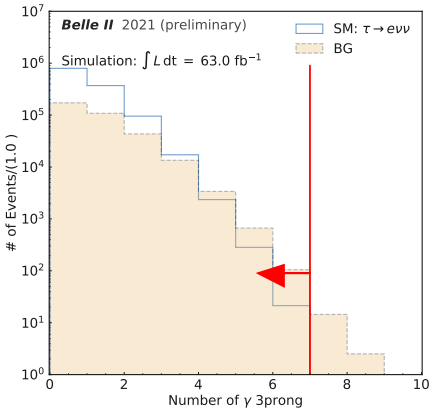
(a)



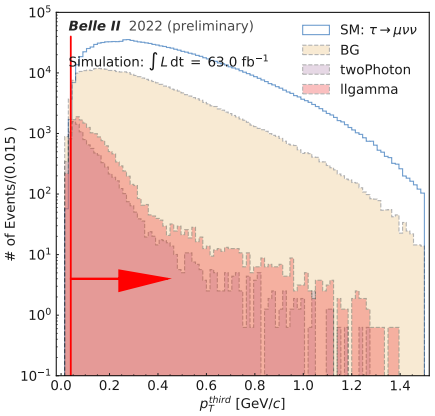
(b)



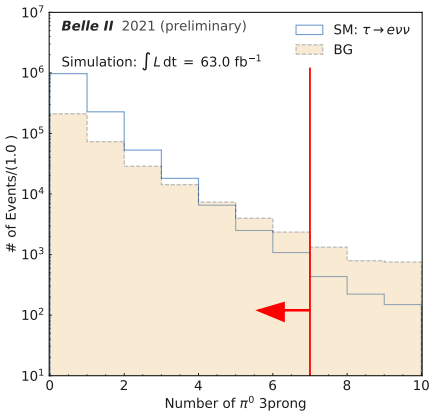
(c)



(d)



(e)



(f)

Figure G.1: $\tau \to \mu \nu_e \nu_\tau$ channel distributions for the ranked p_t of the tag side and neutrals multiplicity of the event. The red marker indicates the FOM optimised selection.

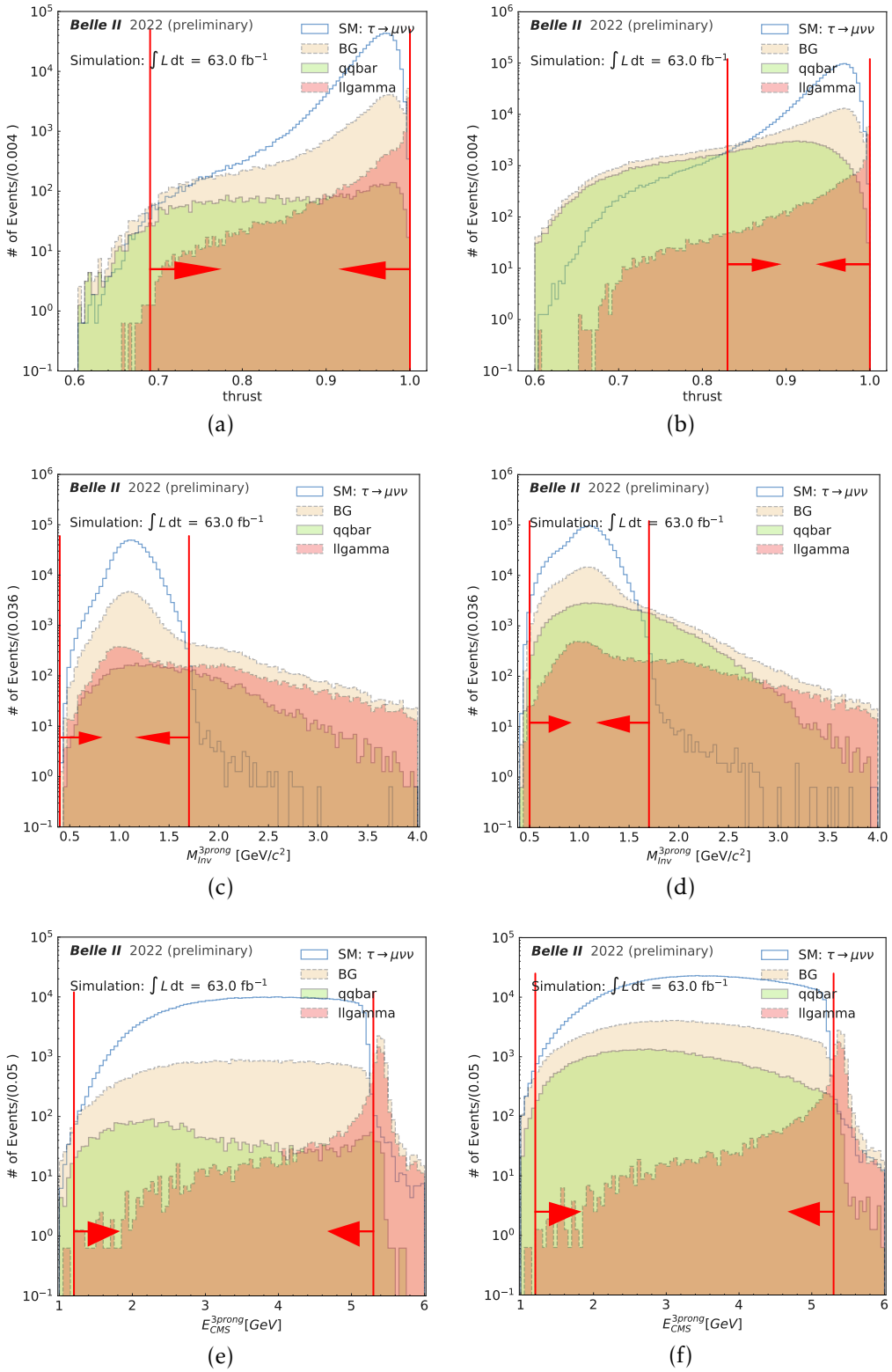
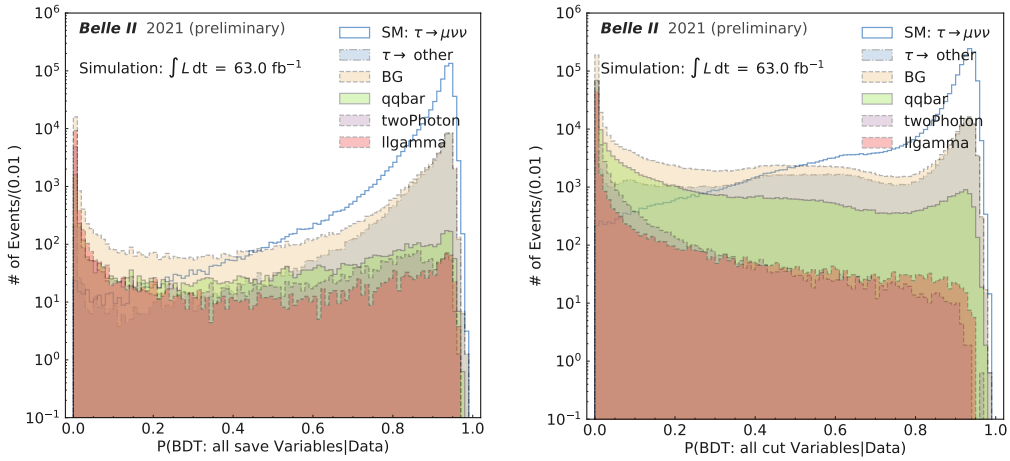


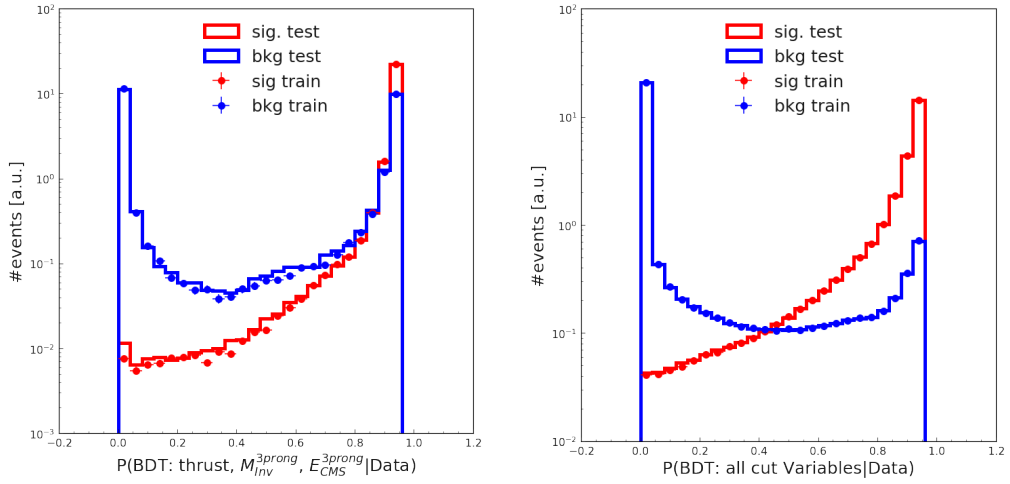
Figure G.2: $\tau \to \mu\nu_\mu\nu_\tau$ channel distributions for the background suppression variables. The left side shows the case with a neutrals veto, the right side with FOM optimised neutrals. Red markers indicate the FOM optimised selection.

G.1.2 BDT Based Event Selection



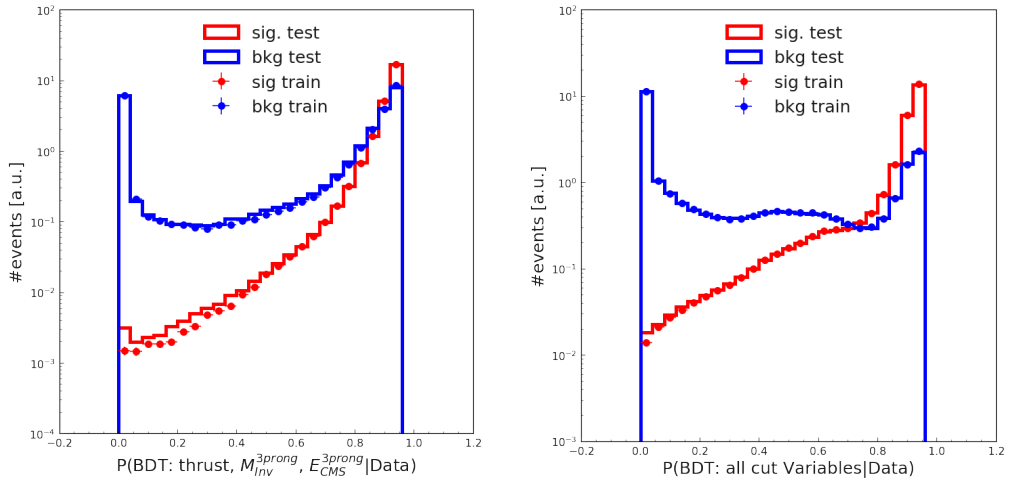
(a) (b)

Figure G.3: $\tau \to \mu\nu_\mu\nu_\tau$ channel distributions for the BDT without Neutrals classification (a) and BDT with Neutrals classification (b). The BDT classification results display the non τ background contributions.



(a) (b)

Figure G.4: $\tau \to e\nu_e\nu_\tau$ channel distributions for the BDT without Neutrals classification (a) and BDT with Neutrals classification(b). The plots compare the BDT ouput for the training and test data samples.



(a) (b)

Figure G.5: $\tau \to \mu\nu_\mu\nu_\tau$ channel distributions for the BDT without Neutrals classification (a) and BDT with Neutrals classification(b)

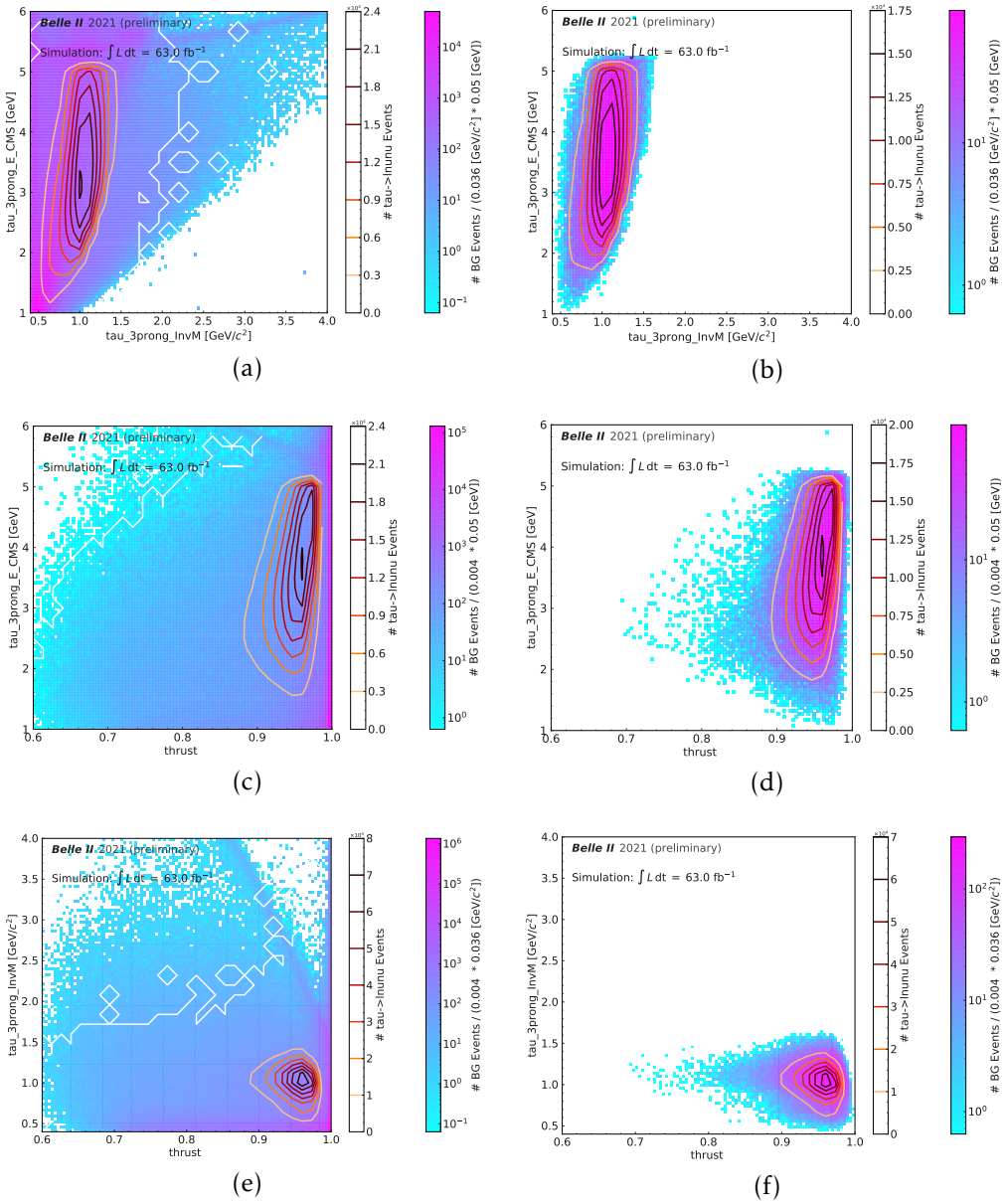


Figure G.6: $\tau \to e\nu_e\nu_\tau$ channel 2D distributions for the background unsuppressed (left) and suppressed with the BDT with Neutrals reproducing the best cut purity.

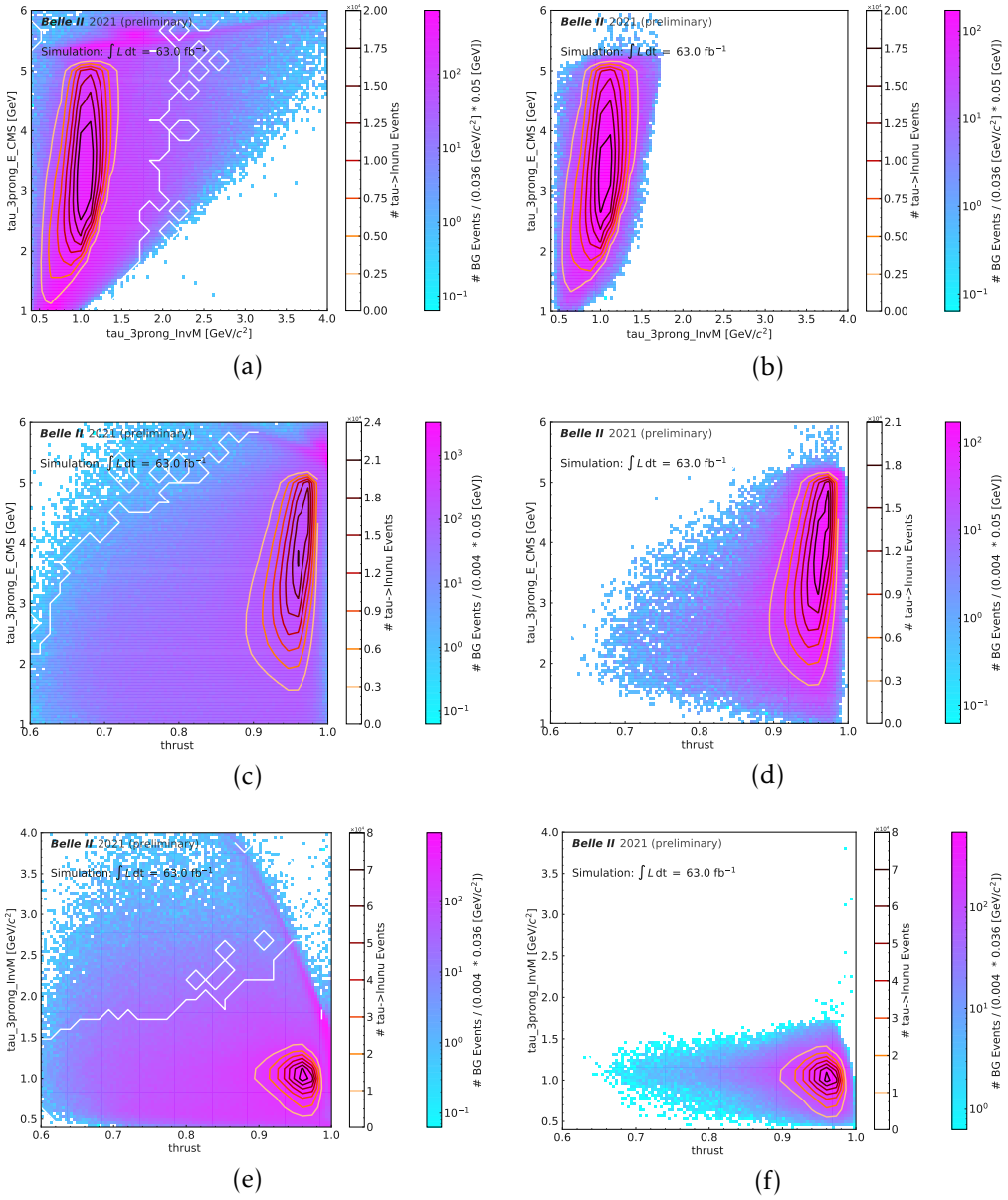


Figure G.7: $\tau \to \mu\nu_\mu\nu_\tau$ channel 2D distributions for the background unsuppressed (left) and suppressed with the BDT with Neutrals reproducing the best cut purity.

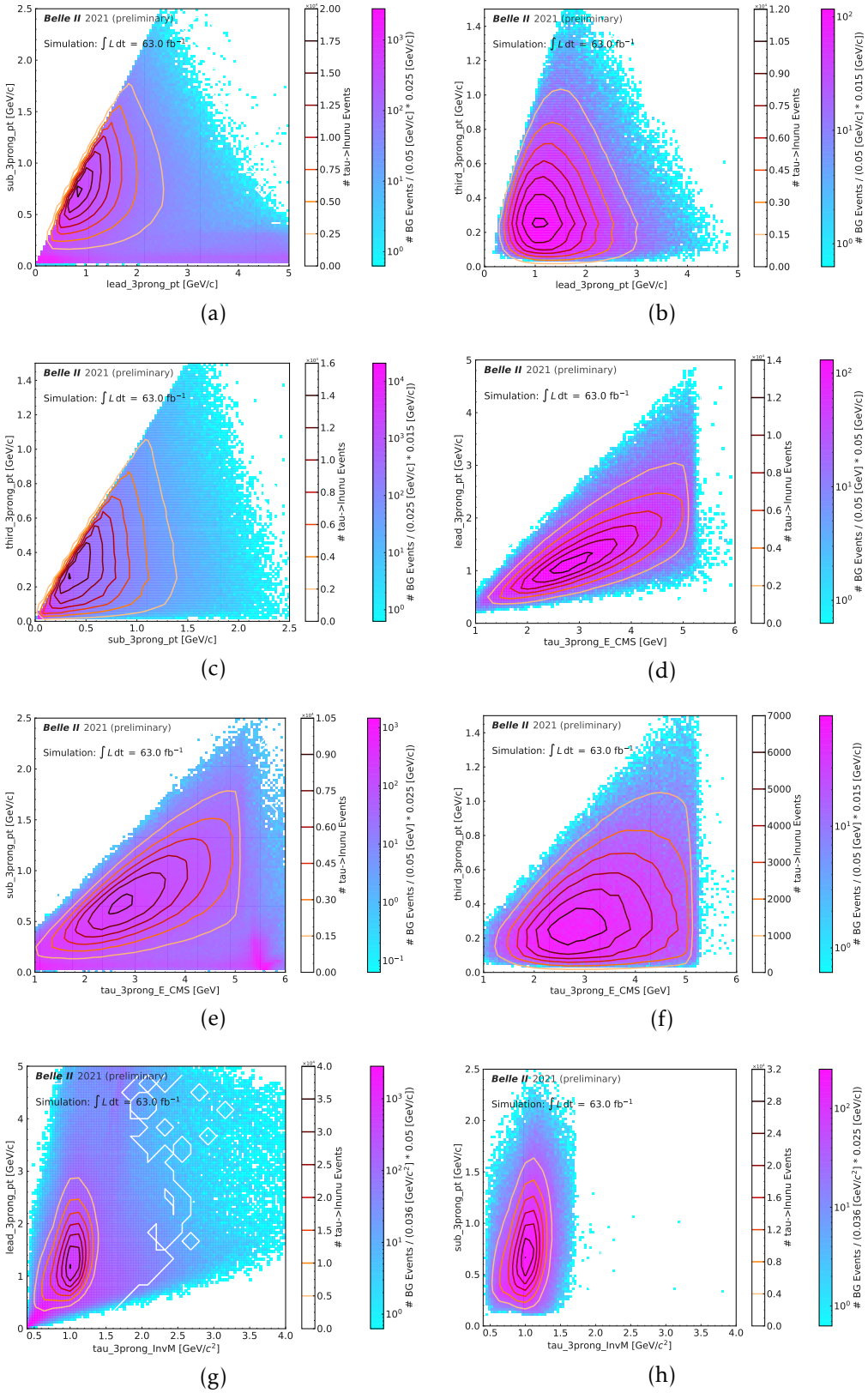


Figure G.8: $\tau \to \mu\nu_\mu\nu_\tau$ channel 2D distributions of background and signal for the ranked p_T variables.

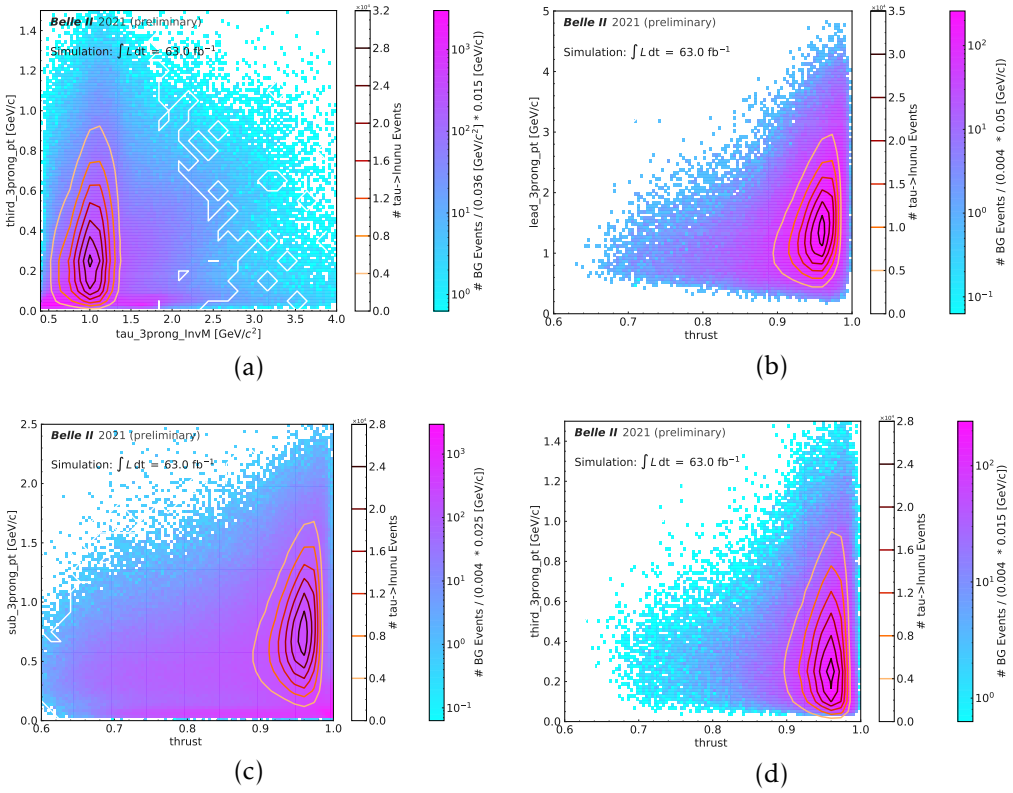
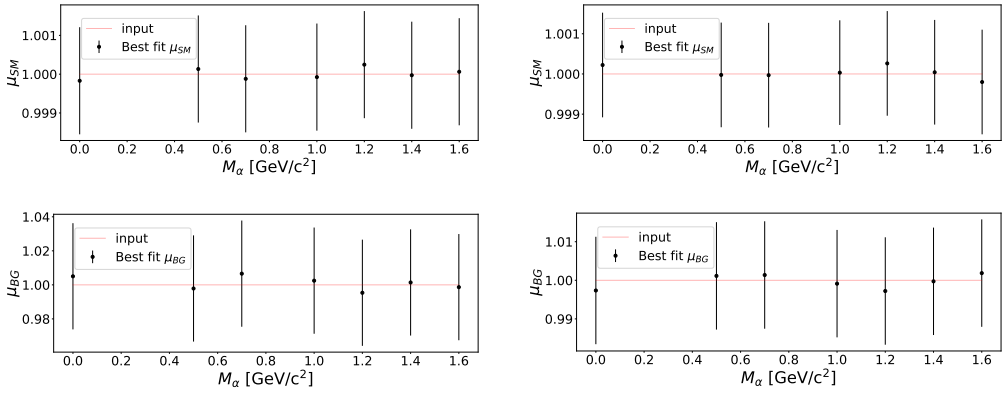


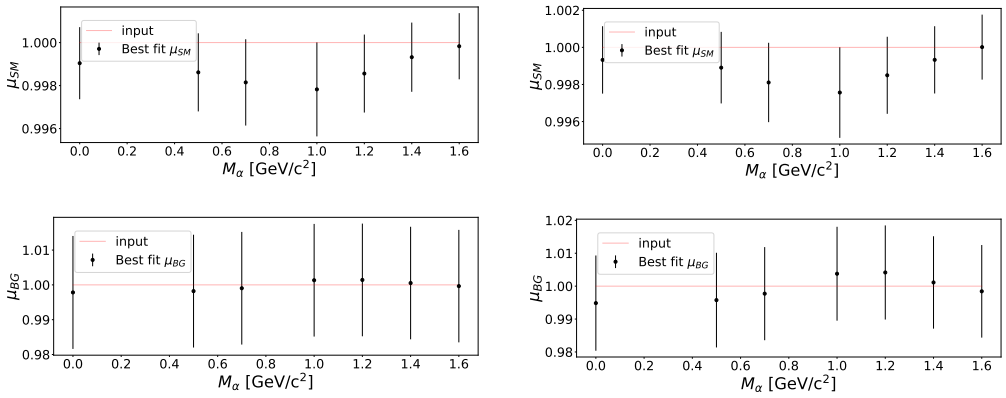
Figure G.9: $\tau \to \mu \nu_\mu \nu_\tau$ channel 2D distributions of background and signal for the ranked p_T variables.

G.2 Search for an Invisible Boson in Tau to one Lepton Decays



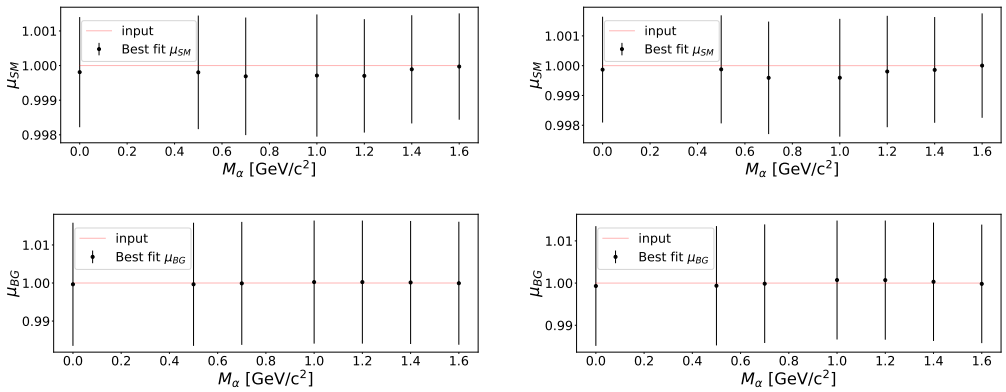
(c) $\tau \to e\alpha$ (d) $\tau \to \mu\alpha$

Figure G.10: Mean fit-parameter results for 1000 toy-data simulations from pyhf, in the nominal fit case, compared to toy-data simulation input.



(c) $\tau \to e\alpha$ (d) $\tau \to \mu\alpha$

Figure G.11: Mean fit-parameter results for 1000 toy-data simulations from Julia BAT using a flat prior in the nominal fit case, compared to toy-data simulation input H_0.



(c) $\tau \to e\alpha$ (d) $\tau \to \mu\alpha$

Figure G.12: Mean fit-parameter results for 1000 toy-data simulations from Julia BAT using a log uniform prior in the nominal fit case, compared to toy-data simulation input H_0.

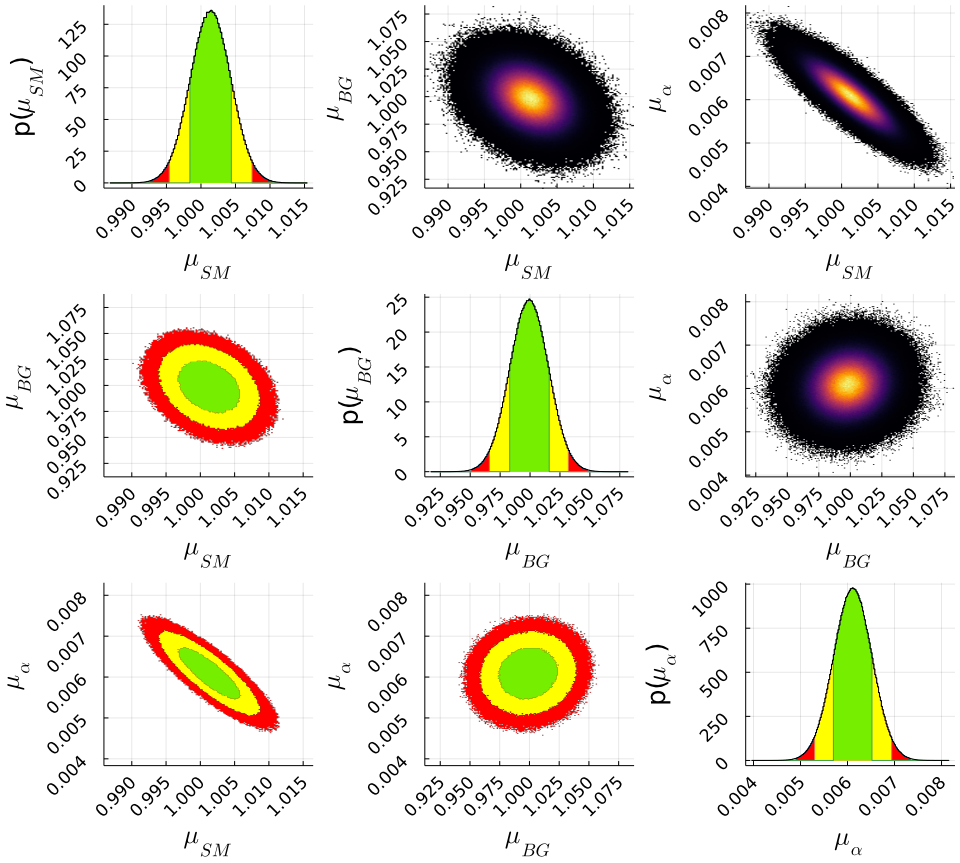


Figure G.13: BAT Marcov Chain Monte Varlo sampling result for the toy-data sample with injected signal using the $m_{\alpha} = 1.0$ GeV/c^2 mass hypothesis for the new physics and a flat prior for μ_{α}. The 68%, 95%, and 99% smallest intervals are green, yellow, and red, respectively.

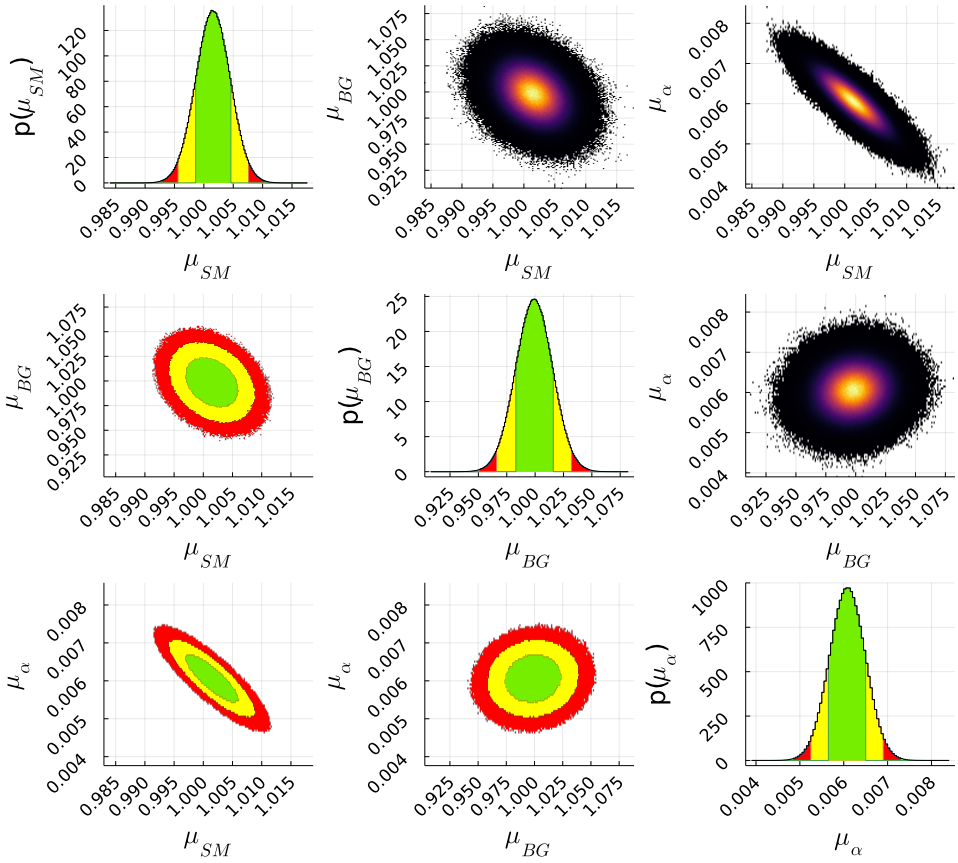
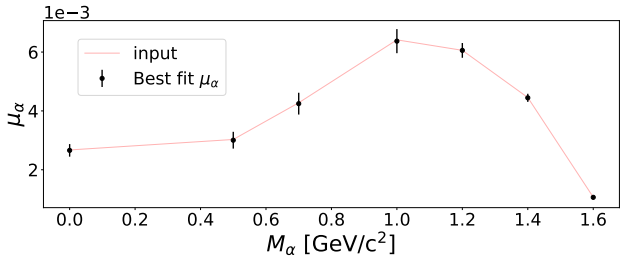
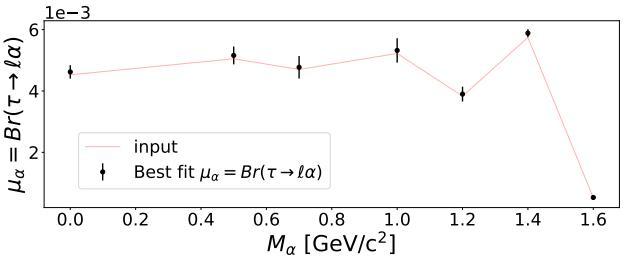


Figure G.14: BAT Marcov Chain Monte Varlo sampling result for toy-data sample with injected signal using the $m_{\alpha} = 1.0$ GeV/c^2 mass hypothesis for the new physics and a log uniform prior for μ_{α}. The 68%, 95%, and 99% smallest intervals are green, yellow, and red, respectively.



(a) $\mu_\alpha = Br(\tau \to e\alpha)$



(b) $\mu_\alpha = Br(\tau \to \mu\alpha)$

Figure G.15: Mean of the sampling result's mean results for 1000 toy-data simulations for Julia BAT with log uniform prior, in the nominal fit case, compared to toy-data simulation input H_S.

G.3 Data Simulation Validation

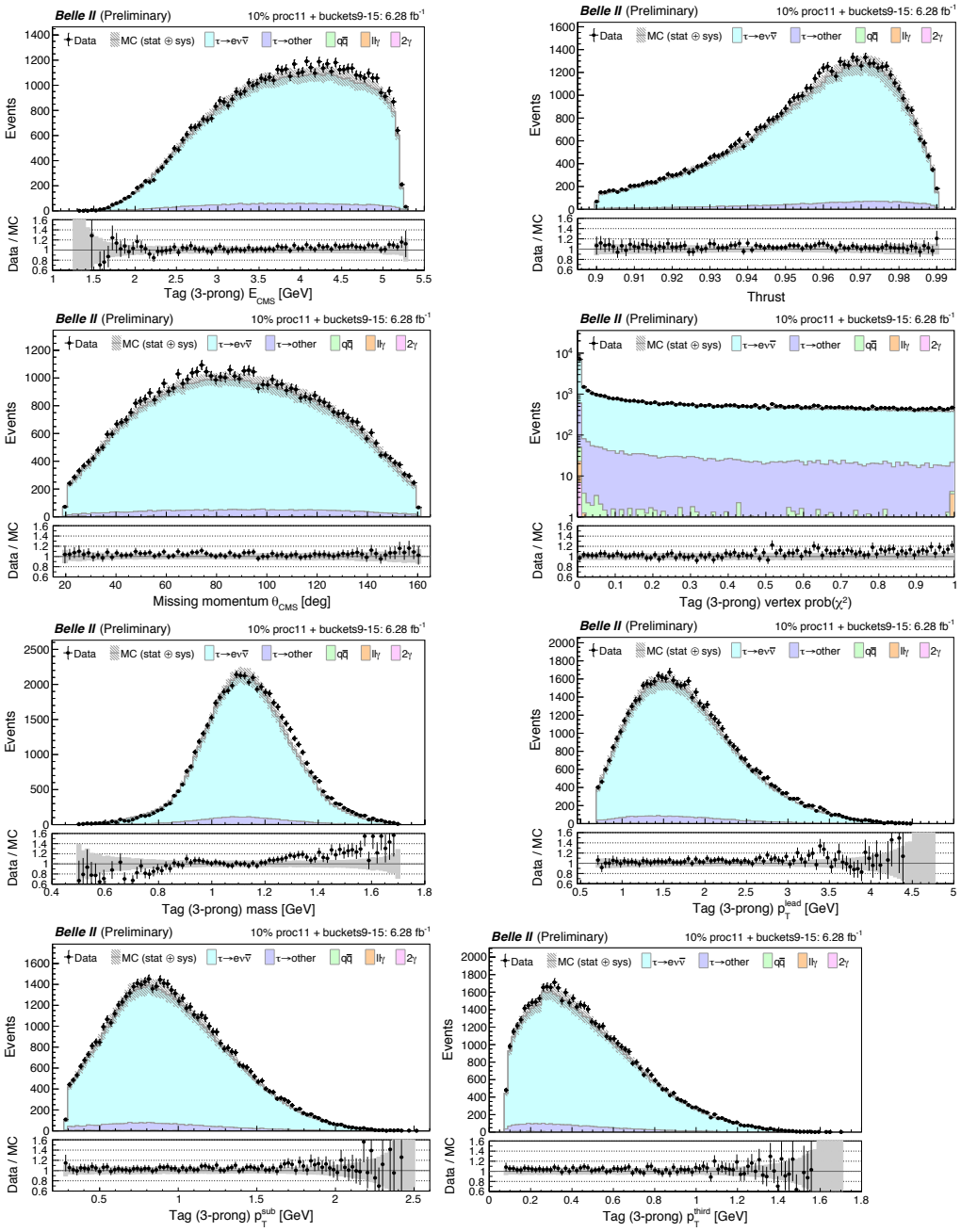


Figure G.16: Measured vs. simulated data distributions in the electron channel signal region. The simulated data error band includes all the uncertainties listed at the start of Section 3.6.2

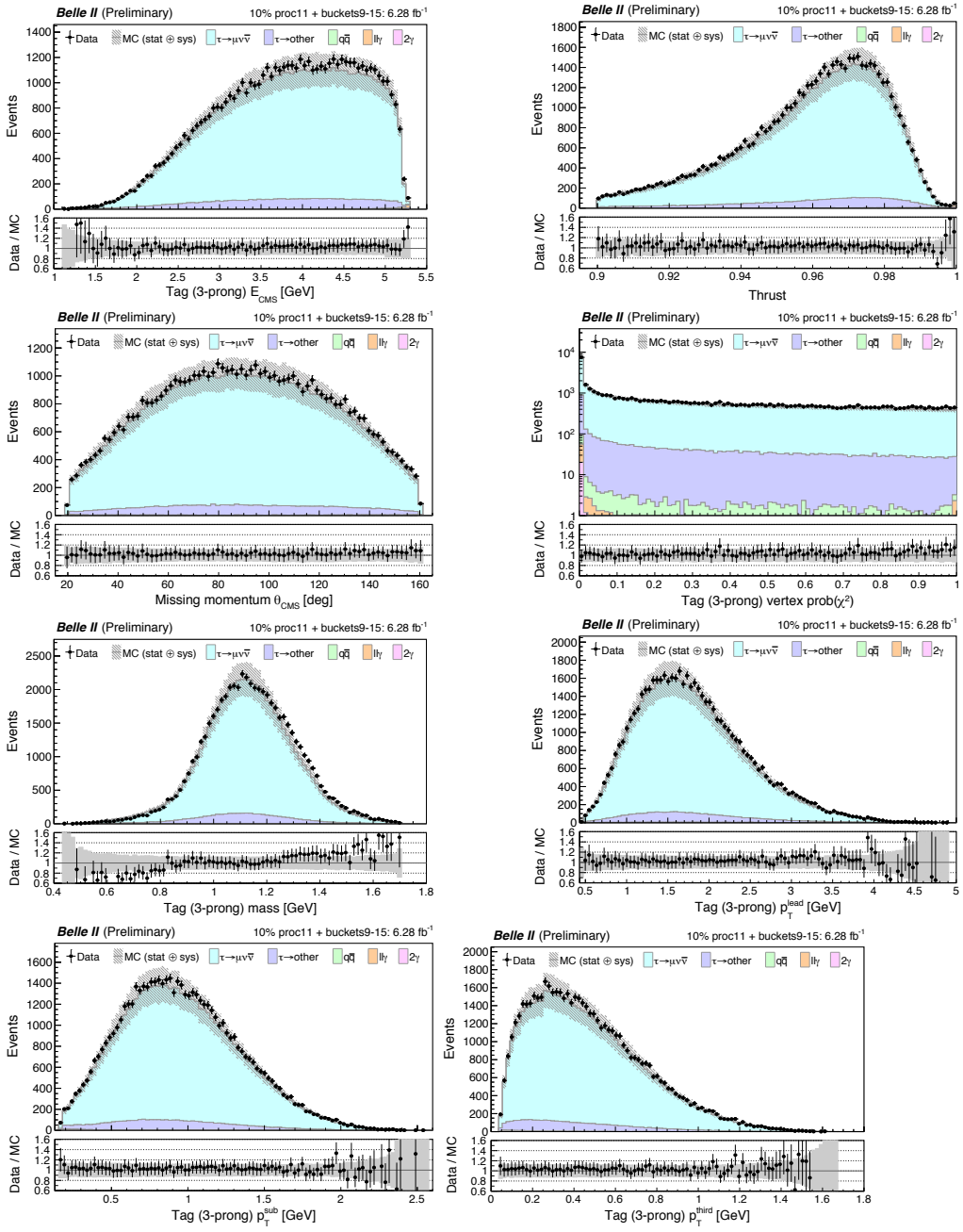


Figure G.17: Measured vs. simulated data distributions in the muon channel signal region. The simulated data error band includes all the uncertainties listed at the start of Section 3.6.2

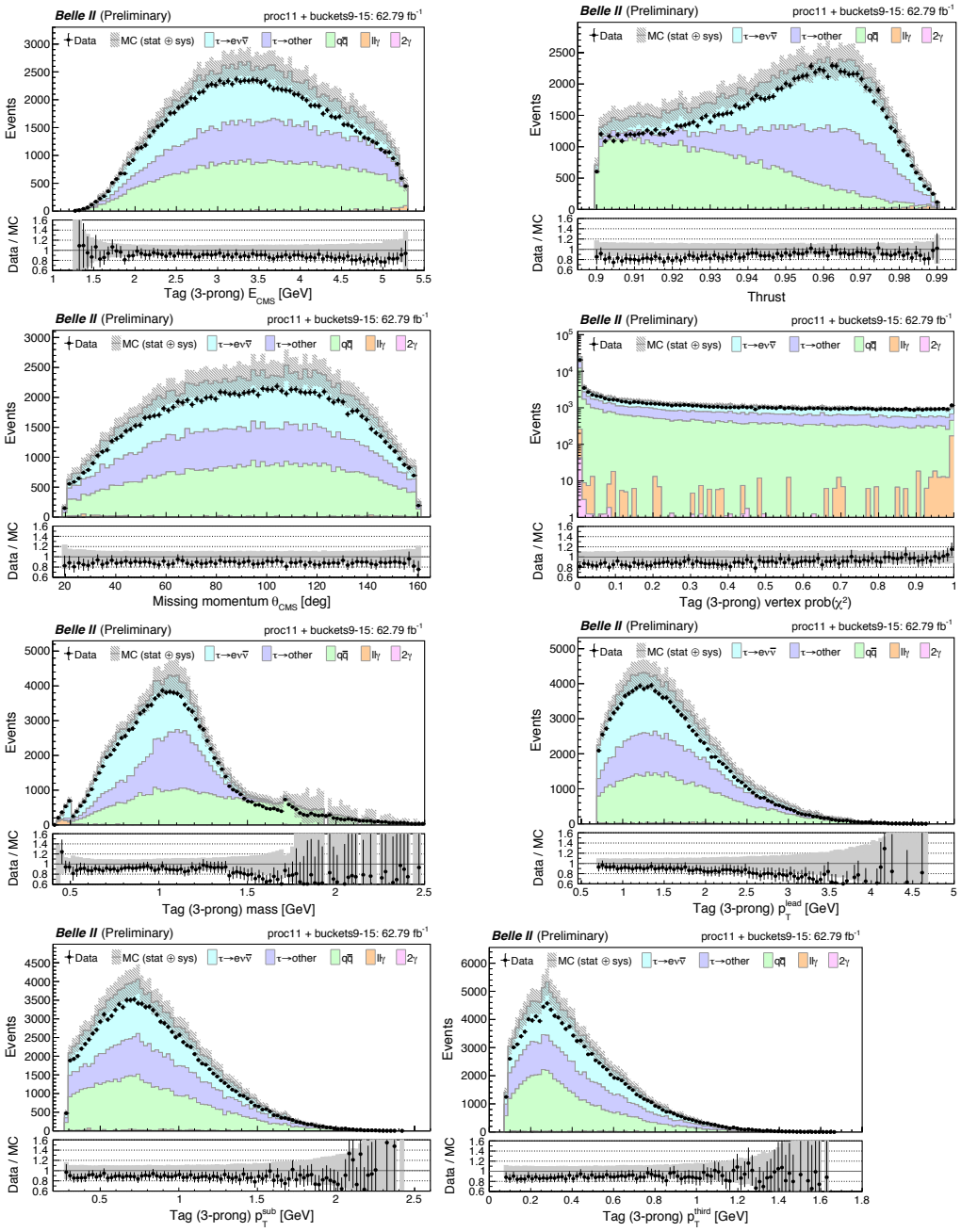


Figure G.18: Measured vs. simulated data distributions in the electron channel sideband. The simulated data error band includes all the uncertainties listed at the start of Section 3.6.2

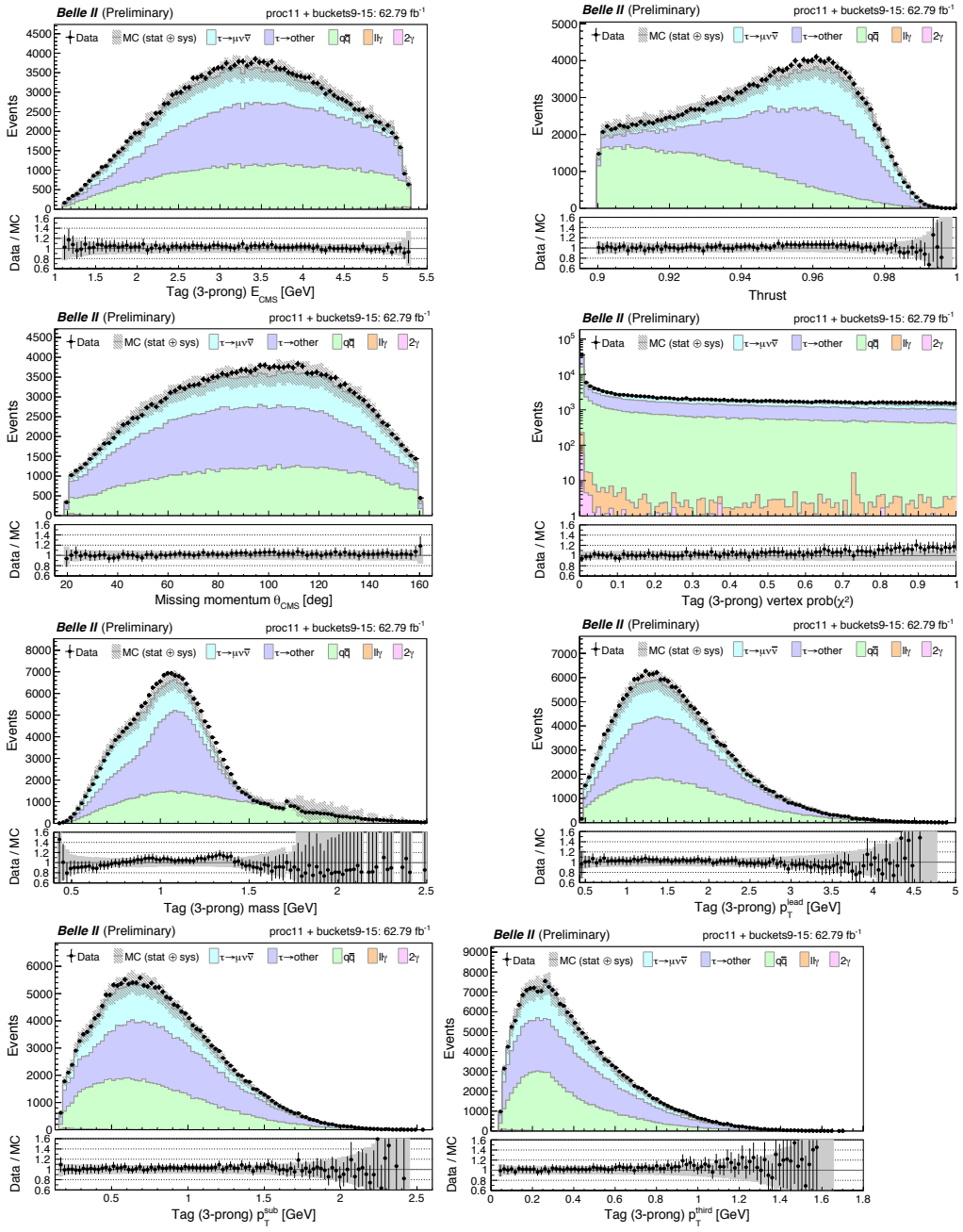
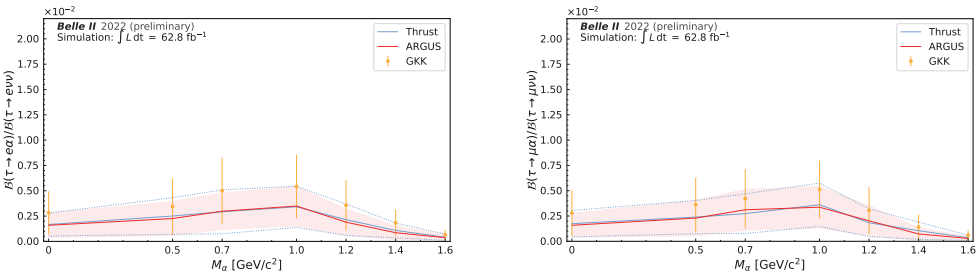


Figure G.19: Measured vs. simulated data distributions in the muon channel sideband. The simulated data error band includes all the uncertainties listed at the start of Section 3.6.2

G.4 Results

Variable Study



(a) Electron channel results.

(b) Muon channel results.

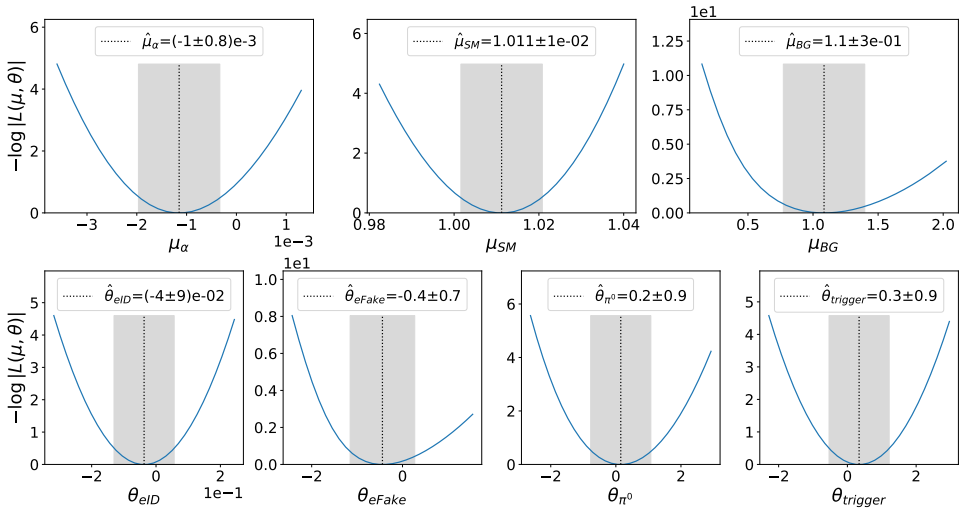
Figure G.20: Upper limit comparison for cifferent τ rest-frame approaches of Chapter 4 for Julia BAT using the log uniform prior.

Results for 62.8 fb^{-1} of Belle II Data

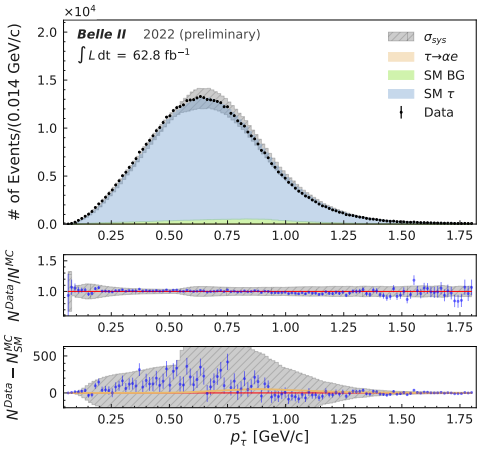
This section presents the summary of the pyhf fit output. Each figure summarises the information for one mass and the lepton hypothesis. At the top in Plot (a), we can consider the profile likelihood plots (blue); they include the best-fit value (black dotted line) and uncertainty (gray band).

Plot (b) compares the best-fit model distribution to data. The topmost part is the actual observed distribution (Data) and fitted template model (blue, green, and orange). Here, *SM* τ denotes the Standard Model $\tau \to \ell\nu_\tau\nu_\ell$-decay and *SM BG* denotes the other Standard Model processes. The gray band represents all uncertainties; for this one-dimensional display, the uncertainties are added in quadrature.

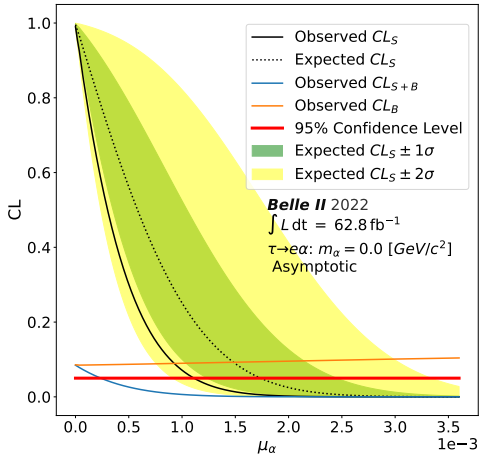
Plot (c) shows the CL_S curve scan distribution for the corresponding $\tau \to \ell\alpha$ channel and mass hypothesis determined on measured data. The *Observed* curves result from the evaluated measured data set. CL_S is the ratio of p-values for the H_0, denoted as CL_B, and the H_S, denoted as CL_{S+B}. The test statistic distributions give the *Expected* CL_S curve. The green and yellow bands denote the 1σ and 2σ variation of the Expected CL_S curve, respectively.



(a) Profile Liklihood Distributions for all Model Parameters

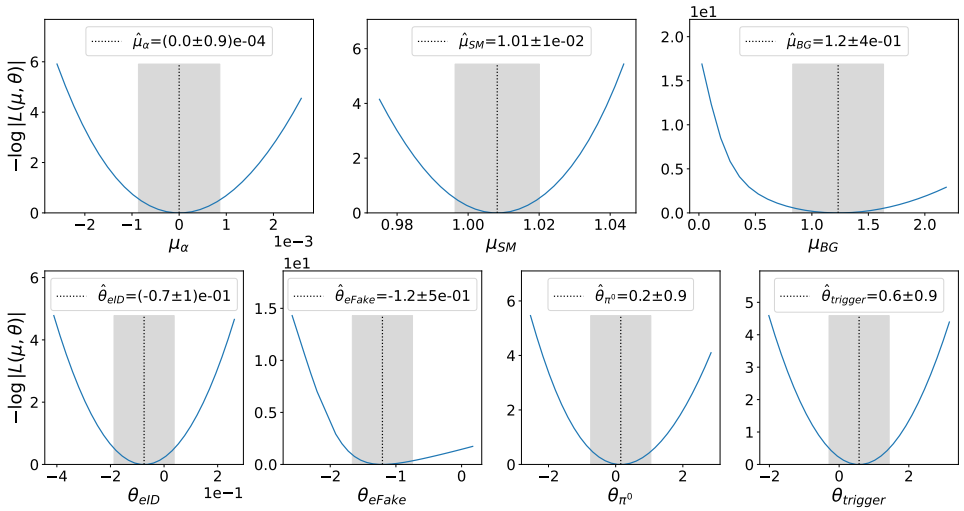


(b) Nominal Best-Fit Distribution

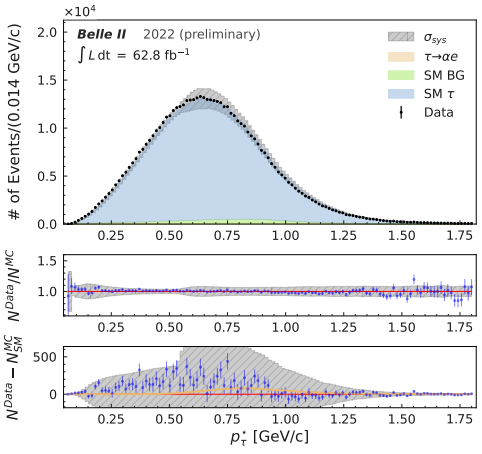


(c) Nominal CL_S Distribution

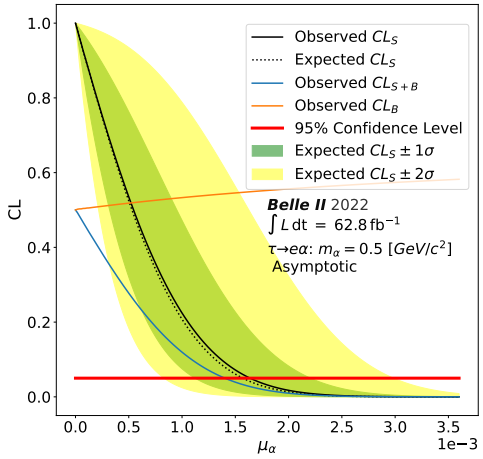
Figure G.21: pyhf fit summary for $\tau \to e\alpha$: $M_\alpha = 0.0\,\mathrm{GeV}/c^2$. A detailed description of the plots is given ablove in the introduction of Section G.4.



(a) Profile Liklihood Distributions for all Model Parameters

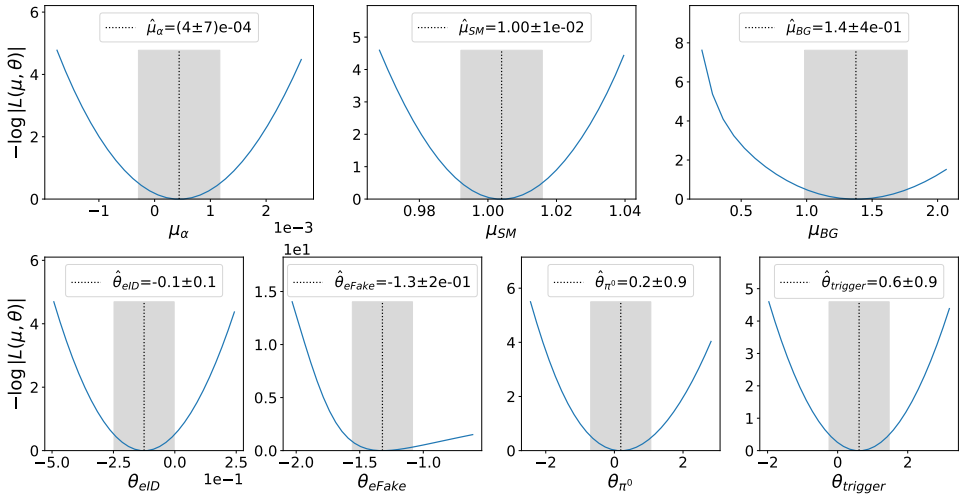


(b) Nominal Best-Fit Distribution

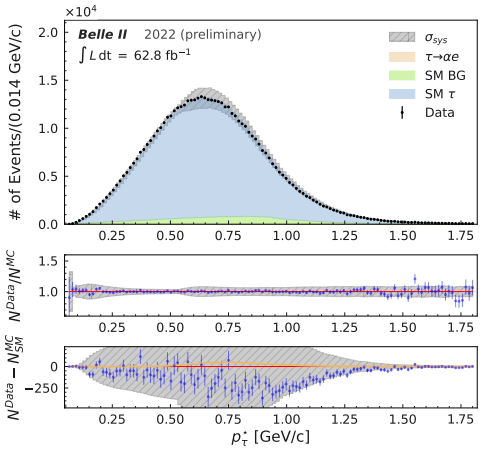


(c) Nominal CL_S Distribution

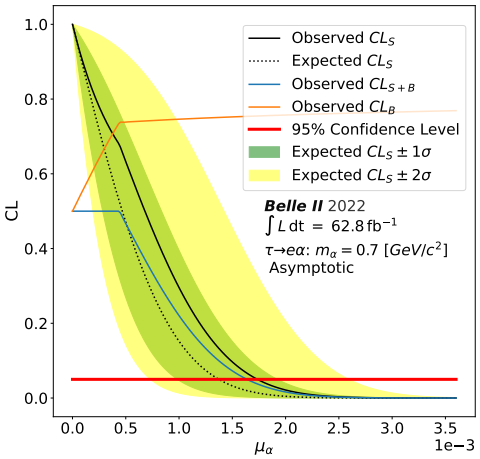
Figure G.22: pyhf fit summary for $\tau \to e\alpha$: $M_\alpha = 0.5\,\mathrm{GeV}/c^2$. A detailed description of the plots is given ablove in the introduction of Section G.4.



(a) Profile Liklihood Distributions for all Model Parameters

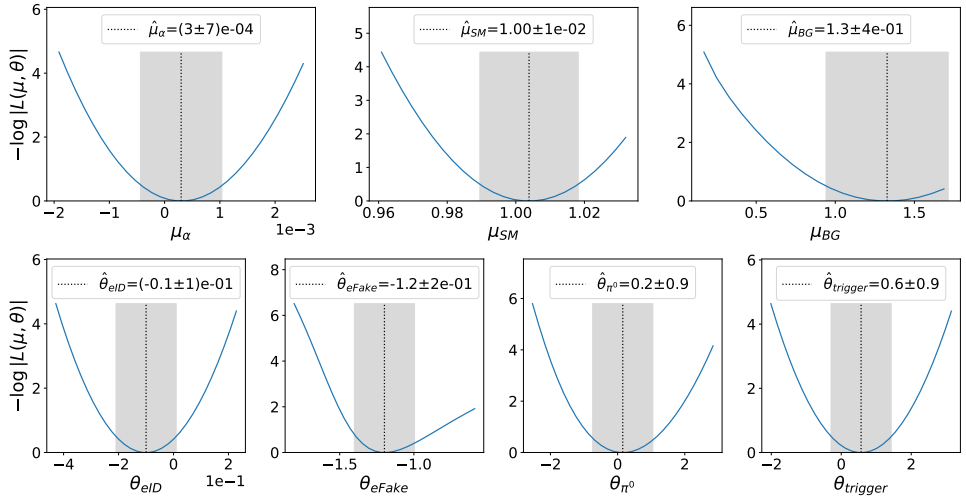


(b) Nominal Best-Fit Distribution

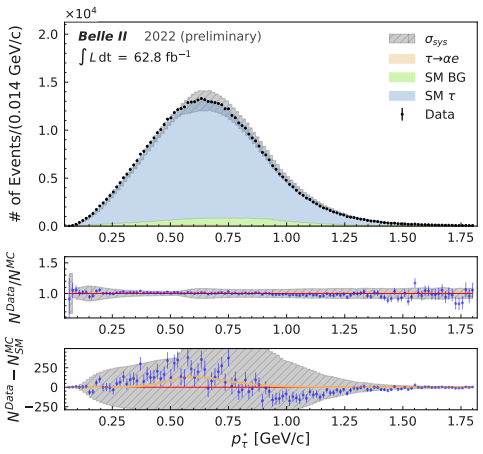


(c) Nominal CL_S Distribution

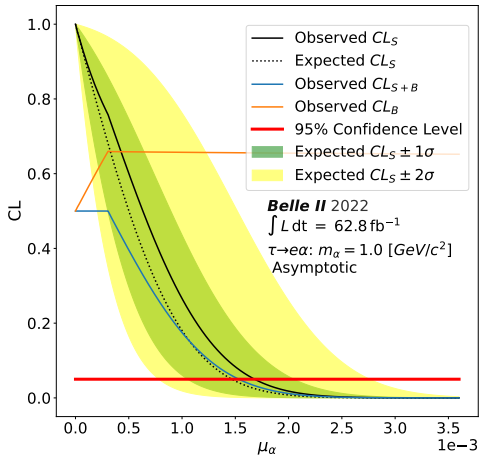
Figure G.23: pyhf fit summary for $\tau \to e\alpha$: $M_\alpha = 0.7\,\mathrm{GeV}/c^2$. A detailed description of the plots is given ablove in the introduction of Section G.4.



(a) Profile Liklihood Distributions for all Model Parameters

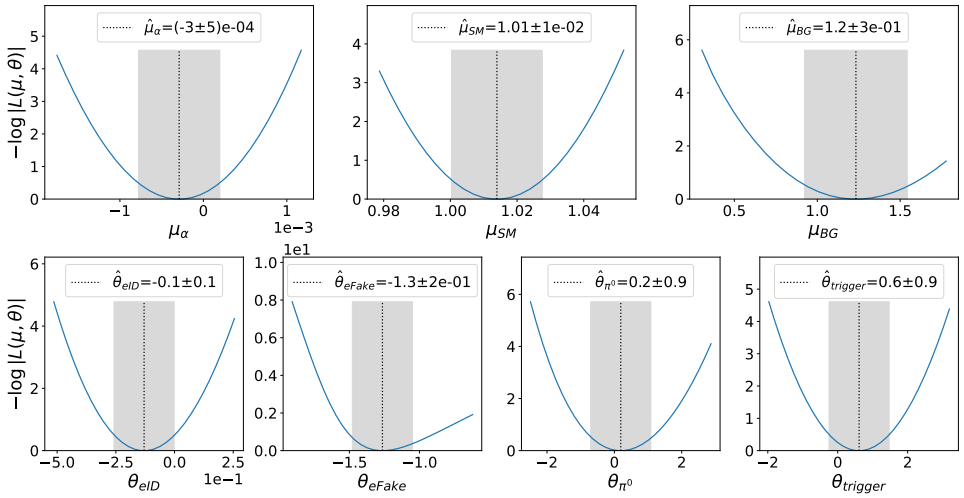


(b) Nominal Best-Fit Distribution

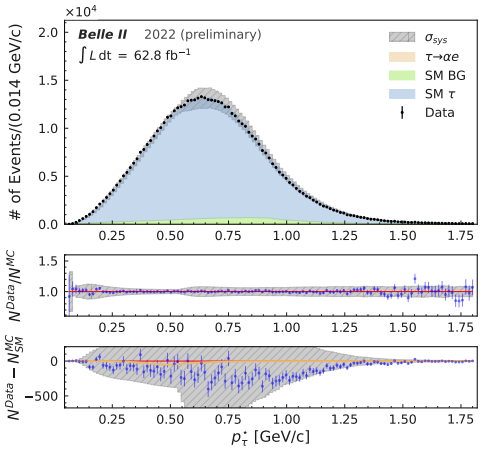


(c) Nominal CL_S Distribution

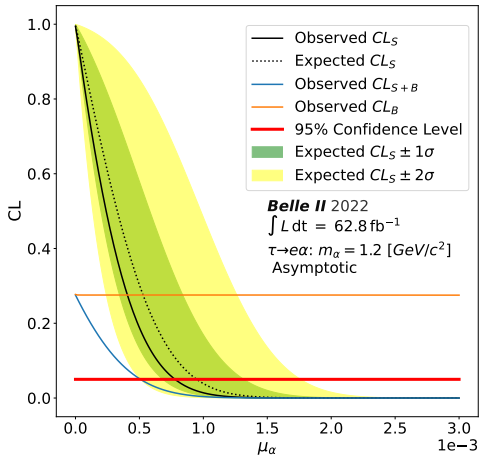
Figure G.24: pyhf fit summary for $\tau \to e\alpha$: $M_\alpha = 1.0\,\mathrm{GeV}/c^2$. A detailed description of the plots is given ablove in the introduction of Section G.4.



(a) Profile Liklihood Distributions for all Model Parameters

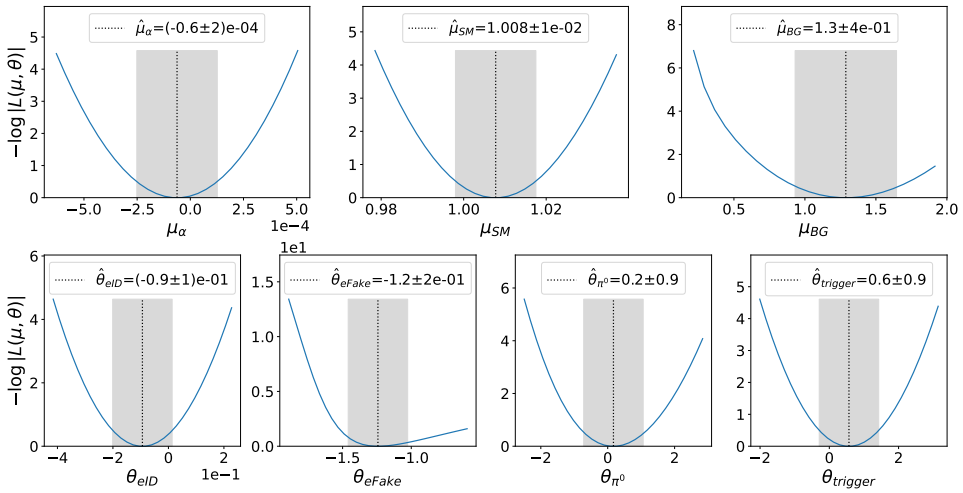


(b) Nominal Best-Fit Distribution

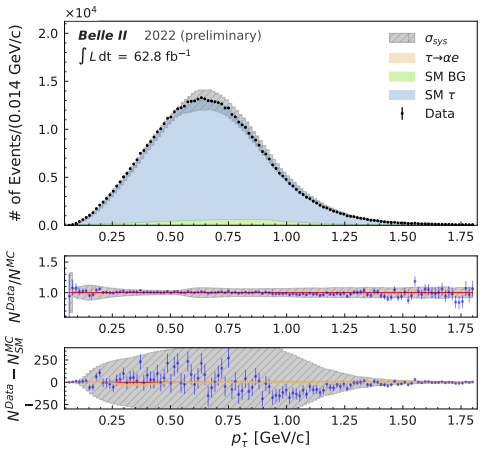


(c) Nominal CL_S Distribution

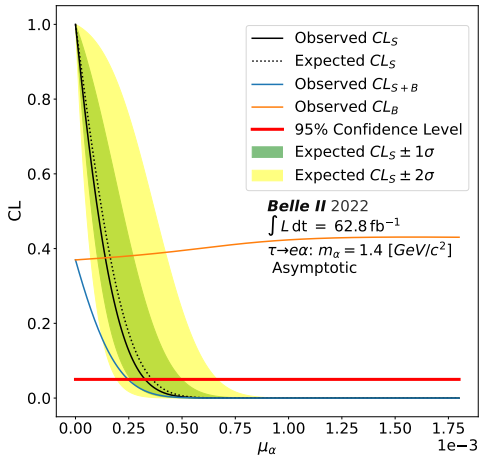
Figure G.25: pyhf fit summary for $\tau \to e\alpha$: $M_\alpha = 1.2\,\mathrm{GeV}/c^2$. A detailed description of the plots is given ablove in the introduction of Section G.4.



(a) Profile Liklihood Distributions for all Model Parameters

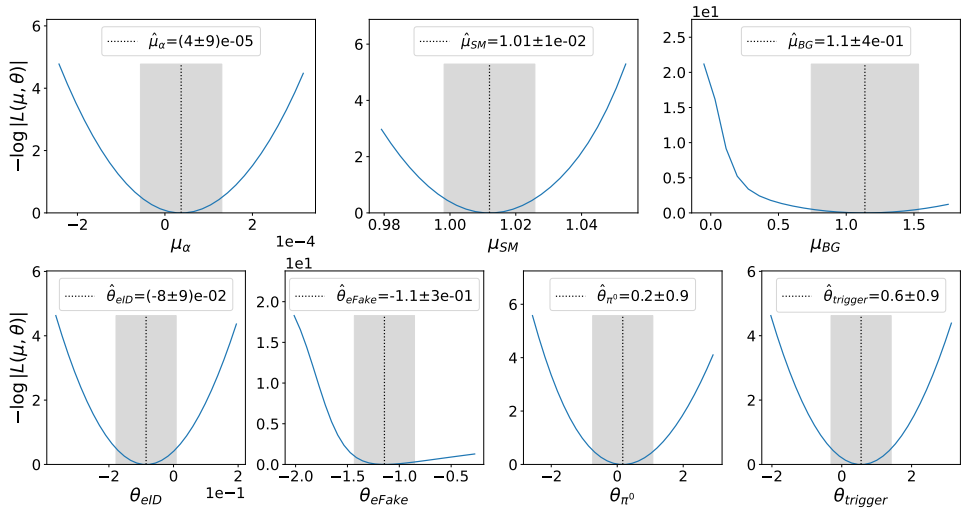


(b) Nominal Best-Fit Distribution

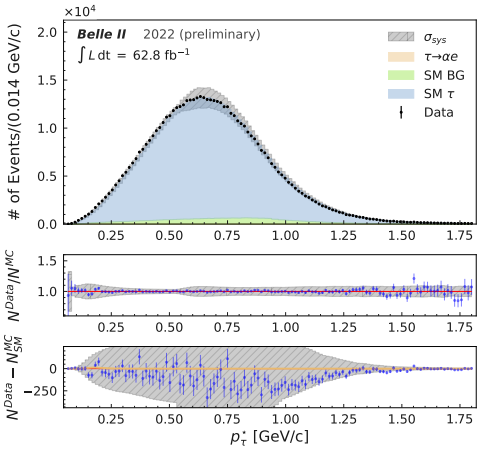


(c) Nominal CL_S Distribution

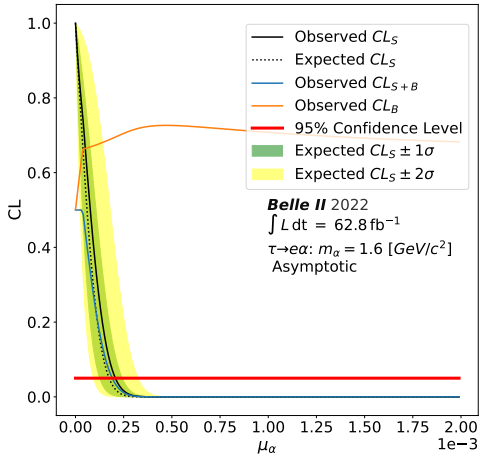
Figure G.26: pyhf fit summary for $\tau \to e\alpha$: $M_\alpha = 1.4\text{GeV}/c^2$. A detailed description of the plots is given ablove in the introduction of Section G.4.



(a) Profile Liklihood Distributions for all Model Parameters

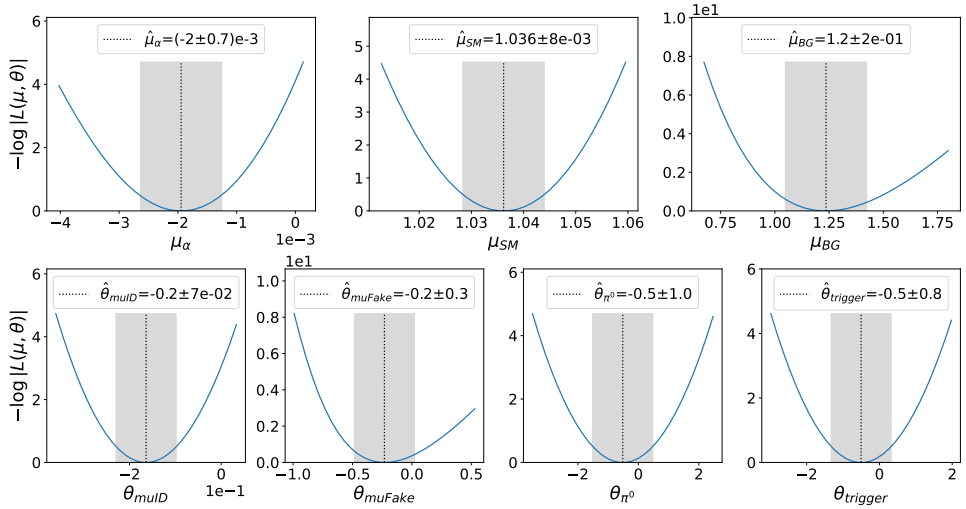


(b) Nominal Best-Fit Distribution

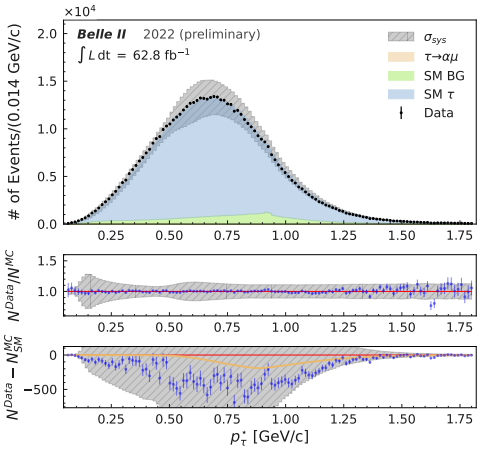


(c) Nominal CL_S Distribution

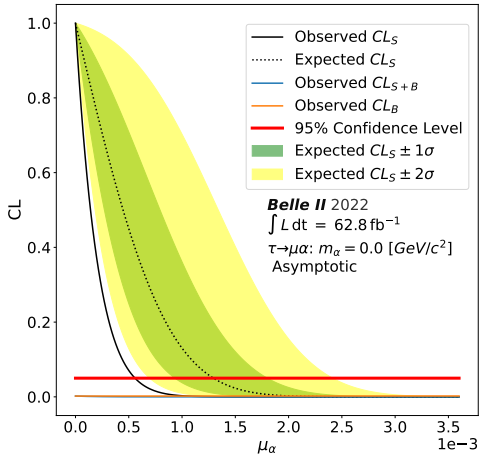
Figure G.27: pyhf fit summary for $\tau \to e\alpha$: $M_\alpha = 1.6\,\mathrm{GeV}/c^2$. A detailed description of the plots is given ablove in the introduction of Section G.4.



(a) Profile Liklihood Distributions for all Model Parameters

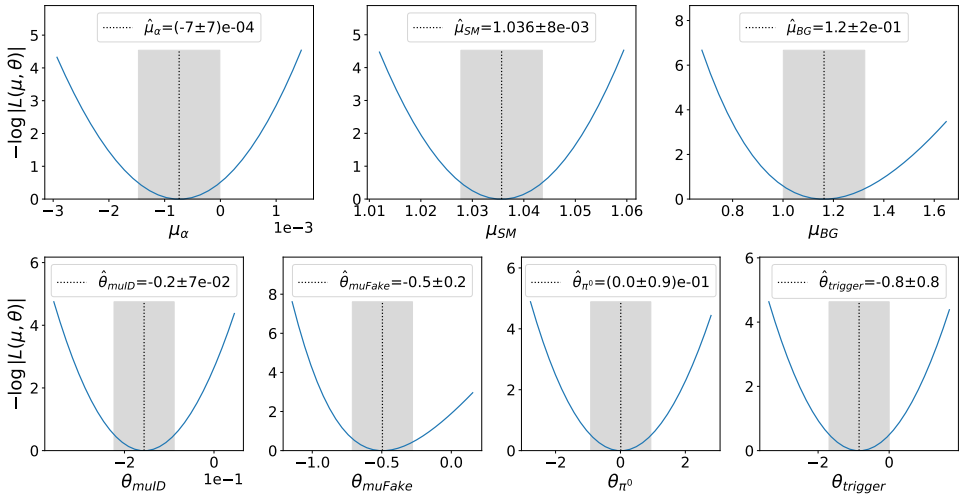


(b) Nominal Best-Fit Distribution

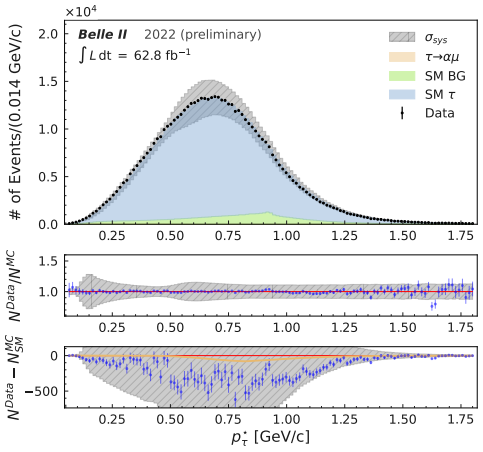


(c) Nominal CL_S Distribution

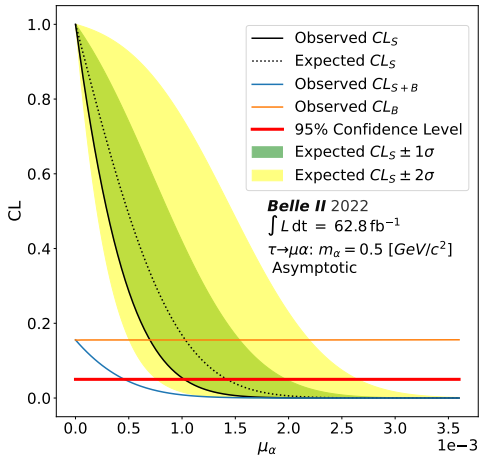
Figure G.28: pyhf fit summary for $\tau \to \mu\alpha$: $M_\alpha = 0.0\,\mathrm{GeV}/c^2$. A detailed description of the plots is given ablove in the introduction of Section G.4.



(a) Profile Liklihood Distributions for all Model Parameters

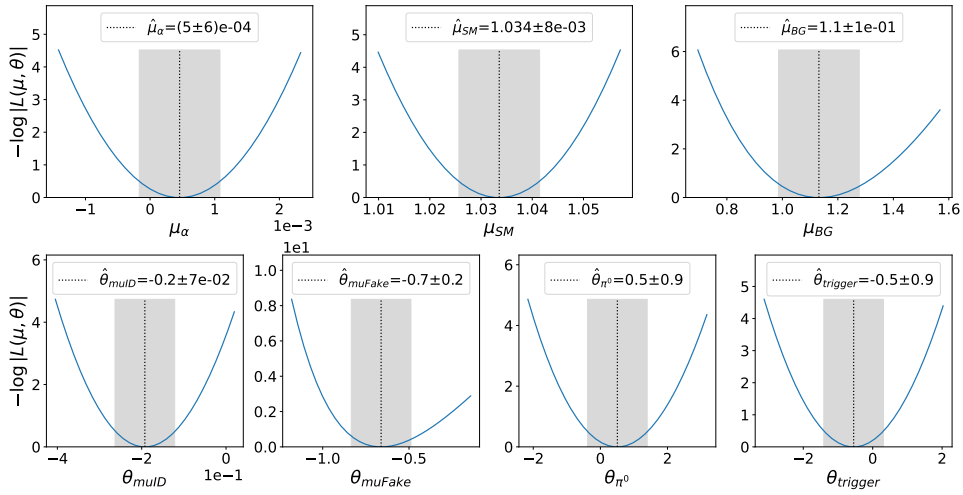


(b) Nominal Best-Fit Distribution

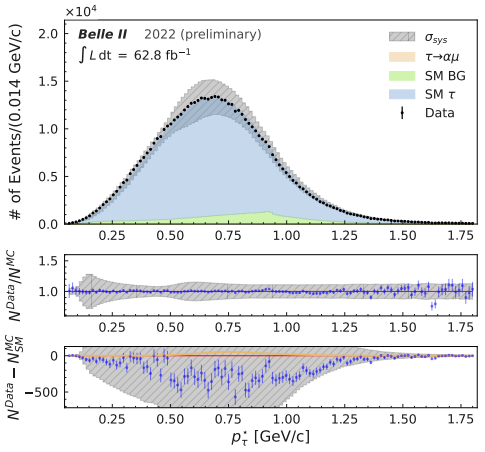


(c) Nominal CL_S Distribution

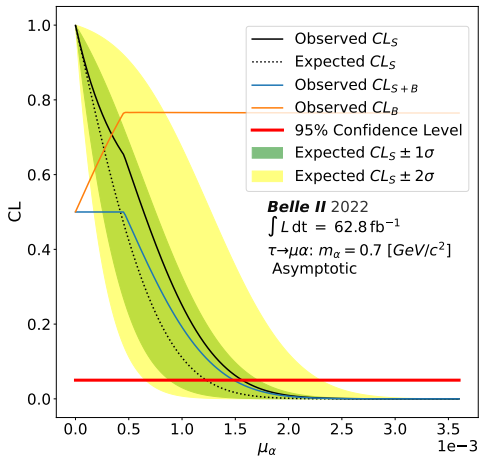
Figure G.29: pyhf fit summary for $\tau \to \mu\alpha$: $M_\alpha = 0.5\,\mathrm{GeV}/c^2$. A detailed description of the plots is given ablove in the introduction of Section G.4.



(a) Profile Liklihood Distributions for all Model Parameters

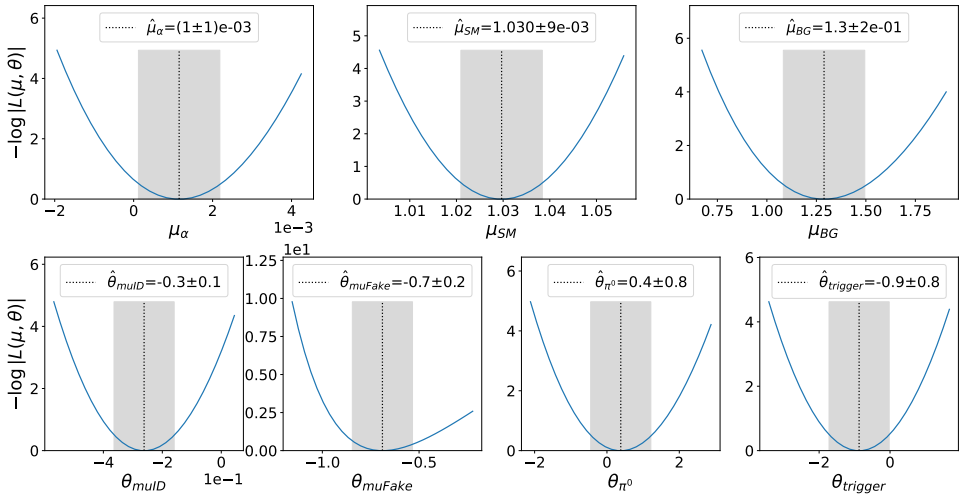


(b) Nominal Best-Fit Distribution

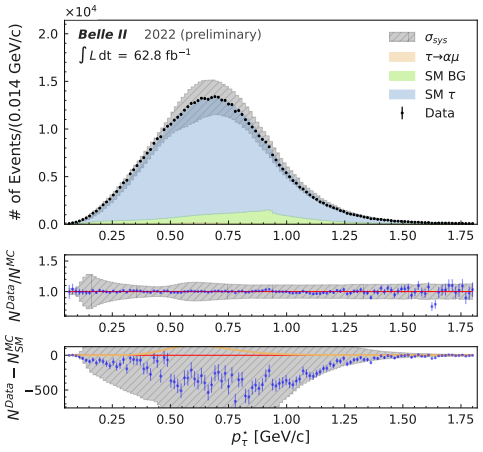


(c) Nominal CL_S Distribution

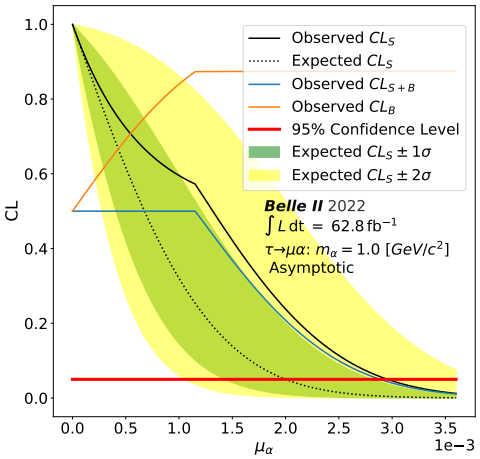
Figure G.30: pyhf fit summary for $\tau \to \mu\alpha$: $M_\alpha = 0.7\,\mathrm{GeV}/c^2$. A detailed description of the plots is given ablove in the introduction of Section G.4.



(a) Profile Liklihood Distributions for all Model Parameters

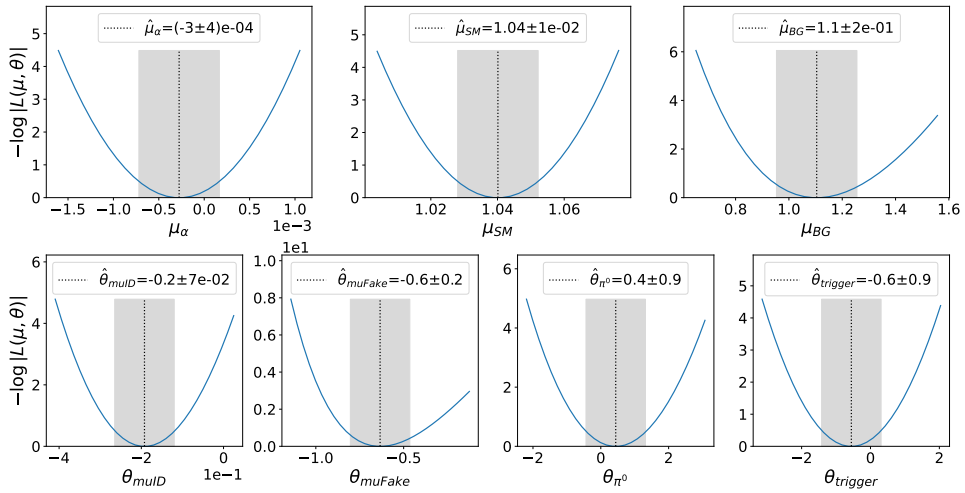


(b) Nominal Best-Fit Distribution

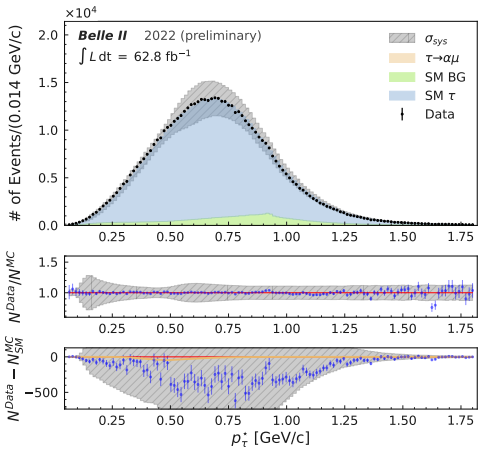


(c) Nominal CL_S Distribution

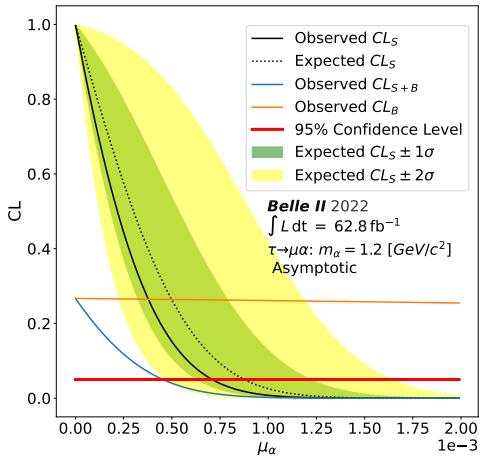
Figure G.31: pyhf fit summary for $\tau \to \mu\alpha$: $M_\alpha = 1.0\,\mathrm{GeV}/c^2$. A detailed description of the plots is given ablove in the introduction of Section G.4.



(a) Profile Liklihood Distributions for all Model Parameters

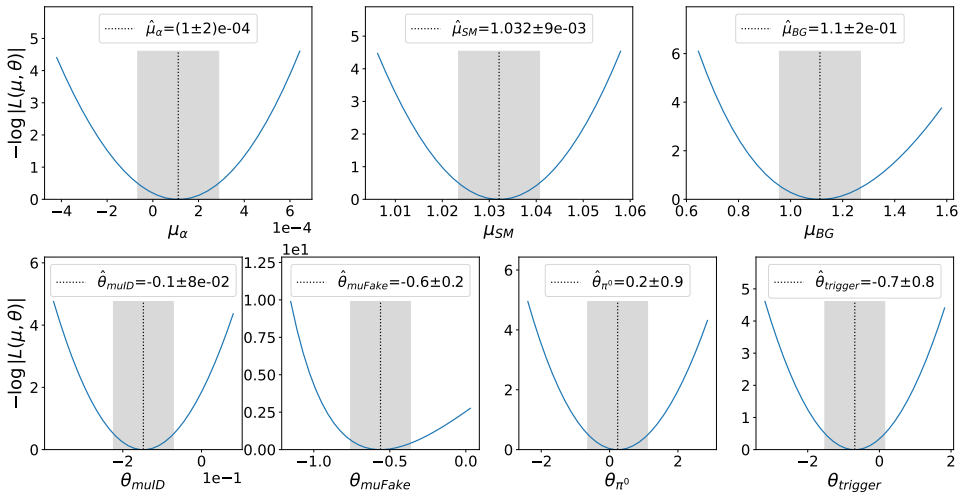


(b) Nominal Best-Fit Distribution

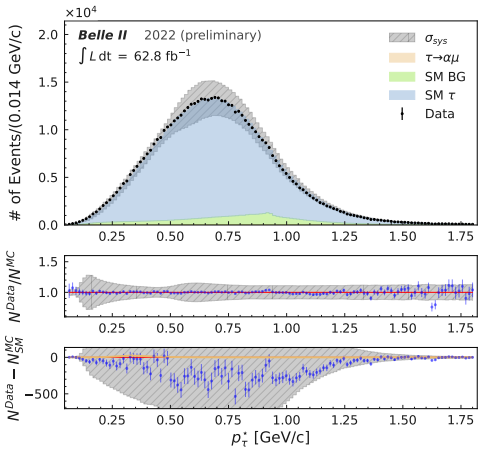


(c) Nominal CL_S Distribution

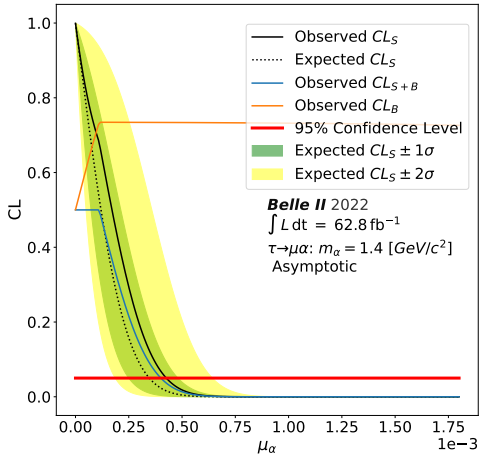
Figure G.32: pyhf fit summary for $\tau \to \mu\alpha$: $M_\alpha = 1.2\,\mathrm{GeV}/c^2$. A detailed description of the plots is given ablove in the introduction of Section G.4.



(a) Profile Liklihood Distributions for all Model Parameters

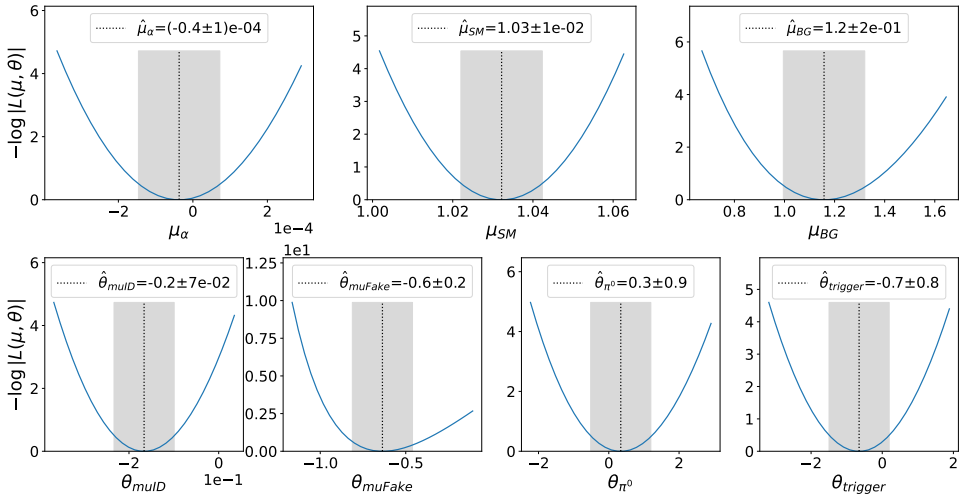


(b) Nominal Best-Fit Distribution

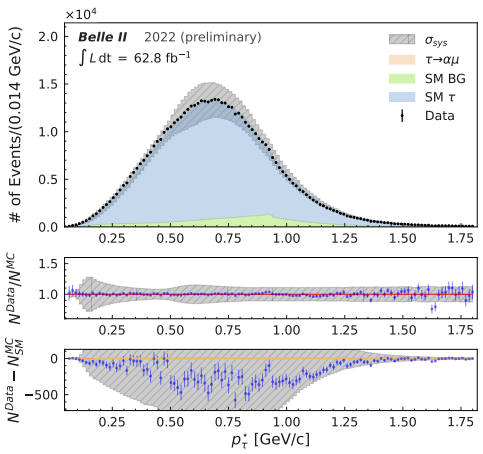


(c) Nominal CL_S Distribution

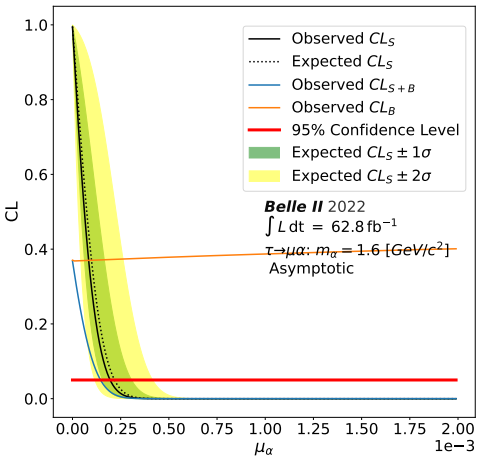
Figure G.33: pyhf fit summary for $\tau \to \mu\alpha$: $M_\alpha = 1.4\text{GeV}/c^2$. A detailed description of the plots is given ablove in the introduction of Section G.4.



(a) Profile Liklihood Distributions for all Model Parameters



(b) Nominal Best-Fit Distribution



(c) Nominal CL_S Distribution

Figure G.34: pyhf fit summary for $\tau \to \mu\alpha$: $M_\alpha = 1.6\,\mathrm{GeV}/c^2$. A detailed description of the plots is given ablove in the introduction of Section G.4.

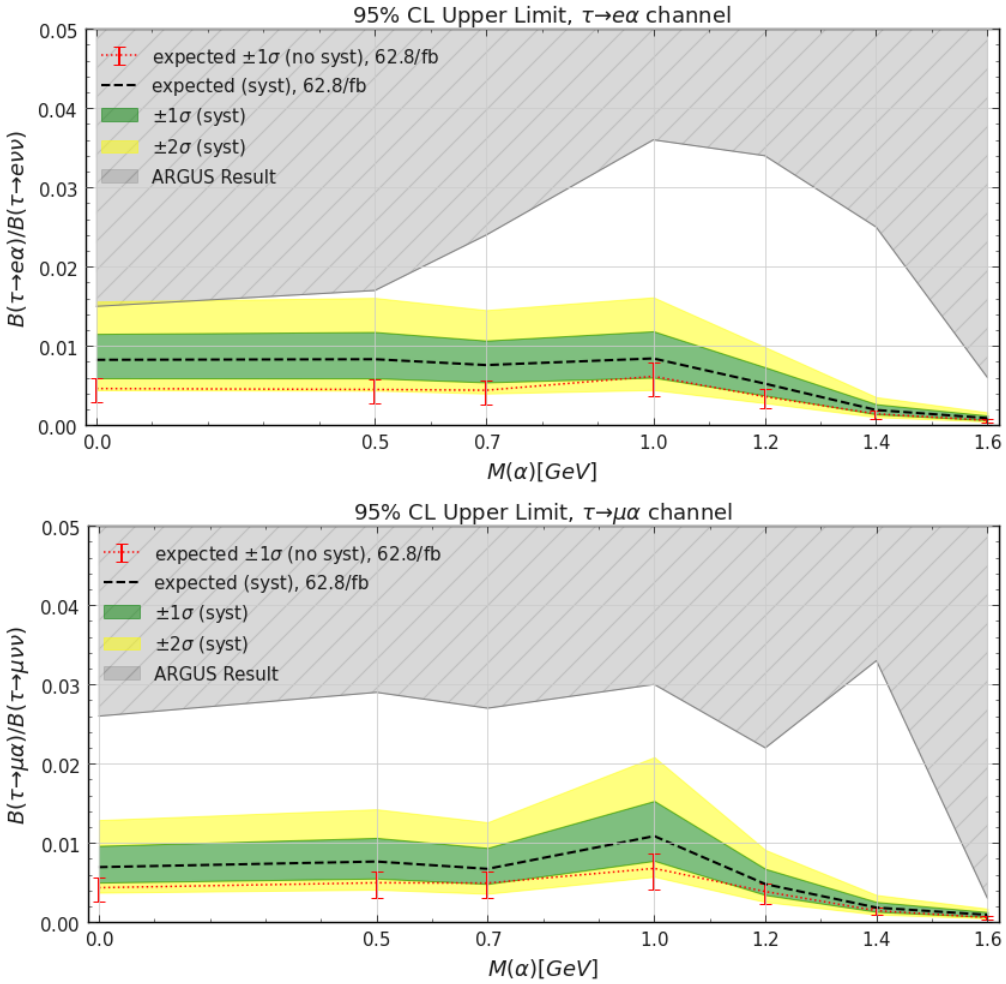


Figure G.35: The impact of the shape variation systematics on the upper limit expectation for the e channel (top) and μ channel (bottom) after increasing the luminosity of the background samples, re-binning and symmetrising. The expected upper limit of our fit model without smoothing and symmetrization and the corresponding uncertainty (red) is compared to the expected upper limit with the approved Belle II fit model (expectation as black dotted line, uncertainty indicated by the Brazilian band) and the previouse most stringent upper limit provided by the ARGUS collaboration (gray). This is a result obtained with the approved Belle II $\tau \to \ell\alpha$ search approach using RooStats. It is included in this thesis for illustration purposes.

Appendix H

Tables

H.1 Event Selection

H.1.1 Cut Based Event Selection

Table H.1: Remaining background events after cut based selection without neutrals. Here $\tau^+\tau^-$ (other) refers to tau pair events that are not signal, this is, the events that do not satisfy the conditions of the $\tau \to \ell\nu_\tau\nu_\ell$ signal definition in MC (see Appendix D).

E(Remaining Events)	(62.8 fb^{-1})	
BG Samples	e	μ
$\tau^+\tau^-$ (other)	17045.18	43575.04
$c\bar{c}$	521.24	1029.29
$s\bar{s}$	53.38	433.32
$d\bar{d}$	50.24	293.90
$u\bar{u}$	214.78	1358.99
$B^0\bar{B}^0$	43.33	72.85
B^+B^-	28.26	39.56
	low multiplicity	
$e^+e^-(\gamma)$	1073.88	0.0
$\mu^+\mu^-(\gamma)$	0.0	1820.57
$e^+e^-\,e^+e^-$	40.19	0.0
$e^+e^-\,\mu^+\mu^-$	47.1	2147.76
$e^+e^-\,K^+K^-$	0.0	0.57
$e^+e^-\,p\,\bar{p}$	0.0	0.13
$e^+e^-\,\pi^+\pi^-$	1.00	12.75
TOTAL BG	19118.58	50784.73

Table H.2: Results listed for cut based selection with neutrals, remaining events efficiencies and purities are stated for signal samples. The total remaining BG events and selection efficiency is stated for comparison.

Channel		e			μ	
Sample $\tau \to l\alpha$	Events	Efficiency	Purity	Events	Efficiency	Purity
$M_\alpha = 0$	1181986	38.88 %		1357376	44.65 %	
$M_\alpha = 0.5$	1181751	38.87 %		1346387	44.29 %	
$M_\alpha = 0.7$	1178226	38.76 %		1334325	43.89 %	
$M_\alpha = 1.0$	1163452	38.27 %		1292767	42.53 %	
$M_\alpha = 1.2$	1136448	37.38 %		1237895	40.72 %	
$M_\alpha = 1.4$	1031645	33.94 %		1110794	36.54 %	
$M_\alpha = 1.6$	704691	23.18 %		675309	22.21 %	
$\tau \to \ell\nu_\tau\nu_\ell$	1112177	44.679 %	92.30 %	1206569	48.471 %	84.70 %
BG	92772	$8.11{\cdot}10^{-7}$		217882	$1.90{\cdot}10^{-6}$	

Table H.3: Remaining background events after cut based selection with neutrals. Here $\tau^+\tau^-$ (other) refers to tau pair events that are not signal, this is, the events that do not satisfy the conditions of the $\tau \to \ell\nu_\tau\nu_\ell$ signal definition in MC (see Appendix D).

E(Remaining Events)	($62.8\ \text{fb}^{-1}$)	
BKG Samples	e	μ
$\tau^+\tau^-$ (other)	55809.73	148181.00
$c\bar{c}$	22863.60	27262.11
$s\bar{s}$	1382.86	5910.11
$d\bar{d}$	1564.98	5049.12
$u\bar{u}$	6729.65	22821.52
$B^0\bar{B}^0$	838.38	543.85
$B^+ B^-$	843.40	440.86
	low multiplicity	
e^+e^- (γ)	2455.48	0.0
$\mu^+\mu^-$ (γ)	1.884	5138.924
e^+e^- e^+e^-	104.248	0.0
e^+e^- $\mu^+\mu^-$	175.21	2527.7
e^+e^- K^+K^-	0.0	0.19
e^+e^- $p\,\bar{p}$	0.0	0.06
e^+e^- $\pi^+\pi^-$	2.64	6.22
TOTAL BG	92772.05	217881.65

H.1.2 BDT Based Event Selection

Table H.4: Results listed for BDT without Neutrals, remaining events efficiencies and purities are stated for signal samples. The total remaining BG events and selection efficiency is stated for comparison.

Channel		e			μ	
Sample $\tau \to l\alpha$	Events	Efficiency	Purity	Events	Efficiency	Purity
$M_\alpha = 0$	486021	15.99 %		608985	20.03 %	
$M_\alpha = 0.5$	488413	16.07 %		603602	19.86 %	
$M_\alpha = 0.7$	487999	16.05 %		600146	19.74 %	
$M_\alpha = 1.0$	485887	15.98 %		583588	19.2 %	
$M_\alpha = 1.2$	482523	15.87 %		560710	18.44 %	
$M_\alpha = 1.4$	464463	15.28 %		511195	16.82 %	
$M_\alpha = 1.6$	398081	13.09 %		355228	11.69 %	
$\tau \to \ell\nu_\tau\nu_\ell$	455465	18.297 %	95.69 %	546144	21.94 %	91.18 %
BG	20534	$1.79{\cdot}10^{-7}$		52803	$4.61{\cdot}10^{-7}$	

Table H.5: Remaining background events after selection with BDT without Neutrals. Here $\tau^+\tau^-$ (other) refers to tau pair events that are not signal, this is, the events that do not satisfy the conditions of the $\tau \to \ell\nu_\tau\nu_\ell$ signal definition in MC (see Appendix D).

E(Remaining Events)	(62.8 fb^{-1})	
BKG Samples	e	μ
$\tau^+\tau^-$ (other)	19276.46	45909.94
$c\bar{c}$	368.64	1153.01
$s\bar{s}$	42.08	476.02
$d\bar{d}$	42.70	336.61
$u\bar{u}$	170.82	1524.16
$B^0\bar{B}^0$	23.24	70.34
$B^+ B^-$	12.56	30.77
	low multiplicity	
e^+e^-	(γ) 458.44	6.28
$\mu^+\mu^-$	(γ) 0.0	1237.79
e^+e^- e^+e^-	24.49	1.26
e^+e^- $\mu^+\mu^-$	114.92	2042.26
e^+e^- K^+K^-	0.0	0.44
e^+e^- $p\,\bar{p}$	0.0	0.19
e^+e^- $\pi^+\pi^-$	0.06	14.13
TOTAL BG	20534.41	52803.18

Table H.6: Remaining background events after selection with BDT with Neutrals. Here $\tau^+\tau^-$ (other) refers to tau pair events that are not signal, this is, the events that do not satisfy the conditions of the $\tau \to \ell\nu_\tau\nu_\ell$ signal definition in MC (see Appendix D).

E(Remaining Events)	(62.8 fb^{-1})	
BKG Samples	e	μ
$\tau^+\tau^-$ (other)	41280.95	102505.3
$c\bar{c}$	1296.82	4251.56
$s\bar{s}$	155.74	1654.78
$d\bar{d}$	206.61	1102.14
$u\bar{u}$	825.82	5300.95
$B^0\bar{B}^0$	41.44	151.98
B^+B^-	18.21	116.81
	low multiplicity	
$e^+e^-(\gamma)$	207.24	0.0
$\mu^+\mu^-(\gamma)$	0.0	577.13
$e^+e^-\,e^+e^-$	1.26	0.0
$e^+e^-\,\mu^+\mu^-$	77.24	315.25
$e^+e^-\,K^+K^-$	0.0	0.06
$e^+e^-\,p\,\bar{p}$	0.0	0.06
$e^+e^-\,\pi^+\pi^-$	0.0	2.76
TOTAL BG	44111.35	115978.79

H.1.3 Data Simulation Comparison

Table H.7: Kolmogorov test results.

Variable	Channel	M_α [GeV] 0	0.5	0.7	1.0	1.2	1.4	1.6	Safe to look at?
Signal (1-prong) E_{CMS}	$\tau \to e\nu\bar{\nu}$	0.73	0.92	1.00	1.00	0.81	0.11	0.02	No
	$\tau \to \mu\nu\bar{\nu}$	0.06	0.24	0.65	0.67	0.03	0.00	0.00	
$E_{\mathrm{CMS}}^{\mathrm{3prong}}$	$\tau \to e\nu\bar{\nu}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	Yes
	$\tau \to \mu\nu\bar{\nu}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
$M_{\mathrm{Inv}}^{\mathrm{3prong}}$	$\tau \to e\nu\bar{\nu}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	Yes
	$\tau \to \mu\nu\bar{\nu}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
thrust	$\tau \to e\nu\bar{\nu}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	Yes
	$\tau \to \mu\nu\bar{\nu}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
Total E_{CMS} of event	$\tau \to e\nu\bar{\nu}$	0.88	0.99	1.00	1.00	1.00	0.68	0.47	No
	$\tau \to \mu\nu\bar{\nu}$	0.18	0.48	0.92	1.00	0.56	0.05	0.05	
ChiProb	$\tau \to e\nu\bar{\nu}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	Yes
	$\tau \to \mu\nu\bar{\nu}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
Missing momentum	$\tau \to e\nu\bar{\nu}$	1.00	1.00	1.00	1.00	1.00	0.94	0.78	No
	$\tau \to \mu\nu\bar{\nu}$	0.99	0.98	0.99	1.00	0.98	0.23	0.18	
Missing momentum θ_{CMS}	$\tau \to e\nu\bar{\nu}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	Yes
	$\tau \to \mu\nu\bar{\nu}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
Missing M^2	$\tau \to e\nu\bar{\nu}$	0.96	1.00	1.00	1.00	1.00	0.97	0.97	No
	$\tau \to \mu\nu\bar{\nu}$	0.31	0.74	1.00	1.00	0.81	0.35	0.58	
Tag (3-prong) $\#\gamma$	$\tau \to e\nu\bar{\nu}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	Yes
	$\tau \to \mu\nu\bar{\nu}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
Tag (3-prong) $\#\pi^0$	$\tau \to e\nu\bar{\nu}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	Yes
	$\tau \to \mu\nu\bar{\nu}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
Signal (1-prong) p_T	$\tau \to e\nu\bar{\nu}$	0.76	0.93	1.00	1.00	0.94	0.23	0.04	No
	$\tau \to \mu\nu\bar{\nu}$	0.10	0.30	0.73	0.99	0.21	0.00	0.00	
Tag (3-prong) p_T^{lead}	$\tau \to e\nu\bar{\nu}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	Yes
	$\tau \to \mu\nu\bar{\nu}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
Tag (3-prong) p_T^{sub}	$\tau \to e\nu\bar{\nu}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	Yes
	$\tau \to \mu\nu\bar{\nu}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
Tag (3-prong) p_T^{third}	$\tau \to e\nu\bar{\nu}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	Yes
	$\tau \to \mu\nu\bar{\nu}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	

List of Figures

List of Acronyms

ARGUS	A Russian-German-United States-Swedish Collaboration
ALP	axion like particle
ARICH	aerosol ring imaging Cherenkov detector
BDT	boosted decision tree
CDC	central drift chamber
CP	charged conjugation and parity
CPT	charged conjugation, parity, and time reversal
DAQ	Data Acquisition
DEPFET	depleted field-effect transistors
ECL	electromagnetic calorimeter
electronID	electron identifcation
EM	Electromagnetic
FET	field-effect transistor
GKK	generalised known kinematics
KLM	K_L^0 and μ detector
MLA	machine learning algorithm
MOSFET	Metal Oxide Semiconductor
MPP	Max-Planck-Insitut für Physik
NC	neutral current
pdf	probability density function
PID	particle identification
PXD	pixel detector
SiPM	Silicon Photomultiplier
SVD	silicon vertex detector
TOP	time of propagation counter

TUM Technische University Munich

VXD vertex detector

List of Terms

boost	Concept in relativity: The transfer from one reference frame to another reference frame.
collider	A particle accelerator that collides two particles beams.
cut	Is the process of selecting and rejecting parts of a distribution.
daughter	Is the particle that resulted from the decay of a heavier particle in the detector.
event	The collission of two particles in the collider, which leads to a detector signal.
hadronic	All constituents are hadrons – particles made from quarks.
interaction point	Is the point of the collider at which the particle beams collide.
likelihood	Is a short version for a single value of the likelihood function.
likelihood function	Is the joined probability of a specific model to describe data.
mother	is a particle that decays within the detector.
neutrals	Short hand for "neutral particles". All particles that have no electromagnetic charge.
preselection	Is the process of applying cuts on the reconstruction level.
probe	Is the name for the particle or group of particles in the event to measure something.
prong	Is the name for types of decays categorised by the number of chagred particles. That means that n-prong refers to a decay with n charged particles in the final state.

radiative Bhabha	Is an event in which e^+e^- scatter and may emit a photon.
reconstruction level	The process of reconstructing particle events from detector signals.
sideband	A sideband is a part of data that emulates properties of the signal itself or the signal region but is not part of the signal or signal region. The sideband allows for studying important properties of measured data.
statistic	Is defined as: "A function of observable random variables, which is itself an observable random variable, which does not contain any unknown parameters." [80].
sub-detector	One detector system of the Belle II-detector, e.g. the PXD.
tag	Is the name for the particle or group of particles which indicate a specific event.
threshold	Is the minimum energy required to produce a specific particle-pair event.
track	Is the path of a charged particle in the detector.
Upper Limit	The highest posible value of a statistically relevant sequence, e.g. a confidence level.
vertex	The reconstructed point of a mother, which decayed into charged daughter particles. estimation.