

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Search for Lepton Flavor Violating τ decays to a Lepton and an Invisible Boson at Belle II using Machine Learning Methods

verfasst von | submitted by

Benjamin Bacher BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt |
Degree programme code as it appears on the
student record sheet:

UA 066 876

Studienrichtung lt. Studienblatt | Degree pro-
gramme as it appears on the student record
sheet:

Masterstudium Physics

Betreut von | Supervisor:

Hon.-Prof. Dipl.-Phys. Dr. Eberhard Widmann

Acknowledgements

I declare special thanks to my co-supervisor Petar Rados for all his time and effort. He introduced me to and guided me through this topic with the most care. I am grateful for all his patience and the tremendous amount of feedback he provided.

I thank Prof. Widmann Eberhard for his feedback and the time he dedicated to my supervision.

I also want to thank Geraldine Räuber and Zuzana Gruberová for their huge support and time in answering all my questions. They also gave me the opportunity to speed up my work since they provided me with some of their code to start with.

I am grateful for my parents and my grandma for a lifetime of support, patience and trust in me. I would not be the same and this thesis would not have been possible without them.

Abstract The Standard Model of Particle Physics (SM) does not explain or even forbids concrete phenomena like Lepton Flavor Violation (LFV) and Dark Matter. Therefore, numerous New Physics (NP) concepts beyond the SM have been brought up for theoretical and experimental research.

A few of those theories predict the existence of an additional boson. Let's call it the α particle. Though it is impossible to detect it directly with current methods, due to non-existing or very weak interactions with known particles, we can calculate the cross-section of the predicted decay and compare the traces of the detectable particles, in this case, τ , e , and μ . LFV for charged particles is called charged LFV. Charged LFV decays are rare events if they exist in the first place thus having a small signal-to-noise ratio.

So we want to focus on improving statistics in the search for such decays. The main point of this thesis is therefore the separation of signal (NP interactions) and background (SM interactions). We use Monte Carlo samples and apply rectangular cuts on a set of discriminating variables. In combination with Boosted Decision Tree training (BDT) to find the optimal cutting point we can improve the signal-to-background ratio for data.

Zusammenfassung Das Standard Model der Teilchenphysik kann bestimmte Phänomene wie Lepton Flavor Violation (LFV) und Dunkle Materie nicht erklären oder verbietet diese zum Teil sogar. Deswegen tauchen immer mehr Konzepte, die über das Standard Models hinausgehen, auf und versuchen diese Phänomene zu erklären, sogenannte Neue Physik.

Ein paar dieser Theorien sagen die Existenz von zusätzlichen Bosonen voraus. Nennen wir sie α Teilchen. Obwohl es unmöglich ist, aufgrund von sehr schwacher beziehungsweise nicht vorhandener Wechselwirkung mit bekannten Teilchen diese Teilchen mit bisherigen Methoden direkt zu messen, kann man den Querschnitt des vorhergesagten Zerfalls berechnen und mit den Spuren der messbaren Teilchen vergleichen, in unserem Fall mit τ , e und μ Leptonen. LFV für geladene Teilchen nennt man charged LFV. Diese charged LFV Zerfälle sind, wenn sie überhaupt existieren, äußerst selten und haben deshalb ein sehr kleines Signal-zu-Hintergrund Verhältnis.

Unser Ziel ist die Verbesserung von statistischer Ungenauigkeit auf der Suche nach solchen Zerfällen. Der Fokus dieser Arbeit ist deshalb die Trennung von Signal (Neue Physik Interaktionen) und Hintergrund (Standard Model Interaktionen). Wir nutzen Monte Carlo generierte Datensätze und schneiden bei bestimmten Variablen rechteckige Datensätze daraus. In Kombination mit dem Boosted Decision Tree Algorithmus können wir die vielversprechensten Schnittpunkte finden und so das Signal-zu-Hintergrund Verhältnis gesammelter Daten verbessern.

Contents

1	Introduction and Motivation	6
2	Theory Overview	8
2.1	The Standard Model	8
2.2	Tau leptons	8
2.3	τ physics and golden decay modes	11
2.4	New Physics and invisible bosons	12
3	The Belle II experiment	17
3.1	SuperKEKB	17
3.2	Coordinate System	18
3.3	Detector	18
3.4	Data recording	21
4	τ pair reconstruction	23
4.1	<i>Belle II</i> as a τ factory	23
4.2	3x1-prong topology	24
4.3	Thrust Axis and Split of Hemispheres	24
4.4	New Physics τ pair reconstruction - ARGUS Method	26
4.5	Event Reconstruction	28
4.6	Discriminating Variables	29
4.7	Monte Carlo sample generation	30
5	Current Method: cut-based Analysis	36
5.1	Method	36
5.2	Recent status of <i>Belle II</i>	37

6	New Approach: XGBDT training	40
6.1	Decision Trees	40
6.2	Extreme Gradient Boosted Decision Trees	41
6.3	Methods	43
6.4	Results	53
7	Conclusion	69
8	Appendix	71
	Table of Figures	75
	Bibliography	81

1 Introduction and Motivation

Particle Physics engages in the most fundamental constituents of our world, elementary particles, and their interactions. The current best theory, the Standard Model (SM) [1], has done a great job in describing fundamental processes with great precision so far. Although there are certain areas where small deviations with experimental results appear, e.g. $(g_\mu - 2)$, the anomalous muon magnetic moment [2] and the strong CP problem [3]. This gave rise to a number of new experiments trying to measure as well as theoretical models trying to explain physics beyond the SM, also known as New Physics (NP).

Experiments to study elementary particles are mostly done in particle accelerators. In the search for NP some accelerator experiments try to detect heavy NP particles at high energy frontiers (e.g. *LHCb* [4–10]) while others look for lighter NP particles or quantum fingerprints of heavier particles at the intensity frontier using lower energies and cleaner conditions (e.g. *Belle* [11], *BaBar* [12]).

One such high intensity particle accelerators is located in Tsukuba, Japan and the main point of interest for this study, the SuperKEKB. As a sequel to the *Belle* experiment KEK, the Japanese High-Energy-Accelerator Research Organisation, started the *Belle II* experiment with a first physics run in 2018. While the *Belle* experiment made important contributions to Particle Physics, especially the discovery of CP asymmetry in the B-sector (Nobel Prize 2008, [13]), the ongoing *Belle II* is dedicated to NP discoveries in B meson, D meson and τ lepton decays.

LFV has been established for the neutrino sector but concerning their charged partners (e , μ and τ leptons) there is no measurable sign of LFV. Like in the SM, charged LFV decays via neutrino oscillation are highly suppressed and immeasurably small. So, if any charged LFV is observed it would indicate the existence of NP.

The $\mu \rightarrow e$ transitions have stringent bounds from previous experiments, such as *MEG* [14]. On the other hand, the transitions $\tau \rightarrow e$ and $\tau \rightarrow \mu$ have weaker bounds coming mainly from

Belle, *BaBar* and *CLEO* [15] and, since τ is the heaviest lepton, NP can couple preferentially to them. Due to their large mass, τ leptons also provide a wide variety of LFV decay modes to study. So searching for evidence of charged LFV in tau decays is of particular interest at *Belle II*. There are LFV tau decays in which the decay products include invisible NP particles. The golden mode for invisible LFV decays is $\tau^\pm \rightarrow \ell^\pm \alpha$, where α is a boson which interacts at most very weakly with SM particles and $\ell^\pm$ denotes $e^\pm$ or $\mu^\pm$. A search for this mode has already been conducted at *Belle II* on a data set of 62.8 fb^{-1} . This thesis presents the possibility of improving the sensitivity of this search using a Machine Learning technique called Extreme Gradient Boosted Decision Tree Method (or short Boosted Decision Tree (BDT)).

Section 2 is a summary of theoretical background, including the SM and physics beyond the SM with emphasis on tau physics and NP. Section 3 gives an overview of the *Belle II* experiment with focus on overall information to KEK, the experimental setup and data recording. In Section 4 we discuss τ pair reconstruction at *Belle II* and introduce the pseudo rest frame. Section 5 describes the current method used by the *Belle II* analysis as well as the recent status of data collection at *Belle II*. In Section 6 we give an insight into Extreme Gradient Decision Tree training and its implications for the *Belle II* analysis of $\tau^\pm \rightarrow \ell^\pm + \text{invisible}$ decays. We present the results of this study and analyze them using the Punzi Figure Of Merit (Punzi FOM) as an independent measure of quality. Section 7 concludes with the overall improvement and prospects of future applications of Machine Learning Techniques in the field of data analysis for collider experiments.

2 Theory Overview

2.1 The Standard Model

In the SM (Figure 1) elementary particles are classified into fermions, with half-integer spin numbers, and bosons, with integer spin numbers. Elementary particles can not be subdivided into smaller particles and form the building blocks our universe. While bosons yield the exchange of fundamental forces fermions are the elementary constituents of matter. Thanks to group theory considerations and the symmetry group of the Lagrangian $SU(3)_c \times U(1)_{em}$ fermions can be further classified into up-type quarks (u), down-type quarks (d), charged leptons (ℓ) and neutrinos (ν) [16]. The first two classes refer to quarks which build up the elementary constituents of hadrons while the latter refer to leptons and their corresponding neutrino flavors. Since the Lagrangian of fermions is unaffected by three distinct spacial rotations in the $U(1)$ group there are three distinct sub-classes called "flavors" for each fermion class. The three flavors for charged leptons are electrons (e), muons (μ), and tauons (τ) with their neutrino counterpart while the flavors for up-type quarks are up, charm, and top. The flavors for down-type quarks are down, strange, and bottom [16].

2.2 Tau leptons

The third generation leptons are called tau leptons and are the heaviest leptons with a mass of $m_\tau = 1777.28 \pm 0.75 \pm 0.33 \text{ GeV}/c^2$ [1]. Tauons carry the charge of -1 . These massive leptons τ^- and their anti-particles τ^+ are the only leptons capable of decaying into lighter hadrons and other lepton families. Tau leptons are unstable and will typically decay in a few picoseconds, with a mean lifetime of $(2.903 \pm 0.005) \times 10^{-12} \text{ s}$ [1]. The most common decay modes for tauons are given in Figure 2 where percentages in the diagram are proportional to the

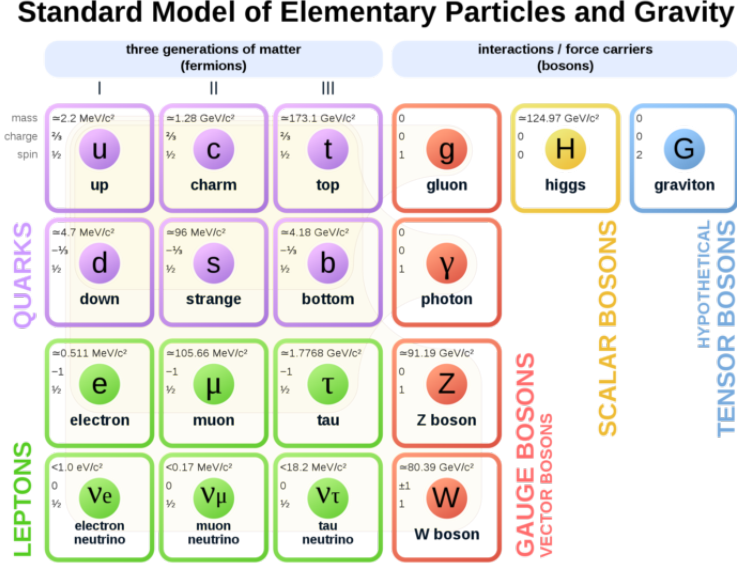


Figure 1: Particle content of the Standard Model of Particle Physics. From [17].

corresponding branching ratio $\mathcal{B}$. The full set of decay modes including their exact branching ratios can be seen in Table 1.

Due to their short lifetime, we do not detect tau leptons directly but through their decay products. About one-third of decays occur in the leptonic mode $\tau^- \rightarrow \ell^- \nu_\tau$, where ℓ^- is an electron or muon. The tau lepton is the only lepton heavy enough to decay hadronically $\tau^- \rightarrow \nu_\tau + \text{hadrons}$, and does so in about two-thirds of all cases. The leptonic and hadronic tau decay modes are illustrated in Figures 3(a) and 3(b), respectively.

A number of neutral and charged hadrons will be produced in hadronic tau decays. One or three charged pions ($\pi^\pm$) are produced in 70% and 21% of the decays, respectively. These will be referred to as 1-prong and 3-prong decays, hereafter. Charged kaons ($K^\pm$) are present in the majority of the remaining 9% of hadronic decays. In 60% of all decays, up to two associated neutral pions (π_0) are also produced. Tau leptons play a crucial role in particle



Figure 2: Most common tau lepton decay modes with corresponding branching ratio $\mathcal{B}$. From [18].

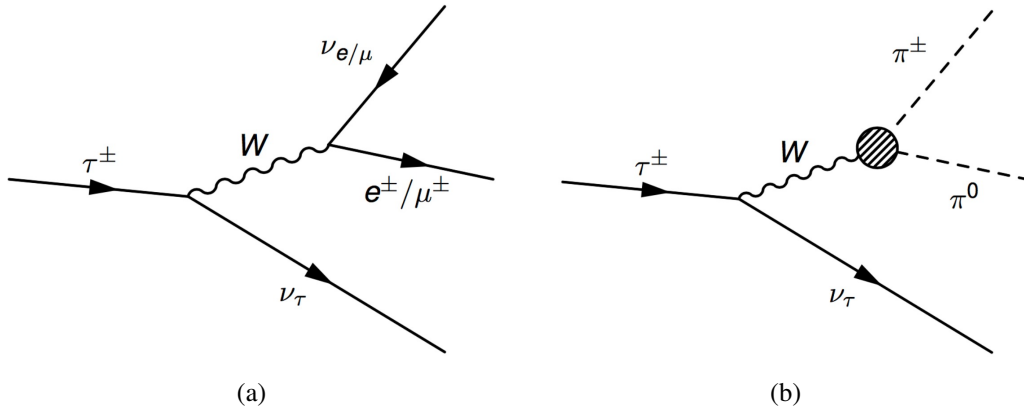


Figure 3: (a) leptonic τ decay. (b) hadronic τ decay.

physics research. Their study helps testing the predictions of the SM, exploring the properties of the weak interaction, and searching for physics beyond the SM, such as evidence of new particles or forces.

Decay Mode	Branching Fraction (%)	Type	n-prong
$\tau^- \rightarrow e^- \bar{\nu}_e \nu_\tau$	17.82 ± 0.04	leptonic	1-prong
$\tau^- \rightarrow e^- \bar{\nu}_e \nu_\tau \gamma$	1.83 ± 0.05	leptonic	1-prong
$\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$	17.39 ± 0.04	leptonic	1-prong
$\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau \gamma$	0.37 ± 0.01	leptonic	1-prong
$\tau^- \rightarrow \pi^- \nu_\tau$	10.82 ± 0.05	hadronic	1-prong
$\tau^- \rightarrow \pi^- \pi^0 \nu_\tau$	25.49 ± 0.09	hadronic	1-prong
$\tau^- \rightarrow \pi^- \pi^0 \pi^0 \nu_\tau$	9.26 ± 0.10	hadronic	1-prong
$\tau^- \rightarrow \pi^- \pi^0 \pi^0 \pi^0 \nu_\tau$	1.04 ± 0.07	hadronic	1-prong
$\tau^- \rightarrow \pi^- \pi^+ \pi^- \nu_\tau$	8.99 ± 0.05	hadronic	3-prong
$\tau^- \rightarrow \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$	2.74 ± 0.07	hadronic	3-prong
$\tau^- \rightarrow h^- h^+ h^- \pi^0 \pi^0 \nu_\tau$	0.51 ± 0.03	hadronic	3-prong
$\tau^- \rightarrow K^- \nu_\tau$	0.70 ± 0.01	hadronic	1-prong
$\tau^- \rightarrow K^- K^0 \nu_\tau$	0.15 ± 0.01	hadronic	1-prong
$\tau^- \rightarrow K^- K^0 K^0 \nu_\tau$	$< 6.3 \cdot 10^{-7}$ CL=90%	hadronic	1-prong
$\tau^- \rightarrow K^- \pi^0 \nu_\tau$	0.43 ± 0.01	hadronic	1-prong
$\tau^- \rightarrow K^- \pi^0 \pi^0 \nu_\tau$	0.07 ± 0.02	hadronic	1-prong

Table 1: Tau lepton decay modes. From [1].

2.3 τ physics and golden decay modes

In the SM, lepton flavor conservation is bound to the existence of only left-handed neutrinos. We already know that this conservation is broken thanks to experimental observations of neutrino oscillations which proves Lepton Flavor Violation (LFV) at least in the neutrino sector. This, however, doesn't not necessarily imply LFV for their charged partners (e , μ or τ). Therefore, further investigation is needed.

There are several advantages when searching for charged LFV in τ decays. As the heaviest lepton, many New Physics (NP) models have enhanced LFV couplings to tau leptons. Moreover, the higher mass gives rise to a wide variety of radiative ($\tau \rightarrow \ell\gamma$, where $\ell = e$ or μ), leptonic ($\tau \rightarrow \ell\ell\ell$) and semileptonic ($\tau \rightarrow \ell h(h)$, where $h = \text{hadrons}$) LFV decay modes to help discriminate between NP models. The golden modes for studying charged LFV in a fully visible decay mode are $\tau \rightarrow 3\mu$ and $\tau \rightarrow \mu\gamma$. Especially the first decay is expected to be free of background while the latter has the largest branching ratio in models where LFV processes are induced by one-loop diagrams including baryons. If the mode $\tau \rightarrow 3\mu$ is observed, this gives the normalization for the LFV couplings, and, further, one can eliminate possible theoretical models by comparing its branching ratio with the branching ratios of semi-leptonic LFV processes. If semi-leptonic decays are observed while $\tau \rightarrow 3\mu$ is not it would indicate a more exotic model for generating LFV couplings. Either way, *Belle II* will push the current bounds by at least one order of magnitude (see Figure 4). [19]

There are models for charged LFV decays with an invisible NP particle. The golden mode for this process is $\tau^- \rightarrow \ell^- \alpha$ and the search for it is the key point of interest for this thesis.

2.4 New Physics and invisible bosons

The SM of particle physics has its limits. Boundaries at which theoretical and experimental values don't coincide. Therefore, theoretical expansions of the SM like strong CP symmetry breaking [3] and neutrino masses [20] as well as observed phenomena which can't be explained with SM physics such as Charged LFV [16], $(g_\mu - 2)$ the anomalous magnetic moment [2] and Dark Matter [21] give rise to possible new theories beyond the Standard Model (SM). Those New Physics (NP) models build upon the SM and resolve discrepancies by introducing new forces or particles. One such particle is the so called axion, a Pseudo-Nambu-Goldstone-Boson (PNGB) originating from spontaneous symmetry breaking which couples derivatively to SM

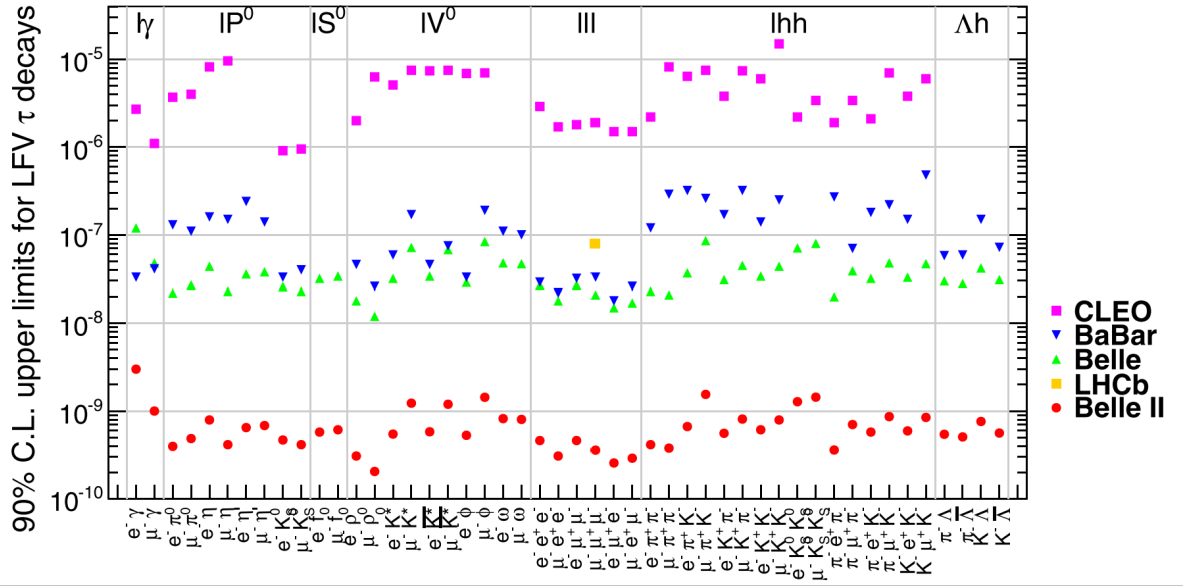


Figure 4: Current 90% CL upper limits for the branching fraction of τ LFV decays obtained in the *CLEO*, *BaBar*, and *Belle* experiments. The purple boxes, blue inverted triangles, green triangles, and yellow boxes show *CLEO*, *BaBar*, *Belle*, and *LHCb* results, respectively, while the red circles express the *Belle II* future prospects, where they are extrapolated from Belle results assuming the integrated luminosity of 50 ab^{-1} . From [19].

fermions. Axions and axion-like-particles (ALP) are hard-to-detect particles since they have either no or rather weak interaction with known forces and are therefore considered "invisible". However, introducing light ALPs with LFV coupling can offer potential solutions to various known issues, as mentioned above. E.g. the LFV QCD axion [22, 23] or the LFV axiflavor [24, 25] would arise as natural solutions for the strong CP problem while the latter could even explain the mass hierarchy of charged leptons. Another potential LFV axion model introduces the familon [26], a PNCB associated with the spontaneous lepton flavor symmetry breaking. [27]

Invisible particles are conventionally labeled with α . The α boson is supposed to be a decay product of possible charged LFV decays, explicitly in $\tau^\pm \rightarrow \ell^\pm \alpha$ and $\mu^\pm \rightarrow e^\pm \alpha$, where $\ell = e, \mu$. The light ALPs branching ratios $\mathcal{BR}$ are reciprocal to the ALP decay constant f_α (Equation 2.1).

$$\mathcal{BR} \propto \frac{1}{f_\alpha} \quad (2.1)$$

We may point at Figure 5(a) where the most recent boundaries on f_α are set by the experiments ARGUS [28], MEGII-fwd [14], Mu3e [29], Jodidio et al. [30], TWIST [31] and *BelleII* [32]. *MEGII-fwd* denotes a study of a detector pointing in the forward beam direction for the *MEG II* experiment which allows to detection of possible $\mu^+ \rightarrow e^+ \alpha$ decays by searching for a positron of maximal energy in the opposite direction to the polarization beam.

The boundaries on the search on LFV decays are directly related to the α decay constant and, therefore, tied to the experimental limitations given by the detection of α decay modes. Such experiments depend heavily on the mass of the ALP particles as well as on the chiral structure of ALP couplings to the SM.

Another interpretation of an α boson is the LFV Z' . This NP vector boson originates from an extension of the SM which tries to explain the $(g_\mu - 2)$ anomalous magnetic moment. Assuming that Z' is a PNGB of an extra $U(1)'$ symmetry, it could couple differently to various SM fermion families. The theoretical mass range of Z' mostly depends on its left- and right-handed coupling constants, g'_L and g'_R , respectively. A purely left- or right-handed coupling would enhance the SM discrepancy of the muon magnetic moment [33]. For the set relation of the coupling constants, $g_L = g_R/10$ Figure 5(b) displays Z' 's experimental constraints on its mass $m_{Z'}$. As a result of experiments such as $e^+e^- \rightarrow e^+e^-$ measurements at LEP [34], electroweak precision tests [35, 36] and $e^+e^- \rightarrow \gamma\ell^+\ell^-$ at BaBar [12] the Z' couplings to $e^\pm$ are heavily constrained while LFV decays, such as $\mu \rightarrow e\gamma$, $\mu \rightarrow 3e$, $\tau \rightarrow \mu\gamma$, $\tau \rightarrow 3e$ and $\tau \rightarrow ee\mu$ in combination with LHC di-lepton resonance searches $pp \rightarrow Z' \rightarrow ee, \mu\mu$ [4, 5], $\tau\tau$ [6], $e\mu$ [7–9], $e\mu, \mu\tau$ [9] provide strong constraints for Z' couplings in general. This either leads to very small couplings to electrons, the first lepton family, or could in a similar fashion blow up the mass scale of Z' to exceed the multi-TeV range [10].

Again, we label all LFV decays with missing energy like $\tau^\pm \rightarrow \ell^\pm + \text{invisible}$ as $\tau^\pm \rightarrow \ell^\pm \alpha$ where α denotes an invisible boson. This can fall into one of the before mentioned categories of NP particles.

The studies presented in this thesis aim to maximize the sensitivity of the *Belle II* experiment in the search for $\tau^\pm \rightarrow \ell^\pm \alpha$. This work will allow *Belle II* to far exceed what was previously achieved by the ARGUS experiment in 1995 [28], significantly improving on the upper limits set by ARGUS on light ALP and Z' models (Figure 5(a) - bottom purple line, 5(b) - blue area).

3 The Belle II experiment

Belle II is a particle accelerator experiment located at the KEK research facility in Tsukuba, Japan. It's the successor of the *Belle* experiment [11] and is designed to study the decays of B mesons, D mesons, and τ leptons. With both precision measurements of SM decays and searches for rare decays, *Belle II* can shed light on fundamental processes of Particle Physics beyond the SM. *Belle II* started detecting collisions in spring 2018. In this section we will provide an overview of the *Belle II* experiment including the SuperKEKB accelerator and the new detector system.

The *Belle* experiment performed LFV searches for 46 τ decay modes using almost the entire data set of about $1000 fb^{-1}$ at the $O(10^{-8})$ level and 90% CL upper limits on the branching ratios but no charged LFV processes were found. Considering the current status of *Belle II*, a sensitivity level of $O(10^{-9}) - O(10^{-10})$ is expected [19]. This motivates further investigations in the LFV sector in the *Belle II* experiment.

3.1 SuperKEKB

SuperKEKB [37] is an anti-symmetric electron-positron collider that produces huge amounts of data with an expected integrated luminosity over its lifetime of $50 ab^{-1}$. The e^-/e^+ get accelerated to energies of up to 7/4 GeV, respectively, in a 3 km circumference ring before they collide with a center of mass energy of $\sqrt{s} = 10.58$ GeV. The expected luminosity corresponds to 57.7 million $e^+e^- \rightarrow \tau^+\tau^-$ events [32]. Therefore, the *Belle II* experiment will provide excellent possibilities to study low multiplicity final states with high precision such as tau decays. [19]

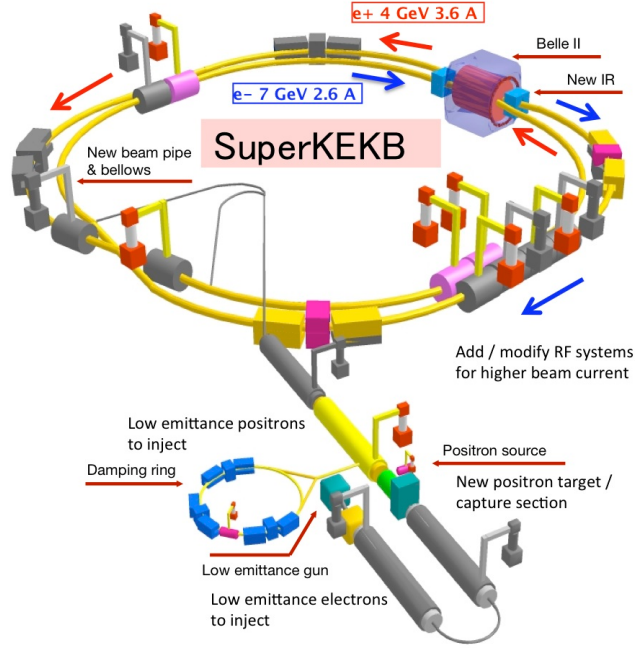


Figure 6: The SuperKEKB accelerator with all its sub-components incl. the *Belle II* detector system. From [38].

3.2 Coordinate System

The *Belle II* detector is comprised of various subdetectors positioned in a cylindrical configuration surrounding the interaction point of electron-positron collisions. The laboratory frame's z -axis corresponds to the solenoid axis of the detector, oriented in the positive direction along the electron beam. The polar angle θ and the transverse plane are defined relative to this axis.

3.3 Detector

The *Belle II* detector is divided into several sub-detectors, each of which is designed to measure specific properties of the particles. These sub-detectors include the Vertex Detector, the Silicon

Strip Detector, the Central Drift Chamber, the Time-of-Flight Detector, the Aerogel Cherenkov Counter, the Imaging Cherenkov Detector, and the Electromagnetic Calorimeter. For further insight, we refer to the *Belle II* Physics Book [19] and the *Belle II* Technical Design Report [39].

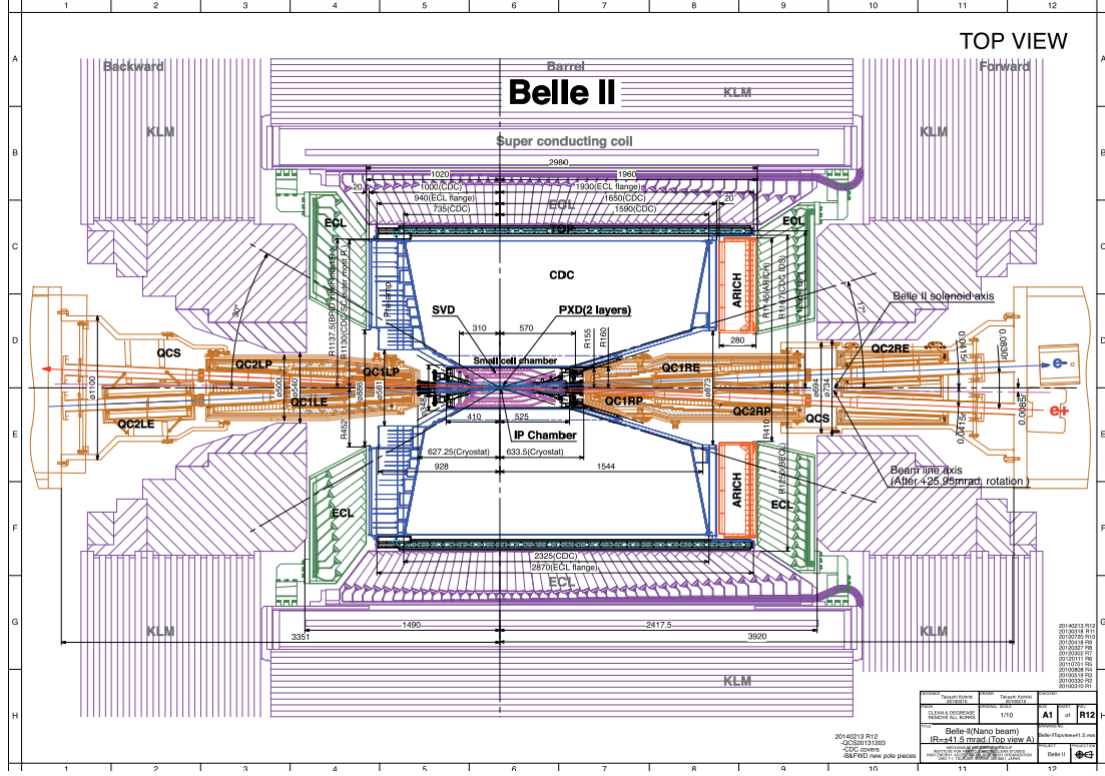


Figure 7: Belle II detectors. From [19].

3.3.1 Sub-detectors

VXD: The Vertex Detector (VXD) is the first sub-detector that the particles encounter as they pass through the *Belle II* detector. It is designed to precisely measure the particles' origin and momenta. The Vertex Detector consists of two layers of silicon pixel detectors (**PXD**), surrounded by four layers of silicon strip vertex detectors (**SVD**) which can detect the particles' origin with high precision. The layers are placed at radii of 14 mm - 135 mm. Compared to

Belle the innermost layers is much closer to the interaction point while the outermost layer has a much greater radius. A significant improvement of vertex resolution is therefore expected with respect to *Belle*. For this study only 15% of the second PXD layer was operational during the data collection period [32].

CDC: The Central Drift Chamber (CDC) is located outside of the Silicon Strip Detector and is designed to measure the momentum and charge of the charged particles. The CDC is made up of 143 360 sense wires arranged in 56 layers in axial and stereo orientation. Combining these gives a full-3D helix track. The large-volume drift chamber consists of small drift cells and can operate at high event rates with increased background levels. The chamber is filled with a He-C₂H₆ mixture (50 : 50). The average drift velocity is $3.3 \text{ cm } \mu \text{ s}^{-1}$. The CDC was commissioned using cosmic rays.

PID system: The Particle Identification system (PID) consists of a Time-of-Propagation counter (**TOP**) in the barrel region and an Aerogel Ring Imaging Cherenkov Detector (**ARICH**) in the forward region.

TOP is a basic Cherenkov detector that gives 2-dimensional information of arrival time and impact position with a Cherenkov ring image. It consists of 128 scintillation counters and 64 aerogel blocks and allows *Belle II* to identify particle types based on their velocity. For this, the particle production time has to be known. As the detector module a thick quartz bar is used which, furthermore, requires photo-sensors with a single-photon time resolution of approximately 100 ps.

ARICH is a proximity focusing Cherenkov ring imaging detector with aerogel as Cherenkov radiator and is located after TOP. It is designed to detect charged particles near the speed of light. The aerogel has two different refracting indices to increase the yield without degrading the Cherenkov angle resolution. It features a barrel-shaped array of 448 photomultiplier tubes that detect the Cherenkov radiation.

TOP and ARICH cover the polar ranges of $31^\circ < \theta < 128^\circ$ and $14^\circ < \theta < 30^\circ$, respectively, where θ is the polar angle relative to the solenoid angle, as described in earlier in Chapter

3.2.

ECL: The Electromagnetic Calorimeter (ECL) occupies the remaining volume within a superconducting solenoid that generates a uniform axial magnetic field of 1.5 T. It's designed to measure the energy of electrons and photons and separate electrons from hadrons. It's made up of 8736 thallium-doped cesium iodide CsI(Tl) crystals that are arranged in a barrel shape around the beam pipe which work as scintillators covering about 90% of the solid angle of in the center-of-mass frame. Compared to *Belle* the electronics and reconstruction software have been updated to cope with an increase in background levels.

KLM: The K_L - μ -detector is built up from alternating layers of 4.7cm-thick iron plates and active detector elements located outside the superconducting solenoid. In comparison to *Belle* the glass electrode resistive plate chambers (RPC) were replaced by layers of scintillator strips with wavelength-shifting fibers since larger background rates reduce detection efficiency, especially for muons.

Trigger System: The Trigger System identifies events of interest. This includes triggers for new phenomena like a single photon trigger for dark sector searches and a 2-photon or 3-photon trigger for axion-like particle searches. *Belle II* first has a hardware-based level 1 trigger (L1 trigger) followed by a software-based high-level trigger (HLT trigger). These are particularly important for low multiplicity events which are typically at or near the trigger thresholds.

3.4 Data recording

The initial data collection for *Belle II* began in the first half of 2018 during the detector commissioning phase (Phase 2). Subsequently, collision detection using the complete *Belle II* detector commenced in March 2019 (Phase 3). The Phase 3 data-taking period is further

subdivided into smaller intervals referred to as experiments (Exp), which are sequentially numbered and occasionally broken down into units called buckets.

In 2020, SuperKEKB achieved a world record for the highest instantaneous luminosity, reaching a value of $2.4 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ [40]. Figure 8(a) illustrates the cumulative recorded luminosity from early 2019 till the Long Shutdown 1 (LS1) in summer 2022. LS1 lasted until Fall 2023 to replace the VXD and further improve detector performance. This performance improvement is expected to result in much higher integrated luminosity (Figure 8(b)). Another break in data taking between 2027-2028 is in consideration, the LS2, during which machine improvements will be done. The overarching objective of the *Belle II* experiment is to accumulate a substantial data set at the $\Upsilon(4S)$ resonance, amounting to an integrated luminosity of 50 ab^{-1} until 2035. Additionally, *Belle II* also records smaller amounts of data from off-resonance collisions. [41]

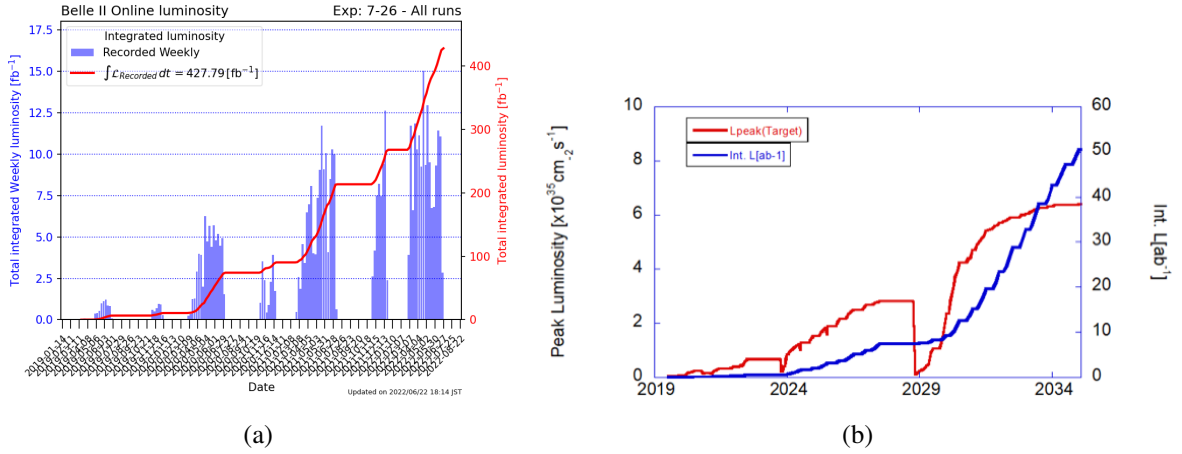


Figure 8: Current and future luminosity of *Belle II* since data taking; (a) Development and current luminosity state. (b) Future projection of luminosity with peak luminosity (red) and integrated luminosity (blue). Second Long Shutdown (LS2) between 2027 and 2028 for machine improvements possible. First Long Shutdown (LS1) and LS2 are recognizable by the steep of peak luminosity curve at 2022 and 2028. From [41].

4 τ pair reconstruction

In the following subsection we discuss the common tau reconstruction method used in *Belle*, *BaBar*, and *Belle II*. We will see why *B-factories* can also be considered as τ -factories. We further look at the desired 3x1-prong topology and how this can help to separate hemispheres and therefore allow to setting of efficient event selection triggers.

4.1 *Belle II* as a τ factory

At e^+e^- -colliders tau pairs are produced via a pair annihilation process. The electron-positron pair gets annihilated which produces a tau pair via a virtual photon, see Figure 9. *Belle II* is considered a B-factory due to its collision energy designed close to the $\Upsilon(4S)$ resonance which then decays into B mesons. However, the cross-section for B-pair production is very similar to that of tau-pair production. Therefore, *Belle II* is just as well a ' τ -factory' as it is a B-factory. This gives us the possibility to study tau lepton decays at B-factories with great precision.

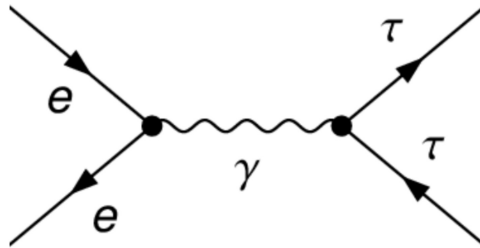


Figure 9: tau pair production at e^+e^- colliders

4.2 3x1-prong topology

In this thesis, we consider tau-pair events where one tau lepton decays into one charged particle (1-prong) and the other decays into three charged particles (3-prong). More specifically we are interested in 3x1-prong tau pair events such that one decays hadronically (3-prong), mainly via $\tau^- \rightarrow h^- h^+ h^- \nu_\tau + n\pi^0$, and the other decays leptonically (1-prong) via $\tau^- \rightarrow \ell^- \bar{\nu}_\ell \nu_\tau$ where $\ell = e/\mu$, with a BR of 8.99 ± 0.05 ($n\pi^0 = 0$) and 17.82 ± 0.04 (e) or 17.39 ± 0.04 (μ), respectively (see Table 1).

The decay modes $\tau^- \rightarrow e^- \bar{\nu}_e \nu_\tau$ and $\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$ and their anti-particle counterparts are crucial for our study because they make up most of the background noise. Considering the wanted New Physics decay mode $\tau^\pm \rightarrow \ell^\pm \alpha$ those modes yield the same detector signature. Our main task will be to minimize said background.

Also the suppression of pions π and kaons K is an important matter because they make up the rest of the SM model signature, see Table 1. It is essential to stress that π^0 mesons can decay into photons which must be separated from general photon detection to establish a good neutral pion reconstruction.

4.3 Thrust Axis and Split of Hemispheres

In the $e^+e^- \rightarrow \tau^+\tau^-$ process, the tau leptons are produced back-to-back in the centre of mass frame. So that, after boosting to the laboratory frame, the decay products of each tau candidate are confined to opposite hemispheres and isolated from one another. The theoretical boundary separating these hemispheres corresponds to a plane perpendicular to the trajectory of the τ particles. The trajectory of the τ pair is approximated by the thrust axis $\hat{t}$ – the vector maximizing the *thrust* value, Equation 4.1. Here, $\vec{p}_i^{CM}$ is the momentum of each final-state

particle in CM frame. [11, 19]

$$thrust = \sum |\hat{t} \cdot \vec{p}_i^{CM}| / \sum |\vec{p}_i^{CM}| \quad (4.1)$$

The thrust axis can be thought of as the axis of maximum elongation of the event. Figure 10 displays a high thrust event which stretches out and elongates the thrust axis and a low thrust event with a more spherical shape. Typical high thrust events are tau lepton pair productions while a typical low thrust event would be for example B meson pair production. Recall that

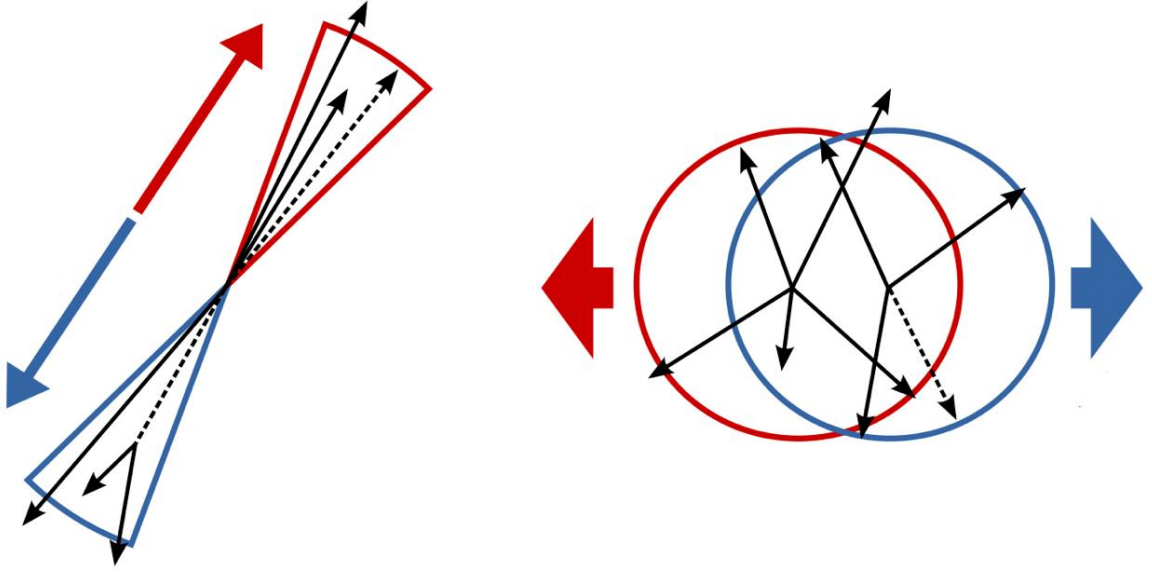


Figure 10: (Left) high thrust event, e.g. τ pair. (Right) low thrust event, e.g. BB. The thrust axis is represented by the red and blue arrows. From [19].

the decay modes of tau leptons consist of mostly 1 or 3 charged particles, that we call 1- or 3-prong decays. Decay modes including more decay particles are energetically unfavorable and, therefore, statistically insignificant (τ decay modes, Table 1).

Specifically, we define the tag side as the one containing three charged particles, namely $\tau^- \rightarrow h^- h^+ h^- \nu_\tau$ with $h = \pi, K$, while the signal side is required to have only one charged particle (Figure 11). This helps to suppress background. Especially since the hadronic 3-prong tag-side decay offers great trigger quality with three charged particles to reconstruct the event.

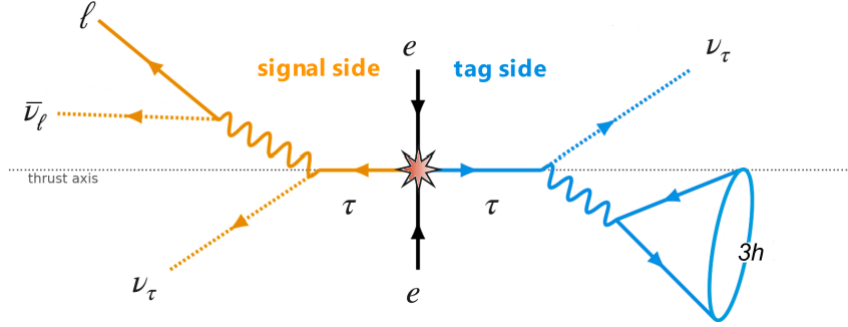


Figure 11: 3x1-topology of τ pair decay with thrust axis indicating signal- and tag-side. From [18]

4.4 New Physics τ pair reconstruction - ARGUS Method

One cannot measure the τ momentum directly since the detection of the neutrino is not feasible, we can approximate its Center-Of-Mass Energy $E_{CM} \approx \sqrt{s}/2$ (s is the beam energy) with its direction being opposite to the tag side hadrons $\hat{p}_\tau \approx -\vec{p}_{3h}/|\vec{p}_{3h}|$. Using those assumptions we can construct the τ *pseudo rest frame*, see Figure 12.

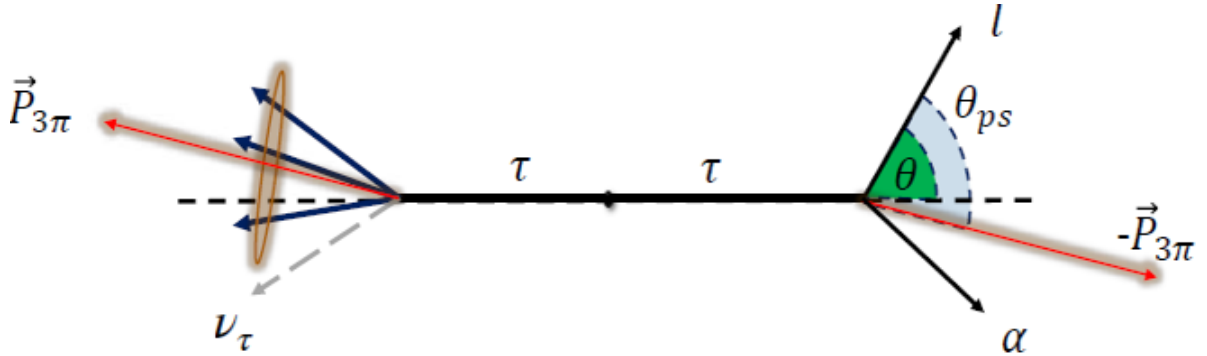


Figure 12: *pseudo rest frame*, separates tag side $\tau^- \rightarrow \pi^- \pi^+ \pi^- \nu_\tau$ and signal side $\tau^+ \rightarrow \ell^+ \alpha$, where $\ell = e/\mu$, to estimate the leptons' momentum p_ℓ ; the dotted line indicates the thrust axis. From [42]

Although there exists an irreducible background $\tau^+ \rightarrow \ell^+ \bar{\nu}_\ell \nu_\tau$ in the search for α particles via tau decays due to an identical visible topology to the alpha decay $\tau^+ \rightarrow \ell^+ \alpha$, the two-body signal decay holds a distinct form compared to the broad distribution of our background, the SM three-body decay. One can find a $\tau^+ \rightarrow \ell^+ \alpha$ peak in the momentum distribution

for the background decay $\tau^+ \rightarrow \ell^+ \bar{\nu}_\ell \nu_\tau$ depending on the α mass m_α , see Figure 13. By estimating the τ momentum in pseudo rest frame and measuring the lepton momentum p_ℓ of the decay $\tau^+ \rightarrow \ell^+ \alpha$ we can reconstruct the mass of the α particle since this is the only variable influencing p_ℓ . This is called the ARGUS Method, used by the ARGUS experiment during their data analysis phase [28, 32].

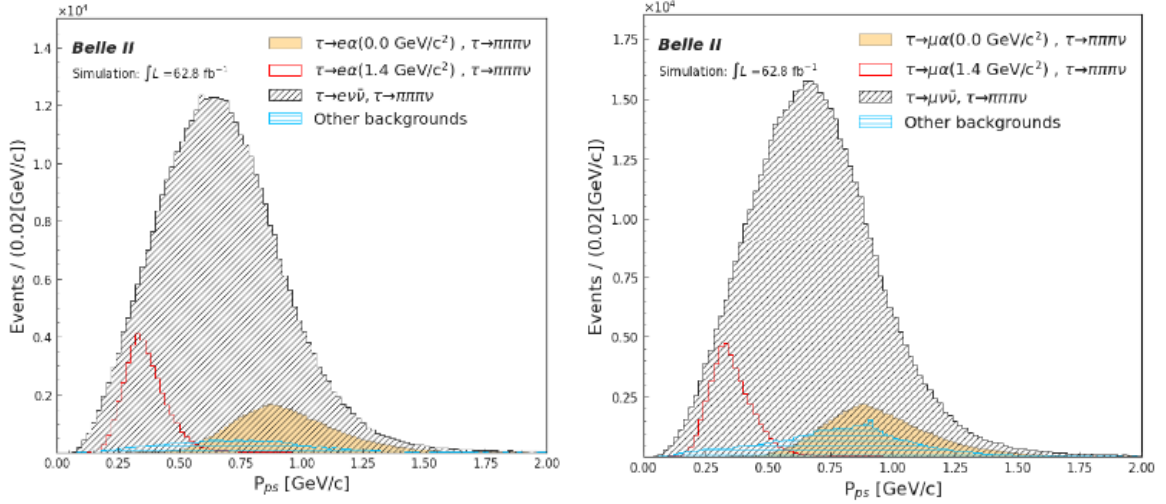


Figure 13: pseudo-rest frame τ momentum spectrum with signal $\tau^\pm \rightarrow \ell^\pm \alpha$ peak for α mass $m_\alpha = [0, 14] \text{ GeV}/c^2$; the signals' $\mathcal{BR}$ was scaled to a constant fraction 0.1; e -channel (left) and μ -channel (right). From [32].

This leads to a peak in the energy distribution and we, therefore, look at the number of events x_ℓ with excess energy on the signal side $\tau^- \rightarrow \ell^- \alpha$ compared to the spectrum of SM decay $\tau^- \rightarrow \ell^- \bar{\nu}_\ell \nu_\tau$ normalized to the SM background as shown in Equation 4.2. E_ℓ^* corresponds to the lepton energy in the pseudo rest frame.

$$x_\ell = \frac{E_\ell^*}{m_\tau c^2 / 2} \quad (4.2)$$

4.5 Event Reconstruction

Trigger: This online event selection only triggers for events with three or more localized energy cluster deposits in the ECL, while one event requires energy above 0.3 GeV and the other two above 0.1 GeV. Furthermore, a topology related to Bhabha scattering is excluded by the trigger system. The system was tested using independent triggers on experimental data gathered by the CDC. The trigger efficiencies are 97% and 86% for e^- and μ^- -channel, respectively.

Track Selection: We ensure the detection of possible $\tau^-\tau^+$ by requiring only events with exactly four charged particles with zero total charge on the tag side. The track associated with each charged particle is located within a 3 cm radius of the z-axis and within a 1 cm radius in the transverse plane.

Particle Identification: The charged particle on the signal side needs to be an electron or muon. The electron and muon identification is achieved by combining data collected from all subdetectors into a global discriminator similar to a likelihood ratio, called *track signal side Lepton ID*. To identify electrons and muons we set *track signal side Electron ID* > 0.9 and *track signal side Muon ID* > 0.9 , respectively. Hadrons are identified by requiring $E_{\text{ECL}}/p < 0.8$ for tag side particles (in this case $\tau^\pm \rightarrow h^\pm 3\pi^0$, where E_{ECL} and p are the particles energy measured in the ECL and its momentum, respectively. Both quantities are chosen to be in the laboratory frame.

Neutrals Reconstruction [π^0 - and γ -multiplicity]: To suppress hadronic background $e^+e^- \rightarrow q\bar{q}$, events with neutral pions or photons are completely rejected. π^0 s are identified via γ pairs within the mass range of [115, 152] MeV/c². Displaced clusters from secondary hadronic interactions and multiple clusters from low-momentum charged particles are a challenge though, since they are hard to simulate. Therefore, we require photons at the inner surface of the ECL with energies less than 0.4 GeV to be at least 40 cm from the next charged particle.

Photons used in neutral pion reconstruction are ECL energy clusters with at least 0.1 GeV which need to be matched by the CDC acceptance. Photons excluded by pion reconstruction with less than 0.2 GeV are rejected, even if the upper criteria were satisfied. The corresponding distribution can be seen in Figure 16(e) - 17(b) for the signal side and in Figure 15(c)-15(d) for the tag side.

4.6 Discriminating Variables

Tag Side p_T : The partial transverse momentum p_T of the tag side particles lets us discriminate background events of the type $e^+e^- \rightarrow \ell^+\ell^-\gamma$, $e^+e^- \rightarrow e^+e^-\ell^+\ell^-$ and $e^+e^- \rightarrow e^+e^-h^+h^-$. Those events are characterized by low momentum. We can require a lower threshold for partial transverse momentum to discriminate given background events. The distribution for a given luminosity can be seen in Figure 14(a)-14(f).

Thrust: We set bounds on the *thrust* values consistent with that of tau pair events to restrict $e^+e^- \rightarrow e^+e^-\gamma$ and $e^+e^- \rightarrow q\bar{q}$ even more. The distribution is given in Figure 17(c) and Figure 17(d).

Missing Momentum of Event in the CMS frame: We further discriminate the $e^+e^- \rightarrow e^+e^-\ell^+\ell^-$ events when requiring the negative vector sum of every reconstructed particle in one event, the *Missing Momentum of Event in the CMS frame*, to be within the polar angle range θ_{miss} where events can be accurately simulated. Figure 17(e) and Figure 17(f) show the corresponding distribution.

χ^2 Vertex Probability: We set restrictions on every background decay by implying a common origin of tag side particles. The probability of such vertex reconstruction is given as the χ^2 *Vertex Probability* and is defined as -1 for events where the vertex reconstruction failed. The feature distribution is shown in Figure 15(a) and Figure 15(b) for e-channel and μ -channel,

respectively.

τ Decay Kinematics: Further restrictions on the tag side are done by choosing only events where the tag side *particles mass* M_{3h} and *Center-of-Mass energy* E_{3h}^{CM} coincide with kinematics of the τ particle decay[32?]. The distribution of M_{3h} is given in Figure 16(a) and Figure 16(b) while the distribution of E_{3h}^{CM} is shown in Figure 16(c) and 16(d).

4.7 Monte Carlo sample generation

Signal: The software used for the production of $\tau^- \rightarrow \ell^- \alpha$ events is the official *Early Phase 3 MC production* of the Belle II experiment. We generate 10^7 events with nominal, run-independent beam background overlays (**BGxI**).

Background: For the background we use generic and low-multiplicity samples by *Early Phase 3 MC 13a production nominal beam backgrounds* (**BGxI**). We produce two batches with the internal *Belle II* software version *BASF2 release-04-00-03*, each of them re-scaled to 100 fb^{-1} , as displayed in Table2.

All ntuples for MC and data were produced using *basf2 release light-2104-poseidon* (*Belle II Analysis Software Framework*) over the full event. For further insight, we refer to [43, 44].

The different background contributions consist of several electron-positron final states each produced by their own software. *KKMC* [45, 46] is used to simulate $\tau^+ \tau^-$, $q \bar{q}$ and $\mu^+ \mu^- (\gamma)$

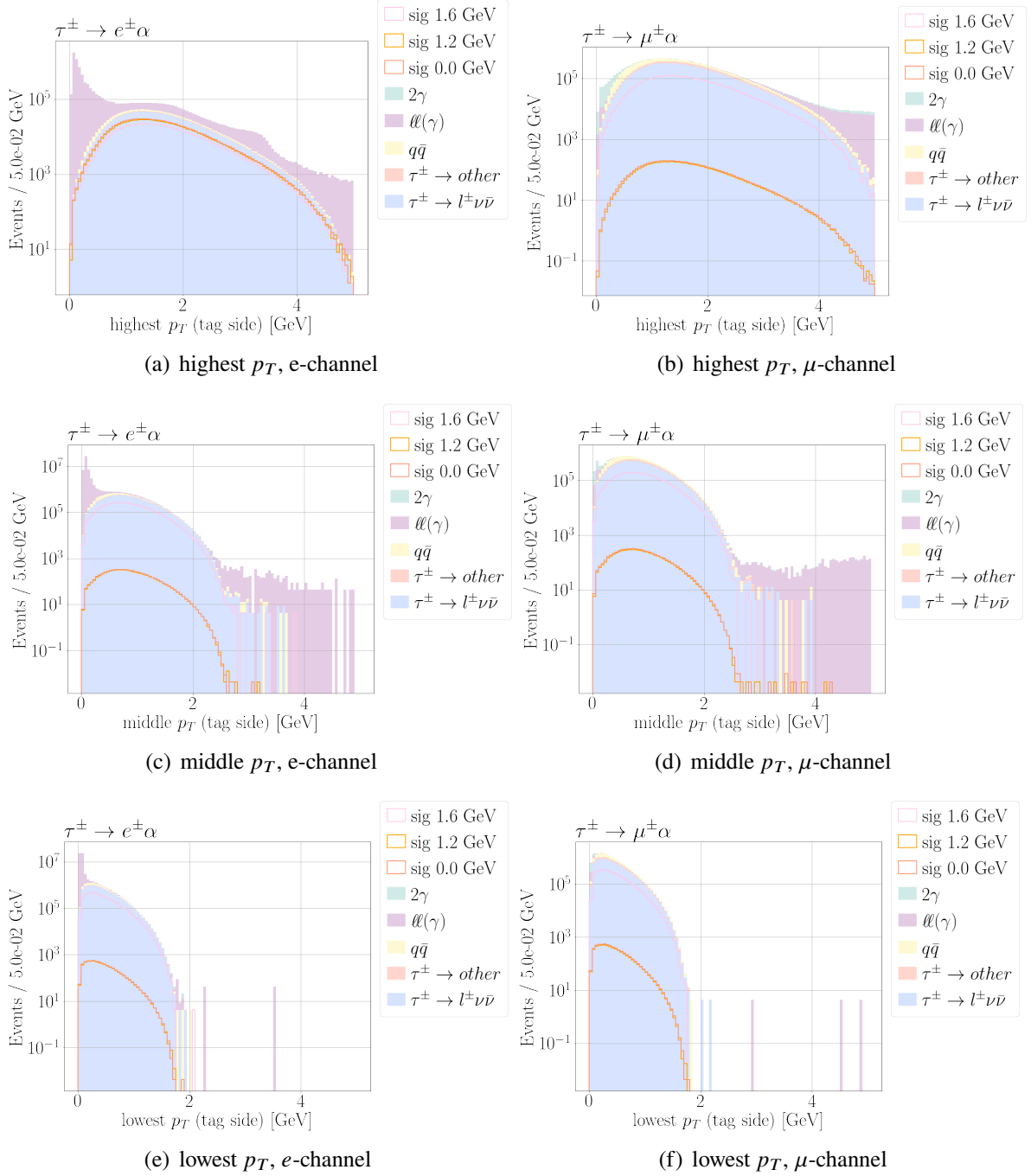


Figure 14: feature distribution of MC data weighted according to sample luminosity given in Table 2 and signal contributions for the alpha masses $m_\alpha = [0.0, 1.2, 1.6] \text{ GeV}/c^2$

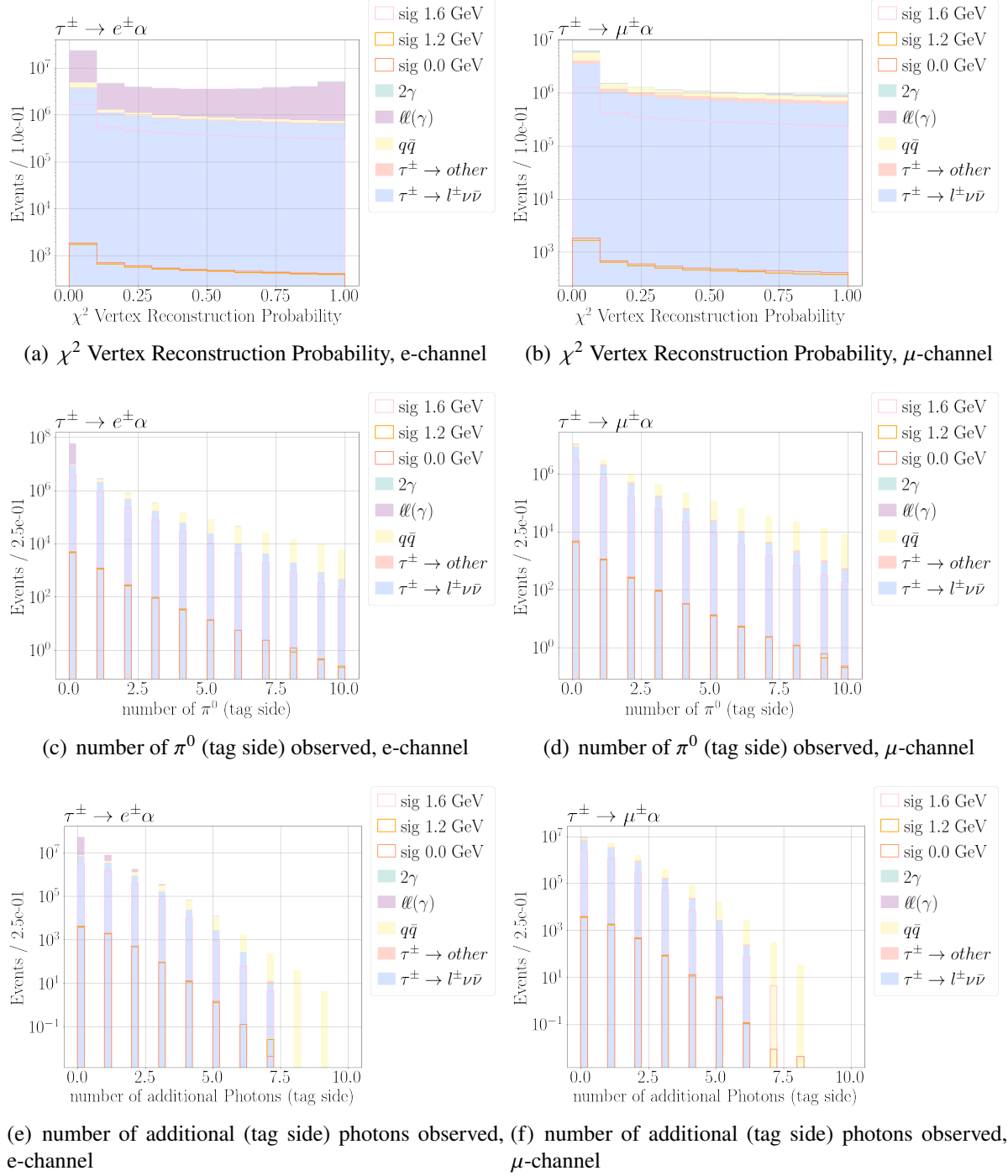


Figure 15: feature distribution of MC data weighted according to sample luminosity given in Table 2 and signal contributions for the alpha masses $m_\alpha = [0.0, 1.2, 1.6] \text{ GeV}/c^2$

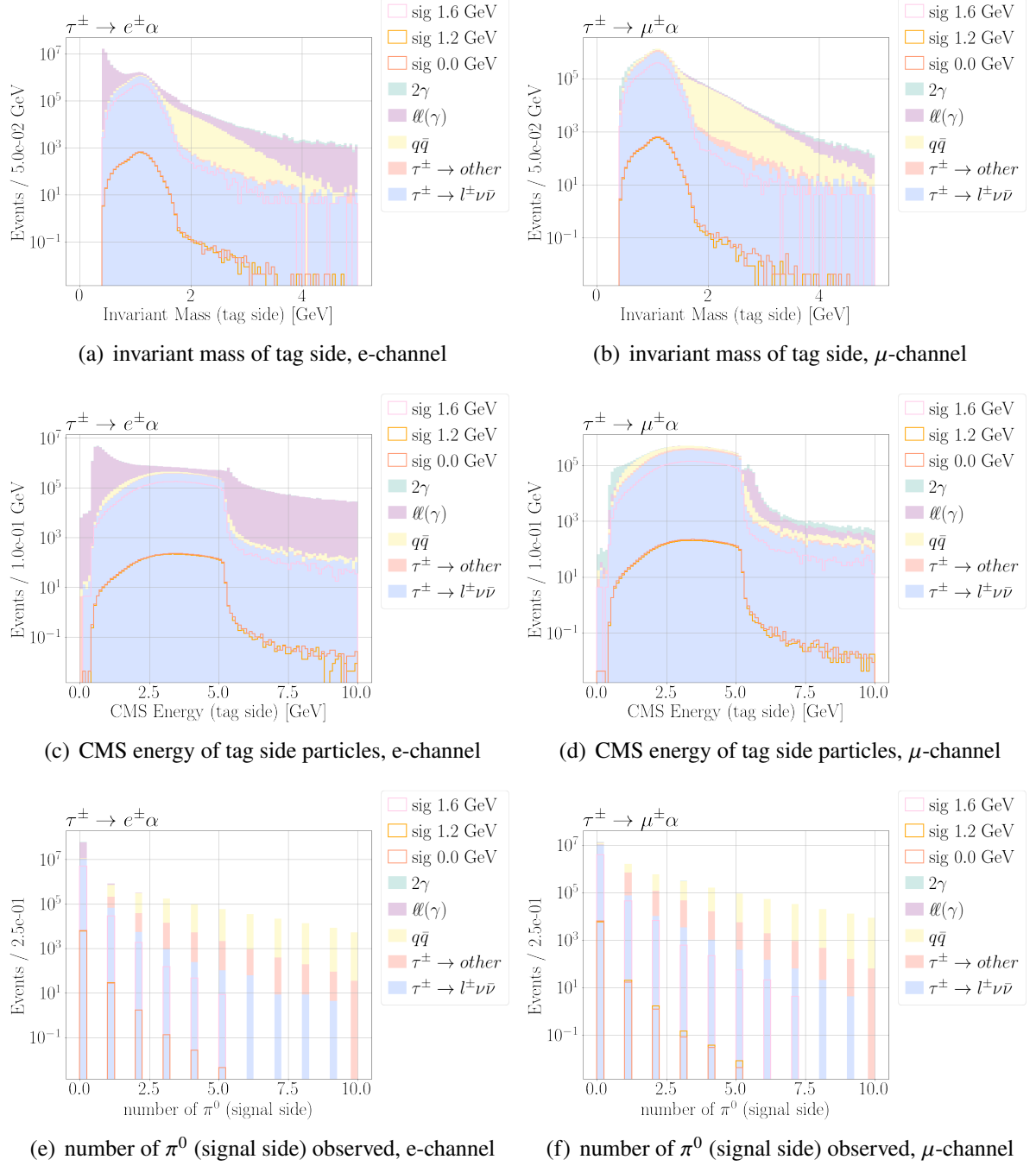
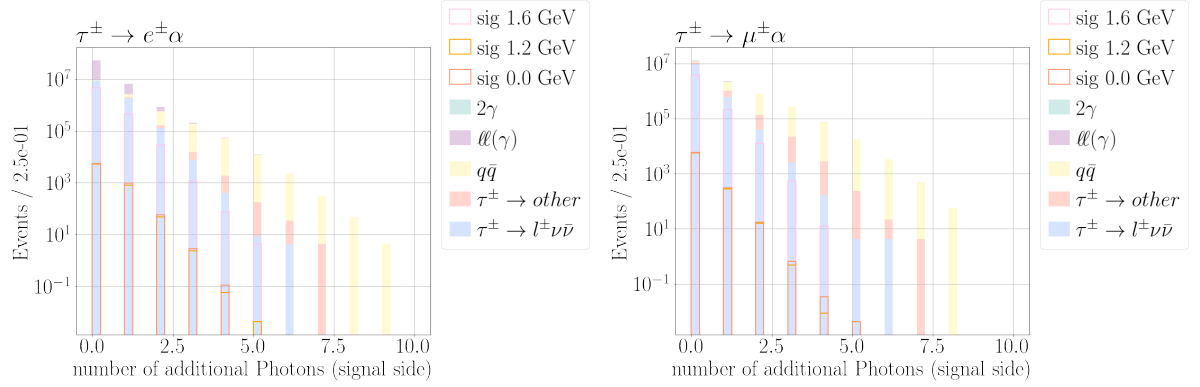
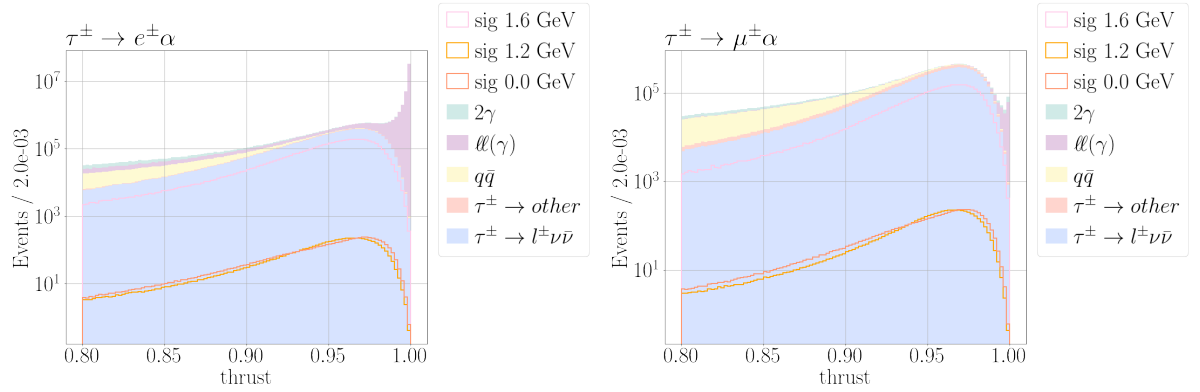


Figure 16: feature distribution of MC data weighted according to sample luminosity given in Table 2 and signal contributions for the alpha masses $m_\alpha = [0.0, 1.2, 1.6]$ GeV/c²; electron-channel (left) and muon-channel (right)

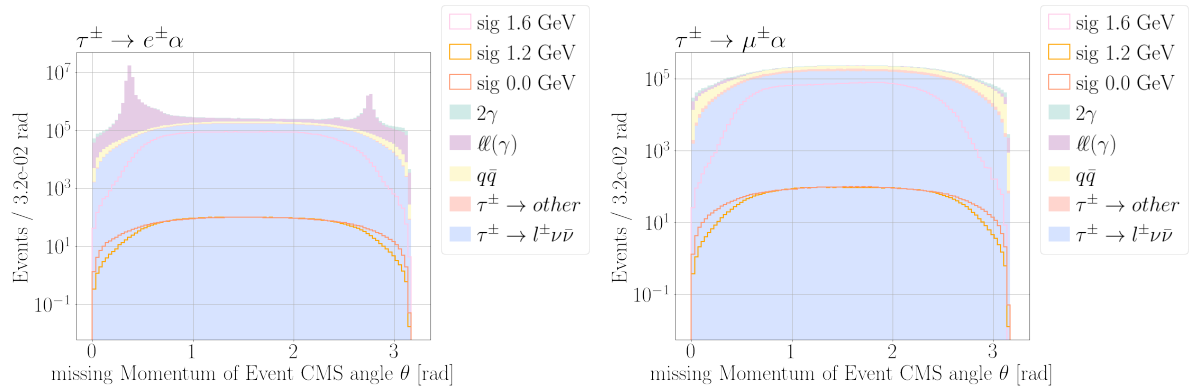


(a) number of additional photons (signal side) observed, e-channel (b) number of additional photons (signal side) observed, μ -channel



(c) thrust value, e-channel

(d) thrust value, μ -channel



(e) angle θ of missing momentum of event in CMS rest frame, e-channel (f) angle θ of missing momentum of event in CMS rest frame, μ -channel

Figure 17: feature distribution of MC data weighted according to sample luminosity given in Table 2 and signal contributions for the alpha masses $m_\alpha = [0.0, 1.2, 1.6]$ GeV/ c^2 ; electron-channel (left) and muon-channel (right)

Processes	Name of the production	Integrated luminosity (per batch) [fb ⁻¹]
$e^+e^- \rightarrow \tau\tau$	‘taupair’	100
$e^+e^- \rightarrow u\bar{u}$	‘uubar’	100
$e^+e^- \rightarrow d\bar{d}$	‘ddar’	100
$e^+e^- \rightarrow s\bar{s}$	‘ssbar’	100
$e^+e^- \rightarrow c\bar{c}$	‘ccbar’	100
$e^+e^- \rightarrow B\bar{B}$	‘mixed’, ‘charged’	100
$e^+e^- \rightarrow e^+e^-\gamma$	‘ee’	10
$e^+e^- \rightarrow \mu^+\mu^-\gamma$	‘mumu’	100
$e^+e^- \rightarrow e^+e^-e^+e^-$	‘eeee’	100
$e^+e^- \rightarrow e^+e^-\mu^+\mu^-$	‘eemumu’	100
$e^+e^- \rightarrow e^+e^-\pi^+\pi^-$	‘eepipi’	1000
$e^+e^- \rightarrow e^+e^-K^+K^-$	‘eeKK’	1000
$e^+e^- \rightarrow e^+e^-p^+p^-$	‘eep p’	1000

Table 2: MC 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; it’s scaled down 10x and replicated for each batch.

production, *BaBaYaga@NLO* [47–51] for e^+e^- and *AAFH* and *TREPS* for non-radiative $\gamma\gamma$ production [52–55]. *TAUOLA* [56] is used for SM τ decays. Their radiative corrections are simulated by *PHOTOS* [57] and the NP signal $\tau^- \rightarrow \ell^- \alpha$ with 0-spin alpha particles are produced by *PYTHIA8.2* [58] for the alpha masses [0.0, 0.5, 0.7, 1.0, 1.2, 1.4, 1.6] GeV/c². To simulate the detector response The *Belle II Analysis Software Framework* [59, 60] and *Geant4* [61] were used. From [32].

5 Current Method: cut-based Analysis

In order to perform a world-leading search for the LFV decay $\tau^- \rightarrow \ell^- \alpha$ it is essential to have excellent signal-background separation. So far at *Belle II* this was achieved using a cut-based approach where a set of rectangular cuts were applied on the discriminating variables listed in Table 3. The variables and their cuts were chosen so that good background discrimination is possible while still keeping most of the signal. These background sensitive variables are described in section 5.1.

5.1 Method

A good choice of background discriminating variables used for a cut-based approach on *Belle II* data is crucial when exploring low-multiplicity physic processes such as $\tau^- \rightarrow \ell^- \alpha$. When looking at the feature distributions in Figures 14-17, where the full MC data sample is presented, one can see that for such a choice the main background contributions of the normalizing channel can be cut-off quite easily without a general loss of signal events using rectangular cuts. The core of this procedure is to find a suitable threshold for the applied cuts. At *Belle II* an optimal signal to background reduction is achieved by maximizing the Figure of Merit (FOM, See Equation 5.1), where S and B stand for signal and background, respectively. We set S to the SM decay $\tau^- \rightarrow \ell^- \bar{\nu}_\ell \nu_\tau$ due to its similar spectrum. This leads to a FOM which is independent on the alpha mass m_α since there is no contribution to S or B originating from $\tau^- \rightarrow \ell^- \alpha$.

$$FOM = \frac{S}{\sqrt{S+B}} \quad (5.1)$$

The cuts for the recent Belle II data published [32] are summarized in Table 3.

	$\tau^- \rightarrow e^- \alpha$	$\tau^- \rightarrow \mu^- \alpha$
highest p_T	> 0.70	> 0.47
middle p_T	> 0.03	> 0.17
lowest p_T	> 0.08	> 0.04
thrust	$[0.90, 0.99]$	$[0.90, 1.00]$
θ_{miss}	$[20^\circ, 160^\circ]$	$[20^\circ, 160^\circ]$
E_{3h}^{CM}	$[1.2, 5.3] \text{ GeV}$	$[1.1, 5.3] \text{ GeV}$
M_{3h}	$[0.5, 1.7] \text{ GeV}/c^2$	$[0.4, 1.7] \text{ GeV}/c^2$

Table 3: Summary of cuts on features [32]

5.2 Recent status of *Belle II*

Belle II has published a search for invisible spin-0 α bosons using the data collected by the SuperKEKb collider between 2019 and 2020 which corresponds to 62.8 fb^{-1} . The current method for data analysis and foremost background reduction is based on rectangular cuts. This cut-based analysis focuses on a small set of discriminating variables for various background contributions, as discussed in Chapter 5.1. The method was evaluated using a simulation with MC samples.

As described in Section 4.4 *Belle II* counts the number of events with relative excess energy (Equation 4.2) which is displayed in Figure 18 for the experimental and simulated data sets in the signal region.

The current best upper limits (UL) for the branching ratio of $\mathcal{B}(\tau^- \rightarrow \ell^- \alpha) / \mathcal{B}(\tau^- \rightarrow \ell^- \bar{\nu}_\ell - \nu_\tau)$, where ℓ is either an electron or muon, is determined by applying this cut-based approach in combination with the CLs method [62]. When looking at the full data-set of 62.8 fb^{-1} they set the UL with 95% CL between $(1.1 - 9.7) \times 10^{-3}$ for the e -channel and $(0.7 - 12.2) \times 10^{-3}$ for the μ -channel, as shown in Figure 19 and summarized in Table 4 [32].

M_α [GeV/c ²]	$\mathcal{B}_{e\alpha}/\mathcal{B}_{e\bar{\nu}\nu}$ ($\times 10^{-3}$)	UL at 95% CL ($\times 10^{-3}$)	UL at 90% CL ($\times 10^{-3}$)
0.0	-8.1 ± 3.9	5.3 (0.94)	4.3 (0.76)
0.5	-0.9 ± 4.3	7.8 (1.40)	6.5 (1.15)
0.7	1.7 ± 4.0	9.0 (1.61)	7.6 (1.36)
1.0	1.7 ± 4.2	9.7 (1.73)	8.2 (1.47)
1.2	-1.1 ± 2.6	4.5 (0.80)	3.7 (0.66)
1.4	-0.3 ± 2.6	1.8 (0.32)	1.5 (0.26)
1.6	0.2 ± 0.5	1.1 (0.19)	0.9 (0.16)
M_α [GeV/c ²]	$\mathcal{B}_{\mu\alpha}/\mathcal{B}_{\mu\bar{\nu}\nu}$ ($\times 10^{-3}$)	UL at 95% CL ($\times 10^{-3}$)	UL at 90% CL ($\times 10^{-3}$)
0.0	-9.4 ± 3.7	3.4 (0.59)	2.7 (0.47)
0.5	-32 ± 3.9	6.2 (1.07)	5.1 (0.88)
0.7	2.7 ± 3.4	9.0 (1.56)	7.8 (1.35)
1.0	1.7 ± 5.4	12.2 (2.13)	10.3 (1.80)
1.2	-0.2 ± 2.4	3.6 (0.62)	2.9 (0.51)
1.4	0.9 ± 0.9	2.5 (0.44)	2.2 (0.38)
1.6	-0.3 ± 0.5	0.7 (0.13)	0.6 (0.10)

Table 4: recent Belle Data with UL, complementary Table to Figure 19; UL limits in brackets are calculated using SM branching fractions; electron channel and muon channel are represented in the upper and lower table, respectively. From [32].

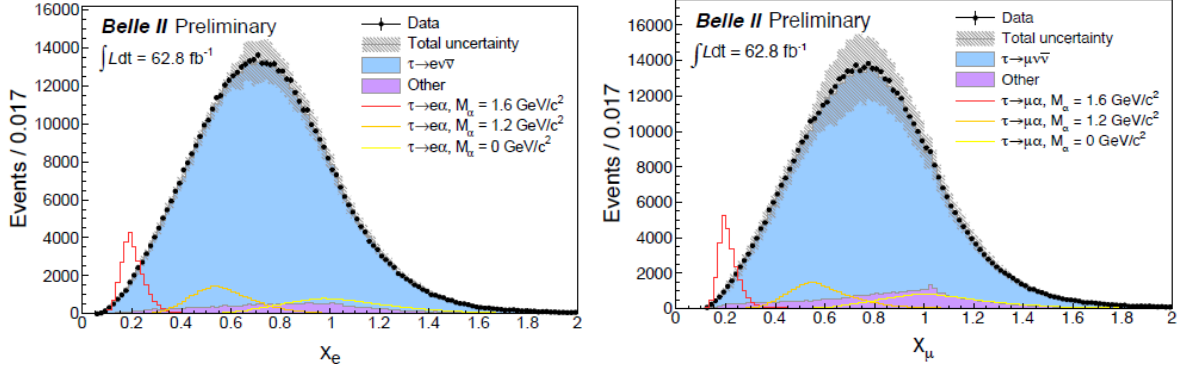


Figure 18: Spectra of x_ℓ (Eq. 4.2) for e -channel (left) and μ -channel (right) in experimental and MC data; SM background apart from main background $\tau \rightarrow \ell \bar{\nu} \nu$ is combined into "Other"; uncertainty (grey area) is dominated by lepton-identification efficiency; assumed $\mathcal{BR}$ fraction of $\tau^- \rightarrow \ell^- \alpha$ set to 5%. From [32].

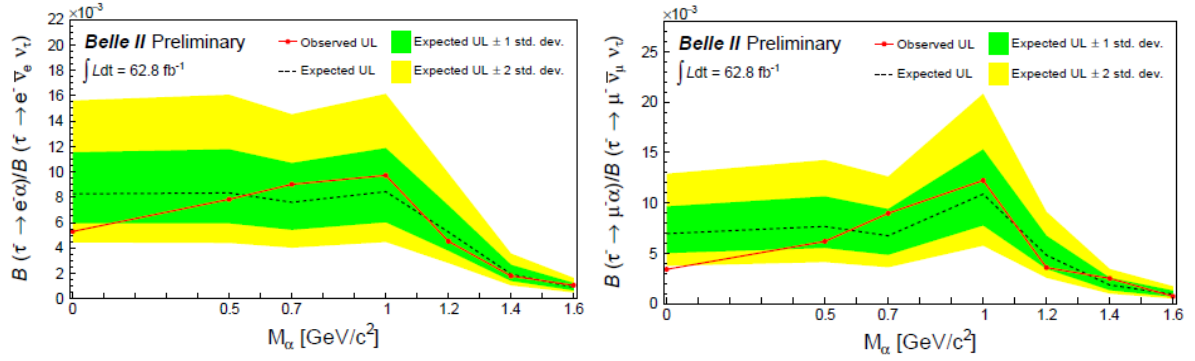


Figure 19: Recent Belle II upper limits within 95% CL for the branching fraction of $\mathcal{B}(\tau^- \rightarrow \ell^- \alpha) / \mathcal{B}(\tau^- \rightarrow \ell^- \bar{\nu}_\ell - \nu_\tau)$ and expectation values from background-only hypothesis for α mass: [0 - 1.6] GeV in e -channel (left) and μ -channel (right). From [32].

6 New Approach: XGBDT training

While the simple cut-based approach for background rejection described in the last chapter was sufficient for delivering a world-leading result for *Belle II* on the early 62.8 fb^{-1} of data, we expect a large space for improvement when using more modern machine learning techniques (ML) instead. The development and testing of such a ML technique is the main topic of this thesis. To achieve a signal-background separation a ML algorithm, namely Extreme Gradient Boosted Decision Tree training (*XGBDT*), to which we refer in this article with *BDT*, was used on MC sample data in combination with rectangular cuts on various physics variables. The BDT training features (variables) were chosen carefully so that a good signal-background separation is given. Those discriminating variables are taken from the *Belle II* cut-based analysis approach (Chapter 5.1) in combination with further background discriminating features, listed as ‘additional variables’ in Table 6, carefully chosen to give the best BDT performance.

6.1 Decision Trees

At the basis of many ML Algorithms we find the Decision Trees using the Gradient Tree Algorithm. We will look at the concept of Decision Trees and what lies beyond when further exploring Extreme Gradient Boosted Decision Trees, the algorithm which was used for this thesis.

Each tree consists of branches and leaves, see Figure 20. Branches characterize Yes-or-No-questions. When passing a branch, prediction values between -1 and 1 are assigned to each constituent whether they pass or fail. For the case that there is no further branch we end up in a leaf and add up their gathered prediction values. Combining several decision trees gives a statistical advantage, especially when parameters were exchanged. A combination of several trees is possible by adding the prediction values for each tree, as it can be seen in Figure

20.

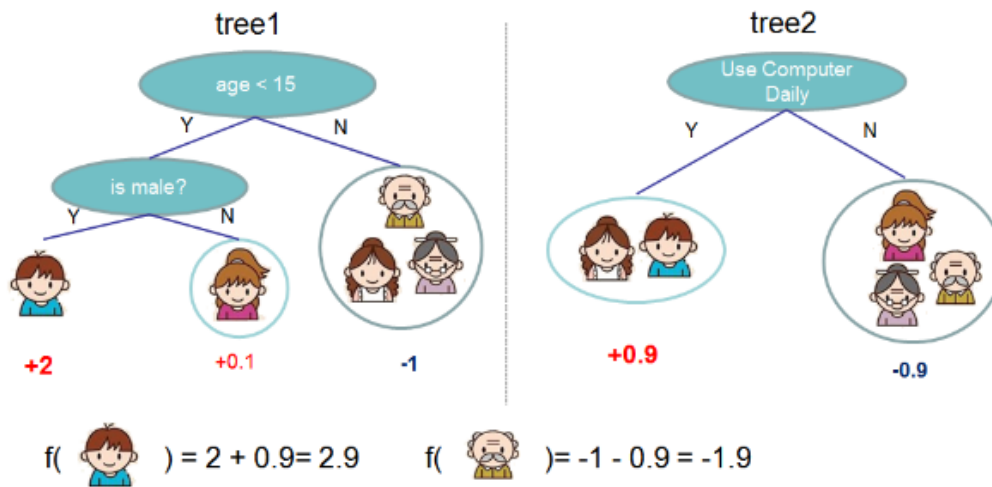


Figure 20: Decision tree scheme to illustrate the method of combining trees by looking at the leaves. We calculate their corresponding prediction values and add them together. From [63].

The concept of Decision Trees has a wide spectrum of possibilities when concerning Machine Learning Algorithms. Such algorithms enhance the functionality by relying on extrapolations in form of e.g. 'Loss Function' or 'Learning Rate'. We will investigate further on this topic in the next section where the principle of Extreme Gradient Boosting is explained.

6.2 Extreme Gradient Boosted Decision Trees

The Extreme Gradient Boosted Decision Tree method is a development originating from Decision Tree Algorithms and a "scalable end-to-end tree boosting system" [63]. It has several advantages compared to other known Machine Learning methods. The main improvements lie in the novel sparsity-aware algorithm for handling sparse data sets and weighted quantile sketch for better approximate tree learning. These concepts are explained in the following paragraphs.

A general **Gradient Boosted Decision Tree** algorithm is a flowchart-like structure where each

internal node represents a test on a feature, each branch represents an outcome of the test, and each leaf node represents a class label or a numerical value. Decision trees recursively split the data based on training features to create a hierarchical structure that helps in making predictions. *Boosting* combines multiple weak models (base learners) to create a strong predictive model. Each base learner focuses on the instances that were previously misclassified by the preceding learners. This iterative process improves the overall predictive accuracy.

Gradient boosting is a specific type of boosting algorithm that minimizes the loss function, which measures the difference between prediction and target value, by iteratively adding weak models to the ensemble. These gradients are calculated for each iteration and represent the difference between the predicted values and the true target values. The loss function is optimized using gradient descent, hence the name "gradient boosting". A weak model is a new decision tree which is fit to the negative gradients (residuals) of the loss function. When training a weak model we use a *Greedy Algorithm* which starts with the residuals/leafs and iteratively adds branches. The algorithms aim to reduce the difference between the previous predictions and the true target values. To avoid overfitting and improve generalization, a learning rate is introduced. This controls the contribution of each weak model to the ensemble. The final prediction is then obtained by combining the predictions of all the weak models in the ensemble. Each model's prediction is weighted by its learning rate, and the sum of these predictions gives the overall prediction of the Gradient Boosted Decision Tree algorithm. [63]

Weighted quantile sketch: To assure an even distribution of feature candidates across the data set one has to introduce (candidate) split points. We, therefore, introduce a function which represents the proportion of instances whose feature value k is smaller than z , the so-called rank function $r_k(z) = \frac{\sum_{(x,h), x < z} h}{\sum_{(x,h)} h}$, where $(x, h) \in D_k = (x_1k, h_1), (x_2k, h_2), \dots, (x_nk, h_n)$. x represents the values of the k -th features and h the second-order gradient statistics of each training instance. The latter can be interpreted as weight. This leads to the split condition

6.1.

$$|r_k(s_{k,j}) - r_k(s_{k,j+1})| < \varepsilon, s_{k1} = \min_i x_{ik}, s_{kl} = \max_i x_{ik} \quad (6.1)$$

To find split criteria which fulfill this condition is no trivial task, especially for large data sets. Therefore, a **Weighted Quantile Sketch Algorithm** is proposed. The algorithm can handle weighted data with a "provable theoretical guarantee". The most basic idea is to introduce a data set with affinity for prune and merge operations. For a detailed description we refer to the *Appendix* of [63].

Sparsity-aware Algorithm: Since the data sets of real-world problems are often sparse, the addition of a default direction for every feature in each tree node can improve BDT statistics. A missing value is classified into a default direction. There are two default directions where the optimal default directions are learned from data. Hence the main improvement is to mostly visit non-missing data entries.

6.3 Methods

We apply rectangular pre-selection cuts before BDT training. Then the BDT training is performed on a carefully chosen set of training features (pre-selection cuts: Table 5, BDT training features: Table 6). Pre-selections are necessary to remove regions where MC simulated data poorly describes actual data. The regions known within the *Belle II* collaboration are selected via the cuts in Table 5. It should be mentioned that the training was done separately for electron- and muon-channel as well as for each alpha mass point.

$$weights = \frac{\text{MC sample luminosity}}{\text{data luminosity}} \quad (6.2)$$

The Figures 14-17 and Figure 21 display feature distributions, e.g. the expected number of events given by MC data including pre-selection cuts for the α masses $m_\alpha = [0.0, 1.2, 1.6]$

GeV/c² and stacked background. Weights applied to the different particle contributions are given by Equation 6.2 with the corresponding MC sample luminosity according to Table 2 which is the number of events generated by the MC simulation. The data luminosity is the current total integrated luminosity of *Belle II*:

$$\int_{tot} \mathcal{L}_{data} = 427.79 \text{ fb}^{-1}$$

The plots are given in log-scale so that alterations in the different signal channels can be observed more easily. We separate the various background contributions to give a clear insight into each and to emphasize the importance of the normalization channel $\tau^- \rightarrow \ell^- \nu_\tau \bar{\nu}_\ell$ as the main background source.

pre-selection variable	cut
<i>thrust</i>	> 0.8
(tag side τ χ -Vertex Prob.)	> 0.0

Table 5: Pre-selection cuts applied before the BDT training.

cut-based features	highest p_T middle p_T lowest p_T thrust θ_{miss} E_{3h}^{CM} M_{3h}
additional variables	momentum in pseudo rest frame (signal side) missing momentum of event (CMS) visible energy of event (CMS)

Table 6: Summary of all features used for BDT training

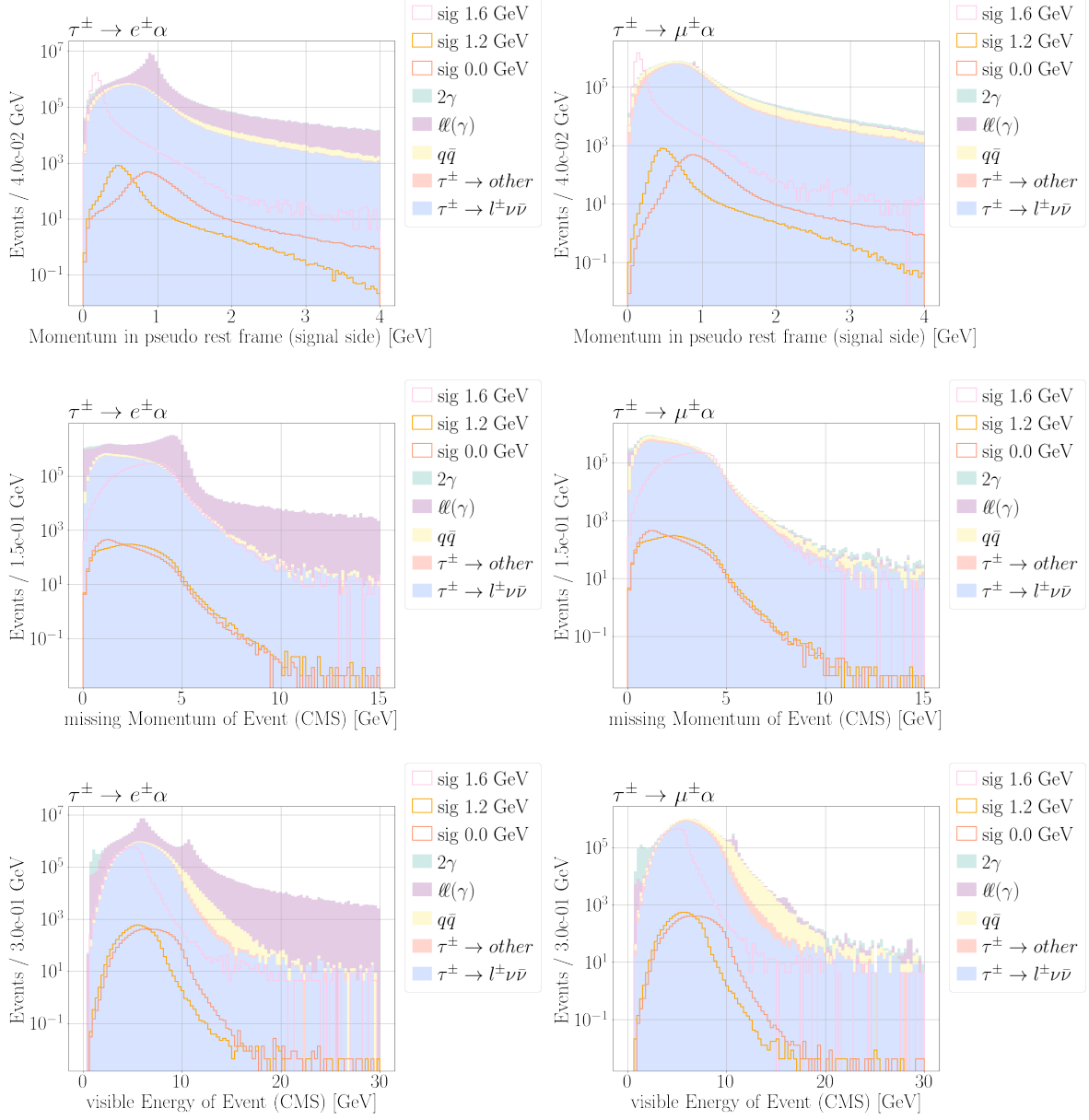


Figure 21: added training features: feature distribution of MC data weighted according to sample luminosity given in Table 2 and signal contributions for the alpha masses $m_\alpha = [0.0, 1.2, 1.6] \text{ GeV}/c^2$; electron-channel (left) and muon-channel (right).

The traditional variables used in the past and described earlier in Chapter 5.1 are great at suppressing all backgrounds except for τ -pair production events with SM $\tau^- \rightarrow \ell^- \bar{\nu}_\ell \nu_\tau$ decay. This is because the $\tau^- \rightarrow \ell^- \alpha$ decay has a very similar visible topology to that of $\tau^- \rightarrow \ell^- \bar{\nu}_\ell \nu_\tau$. To maximize the performance of the BDT and suppress the $\tau^- \rightarrow \ell^- \bar{\nu}_\ell \nu_\tau$ background we include three additional variables to the training.

The first additional variable is the *Lepton Momentum in the τ Pseudo Rest Frame*. For $\tau^- \rightarrow \ell^- \alpha$ decays the magnitude of the lepton momentum depends only on the α mass. This two-body decay provides a distinctive signature for our signal above the spectrum of the three-body SM $\tau^- \rightarrow \ell^- \bar{\nu}_\ell \nu_\tau$ decay, for which the lepton momentum has a broad distribution. This can be fully exploited by going to the τ pseudo rest frame using the *ARGUS* method, as explained earlier in Section 4.4.

Moreover, the higher the mass of the undetected α particle the smaller tend to be the total *Visible Energy* of the event and the larger magnitude of the *Missing Momentum*. Each variable peaks for low and high alpha mass, respectively. Therefore, we also include them as training variables. The addition of those variables highly increases the signal-background separation.

6.3.1 Training Feature Correlations

We look for correlations in the BDT training features for signal and background contributions separately. We choose a low, middle, and high mass point for a general description of the feature correlations and their dependence on alpha mass $m_\alpha = [0.0, 1.2, 1.6] \text{ GeV}/c^2$. For correlation matrices of the other mass points, we refer to Chapter 8 in the Appendix. As one can see in Figure 22-23, the correlations behave in a similar way for different alpha masses and there are no significant correlations between those variables apart from a few exceptions which we will discuss shortly.

There is a small correlation in signal and background for both the electron and the muon channel between *middle* and *lowest partial transverse momentum* p_T . This can be explained

by interchangeable *middle* p_T and *lowest* p_T for values close to each other.

The correlation between the *highest* p_T and *tag side* τ *CMS Energy* should be clear since the energy of the tag side τ has a direct influence on the momentum of its decay products. Although there seems to be a higher correlation for the latter two p_T s, the same is valid for the correlation of *tag side* τ *Invariant Mass* and *highest, middle and low* p_T , especially for the background of the $e^\pm$ -channel.

Whilst there is a stronger correlation between *Missing Momentum of CMS Event*, *Visible Energy of CMS Event*, τ *CMS Energy* and *highest* p_T , the interconnection between the first three features is straight forward. The correlation of the highest p_T on the tag side is due to the fact that the *Visible Energy*, *Missing Momentum* and τ *CMS Energy* of the event are also measured on the tag side (by CMS) and the *highest* p_T has the greatest influence on total energy and momentum, in comparison the other p_T values.

The three p_T values have some correlation which could be explained by the interchangeability of their order for values close to each other.



Figure 22: Correlation matrix of BDT training features with signal (upper row) and background (lower row) events; electron-channel (left) and muon-channel (right); $m_\alpha = 0.0 \text{ GeV}/c^2$ and $m_\alpha = 1.2 \text{ GeV}/c^2$ in first and second two rows, respectively.

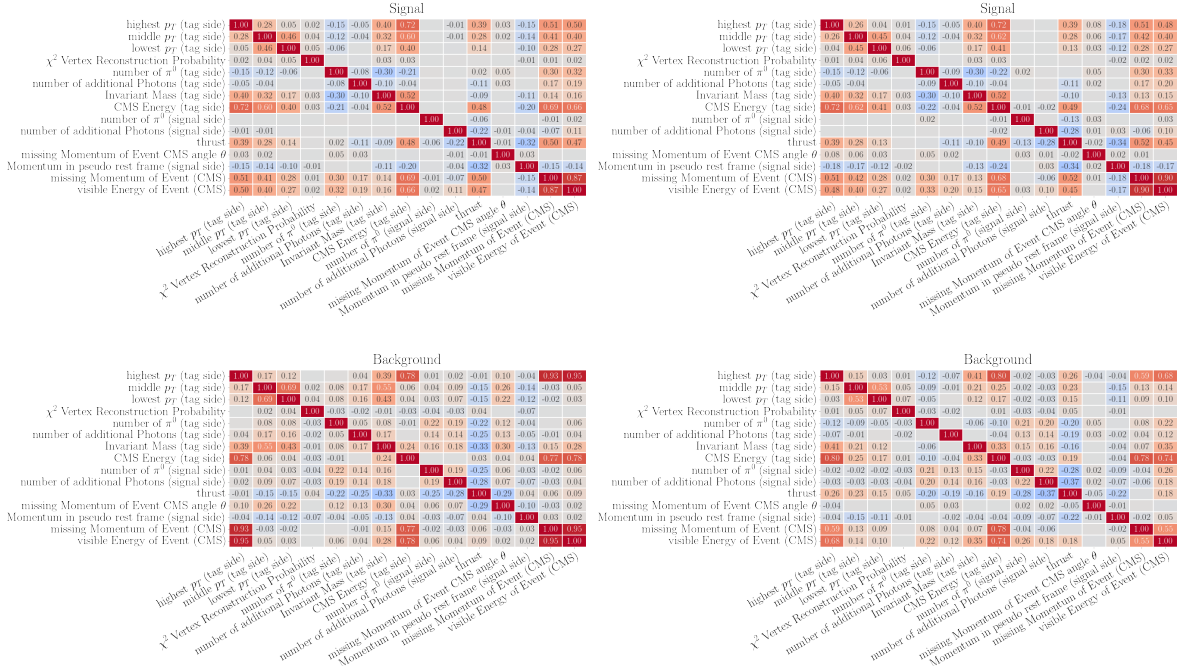


Figure 23: Correlation matrix of BDT training features for $m_\alpha = 1.6$ GeV/ c^2 signal (upper row) and background (lower row) events; electron-channel (left) and muon-channel (right).

6.3.2 Feature Ranking

We look at the BDT feature ranking evaluating the significance of each feature for the BDT training using k-fold cross validation [64] (here $k = 5$). Figure 24-25 displays the ranking for each alpha mass point where electron and muon channels are placed on the left and right-hand side, respectively. When we focus on the similarities of these plots we recognize a certain pattern concerning the relevance of each training feature.

middle p_T and *highest p_T* take turns as one of the most important features to train on. Values for mid and high p_T are in close proximity which leads to classification problems confusing *middle p_T* and *highest p_T* as discussed in the previous subsection 6.3.1. Nevertheless, these training features are of great importance, especially in the e -channel, due to their ability to disregard SM events. The *number of π^0* observed on the signal side is the leading feature in the μ -channel by far as it discards $q\bar{q}$ decays with π^0 in the final decay product. *Momentum in pseudo rest frame (Argus Momentum)* gains importance with raising alpha mass in both, electron- and muon-channel. We already argued that the signal side τ momentum in the pseudo rest frame has a distinct spectrum for 2-body decays (signal) in comparison to 3-body decays (SM background). The momentum distributions develop sharp peaks for larger alpha masses as it solely depends on m_α (Fig. 21).

We also tested a diminished set of training features which can help to avoid overtraining. For the sake of the extent of this thesis, we will only briefly discuss its outcome. We discarded the three worst overall performing features which lead to slightly lower Punzi FOM and therefore worsens the performance. This could be negligible if it wouldn't also lower the stability of the BDT performance as the peaks of signal and background in the BDT output distribution were weakened tremendously for some mass points. Therefore, one can imply that a diminished set of training variables lowers stability and performance and is thus not recommended.

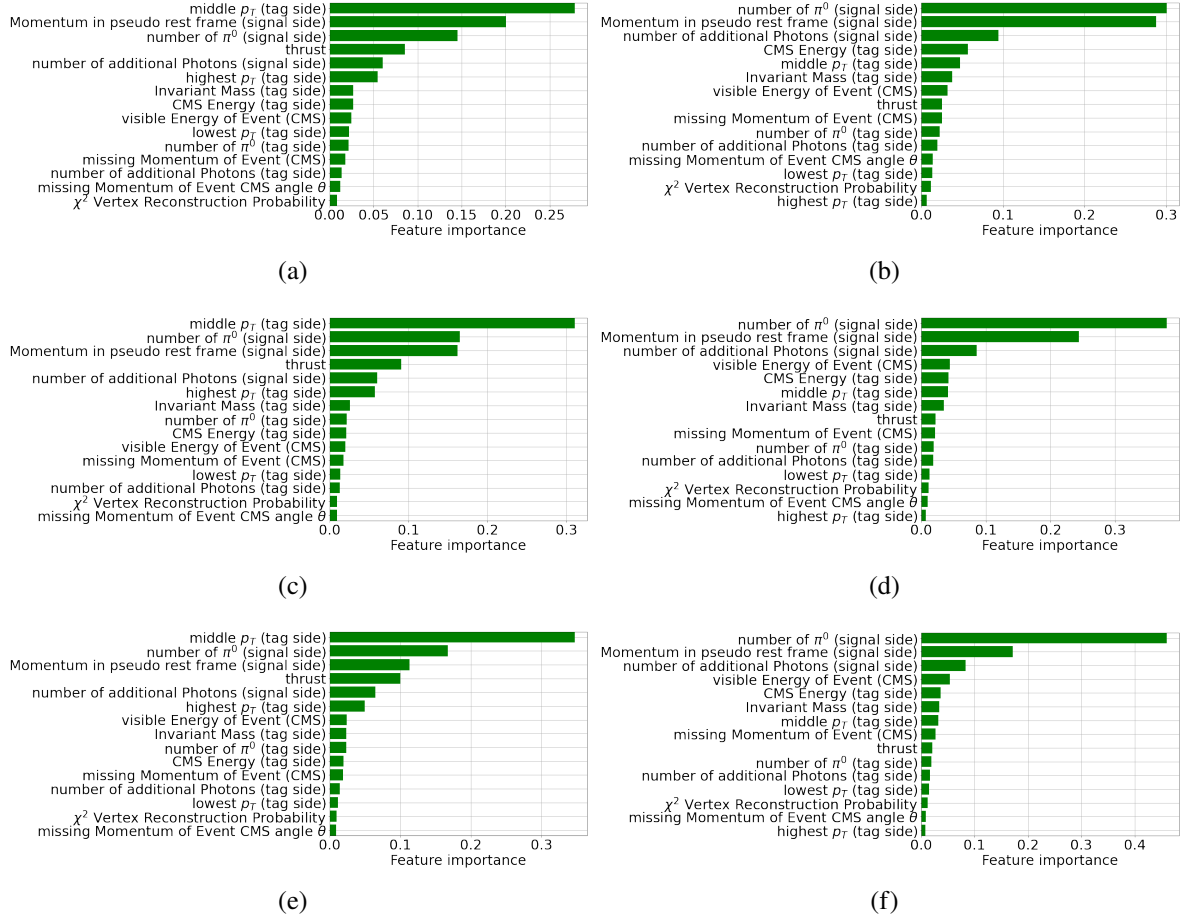


Figure 24: ranking the BDT features according to their significance in BDT training for alpha mass $m_\alpha = [0.0, 0.5, 0.7] \text{ GeV}/c^2$.

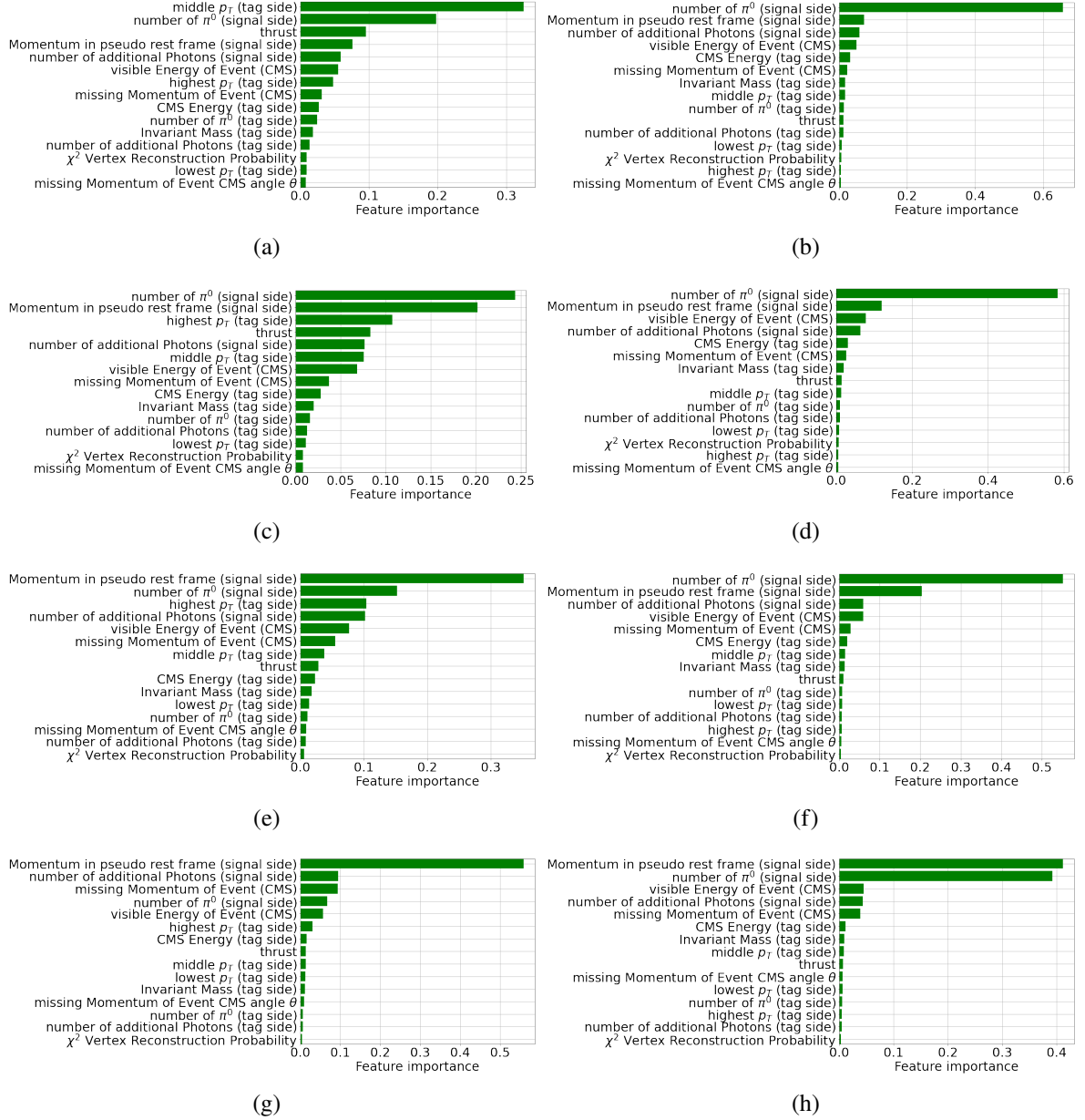


Figure 25: ranking the BDT features according to their significance in BDT training for alpha mass $m_\alpha = [1.0, 1.2, 1.4, 1.6] \text{ GeV}/c^2$.

6.4 Results

First we discuss the benefits of introducing adjusted weights for training and compare the results with the regular approach using sample luminosity. To give an overview of the BDT performance concerning background reduction we look at the training and test results of the BDT training. The BDT output values are given by prediction values normalized to 1 as discussed in Chapter 6.1. We are looking for stability in the training-to-test ratio and an overall good shape for the BDT output distribution with clear peaks at the upper and lower ends.

6.4.1 Test of adjusted weights

There is a well-known approach to weighted BDT training where the total background gets weighted such that signal and background lead to the same number of events, let's call it adjusted weight. This can improve the BDT performance. As shown in Figure 26 for $m_\alpha = 1.6$ GeV/ c^2 within the muon channel the direct comparison of regular weighting with adjusted weighting gives clearer peaks of signal and background at the upper end of the BDT output spectrum, although, its training-to-test ratio is weaker as well as a slightly worse performance when comparing the Punzi FOM values. This is true for every mass point in electron- and muon-channel (see following subsection, Figure 33(a)-35(b)). Since the Punzi FOM values are negligibly lower after the BDT training with adjusted weights and BDT has a well developed peaks we decide to stay with this method for more stability while training. For a comparison between those two methods one is invited to look into Chapter 8 where the BDT output plots with the regular BDT training are given for all α mass points (see Appendix, Chapter 8, Figure 38-40).

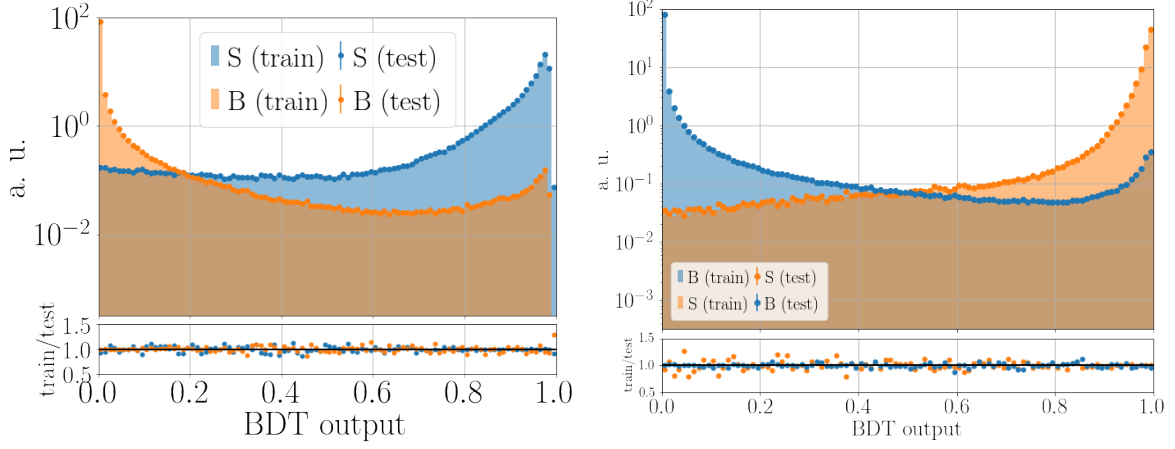


Figure 26: Direct comparison of BDT training methods (log-scale), regular method (left) and adjusted weighting (right) for $m_\alpha = 1.6 \text{ GeV}/c^2$. The train/test ratios are given in the plots below.

Punzi Figure of Merit

The Punzi FOM [65] is defined in Equation 6.3, where ε , a , B and t are efficiency, number of sigmas (as a measure of uncertainty) at a certain significance level, background and the total set of cuts, respectively. a is set to 1.28. L denotes the given luminosity. When the Punzi FOM of rectangular cuts (Table 3) exceeds the Punzi value for BDT training we can consider using the former approach for a better signal-background separation. As a qualitative measure of different data analysis methods, it yields the advantage (compared to the regular FOM, Equation 5.1) of canceling out any dependencies on signal normalization. Since the branching ratio of the NP signal is unknown this method gives us the possibility to quantitatively assess the sensitivity for different analysis approaches and improve optimization procedures. [65]

$$Punzi = \frac{\varepsilon(t)}{a/2 + \sqrt{B(t)}} \cdot L \quad (6.3)$$

$$\varepsilon(t) = \frac{S(t)}{S} \quad (6.4)$$

The efficiency is defined by Equation 6.4 where the dependence on the applied cuts is given in t (e.q. $S(t)$). Therefore, uncut quantities are given without t dependence (e.q. S).

6.4.2 Test for overfitting

Figures 27(a)-29(b) display the BDT training performance for the α mass spectrum [0.0, 0.5, 0.7, 1.0, 1.2, 1.4, 1.6] GeV/c^2 in the e^- and μ^- channels (left and right plots, respectively). The number of events (signal and background) are shown with their respective BDT training output, where an optimal outcome would be a peak at 0 for background and for signal at 1 with a steep slope in-between, as well as the training to test ratio, to represent potential overfitting. Here, S (orange) stands for signal, and B (blue) for background. This was trained with adjusted weights as stated above but plotted with sample weights for comparable outputs.

The first thing we notice is that electron- and muon channels behave almost the same for different mass points and that the BDT training performance is highly mass-dependent. Although the picture of an optimal BDT-training output is never completely satisfied, the training with high α masses, especially $m_\alpha = 1.4 \text{ GeV}/c^2$ (Fig. 28(e), 28(f)) and $m_\alpha = 1.6 \text{ GeV}/c^2$ (Fig. 29) get pretty close to optimal BDT performance. This is due to the introduction of the three additional training features *Momentum in Pseudo Rest Frame*, *Missing Momentum of Event* and *Visible Energy of Event* (Tab. 3). For lower masses the performance gets worse and peaks approximately at 0.6 (Fig. 27(c)-28(d)). There is one exception to this with a massless α particle (Fig. 27(a), 27(b)). In this case the BDT training performs a bit better compared to other low mass points. The train/test ratio is approximately the same as for high mass points which indicates a stable BDT performance. Though the BDT output misses a clear peak at the upper end for lower mass points as signal background separation becomes more efficient with high α masses, the overall BDT performance is still reasonable. The background peaks at 1 because of the similar topology of the main background decay $\tau^- \rightarrow \ell^- \bar{\nu}_\ell \nu_\tau$ to the signal

decay.

In the μ -channel a clear signal-background separation is harder to achieve which can also be seen when comparing the BDT output distributions (following subsection, Figure 27(a)-29) for the e - and μ -channels. This is due to the fact that the SM μ decay spectrum includes $\tau \rightarrow \mu\nu \rightarrow \pi\nu(n\pi^0)$ where $n = \{0, 1, 2, 3, \dots\}$. Such final decay products with exactly one charged particle give similar detector signatures compared to $\tau^\pm \rightarrow \ell^\pm \alpha$. A clear identification of charged pions is therefore not possible. Nevertheless, the Punzi FOM values for the muon channel are slightly better overall. This is caused by higher signal-to-background ratios after cutting. One can argue that the number of signal events in the muon channel exceeds the ones in the electron channel due to the already mentioned decay mode $\tau \rightarrow \mu\nu \rightarrow \pi\nu(n\pi^0)$. As it is unclear for the algorithm if this accounts for signal or background, this can lead to fake rates, slightly increasing the Punzi FOM.

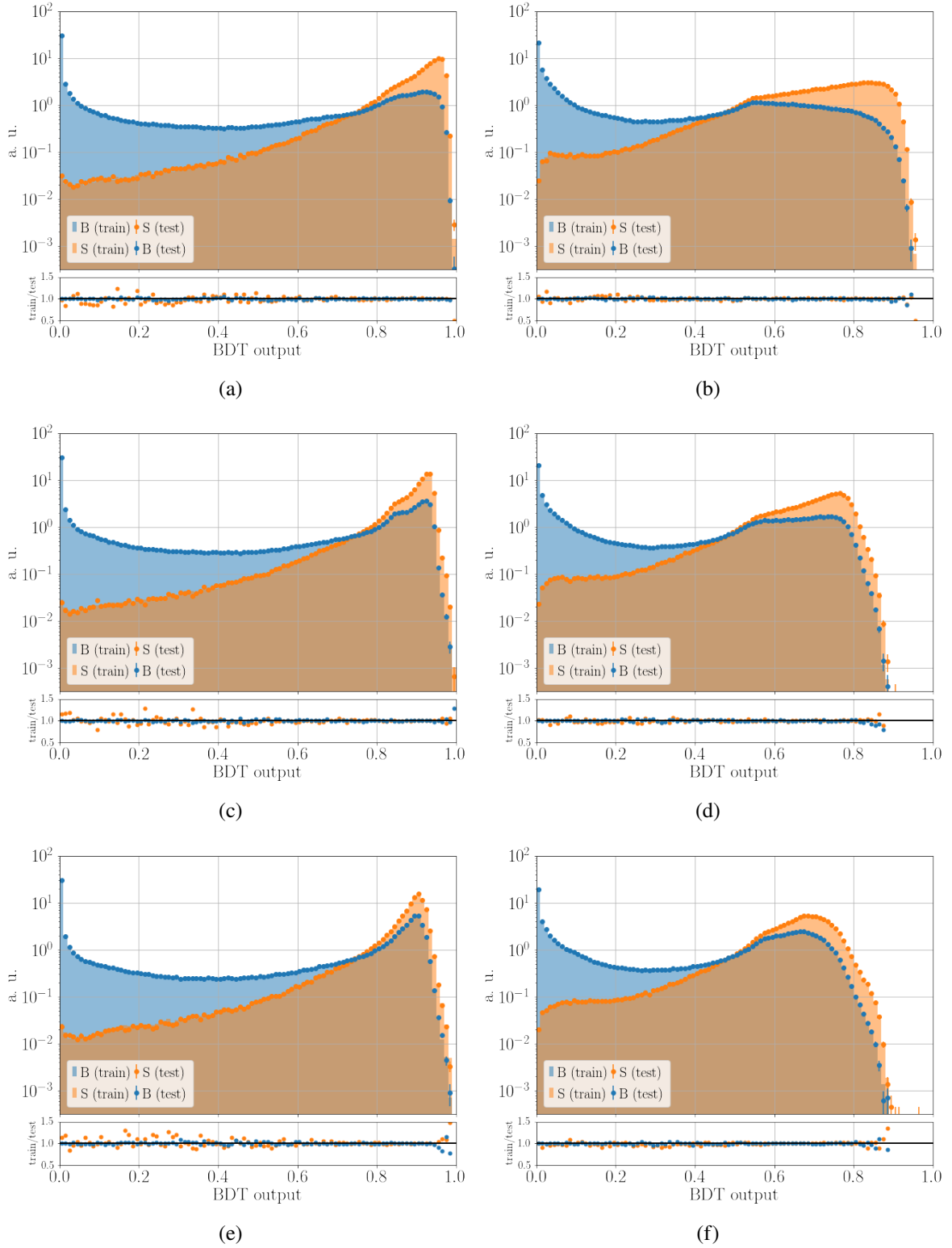


Figure 27: check for BDT overfitting with mass for the alpha particle set to $m_\alpha = [0.0, 0.5, 0.7]$ GeV/c² [top to bottom row]; electron-channel (left) and muon-channel (right).

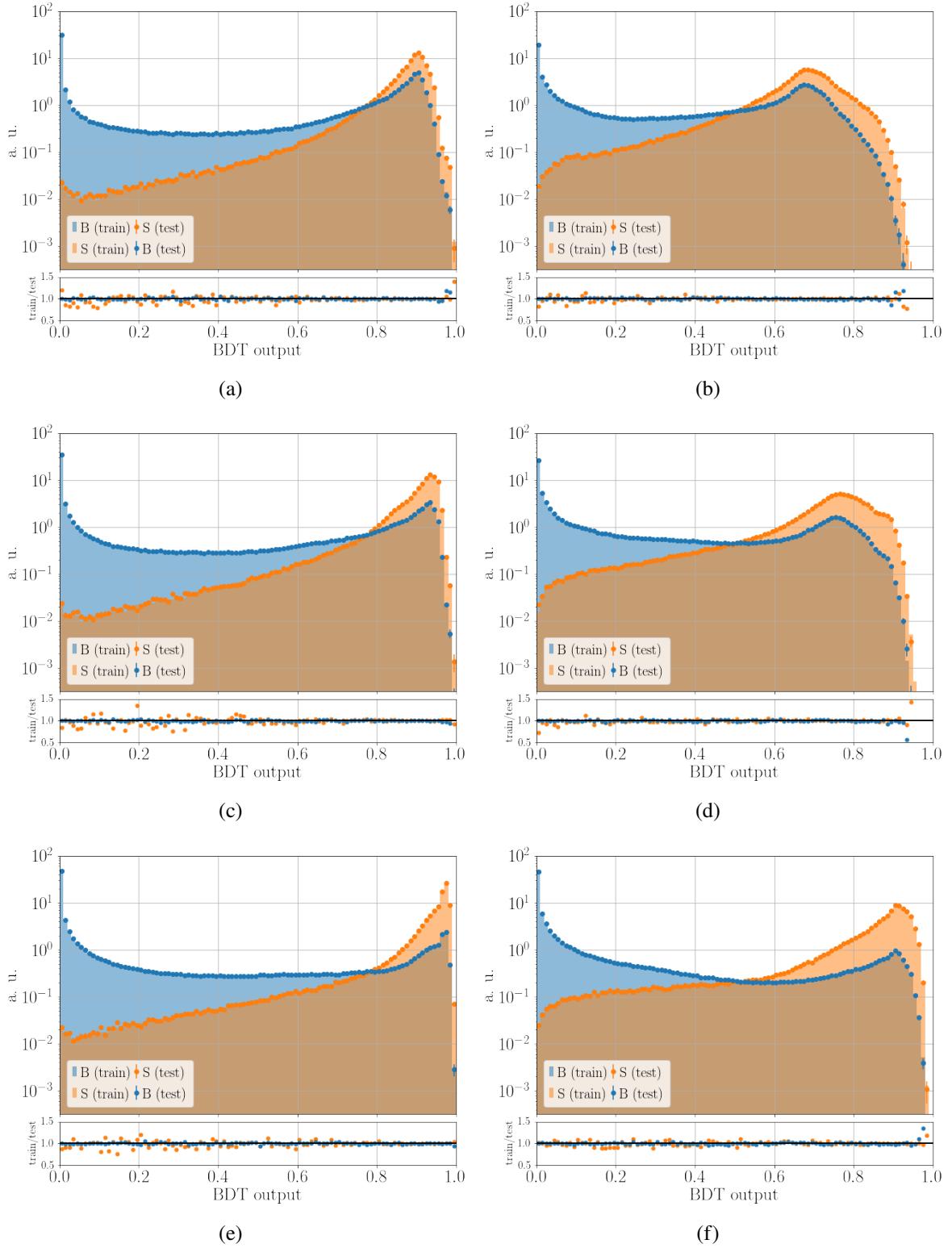


Figure 28: check for BDT overfitting with mass for the alpha particle set to $m_\alpha = [1.0, 1.2, 1.4]$ GeV/c² [top to bottom row]; electron-channel (left) and muon-channel (right).

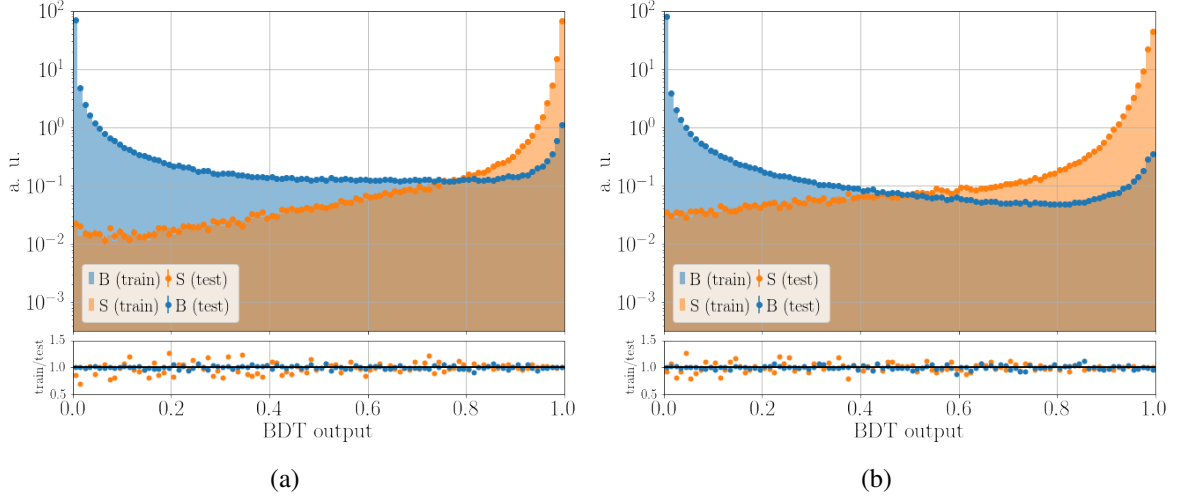


Figure 29: check for BDT overfitting with mass for the alpha particle set to $m_\alpha = 1.6 \text{ GeV}/c^2$; electron-channel (left) and muon-channel (right).

6.4.3 BDT cut optimization

We calculate the *Punzi FOM* values using Equation 6.3 which can give a clearer look at the performance of different analysis methods in the search for NP as the *Punzi FOM* is in-dependable of signal information. It's our measure of choice for signal-to-background separation. Adjusted weights were applied for the BDT training but the plots were made using the sample weights for a relatable comparison. For comparison with the standard training using sample weights, we plot their distributions of the *Punzi FOM* with respect to the corresponding BDT output values for different alpha masses in Appendix, Chapter 8, Figure 38-40, $m_\alpha = 0.0$ - $1.6 \text{ GeV}/c^2$.

The lower plots in Figure 30(a)-32(f) display the *Punzi FOM*-BDT output values for the current training with adjusted weights. In those multi-plot figures, the black vertical line marks the

max. *Punzi FOM* value to clearly indicate the BDT branch with the best performance. The overall performance in the μ -channel is worse than in the e -channel as expected since the separation of signal and background in μ decays yields more complex decay channels (Chapter 7). The BDT performances with adjusted weights stay for lower mass points at approximately 0.8 and 0.5 for e -channel and μ , respectively, and approaches BDT output = 1.0 with higher mass points which indicates a quite solid overall performance.

We see that the maximum *Punzi FOM* value (black vertical line) shifts to higher BDT-output values for higher mass points. This implies an improvement of BDT performance for higher alpha masses while for the mass points [0.0, 0.5, 0.7, 1.0] GeV/c^2 the performance stays constant at a BDT-output value of 0.3. We conclude that the adjusted weights BDT performs better overall which strengthens our decision. However, in the following subsection, we will see that the max. *Punzi FOM* is lower for the adjusted weights method, via direct comparison of the two BDT training methods by a few percent, our choice remains valid as discussed below.

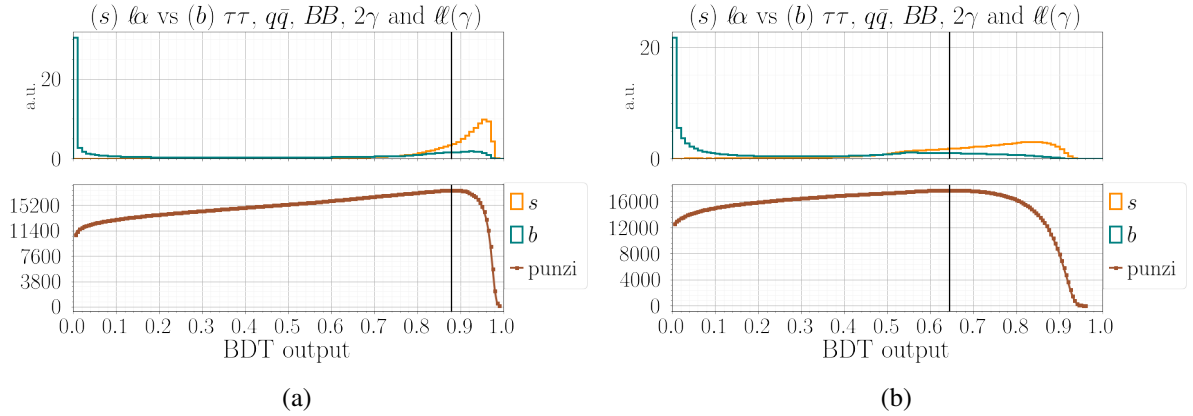


Figure 30: Figure of Merit (FOM) plotted for BDT output spectrum with alpha particle mass set to $m_\alpha = 0.0 \text{ GeV}/c^2$. The lower plot gives the distribution of *Punzi FOM* values corresponding to each BDT output branch. The plots are given in logarithmic scale. The black vertical line marks the max. *Punzi FOM*; electron-channel (left) and muon-channel (right).

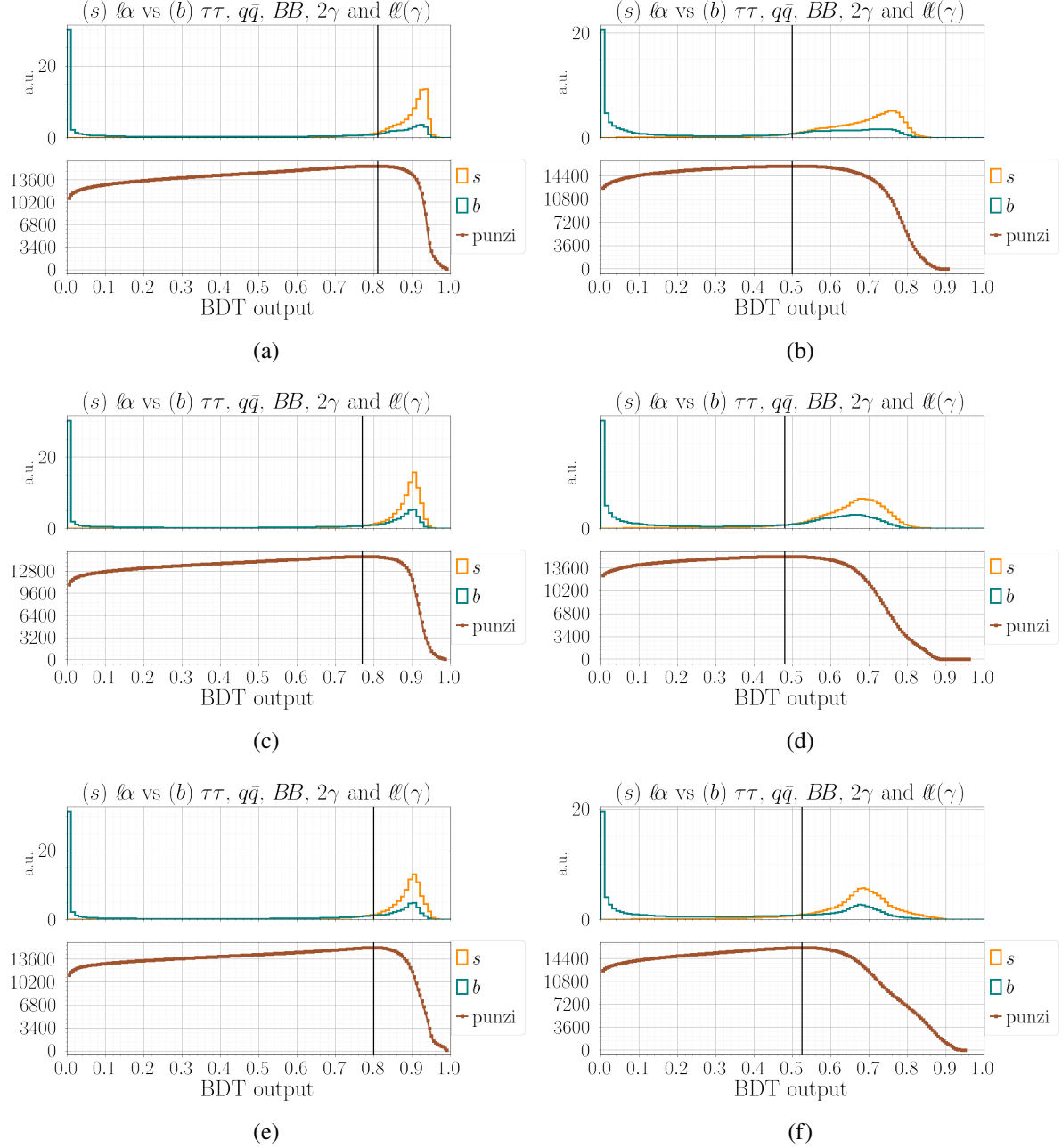


Figure 31: Figure of Merit (FOM) plotted for BDT output spectrum with alpha particle mass set to $m_\alpha = [0.5, 0.7, 1.0]$ GeV/c²; electron-channel (left) and muon-channel (right).

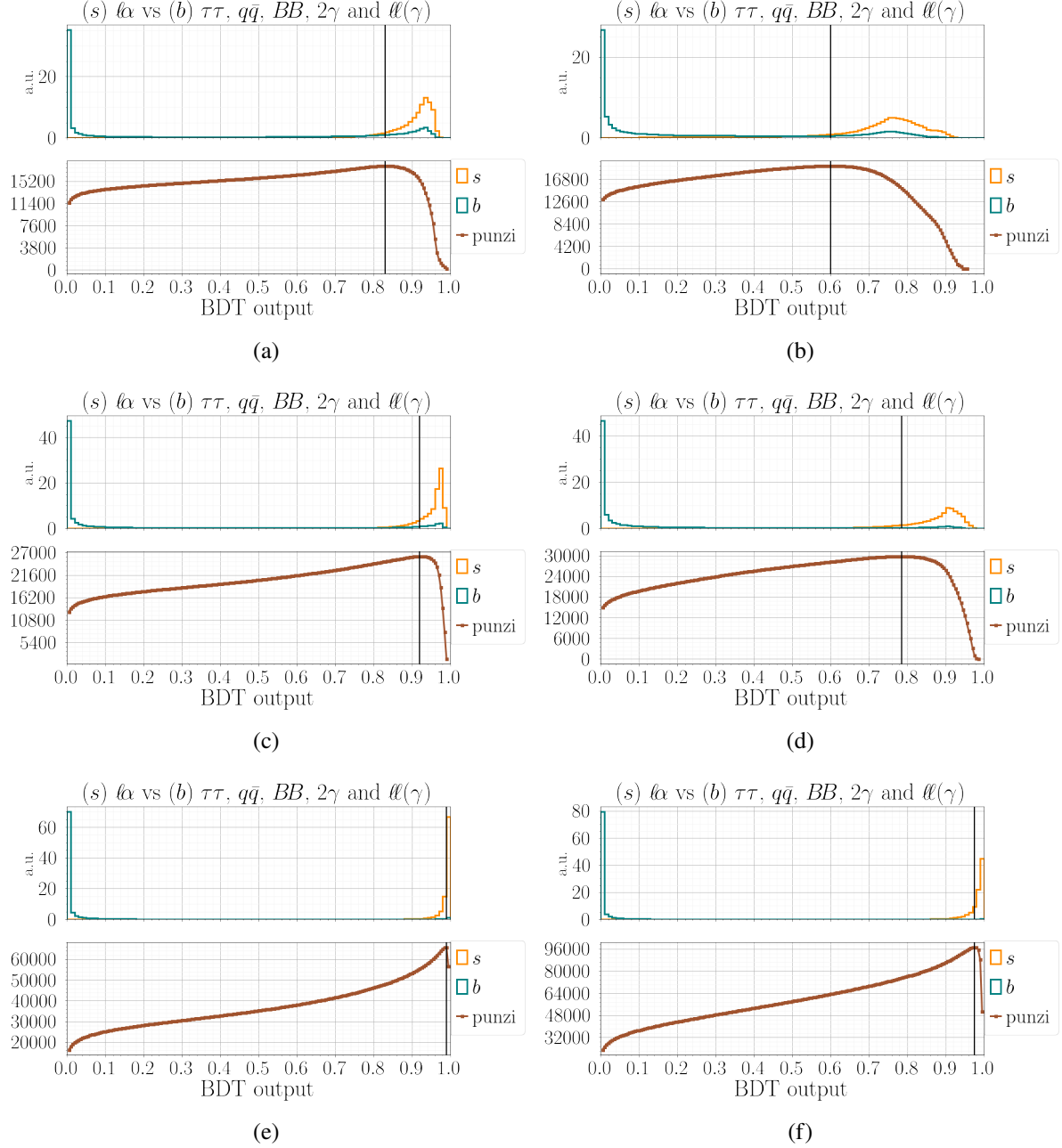


Figure 32: Figure of Merit (FOM) plotted for BDT output spectrum with alpha particle mass set to $m_\alpha = [1.2, 1.4, 1.6]$ GeV/c²; electron-channel (left) and muon-channel (right).

6.4.4 Performance

For a direct and qualitative comparison of the BDT training methods we use the Punzi Figure of Merit (*Punzi FOM*, Equation 6.3) and we compare this quantity to the corresponding efficiency. The plots are shown for the α mass points [0.0, 0.5, 0.7, 1.0, 1.2, 1.4, 1.6] GeV/c² in Figure 33(a)-35(b) for electron- (left) and muon-channel (right). The exact values are given in Table 7 and 8.

Especially, high mass points show very few data points in the low-efficiency region which can explain the odd shape of the adjusted weight Punzi spectrum in Figure 35(a) and 35(b) where the lower end is cut off. Using BDT training leads to a higher *Punzi FOM* for the same efficiency compared to the cut-based approach for every mass point and therefore a overall higher performance. This is true for adjusted weights and sample weights method alike. Therefore, our study seems successful since the goal is to determine a possible improvement in the Belle II analysis compared to the previously used cut-based method for signal-to-background reduction. Now, we have to compare the different BDT training methods. The slightly lower maximum values of the adjusted weight training would imply a worse performance compared to the standard approach using the given sample weights while BDT training. Since the maximum values show only a tiny discrepancy we decide to use adjusted weights for the training which offers a greater training stability (higher BDT output, Figure 27(a)-29(b)).

This is summarized in Figure 36 where we define *Punzi FOM* of the BDT relative to the cut-based *Punzi FOM* for a quantitative comparison. As one can easily see that there exists a rising potential in this method with respect to rising alpha masses. This concludes a tremendous improvement in signal-background separation for the Belle II τ to low multiplicity search and approves our study of BDT training.

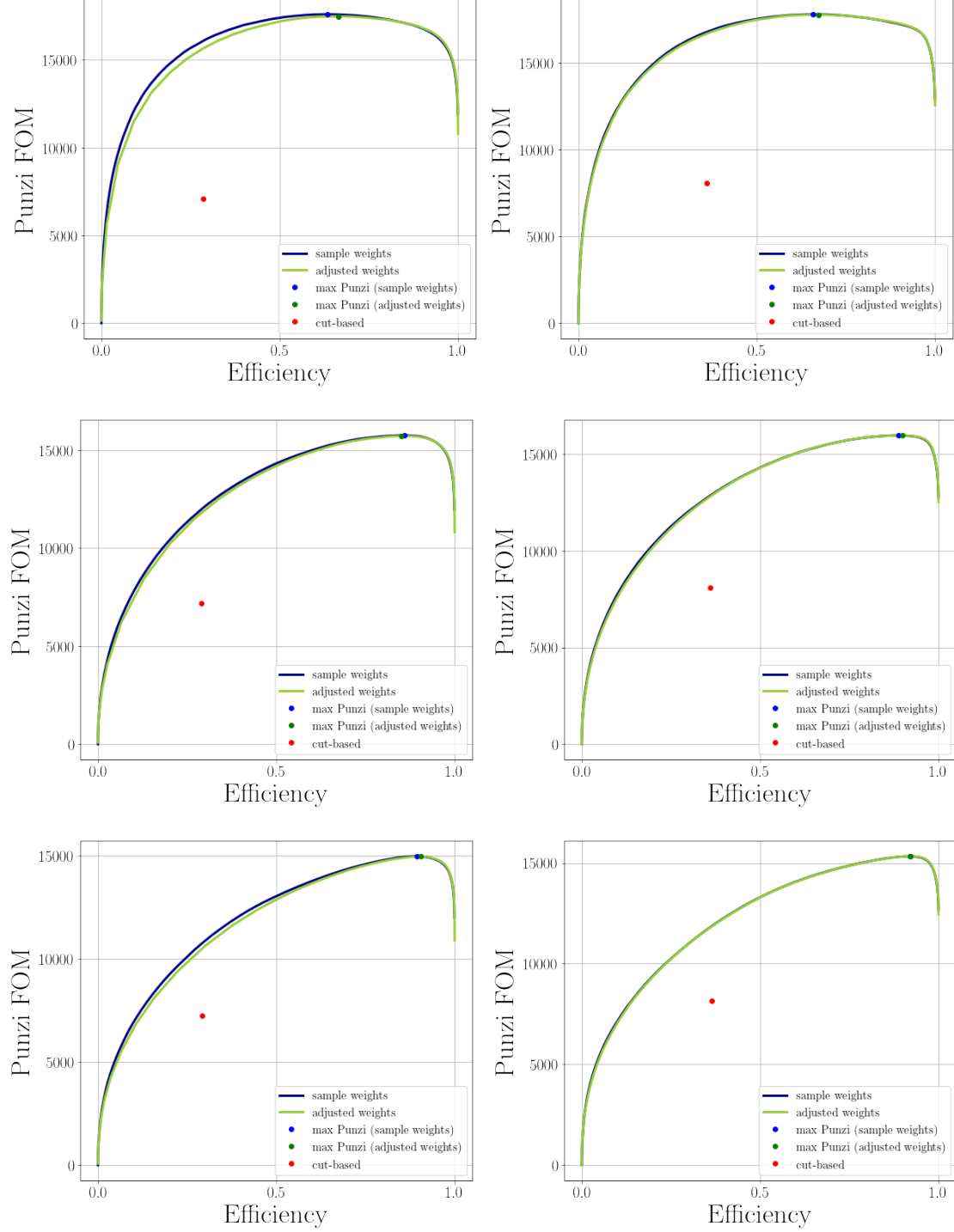


Figure 33: PUNZI plot for the (BDT-training) efficiency spectrum with mass for the alpha particle set to $m_\alpha = [0.0, 0.5, 0.7]$ GeV/c²; The blue line gives the adjusted weights and the green line the sample weights training with dots in the corresponding colors indicating each maximum *Punzi FOM* values. The efficiency and *Punzi FOM* value of the cut-based approach is marked by the red dot.

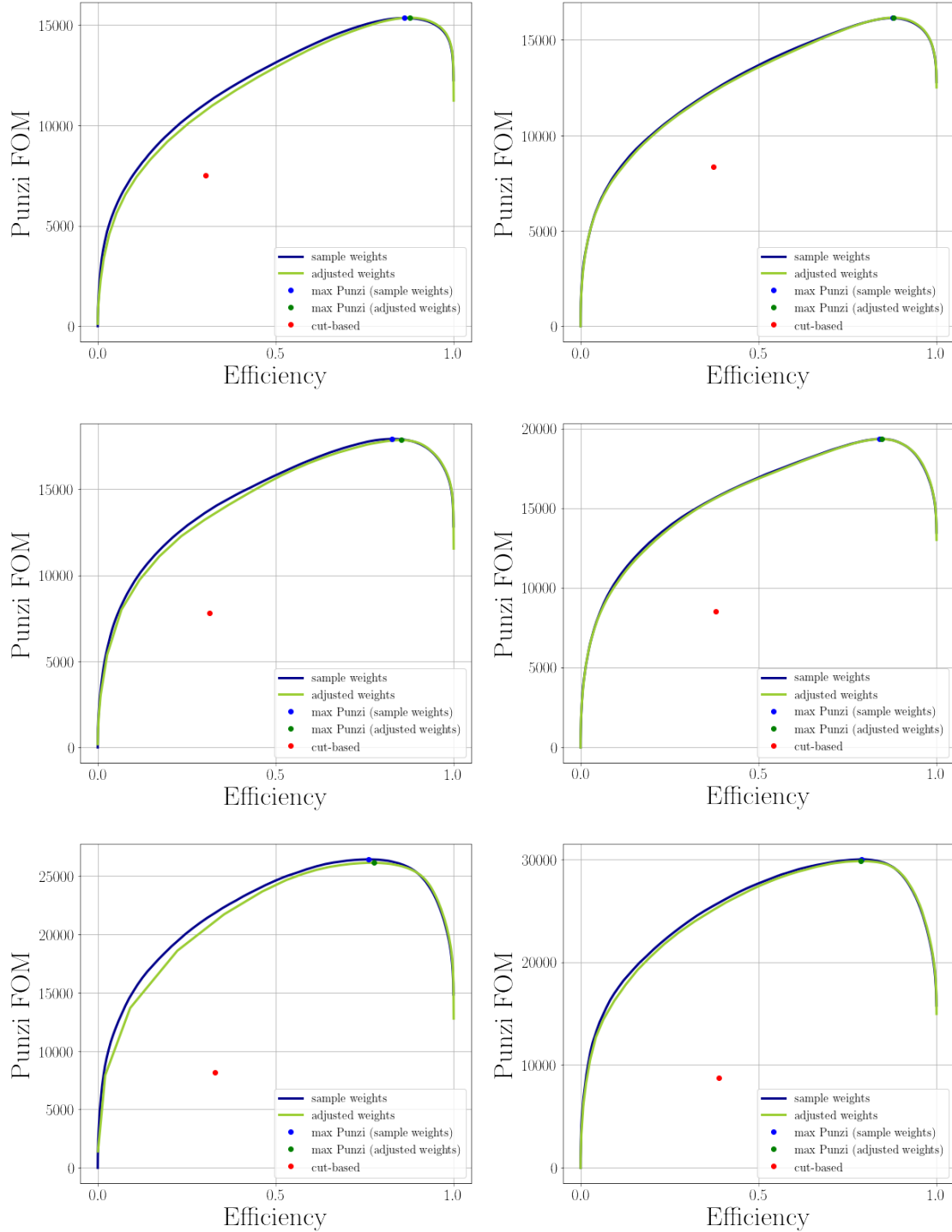


Figure 34: PUNZI plot for the (BDT-training) efficiency spectrum with mass for the alpha particle set to $m_\alpha = [1.0, 1.2, 1.4] \text{ GeV}/c^2$; The blue line gives the adjusted weights and the green line the sample weights training with dots in the corresponding colors indicating each maximum *Punzi FOM* values. The efficiency and *Punzi FOM* value of the cut-based approach is marked by the red dot.

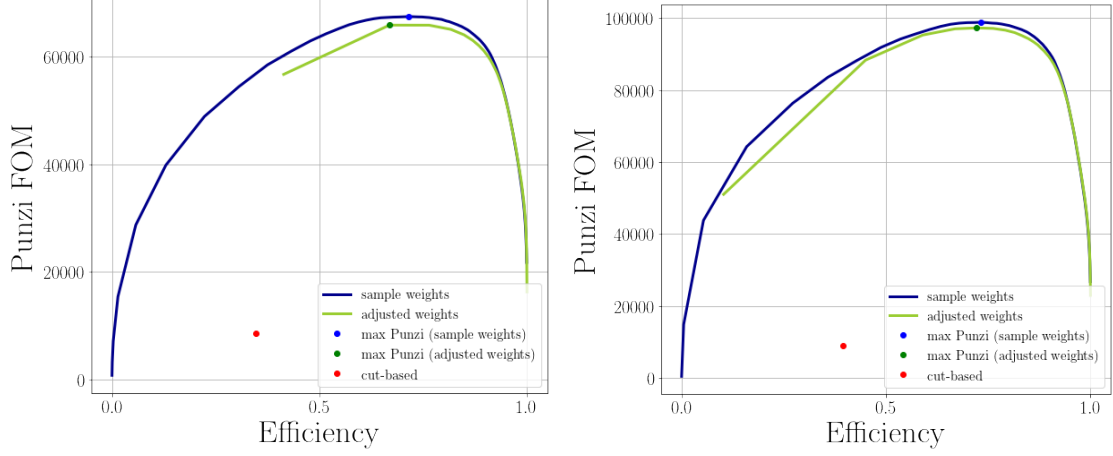


Figure 35: PUNZI plot for the (BDT-training) efficiency spectrum with mass for the alpha particle set to $m_\alpha = 1.6 \text{ GeV}/c^2$; The blue line gives the adjusted weights and the green line the sample weights training with dots in the corresponding colors indicating each maximum *Punzi FOM* values. The efficiency and *Punzi FOM* value of the cut-based approach is marked by the red dot.

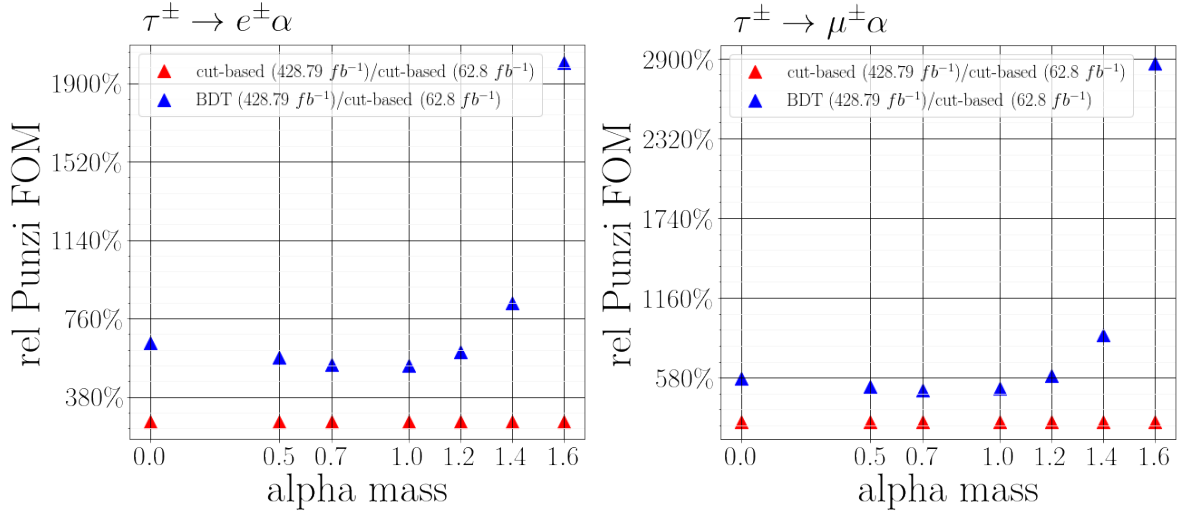


Figure 36: Relative Punzi FOM values for all alpha masses. Direct comparison of Punzi FOM values relative to the recent published cut-based analysis of *Belle II* at 62.8 fb^{-1} (red) relative Punzi FOM of cut-based analysis with the current luminosity of *Belle II* with 428.79 fb^{-1} (blue) rel. Punzi FOM of BDT training with current luminosity.

alpha mass	analyzing procedure	Efficiency	Purity	Punzi FOM
0.0 GeV	Cut-based	0.286	0.385	7100
	BDT (sample lum.)	0.914	0.518	16658
	BDT (adj. weights)	0.983	0.451	15094
0.5 GeV	Cut-based	0.289	0.386	7174
	BDT (sample lum.)	0.964	0.457	15154
	BDT (adj. weights)	0.988	0.425	14381
0.7 GeV	Cut-based	0.292	0.387	7253
	BDT (sample lum.)	0.974	0.431	14485
	BDT (adj. weights)	0.99	0.408	13933
1.0 GeV	Cut-based	0.302	0.391	7504
	BDT (sample lum.)	0.969	0.434	14697
	BDT (adj. weights)	0.991	0.406	14043
1.2 GeV	Cut-based	0.314	0.394	7809
	BDT (sample lum.)	0.954	0.498	16810
	BDT (adj. weights)	0.989	0.447	15427
1.4 GeV	Cut-based	0.329	0.39	8172
	BDT (sample lum.)	0.928	0.665	24187
	BDT (adj. weights)	0.984	0.564	20088
1.6 GeV	Cut-based	0.346	0.358	8600
	BDT (sample lum.)	0.941	0.891	54434
	BDT (adj. weights)	0.989	0.767	35300

Table 7: e -channel: Summary table including efficiency, purity and Punzi FOM values for the given alpha mass spectrum using cut-based approach, BDT training with sample weights or BDT training with adjusted weighting.

alpha mass	analyzing procedure	Efficiency	Purity	Punzi FOM
0.0 GeV	Cut-based	0.359	0.384	8059
	BDT (sample lum.)	0.941	0.513	16961
	BDT (adj. weights)	0.965	0.498	16638
0.5 GeV	Cut-based	0.36	0.383	8082
	BDT (sample lum.)	0.973	0.46	15560
	BDT (adj. weights)	0.979	0.454	15426
0.7 GeV	Cut-based	0.364	0.384	8175
	BDT (sample lum.)	0.978	0.436	14925
	BDT (adj. weights)	0.982	0.432	14829
1.0 GeV	Cut-based	0.372	0.383	8356
	BDT (sample lum.)	0.969	0.446	15376
	BDT (adj. weights)	0.978	0.437	15166
1.2 GeV	Cut-based	0.379	0.378	8513
	BDT (sample lum.)	0.95	0.523	18104
	BDT (adj. weights)	0.965	0.505	17596
1.4 GeV	Cut-based	0.388	0.362	8722
	BDT (sample lum.)	0.928	0.698	27182
	BDT (adj. weights)	0.953	0.665	25520
1.6 GeV	Cut-based	0.395	0.288	8864
	BDT (sample lum.)	0.949	0.921	73846
	BDT (adj. weights)	0.975	0.874	57776

Table 8: μ -channel: Summary table including efficiency, purity and Punzi FOM values for the given alpha mass spectrum using cut-based approach, BDT training with sample weights or BDT training with adjusted weighting.

7 Conclusion

In search for Physics beyond the SM low-multiplicity decays are a special challenge. The low number of potential signal events leads to very high statistics which makes the discovery of new particles, conventionally at the threshold of 5σ , a data-expensive task. Using rectangular cuts, such as the currents analysis at *Belle II* in the search of $\tau^\pm \rightarrow \ell^\pm \alpha$ where α is an invisible NP boson, can diminish the amount of background events and improve its statistics. We propose a new approach to background reduction using ML algorithms with rectangular cuts.

The *Extreme Gradient Boosted Decision Tree* method (BDT) is an improved version of *Boosted Decision Trees* and can provide suitable thresholds for the cut-based approach. With a carefully chosen set of training variables (features) this can push statistics to the next level and narrow down the potential parameter space of NP particles. Furthermore, we use the known approach of adjusted (equal) weighting, such that the numbers of signal and background events are the same. This leads to quite an improvement in stability but slightly lowers statistics of the BDT training and makes the use of BDT training more reliable for overall analysis.

There proves to be an overall improvement of signal-background separation in the *Belle II* $\tau^\pm \rightarrow \ell^\pm \alpha$ search when utilizing *Extreme Gradient Boosted Decision Tree* training instead of fixed rectangular cuts (as shown in Figure 36). When comparing the relative BDT Punzi FOM values as a qualitative measure of performance independent of signal information the relative improvement falls between factor 5.34 and 19.98 in the electron channel and between factor 4.90 and 28.64 in the muon channel. This is a tremendous improvement and exceeds the relative

improvement of solely raising luminosity which lays at factor 2.61 for both, e^- and μ^- -channel. We easily double the relative performance with even our worst mass point. The relative uncertainty is diminishing small and therefore negligible [32].

The choice of using adjusted weighting while training is justified by stability improvements in exchange for slightly worse performance. This is reflected by BDT output values which peak close to 1 and the higher maximum Punzi Figure of Merit, respectively. Compared to BDT training with standard 'sample weights' which shift the peak to lower output values our approach benefits from a higher performance stability and reliability in signal-background separation. We encourage to check the results for BDT training using the given sample rates for a direct comparison (Appendix, Chapter 8, Figure 38-40).

The presented results directly compare the current *Belle II* analysis with a possible extension using *Extreme Gradient Boost Decision Trees* and indicate an improvement in BDT performance. This will increase the statistics of *Belle II* and help lower upper boundaries in the search for NP particles. Machine Learning Algorithms are powerful tools when used carefully and the discussed results open up new possibilities, especially in the search for low multiplicity decays. Previously analyzed data sets of accelerator experiments, such as *Belle* [11] and *ARGUS* [28], can be re-evaluated using BDT analysis which will most likely improve their accuracy and lower their statistics.

8 Appendix

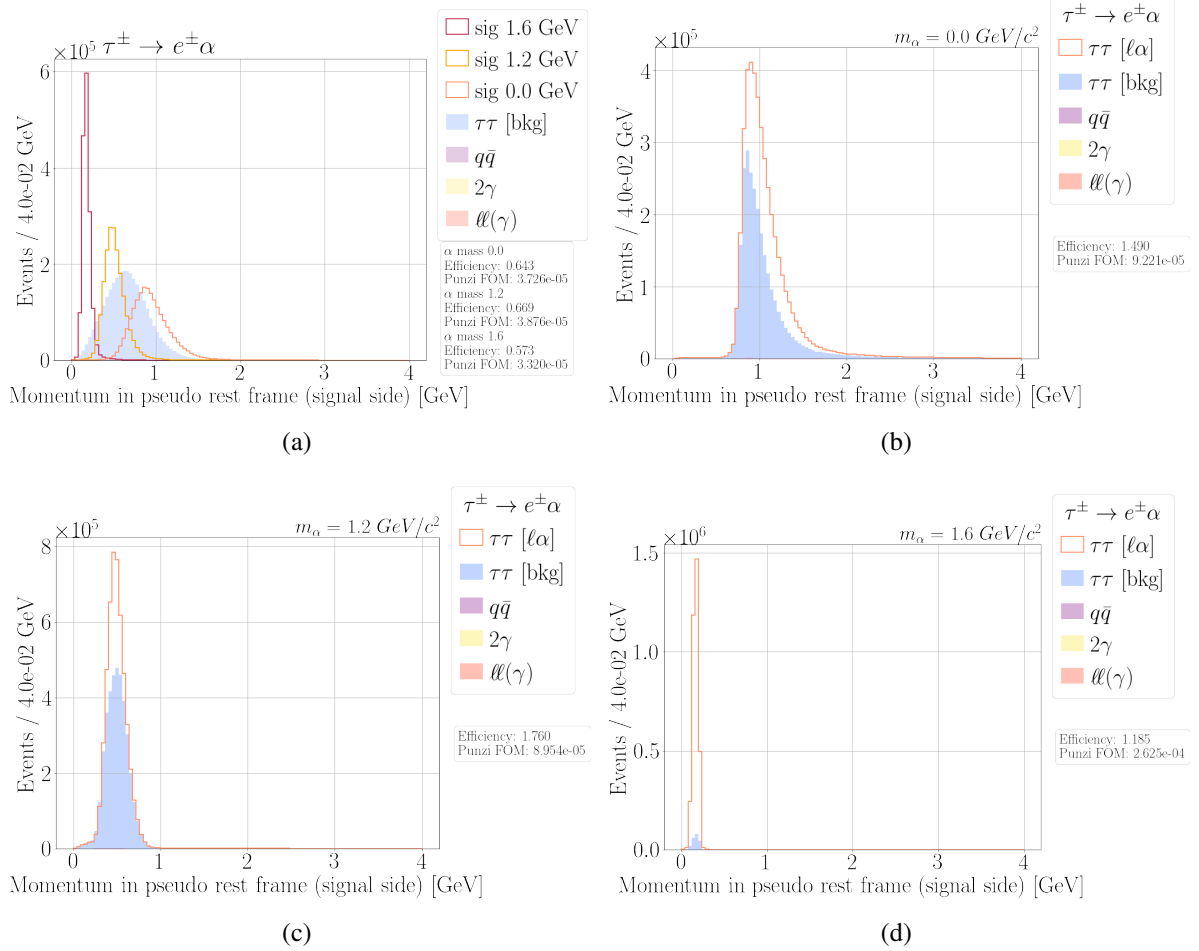


Figure 37: Feature distribution of 'momentum in pseudo rest frame', electron channel, including the cut-based analysis results (a) and BDT training results (b), (c) and (d) for the alpha masses $m_\alpha = [0.0, 1.2, 1.6] \text{ GeV}/c^2$. Distribution of MC data weighted according to sample luminosity given in Table 2. These plots show the signal and background distributions of *Momentum in pseudo rest frame (signal side)* in the electron channel with respect to cut-based and BDT analysis to give a qualitative comparison with the current analysis of *Belle II* using their official plot of the same variable given in Figure 13.

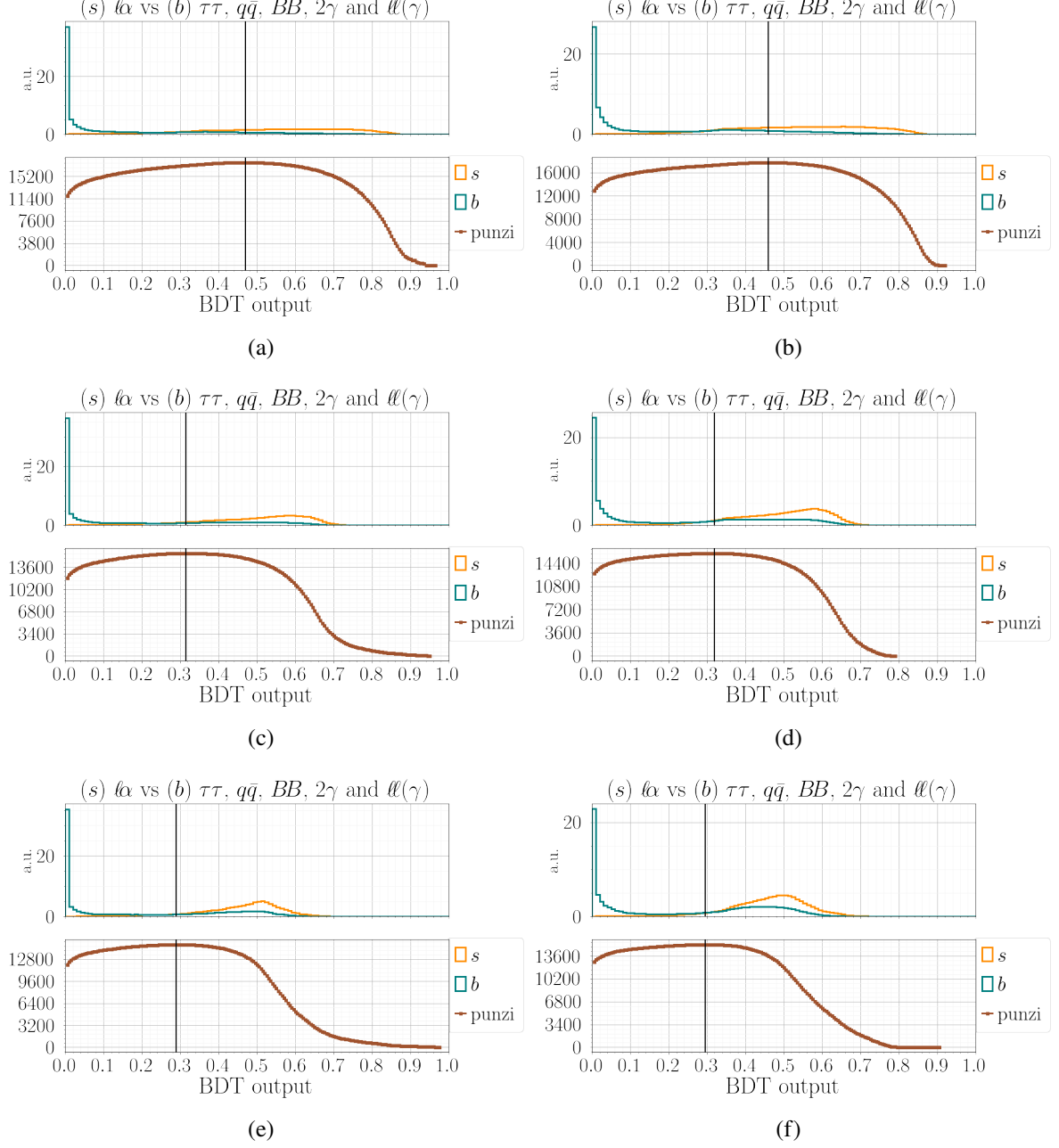


Figure 38: Figure of Merit (FOM) plotted for BDT output spectrum with alpha particle mass set to $m_\alpha = [0.0, 0.5, 0.7] \text{ GeV}/c^2$; electron-channel (left) and muon-channel (right).

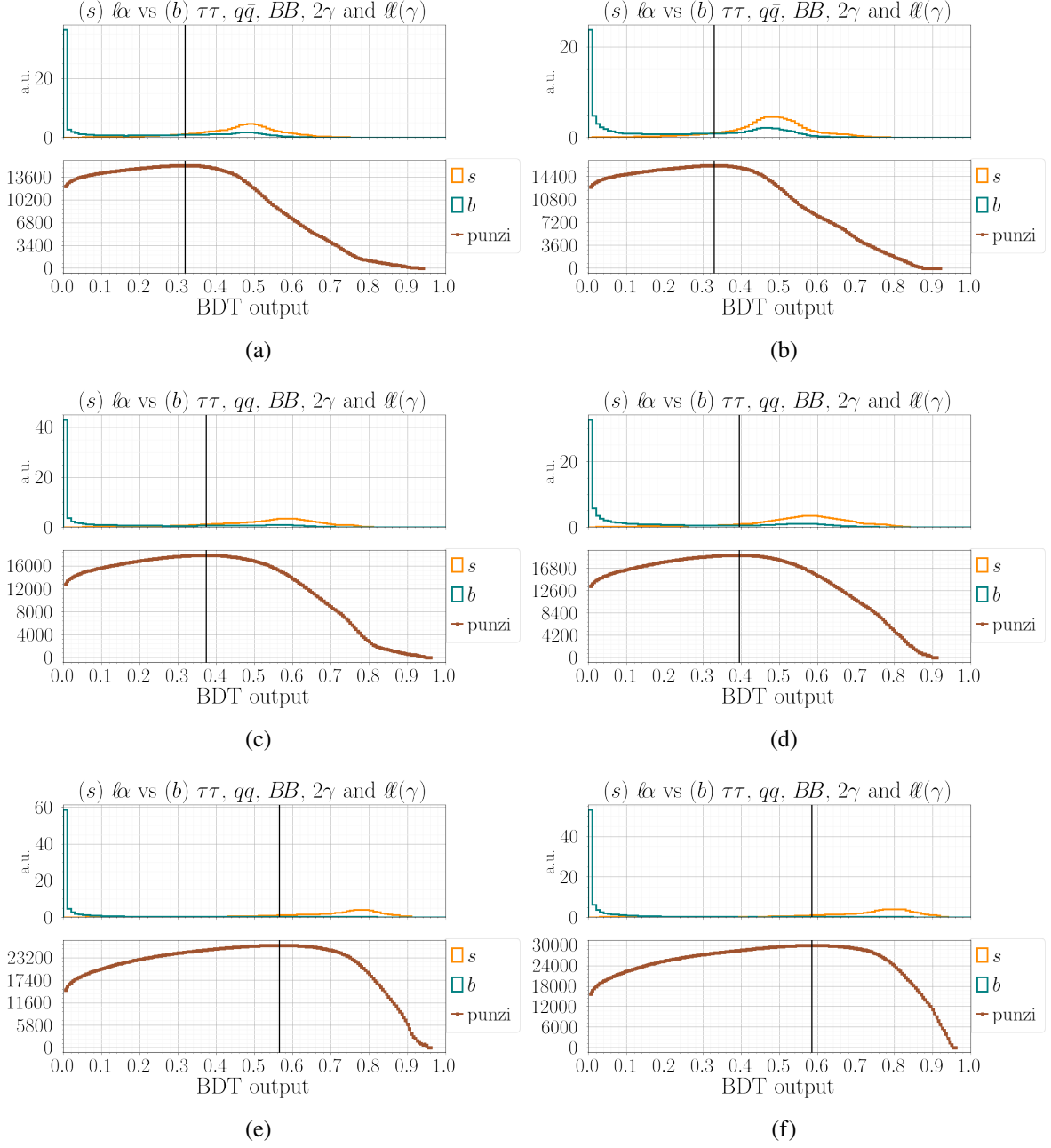


Figure 39: Figure of Merit (FOM) plotted for BDT output spectrum with alpha particle mass set to $m_\alpha = [1.0, 1.2, 1.4] \text{ GeV}/c^2$; electron-channel (left) and muon-channel (right).

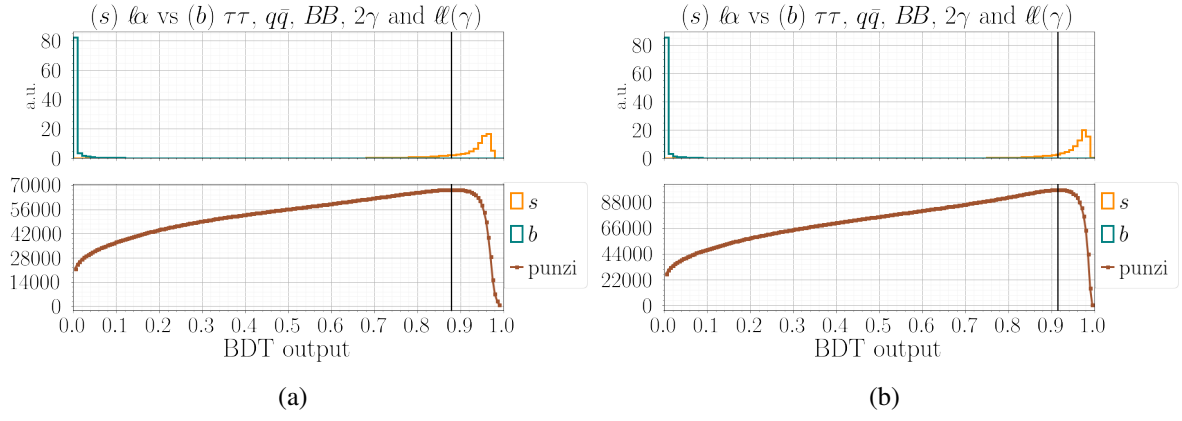


Figure 40: Figure of Merit (FOM) plotted for BDT output spectrum with alpha particle mass set to $m_\alpha = 1.6 \text{ GeV}/c^2$; electron-channel (left) and muon-channel (right).

List of Figures

1	Particle content of the Standard Model of Particle Physics. From [17].	9
2	Most common tau lepton decay modes with corresponding branching ratio $\mathcal{B}$. From [18].	10
3	(a) leptonic τ decay. (b) hadronic τ decay.	10
4	Current 90% CL upper limits for the branching fraction of τ LFV decays obtained in the <i>CLEO</i> , <i>BaBar</i> , and <i>Belle</i> experiments. The purple boxes, blue inverted triangles, green triangles, and yellow boxes show <i>CLEO</i> , <i>BaBar</i> , <i>Belle</i> , and <i>LHCb</i> results, respectively, while the red circles express the <i>Belle II</i> future prospects, where they are extrapolated from Belle results assuming the integrated luminosity of 50 ab^{-1} . From [19].	13
5	Boundaries on the α mass set by various experiments. (a) Experimental upper limits of mass and decay constant for a light ALP. From [27]. (b) Experimental upper limits and exclusion areas of the $m_{Z'}$ vs. g'_R parameter space for Z' . The left and right coupling constants represent the particles coupling to left and right chirality, respectively, and are set to $g'_L = g'_R/10$. The green band is the preferred range by the $(g_\mu - 2)$ anomaly at 2σ . The red region is forbidden by lepton flavor universality in τ decays at 2σ . The blue region is excluded by ARGUS' search for the NP decay $\tau \rightarrow \mu Z'$ at 95% CL. The black dashed line indicates the exclusion for the red region given by the LHC from their searches for leptonic W decays at 95% CL. The purple dashed curve is the exclusion given by LEP Z coupling measurements at 95% CL. The blue and orange dotted lines indicate limits in parameter space at 3σ in searches for the $\mu^\pm \mu^\pm \tau^\pm \tau^\pm$ final state at the high-luminosity LHC and at a future electron- positron collider running at the Z pole. The grey region is excluded at $> 5\sigma$. From [33].	15

6	The SuperKEKB accelerator with all its sub-components incl. the <i>Belle II</i> detector system. From [38].	18
7	Belle II detectors. From [19].	19
8	Current and future luminosity of <i>Belle II</i> since data taking; (a) Development and current luminosity state. (b) Future projection of luminosity with peak luminosity (red) and integrated luminosity (blue). Second Long Shutdown (LS2) between 2027 and 2028 for machine improvements possible. First Long Shutdown (LS1) and LS2 are recognizable by the steep of peak luminosity curve at 2022 and 2028. From [41].	22
9	tau pair production at e^+e^- colliders	23
10	(Left) high thrust event, e.g. τ pair. (Right) low thrust event, e.g. BB. The thrust axis is represented by the red and blue arrows. From [19].	25
11	3x1-topology of τ pair decay with thrust axis indicating signal- and tag-side. From [18]	26
12	<i>pseudo rest frame</i> , separates tag side $\tau^- \rightarrow \pi^- \pi^+ \pi^- \nu_\tau$ and signal side $\tau^+ \rightarrow \ell^+ \alpha$, where $\ell = e/\mu$, to estimate the leptons' momentum p_ℓ ; the dotted line indicates the thrust axis. From [42]	26
13	pseudo-rest frame τ momentum spectrum with signal $\tau^\pm \rightarrow \ell^\pm \alpha$ peak for α mass $m_\alpha = [0, 14] \text{ GeV}/c^2$; the signals' $\mathcal{BR}$ was scaled to a constant fraction 0.1; e -channel (left) and μ -channel (right). From [32].	27
14	feature distribution of MC data weighted according to sample luminosity given in Table 2 and signal contributions for the alpha masses $m_\alpha = [0.0, 1.2, 1.6] \text{ GeV}/c^2$	31
15	feature distribution of MC data weighted according to sample luminosity given in Table 2 and signal contributions for the alpha masses $m_\alpha = [0.0, 1.2, 1.6] \text{ GeV}/c^2$	32

16	feature distribution of MC data weighted according to sample luminosity given in Table 2 and signal contributions for the alpha masses $m_\alpha = [0.0, 1.2, 1.6]$ GeV/c ² ; electron-channel (left) and muon-channel (right)	33
17	feature distribution of MC data weighted according to sample luminosity given in Table 2 and signal contributions for the alpha masses $m_\alpha = [0.0, 1.2, 1.6]$ GeV/c ² ; electron-channel (left) and muon-channel (right)	34
18	Spectra of x_ℓ (Eq. 4.2) for e -channel (left) and μ -channel (right) in experimental and MC data; SM background apart from main background $\tau \rightarrow \ell \bar{\nu} \nu$ is combined into "Other"; uncertainty (grey area) is dominated by lepton-identification efficiency; assumed $\mathcal{BR}$ fraction of $\tau^- \rightarrow \ell^- \alpha$ set to 5%. From [32].	39
19	Recent Belle II upper limits within 95% CL for the branching fraction of $\mathcal{B}(\tau^- \rightarrow \ell^- \alpha)/\mathcal{B}(\tau^- \rightarrow \ell^- \bar{\nu}_\ell - \nu_\tau)$ and expectation values from background-only hypothesis for α mass: [0 - 1.6] GeV in e-channel (left) and μ -channel (right). From [32].	39
20	Decision tree scheme to illustrate the method of combing trees by looking at the leaves. We calculate their corresponding prediction values and add them together. From [63].	41
21	added training features: feature distribution of MC data weighted according to sample luminosity given in Table 2 and signal contributions for the alpha masses $m_\alpha = [0.0, 1.2, 1.6]$ GeV/c ² ; electron-channel (left) and muon-channel (right).	45
22	Correlation matrix of BDT training features with signal (upper row) and background (lower row) events; electron-channel (left) and muon-channel (right); $m_\alpha = 0.0$ GeV/c ² and $m_\alpha = 1.2$ GeV/c ² in first and second two rows, respectively.	48
23	Correlation matrix of BDT training features for $m_\alpha = 1.6$ GeV/c ² signal (upper row) and background (lower row) events; electron-channel (left) and muon-channel (right).	49

24	ranking the BDT features according to their significance in BDT training for alpha mass $m_\alpha = [0.0, 0.5, 0.7]$ GeV/c ²	51
25	ranking the BDT features according to their significance in BDT training for alpha mass $m_\alpha = [1.0, 1.2, 1.4, 1.6]$ GeV/c ²	52
26	Direct comparison of BDT training methods (log-scale), regular method (left) and adjusted weighting (right) for $m_\alpha = 1.6$ GeV/c ² . The train/test ratios are given in the plots below.	54
27	check for BDT overfitting with mass for the alpha particle set to $m_\alpha = [0.0, 0.5, 0.7]$ GeV/c ² [top to bottom row]; electron-channel (left) and muon-channel (right).	57
28	check for BDT overfitting with mass for the alpha particle set to $m_\alpha = [1.0, 1.2, 1.4]$ GeV/c ² [top to bottom row]; electron-channel (left) and muon-channel (right).	58
29	check for BDT overfitting with mass for the alpha particle set to $m_\alpha = 1.6$ GeV/c ² ; electron-channel (left) and muon-channel (right).	59
30	Figure of Merit (FOM) plotted for BDT output spectrum with alpha particle mass set to $m_\alpha = 0.0$ GeV/c ² . The lower plot gives the distribution of <i>Punzi FOM</i> values corresponding to each BDT output branch. The plots are given in logarithmic scale. The black vertical line marks the max. <i>Punzi FOM</i> ; electron-channel (left) and muon-channel (right).	60
31	Figure of Merit (FOM) plotted for BDT output spectrum with alpha particle mass set to $m_\alpha = [0.5, 0.7, 1.0]$ GeV/c ² ; electron-channel (left) and muon-channel (right).	61
32	Figure of Merit (FOM) plotted for BDT output spectrum with alpha particle mass set to $m_\alpha = [1.2, 1.4, 1.6]$ GeV/c ² ; electron-channel (left) and muon-channel (right).	62

- 33 PUNZI plot for the (BDT-training) efficiency spectrum with mass for the alpha particle set to $m_\alpha = [0.0, 0.5, 0.7]$ GeV/c²; The blue line gives the adjusted weights and the green line the sample weights training with dots in the corresponding colors indicating each maximum *Punzi FOM* values. The efficiency and *Punzi FOM* value of the cut-based approach is marked by the red dot. 64
- 34 PUNZI plot for the (BDT-training) efficiency spectrum with mass for the alpha particle set to $m_\alpha = [1.0, 1.2, 1.4]$ GeV/c²; The blue line gives the adjusted weights and the green line the sample weights training with dots in the corresponding colors indicating each maximum *Punzi FOM* values. The efficiency and *Punzi FOM* value of the cut-based approach is marked by the red dot. 65
- 35 PUNZI plot for the (BDT-training) efficiency spectrum with mass for the alpha particle set to $m_\alpha = 1.6$ GeV/c²; The blue line gives the adjusted weights and the green line the sample weights training with dots in the corresponding colors indicating each maximum *Punzi FOM* values. The efficiency and *Punzi FOM* value of the cut-based approach is marked by the red dot. 66
- 36 Relative Punzi FOM values for all alpha masses. Direct comparison of Punzi FOM values relative to the recent published cut-based analysis of *Belle II* at 62.8 fb⁻¹ (red) relative Punzi FOM of cut-based analysis with the current luminosity of *Belle II* with 428.79 fb⁻¹ (blue) rel. Punzi FOM of BDT training with current luminosity. 66

- 37 Feature distribution of 'momentum in pseudo rest frame', electron channel, including the cut-based analysis results (a) and BDT training results (b), (c) and (d) for the alpha masses $m_\alpha = [0.0, 1.2, 1.6] \text{ GeV}/c^2$. Distribution of MC data weighted according to sample luminosity given in Table 2. These plots show the signal and background distributions of *Momentum in pseudo rest frame (signal side)* in the electron channel with respect to cut-based and BDT analysis to give a qualitative comparison with the current analysis of *Belle II* using their official plot of the same variable given in Figure 13. 71
- 38 Figure of Merit (FOM) plotted for BDT output spectrum with alpha particle mass set to $m_\alpha = [0.0, 0.5, 0.7] \text{ GeV}/c^2$; electron-channel (left) and muon-channel (right). 72
- 39 Figure of Merit (FOM) plotted for BDT output spectrum with alpha particle mass set to $m_\alpha = [1.0, 1.2, 1.4] \text{ GeV}/c^2$; electron-channel (left) and muon-channel (right). 73
- 40 Figure of Merit (FOM) plotted for BDT output spectrum with alpha particle mass set to $m_\alpha = 1.6 \text{ GeV}/c^2$; electron-channel (left) and muon-channel (right). 74

References

- [1] M. Tanabashi *et al.*, “Review of Particle Physics 2018,” *Physical Review D*, vol. 98, no. 3, p. 030001, Aug. 2018. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevD.98.030001>
- [2] A. Keshavarzi, K. S. Khaw, and T. Yoshioka, “Muon $g-2$: A review,” *Nuclear Physics B*, vol. 975, p. 115675, Feb. 2022, arXiv:2106.06723 [hep-ex, physics:hep-ph]. [Online]. Available: <http://arxiv.org/abs/2106.06723>
- [3] J. W. Cronin, “CP Symmetry Violation: The Search for Its Origin,” *Science*, vol. 212, no. 4500, pp. 1221–1228, Jun. 1981. [Online]. Available: <https://www.science.org/doi/10.1126/science.212.4500.1221>
- [4] *CMS Collaboration*, *CMS-PAS-EXO-15-005*.
- [5] *M. Aaboud et al.*, *ATLAS Collaboration*, arXiv:1607.03669 [hep-ex].
- [6] *CMS Collaboration*, *CMS-PAS-EXO-16-008*.
- [7] *CMS Collaboration*, *ATLAS-CONF-2015-072*.
- [8] *CMS Collaboration*, *CMS-PAS-EXO-16-001*.
- [9] *M. Aaboud, et al.*, *ATLAS Collaboration*, arXiv:1607.08079 [hep-ex].
- [10] *P. Langacker*, *Rev. Mod. Phys.* 81 (2009) 1199, arXiv:0801.1345 [hep-ph].
- [11] J. Brodzicka, T. Browder, and et al., Belle Collaboration, “Physics achievements from the Belle experiment,” *Progress of Theoretical and Experimental Physics*, vol. 2012, no. 1, pp. 4D001–0, Dec. 2012. [Online]. Available: <https://academic.oup.com/ptep/article-lookup/doi/10.1093/ptep/pts072>

- [12] T. B. Collaboration, “Search for a dark photon in $e+e-$ collisions at BaBar,” *Physical Review Letters*, vol. 113, no. 20, p. 201801, Nov. 2014, arXiv:1406.2980 [hep-ex]. [Online]. Available: <http://arxiv.org/abs/1406.2980>
- [13] “Nobelprize 2008.” [Online]. Available: <https://www.nobelprize.org/prizes/physics/2008/summary/>
- [14] A. M. Baldini *et al.*, “The design of the MEG II experiment,” *The European Physical Journal C*, vol. 78, no. 5, p. 380, May 2018, arXiv:1801.04688 [hep-ex, physics:physics]. [Online]. Available: <http://arxiv.org/abs/1801.04688>
- [15] A. Anastassov *et al.*, “CLEO Collaboration, Phys. Rev. D 55 (1997) 2559, Erratum: Phys. Rev. D 58 (1998) 119904.”
- [16] L. Calibbi and G. Signorelli, “Charged Lepton Flavour Violation: An Experimental and Theoretical Introduction,” *La Rivista del Nuovo Cimento*, vol. 41, no. 2, pp. 71–174, Jan. 2018, arXiv:1709.00294 [hep-ex, physics:hep-ph]. [Online]. Available: <http://arxiv.org/abs/1709.00294>
- [17] “The neutrino: The key to understanding the universe.” [Online]. Available: <https://physicscommunication.ie/the-neutrino-the-key-to-understanding-the-universe/>
- [18] Z. Gruberova, “Study of the lepton decays at the Belle II experiment.” [Online]. Available: <https://docs.belle2.org/record/2419>
- [19] E. Kou, P. Urquijo *et al.*, “The Belle II Physics Book,” *Progress of Theoretical and Experimental Physics*, vol. 2019, no. 12, p. 123C01, Dec. 2019. [Online]. Available: <https://academic.oup.com/ptep/article/doi/10.1093/ptep/ptz106/5685006>
- [20] A. Taroni, “Nobel Prize 2015: Kajita and McDonald,” *Nature Physics*, vol. 11, no. 11, pp. 891–891, Nov. 2015. [Online]. Available: <https://www.nature.com/articles/nphys3543>

- [21] A. Arbey and F. Mahmoudi, “Dark matter and the early Universe: a review,” *Progress in Particle and Nuclear Physics*, vol. 119, p. 103865, Jul. 2021, arXiv:2104.11488 [astro-ph, physics:hep-ph]. [Online]. Available: <http://arxiv.org/abs/2104.11488>
- [22] A. R. Zhitnitskij, “On possible suppression of the axion-hadron interactions,” 1980. [Online]. Available: <https://api.semanticscholar.org/CorpusID:201266379>
- [23] M. Dine, W. Fischler, and M. Srednicki, “A simple solution to the strong CP problem with a harmless axion,” *Physics Letters B*, vol. 104, no. 3, pp. 199–202, Aug. 1981. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/0370269381905906>
- [24] L. Calibbi, F. Goertz, D. Redigolo, R. Ziegler, and J. Zupan, “The Axiflavor,” *Physical Review D*, vol. 95, no. 9, p. 095009, May 2017, arXiv:1612.08040 [hep-ph]. [Online]. Available: <http://arxiv.org/abs/1612.08040>
- [25] Y. Ema, K. Hamaguchi, T. Moroi, and K. Nakayama, “Flaxion: a minimal extension to solve puzzles in the standard model,” *Journal of High Energy Physics*, vol. 2017, no. 1, p. 96, Jan. 2017, arXiv:1612.05492 [hep-ph]. [Online]. Available: <http://arxiv.org/abs/1612.05492>
- [26] C. Froggatt and H. Nielsen, “Hierarchy of quark masses, cabibbo angles and CP violation,” *Nuclear Physics B*, vol. 147, no. 3-4, pp. 277–298, Jan. 1979. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/055032137990316X>
- [27] L. Calibbi, D. Redigolo, R. Ziegler, and J. Zupan, “Looking forward to Lepton-flavor-violating ALPs,” *Journal of High Energy Physics*, vol. 2021, no. 9, p. 173, Sep. 2021, arXiv:2006.04795 [hep-ex, physics:hep-ph]. [Online]. Available: <http://arxiv.org/abs/2006.04795>
- [28] ARGUS Collaboration, H. Albrecht *et al.*, “A search for the lepton-flavour violating decays $\tau \rightarrow e\alpha$, $\tau \rightarrow \mu\alpha$,” *Zeitschrift für Physik C Particles and Fields*, vol. 68, no. 1, pp. 25–28, Mar. 1995. [Online]. Available: <http://link.springer.com/10.1007/BF01579801>

- [29] A.-K. Perrevoort, “Sensitivity Studies on New Physics in the Mu3e Experiment and Development of Firmware for the Front-End of the Mu3e Pixel Detector,” 2018, publisher: Heidelberg University Library. [Online]. Available: <http://archiv.ub.uni-heidelberg.de/volltextserver/id/eprint/24585>
- [30] A. Jodidio, B. Balke, J. Carr, G. Gidal, K. A. Shinsky, H. M. Steiner, D. P. Stoker, M. Strovink, R. D. Tripp, B. Gobbi, and C. J. Oram, “Search for right-handed currents in muon decay,” *Physical Review D*, vol. 34, no. 7, pp. 1967–1990, Oct. 1986. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevD.34.1967>
- [31] R. Bayes, J. Bueno, and et al., “A search for two body muon decay signals,” *Physical Review D*, vol. 91, no. 5, p. 052020, Mar. 2015, arXiv:1409.0638 [hep-ex]. [Online]. Available: <http://arxiv.org/abs/1409.0638>
- [32] B. I. Collaboration, I. Adachi *et al.*, “Search for Lepton-Flavor-Violating τ Decays to a Lepton and an Invisible Boson at Belle II,” May 2023, arXiv:2212.03634 [hep-ex]. [Online]. Available: <http://arxiv.org/abs/2212.03634>
- [33] W. Altmannshofer, C.-Y. Chen, P. B. Dev, and A. Soni, “Lepton flavor violating Z' explanation of the muon anomalous magnetic moment,” *Physics Letters B*, vol. 762, pp. 389–398, Nov. 2016. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0370269316305500>
- [34] The ALEPH Collaboration, The DELPHI Collaboration, The L3 Collaboration, The OPAL Collaboration, and The LEP Electroweak Working Group, “Electroweak Measurements in e^-e^+ Collisions at W-Boson-Pair Energies at LEP,” *Physics Reports*, vol. 532, no. 4, pp. 119–244, Nov. 2013, arXiv:1302.3415 [hep-ex, physics:hep-ph]. [Online]. Available: <http://arxiv.org/abs/1302.3415>
- [35] A. Hook, E. Izaguirre, and J. G. Wacker, “Model Independent Bounds on Kinetic Mixing,” Jun. 2010, arXiv:1006.0973 [hep-ex, physics:hep-ph]. [Online]. Available: <http://arxiv.org/abs/1006.0973>

- [36] D. Curtin, R. Essig, S. Gori, and J. Shelton, “Illuminating Dark Photons with High-Energy Colliders,” *Journal of High Energy Physics*, vol. 2015, no. 2, p. 157, Feb. 2015, arXiv:1412.0018 [hep-ex, physics:hep-ph]. [Online]. Available: <http://arxiv.org/abs/1412.0018>
- [37] K. Akai, K. Furukawa, and H. Koiso (SuperKEKB Collaboration), *Nucl. Instrum. Meth. A* **907**, 188 (2018).
- [38] “Istituto nazionale di fisica nucleare.” [Online]. Available: <https://w3.lnf.infn.it/superkekb-conquista-il-nuovo-record-mondiale-di-luminosita/>
- [39] T. Abe, I. Adachi *et al.*, “Belle II Technical Design Report,” Nov. 2010, arXiv:1011.0352 [hep-ex, physics:physics]. [Online]. Available: <http://arxiv.org/abs/1011.0352>
- [40] *KEK News Article*, <https://www2.kek.jp/accl/eng/topics/topics211224.html>.
- [41] “*Belle II* official website.” [Online]. Available: <https://confluence.desy.de/display/BI/Belle+II+Luminosity>
- [42] P. Rados, “Tau measurements and prospects.” [Online]. Available: <https://docs.belle2.org/record/2577?ln=en>
- [43] T. Kuhr, C. Pulvermacher, M. Ritter *et al.*, *The Belle II Core Software. Comput Softw Big Sci* **3**, 1 (2019).
- [44] “basf2 - documentation,” *Git HUB respository*. [Online]. Available: <https://github.com/belle2/basf2>
- [45] S. Jadach, F. L. Ward, B., and Z. Was, *Comput. Phys. Commun.* **130**, 260 (2000).
- [46] —, *Phys. Rev. D* **63**, 113009 (2001).
- [47] G. Balossini, C. M. Carloni Calame, G. Montagna, O. Nicosini, and F. Piccinini, *Nucl. Phys. B* **758**, 227 (2006).

- [48] —, *Phys. Lett. B* 663, 209 (2008).
- [49] C. M. Carloni Calame, G. Montagna, O. Nicrosini, and F. Piccinini, *Nucl. Phys. B Proc. Suppl.* 131, 48 (2004).
- [50] C. M. Carloni Calame, *Phys. Lett. B* 520, 16 (2001).
- [51] C. M. Carloni Calame, C. Lunardini, G. Montagna, O. Nicrosini, and F. Piccinini, *Nucl. Phys. B* 584, 459 (2000).
- [52] F. Berends, P. Daverveldt, and R. Kleiss, *Nucl. Phys. B* 253, 421 (1985).
- [53] —, *Nucl. Phys. B* 253, 441 (1985).
- [54] —, *Comp. Phys. Commun.* 40, 285 (1986).
- [55] S. Uehara, (1996), *arXiv:1310.015*.
- [56] M. Chrzaszcz, T. Przedzinski, Z. Was, and J. Zaremba, *Comput. Phys. Commun.* 232, 220 (2018).
- [57] E. Barberio, B. van Eijk, and Z. Was, *Comput. Phys. Commun.* 66, 115 (1991).
- [58] T. Sjöstrand, S. Ask, R. Christiansen, J., R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, O. Rasmussen, C., and P. Z. Skands, *Comput. Phys. Commun.* 191, 159 (2015).
- [59] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, and N. Braun, (*Belle II Framework Software Group*), *Comput. Softw. Big Sci.* 3, 1 (2019).
- [60] *Belle II Analysis Software Framework (basf2)*, <https://doi.org/10.5281/zenodo.5574115>.
- [61] S. Agostinelli *et al.*, (*GEANT4 Collaboration*), *Nucl. Instrum. Meth. A* 506, 250 (2003).
- [62] A. L. Read, *J. Phys. G* 28, 2693 (2002).

- [63] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, pp. 785–794, arXiv:1603.02754 [cs]. [Online]. Available: <http://arxiv.org/abs/1603.02754>
- [64] M. Stone, “Cross-validatory choice and assessment of statistical predictions,” *Journal of the Royal Statistical Society: Series B (Methodological)* 36(2), 111–133 (Jan 1974). [Online]. Available: <http://dx.doi.org/10.1111/j.2517-6161.1974.tb00994.x>
- [65] G. Punzi, “Sensitivity of searches for new signals and its optimization,” Dec. 2003, arXiv:physics/0308063. [Online]. Available: <http://arxiv.org/abs/physics/0308063>