Technical University of Munich
Max Planck Institute

Background Modeling in Partial-Wave Analysis of $\tau^- \rightarrow \pi^+\pi^-\pi^-\nu_\tau$ Decays at Belle II

Miriam Weiskopf

Thesis submitted
within the Master's Program
Nuclear-, Particle-, Astrophysics

Supervisor: Prof. Dr. Stephan Paul
Second Examiner: Dr. Oliver Kortner
Advisor: Dr. Stefan Wallner

May 18, 2026

Thesis advisor: Prof. Dr. Stephan Paul *Miriam Weiskopf*

Background Modeling in Partial-Wave Analysis of $\tau^- \to \pi^+\pi^-\pi^-\nu_\tau$ Decays at Belle II

ABSTRACT

The decay $\tau^- \to \pi^+\pi^-\pi^-\nu_\tau$ provides an important laboratory for studying the strong interaction and light-meson resonances using the technique of partial-wave analysis (PWA). A major challenge in this analysis is the treatment of background contributions, which exhibit highly correlated multidimensional phase-space distributions. In this thesis, we investigate neural-network-based background modeling methods and integrate them into the PWA framework. We study different neural network approaches and evaluate their performance in partial-wave decomposition fits on simulated pseudo data. We find that strong regularization and kinematic restrictions to signal-populated phase-space regions are necessary to avoid unstable neural-network predictions. Furthermore, we compare different training strategies and examine the impact of dividing the full invariant-mass range into multiple training regions for the neural networks. We find that training on overlapping invariant-mass regions (so-called sliding-window approach) yields the most stable and accurate background description. In addition, the analyzed data contains several distinct background contributions. To model these, two modeling approaches are investigated: a combined approach, where all background contributions are modeled simultaneously, and a split approach, where individual background contributions are modeled as separate components. Their performance is validated through complementary input-output studies. We find that the combined approach yields stable background descriptions and reproduces the input in the signal partial waves well. In contrast, the split approach exhibits significant leakage into the dominant $\tau^- \to \pi^+\pi^-\pi^-\pi^0\nu_\tau$ background component. However, this effect can be mitigated by fixing its yield to the true Monte Carlo value. Additional supporting studies show that neglecting detector resolution introduces only small biases in the signal description. We further investigate biases caused by overlapping partial waves and find that these effects can be reduced by excluding individual partial waves in weakly constrained invariant-mass regions. Overall, the presented studies demonstrate that neural-network-based background modeling can successfully be integrated into the partial-wave decomposition framework and provide a foundation for future partial-wave analyses at Belle II.

Contents

1 Introduction

The Standard Model of particle physics (SM) represents our current understanding of the fundamental particles and three of the four fundamental interactions: the electromagnetic, weak, and strong interactions [1]. The electromagnetic interaction, mediated by a photon, acts between charged particles, for example, between the positively charged protons and the negatively charged electrons in atoms. Such bound states can be described with very high precision within Quantum Electrodynamics (QED), as it can be calculated perturbatively.
The strong interaction is described by Quantum Chromodynamics (QCD) [2] within the SM and acts between color-charged quarks. It binds them to hadrons, which are the strong analogous to electromagnetically bound states. The mediating particle is the gluon, which, contrary to photons, carries charge (color) and also interacts with other gluons (self-interaction). At hadronic energy scales ($\sim 1\,\mathrm{GeV}$), the strong coupling α_s is large, and quarks are confined in color-neutral bound states, i.e., there are no free quarks. Due to the large α_s, perturbative methods are generally not applicable in this regime, making first-principles calculations of these strongly bound systems challenging.

A phenomenological description of the hadrons is given by the quark model [3, 4]. The quark model classifies hadron properties in terms of their constituent quarks. It predicts their properties, including mass, width, and quantum numbers, i.e., the isospin I, the total angular momentum J, and the parity P, with the latter two typically written as J^P. The isospin represents the approximate symmetry of the strong interaction between up and down quarks, while the parity P describes the behavior of quantum states under spatial inversion.
In the quark model, the simplest strongly bound systems are mesons. These are composed of constituent quark and antiquark pairs and can be further classified as light mesons, which are pairs of the lightest quarks, i.e., up, down, and strange; and heavy mesons, which contain at least one heavy quark.

The quark model is only an effective, phenomenological description of strongly bound systems, which relies on external inputs and model assumptions. A fundamental approach is lattice QCD, a numerical method for calculating non-perturbative hadronic processes on a four-dimensional space-time lattice [5]. Such calculations at physical quark masses require immense computational resources. Consequently, they are typically performed at artificially large quark masses and then extrapolated to the physical point. Despite these limitations, lattice QCD provides first-principle calculations of QCD. To compare such calculations with experimental observations, we require high-precision measurements.

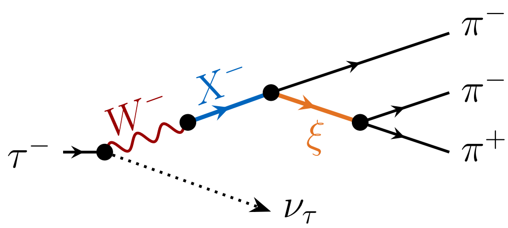


Figure 1.1: Illustration of τ decaying to three pions via the intermediate resonance X^- and the two-body resonance ξ.

One way to study light mesons is the decay of τ leptons. The τ lepton is the heaviest lepton in the SM with a mass of $\approx 1.777\text{GeV}/c^2$ [3], and it is the only lepton that can decay hadronically. In such τ decays, light mesons appear as intermediate resonances, since they have an extremely short lifetime and decay quasi immediately to lighter mesons such as pions and kaons. These intermediate meson resonances are observed in the distribution of their decay products, from which we study their mass, width, and quantum numbers.

This analysis focuses on the decay of $\tau^- \to \pi^+\pi^-\pi^-\nu_\tau$ (shown in Figure 1.1), in τ pairs produced back-to-back in the center-of-mass system (CMS) of the e^+e^- collisions at the Belle II experiment.

Assuming isospin symmetry, the hadronic $\pi^+\pi^-\pi^-$ system is produced from the weak axial-vector current. The weak axial-vector current can produce axial-vector resonances a_1 ($J^P = 1^+$) or pseudo-scalar resonances π ($J^P = 0^-$) [6]. Contributions through the weak vector current to the $\pi^+\pi^-\pi^-$ final state are suppressed by isospin symmetry and correspond to second-class currents. They can lead to resonances with quantum numbers $J^P = 1^-$ (ρ).

The decay of $\tau^- \to \pi^+\pi^-\pi^-\nu_\tau$ is dominated by the production of the meson resonance $a_1(1260)$, which mainly decays via the intermediate $\rho(770)\pi$ state. Since the mass and width of the $a_1(1260)$ meson are not well determined [3, 7, 8], further studies on the $a_1(1260)$ meson are necessary. Furthermore, COMPASS [9] observed contributions from the $a_1(1420)$, whose nature is not fully explained [10].

A previous analysis of the $\tau^- \to \pi^+\pi^-\pi^-\nu_\tau$ decay at Belle, reported in reference [11], shows promising results for the high-precision measurement of light mesons and a possible indication of an isospin-violating process. Thus, we want to study the nature of these meson resonances and confirm their presence using the Belle II dataset. While the currently available dataset size of Belle II is comparable to that of the Belle experiment, Belle II offers an improved detector performance and advanced analysis techniques [12], enabling more precise measurements and improved sensitivity to small isospin-violating effects.

Since the decay involves broad, overlapping resonance states, we perform a partial-wave analysis (PWA). It is based on the principle that such processes can be modeled as a coherent sum of intermediate resonant states and their decay mode (partial waves) with well-defined quantum numbers J^P, as a function of their energies and angular distributions. This enables the modeling of interferences between partial waves and the extraction of their individual contributions to the final state.
A PWA is performed in two steps. The first step is the partial-wave decomposition (PWD), in which resonance contributions are decomposed into individual partial waves. The results from the PWD are then used in the second step, the resonance model fit to determine the masses and widths of the meson resonances.

A major challenge in this partial-wave analysis is the treatment of background contributions. With around 10 percent of background contributions in the data sample, its description is essential for such high-precision measurements. Since these contributions are high-dimensional and non-trivially correlated, we do not have an analytical parametrization.
To address this challenge, we investigate whether neural networks can be used to model background contributions within the partial-wave decomposition framework. As universal approximators [13], neural networks are, in principle, capable of approximating arbitrary continuous functions from a given distribution and may therefore provide a flexible description of complex background structures. The aim of this work is to evaluate the performance and limitations of such background models.
In addition to the background modeling studies, we investigate potential biases of the PWD framework and explore a method to stabilize the fit results in poorly constrained kinematic regions.

The thesis is structured as follows. Chapter 2 introduces the Belle II detector and the SuperKEKB accelerator. This is followed by a discussion of the reconstruction and event selection of the $\tau^- \to \pi^+\pi^-\pi^-\nu_\tau$ decay used in this analysis in chapter 3, established in reference [14] with contributions from references [12, 15]. In Chapter 4, the formalism of the partial-wave decomposition and the fitting framework are introduced, based on the previous Belle analysis [11] and integrated for the Belle II analysis in references [16, 17]. In chapter 5, we investigate threshold effects and ambiguities arising from strongly overlapping partial waves in poorly constrained kinematic regions, and study their mitigation through the introduction of thresholds. Two background modeling methods are presented in Chapter 6, while the corresponding performance studies and results for the chosen method are shown in Chapter 7. These results are further investigated in Chapter 8 via dedicated studies that assess leakage between signal and background contributions, the impact of detector resolution, and the impact of incorrectly estimated background contributions in the description of the signal contributions. Finally, Chapter 9 summarizes the results and provides an outlook for future studies.

2 Belle II Experiment

The Belle II experiment is the successor to the Belle experiment and is located at the SuperKEKB asymmetric energy e^+e^- collider in Japan, which in turn succeeded KEKB. The collider operates at center-of-mass energies at and around the $\Upsilon(4S)$ resonance, which is right above the production threshold of a B-meson pair.
The primary goal of Belle II is to search for new phenomena in flavor physics and the precision measurements of the Standard Model (SM) parameters. While the experiment was originally designed to study B decays, the high production rate of τ-lepton pairs in e^+e^- collisions also enables broad τ-lepton studies [18].
The following introduces the SuperKEKB collider and the major components of the Belle II detector. A detailed overview of the Belle II experiment can be found in [18, 19], on which the following sections are based.

2.1 The SuperKEKB Accelerator

SuperKEKB is a circular electron-positron (e^+e^-) collider. As shown in Figure 2.1, the electrons and positrons are first accelerated in the injector linear accelerator (linac) to energies of 7 GeV and 4 GeV. The positrons are injected into the low-energy ring (LER), while the electrons are injected into the high-energy ring (HER). The particles collide at a center-of-mass energy of $\sqrt{s} = 10.58\,\text{GeV}$ at the interaction point, which is surrounded by the Belle II detector. The detector is discussed in Section 2.2.
In order to assess the performance of the collider, the luminosity, defined as [20]

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

is used. This represents the proportionality factor between the event rate $\frac{dN}{dt}$ and the cross section σ of the process under study. SuperKEKB is designed to achieve a peak luminosity of $8 \times 10^{35}\text{cm}^{-2}s^{-1}$, which is 40 times larger than the previous KEKB peak luminosity [21]. The more relevant quantity is the integrated luminosity, given by

$$\mathcal{L}_{int} = \int_0^T \mathcal{L}(t')dt' \,, \tag{2.2}$$

as this directly relates to the total number of observed events in a given time period T. Belle II aims to collect a total integrated luminosity of $50\,\text{ab}^{-1}$ over its full runtime.

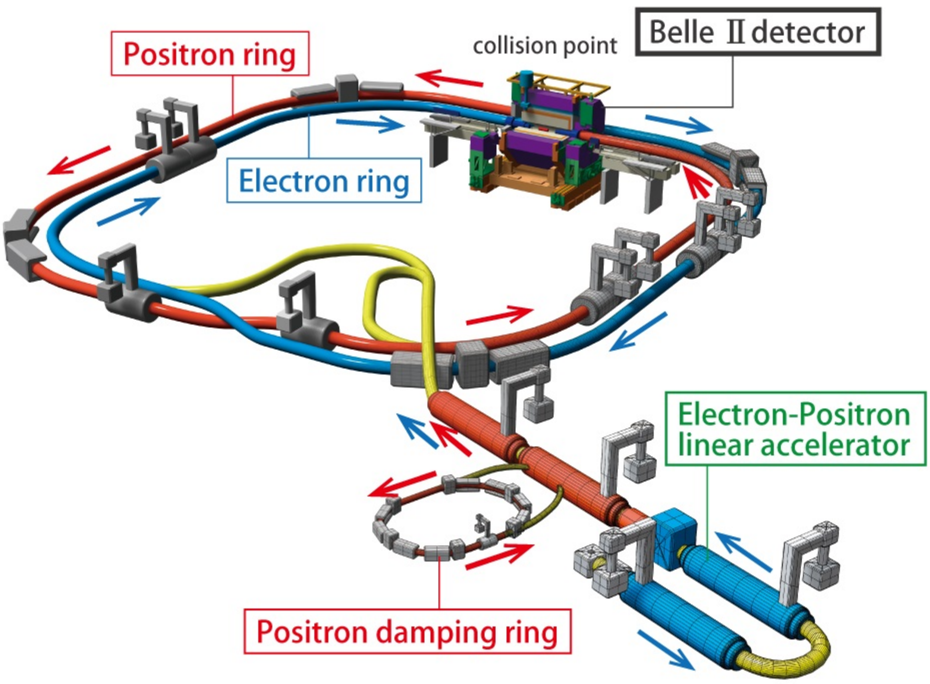


Figure 2.1: A schematic view of the SuperKEKB Collider, taken from [21]

2.2 The Belle II Detector

The Belle II Detector is positioned at the collision point of the SuperKEKB collider. As shown in Figure 2.2, it is composed of several sub-detectors which are arranged in cylindrical layers around the beam pipe, measuring numerous quantities, including the position, momentum, and energy of particles. Belle II uses a right-handed coordinate system [22], with the interaction point defined as the origin. The z-axis is approximately aligned with the electron beam, the x-axis points radially outward from the center of the accelerator ring, and the y-axis points upward. The polar angle θ is measured with respect to the z-axis, while the azimuthal angle ϕ is measured in the transverse (x, y) plane.

The Inner part, consisting of the Vertex Detector (VXD) and the Central Drift Chamber (CDC), is the tracking and vertex detection system, responsible for reconstructed charged-particle tracks and decay vertices, as detailed in Section 2.2.1. While the information from several sub-detectors contributes to the particle identification (PID),i.e., the determination of the particle species such as electrons, muons, pions, kaons, protons, and deuterons, the time-of-propagation (TOP) and Aerogel Ring-Imaging Cherenkov Detector (ARICH) are dedicated sub-detectors for charged particle identification. These are especially important for the distinction of kaons and pions and will be discussed in Section 2.2.2. The outer layers of the detector are the electromagnetic calorimeter (ECL), described in Section 2.2.3, and the K_L^0 and μ detector (KLM) discussed in Section 2.2.4.

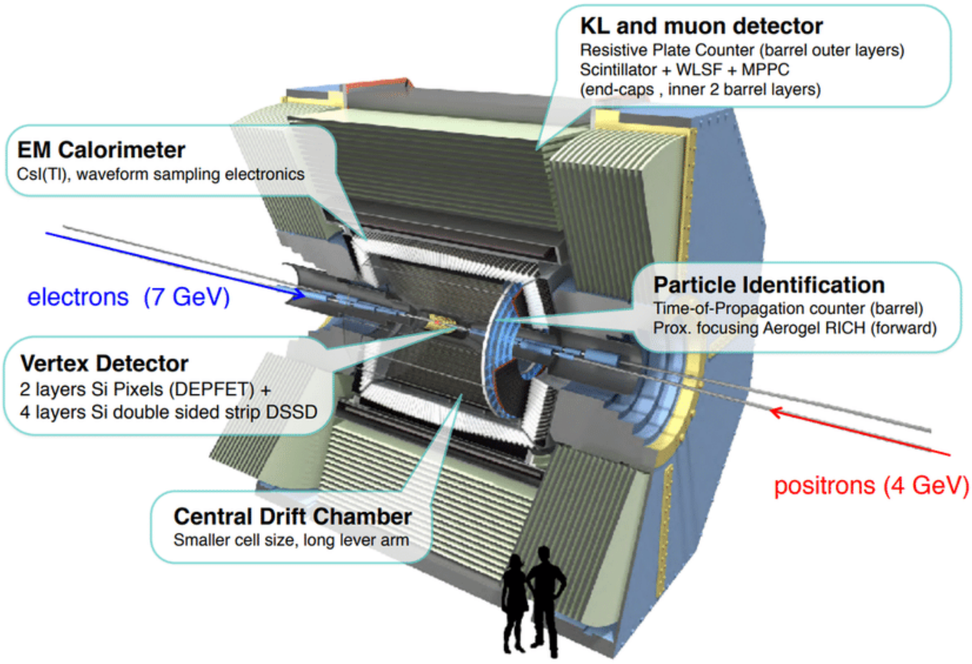


Figure 2.2: A schematic view of the Belle II detector with all subdetectors [23].

2.2.1 Tracking

The vertex reconstruction and tracking of charged particles is done using information from the VXD and the CDC. The tracking sub-detectors are immersed in a magnetic field of 1.5 T, resulting in charged particle tracks represented by helices. Measuring these tracks gives information about the momentum and position of the particles. Both components are discussed in the following sections.
The VXD and CDC cover the full range of the azimuthal angle, as well as a broad coverage of the polar angle from 17° to 150°.

Vertex Detector (VXD)

The VXD is composed of the Pixel Vertex Detector (PXD), which forms the two innermost layers, and the Silicon Vertex Detector (SVD), corresponding to the four outer layers.
The SVD layers consist of double-sided silicon strip detectors, which operate based on charge collection in semiconductors, i.e., when charged particles traverse the silicon, they ionize the material and create electron-hole pairs that are collected, producing an electrical signal [24].
Its layers are located at radii of 39, 80, 104, and 135 mm from the interaction point.

The innermost layers (the PXD), being closest to the interaction point with radii of 14 and 22 mm, are strongly affected by beam-induced background. This results in a high occupancy, with many readout channels being hit in each triggered event. To cope with this, these layers use pixel detectors instead of strips, providing a much finer segmentation due to the larger number of readout channels. The pixel detectors are based on the DEPFET (DEPleted Field Effect Transistor) technology.

Central Drift Chamber (CDC)

The CDC located around the VXD is used for track reconstruction and for measuring particle momenta. The CDC is a cylindrical chamber with 56 total layers of sense wires, grouped into 9 superlayers. These layers alternate in orientation related to the beam axis (axial or stereo orientation), providing 3D reconstruction.
Besides being used for tracking, it is an important part of particle identification. Charged particles traversing the gas-filled chamber ionize, i.e., the particles lose energy. By measuring the energy loss and the momentum of the particle, the species of the transversing particle can be determined. Finally, it provides efficient trigger information for charged particles.

2.2.2 Particle Identification System

The TOP and ARICH detectors are used to identify charged particles, with a focus on differentiating kaons from pions. Both rely on the detection of Cherenkov radiation generated in the radiator of each sub-detector, as detailed in the following sections. The underlying principle is the relation between the emission angle of the Cherenkov radiation and the ratio of the velocity to the speed of light, denoted as β, which is defined as

$$\beta = \frac{1}{\sqrt{1 + \frac{m^2c^2}{p^2}}}\,, \tag{2.3}$$

with m being the mass of the particle and p the momentum. Thus, the relation of

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

,with n being the refractive index of the radiator medium, can be used to uniquely determine the particle's mass.

Time-Of-Propagation (TOP)

The TOP counter, located in the barrel region of the spectrometer, measures the time of propagation, and positions of Cherenkov photons generated in the quartz radiator. The Cherenkov image is reconstructed from the photon's time and position, which are measured by micro-channel plate (MCP) photomultiplier tubes (PMTs).

Aerogel Ring-Imaging Cherenkov Detector (ARICH)

The ARICH detector is designed to distinguish kaons and pions over a large momentum range, as well as between pions, muons, and electrons below 1 GeV/c. Photons are generated from particles interacting with the aerogel radiator, which then propagate through the expansion volume (20 cm), allowing the Cherenkov ring to form. The photons are detected by position-sensitive photon detectors, from which the particles are differentiated.

2.2.3 Electromagnetic Calorimeter (ECL)

The ECL is a 3m long barrel with forward and backward endcaps, designed to detect photons and identify electrons. This is required because of the large number of neutral particles, especially π^0 decaying to two photons, produced in B or τ decays.
The ECL consists of 8736 thallium-doped cesium iodide CsI(TI) crystals. An electromagnetic shower is induced when an incoming photon or electron interacts with a crystal. The resulting scintillating light is read out by two photodiodes, which can be related to the particle's deposited energy.

2.2.4 K_L^0 and μ detector

The KLM is the outermost sub-detector of Belle II, composed of a barrel detector and forward and backward end-cap sections, covering the full azimuthal angle range and polar angles between 25° to 155°. It aims to detect K_L^0 and μ that have not (fully) deposited their energy within the other detectors. The KLM is constructed using an alternating sequence of iron plates and active detectors in order to detect long-lived particles.
The two particles leave different signatures in the detector. While the K_L interacts hadronically in the ECL or the iron plates of the KLM, the μ and other non-showering charged hadrons traverse the KLM via a near-straight line. These signatures are read out by resistive plate chambers (RPCs) and silicon photomultipliers (SiPMs).

3 Event Selection

The Belle II experiment is designed to record an integrated luminosity of 50 ab^{-1}. For this analysis, we use generic run-dependent MC15 data, and the phase-space variables used throughout are introduced in Section 3.1. In order to choose $\tau^- \to \pi^+\pi^-\pi^-\nu_\tau$ events relevant for the analysis, an event reconstruction is required, described in Section 3.2, as well as a pre-selection (see Section 3.3) and fine-selection (see Section 3.4). The resulting sample after the event selection includes a remaining $\approx 9.5\%$ background events not corresponding to the signal process, which are detailed in Section 3.5. The general notation follows $\tau^- \to \pi_1^+\pi_2^-\pi_3^-\nu_\tau$, where the same charged pions are ordered by their momenta such that $p_3 > p_2$.

3.1 Phase Space Variables

The phase space of a four-body decay is described by 12 kinematic variables, of which only 8 are independent, because of energy-momentum conservation. A possible choice of these 8 phase space variables includes three invariant masses: the mass of the three-pion system $m_{3\pi}$ and two two-pion masses m_{12} and m_{23}, as well as 5 angular variables. The mass variables will be discussed in Section 3.1.1, while the angular variables are detailed in Section 3.1.2.

3.1.1 Mass Variables

The mass of the three-pion system $m_{3\pi}$ is calculated as the sum of the four-momenta of the pions as:

$$m_{3\pi} = \sqrt{(\sum_i p_i)^2}\,, \tag{3.1}$$

where p_i denote the momenta in the CMS. Each invariant mass of the two pions is calculated as the sum of the two-pion four-momenta, as:

$$m_{12} = \sqrt{(p_1 + p_2)^2}, \qquad m_{23} = \sqrt{(p_2 + p_3)^2} \tag{3.2}$$

3.1.2 Angular Variables

Furthermore, the decay is described by five angles, three Euler angles (α, $\cos(\beta)$, γ) and two helicity angles ($\cos(\theta)$, ϕ). The helicity angles describe the orientation of the three-pion subsystem X in the τ rest frame relative to the τ flight direction in the CMS. The Euler angles describe the orientation of the three-pion decay plane in the three-pion rest frame. We follow the definition detailed in ref. [25, 26].

The Euler angles are determined geometrically from measured pion momenta in the three-pion rest frame. We define the direction of the CMS $\vec{n}_{\text{CMS}}$ in the three-pion rest frame to be exactly opposite to the sum of the three-pion momenta in the CMS (denoted with $\vec{p}$)

$$\vec{n}_{\text{CMS}} = -\frac{\vec{p}_1 + \vec{p}_2 + \vec{p}_3}{|\vec{p}_1 + \vec{p}_2 + \vec{p}_3|}. \tag{3.3}$$

Additionally, we define the normal vector to the three-pion decay plane as

$$\vec{n}_{\perp} = \frac{\vec{q}_2 \times \vec{q}_3}{|\vec{q}_2 \times \vec{q}_3|}, \tag{3.4}$$

where q_i are the four-momenta boosted to the three-pion rest frame. Finally, the unit vector along π_1^+ is defined as:

$$\hat{q}_1 = \frac{\vec{q}_1}{|\vec{q}_1|}. \tag{3.5}$$

Using these definitions, we can calculate the Euler angles.
The angle $\cos(\beta)$ is defined as the angle between the CMS direction and the normal to the hadron decay plane $n_{\perp}$ as

$$\cos\beta = \vec{n}_{\text{CMS}} \cdot \vec{n}_{\perp} \,. \tag{3.6}$$

The Euler angle γ is defined via its sin and cos values as

$$\gamma = \text{atan2}(\sin\gamma, \cos\gamma)\,, \tag{3.7}$$

where $\sin(\gamma)$ is calculated as

$$\sin\gamma = \frac{(\vec{n}_{\text{CMS}} \times \vec{n}_{\perp}) \cdot \hat{q}_1}{|\vec{n}_{\text{CMS}} \times \vec{n}_{\perp}|}, \tag{3.8}$$

and $\cos(\gamma)$ as

$$\cos\gamma = -\frac{\vec{n}_{\text{CMS}} \cdot \hat{q}_1}{|\vec{n}_{\text{CMS}} \times \vec{n}_{\perp}|}. \tag{3.9}$$

The third Euler angle α is the angle between the two planes spanned by $[\vec{n}_{\text{CMS}}, \vec{n}_{\tau}]$, with $\vec{n}_{\tau}$ being the τ flight direction in the hadronic rest frame, and $[\vec{n}_{\perp}, \vec{q}_1]$. Thus, the respective sin values are given by

$$\cos(\alpha) = \frac{(\vec{n}_{\text{CMS}} \times \vec{n}_{\tau}) \cdot (n_{\text{CMS}} \times \vec{n}_{\perp})}{|\vec{n}_{\text{CMS}} \times \vec{n}_{\tau}|\,|\vec{n}_{\text{CMS}} \times \vec{n}_{\perp}|} \tag{3.10}$$

and the cos values as

$$\sin(\alpha) = -\frac{\vec{n}_{\tau} \cdot (n_{\text{CMS}} \times \vec{n}_{\perp})}{|\vec{n}_{\text{CMS}} \times \vec{n}_{\tau}|\,|\vec{n}_{\text{CMS}} \times \vec{n}_{\perp}|}\,. \tag{3.11}$$

However, since we do not fully reconstruct the τ direction due to the missing ν_{τ}, this angle cannot be calculated directly.

Such a geometric calculation is not possible for the helicity angles, since the τ neutrino escapes detection at the Belle II experiment and therefore the τ rest frame is not reconstructed. Thus, the helicity angle θ_H, defined as the angle between the direction of the three-pion system in the τ rest frame, and the τ flight direction in CMS, is kinematically reconstructed from the energy of the hadronic system as

$$\cos\theta_H = \frac{2E_{3\pi}\frac{m_\tau^2}{E_\tau} - m_\tau^2 - m_{3\pi}^2}{(m_\tau^2 - m_{3\pi}^2)\sqrt{1 - \frac{m_\tau^2}{E_\tau^2}}}\,, \tag{3.12}$$

where $E_{3\pi}$ and $m_{3\pi}$ are the energy and mass of the three-pion system, and E_τ and m_τ the energy and mass of the τ lepton. We integrate over the second helicity angle, since we do not measure the neutrino.

3.2 Event Reconstruction

Belle II operates at a center-of-mass energy $E_{\text{CMS}} = 10.58$ GeV$/c^2$, corresponding to the $\Upsilon(4S)$ resonance. At this energy, the e^+e^- collision produces highly boosted τ-pairs, such that the two τ leptons are back-to-back. The thrust axis of this decay (gray dashed line in Figure 3.1) is the unit vector where the projection of the momenta of the measured final state particles is maximal [18]. Based on the plane perpendicular to this thrust axis, the decay is split into two hemispheres, with one side denoted the signal side and the other the tag side. τ leptons have an average lifetime of 290×10^{-15} s [3], thus decay before detection, and the event topology seen for the τ pair decay has either 2, 4, or 6 charged tracks.
The decay of $\tau^- \to \pi^-\pi^-\pi^+\nu_\tau$ occurs with a branching fraction of ≈ 0.09 [3] and produces 3 charged tracks on one side.
In this analysis, the 3x1 prong topology is used, visualized in 3.1. It requires 3 good charged tracks on the signal side corresponding to the signal decay and 1 good charged track on the tag side.
The definition of a good track is that the radius from the interaction point in the transverse plane is below 1 cm and the distance from the interaction point along the longitudinal coordinate is less than 3 cm. This definition comes from the short lifetime of the τ leptons, i.e., the tracks from the decay products of the τ lepton should be close to the interaction point.

3.3 Preselection

Reducing the size of the Belle II dataset requires filtering based on specific trigger configurations and using the tau-thrust skim.
The exact values for all kinematic cuts are specified in table 3.1.
Events are selected if the configuration for one of the low-multiplicity prescaled trigger bits ”lml” 2, 6-10, or 12 is satisfied, or alternatively, the trigger bit ”hie”.

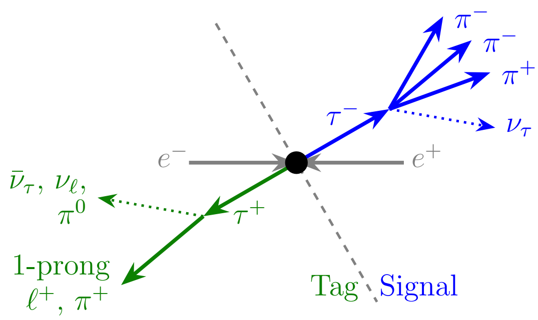


Figure 3.1: Illustration of the tag and signal side decay of the τ leptons produced in the $e^+e^- \to \tau^+\tau^-$ process, separated by the thrust axis (gray).

The tauThrust skim sets additional criteria for accepted events. The number of tracks N_t is limited to $1 < N_t < 7$ due to the decay topology of the τ. Kinematically, we require the thrust to be greater than 0.8 and the visible energy to be below 10.4 GeV.
Beyond that, additional kinematic preselections are introduced. First, a lower limit on the visible energy and an upper limit on the thrust T are set. The thrust angle $\cos\theta_{\text{thrust}}$, i.e., the angle between the thrust axis and the beam axis, is also constrained, including both a direct range requirement and an additional correlation with the missing momentum angle θ_{miss}. Similarly, the squared missing mass, m^2_{miss}, is restricted by both a fixed interval and a θ_{miss}-dependent condition. Further topological constraints are imposed through an upper bound on the invariant mass of the three-prong system. Finally, photo and π^0 multiplicities are limited in the one-prong and three-prong hemispheres.

3.4 Fine Selection

The fine selection is done in three steps. First, a boosted decision tree (BDT) is trained, separating signal from background events, and final state kaons are rejected using a neural-network based PID variable, as discussed in Section 3.4.1. The MC data requires corrections due to discrepancies in the simulation of the Belle II detector and the real detector [12]. Therefore, we need to limit the range to regions where these are valid, as detailed in Section 3.4.2, and finally restrict the range based on the kinematically allowed phase-space borders, described in Section 3.4.3.

3.4.1 Selection using a Boosted Decision Tree

After preselection, the data sample still contains a large fraction of background events. Therefore, we train a Boosted Decision Tree (BDT) using lightGBM, separating background from signal and assigning a score between 0 and 1 to each event. Only events above a threshold of $T_{\text{BDT}} = 0.633$ are kept, which corresponds to a sample purity of 90 %.
This selection follows the implementation introduced in ref. [15] and is discussed for the

Table 3.1: Preselection cuts

Variable	**Selection cut**
Visible energy (CMS)	$3.5\ \text{GeV} \leq E_{\text{vis}}^{\text{CMS}} < 10.5\ \text{GeV}$
Thrust	$0.8 \leq T < 0.99$
Thrust angle I	$-0.7 \leq \cos\theta_{\text{thrust}} < 0.9$
Thrust angle II	$\cos\theta_{\text{thrust}} > 0.9\,\theta_{\text{miss}} - 1.575$ **or** $\cos\theta_{\text{thrust}} < -1.4\,\theta_{\text{miss}} + 2.45$
Missing mass squared I	$-3\ \text{GeV}^2/c^4 < m^2_{\text{miss}} < 45\ \text{GeV}^2/c^4$
Missing mass squared II	$m^2_{\text{miss}} < \dfrac{-35\ \text{GeV}^2/c^4}{\pi - 1.1}(\theta_{\text{miss}} - 1.1) + 45\ \text{GeV}^2/c^4$
Three-prong invariant mass	$m_{\text{3prong}} < 2.0\ \text{GeV}/c^2$
Photon multiplicity (1-prong)	$N_{\gamma}^{\text{1prong}} < 5$
Photon multiplicity (3-prong)	$N_{\gamma}^{\text{3prong}} < 5$
π^0 multiplicity (1-prong)	$N_{\pi^0}^{\text{1prong}} < 2$
π^0 multiplicity (3-prong)	$N_{\pi^0}^{\text{3prong}} < 2$

$\tau^- \to \pi^+\pi^-\pi^-$ decay in ref. [27]. The BDT is trained on 16 variables listed in table 3.2. These include the multivariate analysis (MVA) variables concerning photon selection [28], which improve the purity of the sample, especially for the $\tau^- \to \pi^+\pi^-\pi^-\pi^0\nu_\tau$ background.

Table 3.2: Input variables used for the BDT.

Kinematic variables and π^0 rejection			
thrust	$E_{\text{vis}}^{\text{CMS}}$	m_{miss}	$\theta_{\text{miss}}^{\text{CMS}}$
$\lvert\vec{p}_{\text{miss}}^{\,\text{CMS}}\rvert$	$E_{\text{3prong}}^{\text{CMS}}$	$p_{t,\text{3prong}}^{\text{CMS}}$	$N_{\pi^0}^{\text{3prong}}$
$N_{\pi^0}^{\text{1prong}}$	dr_1^{3prong}	dr_2^{3prong}	dr_3^{3prong}
$\chi^2_{\text{vtx, 3prong}}$	$\Delta_{\text{vtx, 3prong}}$		
MVA γ-rejection variables			
$N_{\gamma,\text{MVA,good}}^{\text{3prong}}$	$N_{\gamma,\text{MVA,good}}^{\text{1prong}}$	$E_{\gamma,\text{MVA,good}}^{\text{3prong}}$	$E_{\gamma,\text{MVA,good}}^{\text{1prong}}$
$N_{\gamma,\text{MVA,loose}}^{\text{3prong}}$	$N_{\gamma,\text{MVA,loose}}^{\text{1prong}}$	$E_{\gamma,\text{MVA,loose}}^{\text{3prong}}$	$E_{\gamma,\text{MVA,loose}}^{\text{1prong}}$

We further use neural-network-based PID variables for a binary separation between kaons and pions in the final state called kaonIDNN [29], which follow kaonIDNN + pionIDNN = 1. Specifically, we set a threshold of kaonIDNN < 0.6 for any of the 3-prong side tracks.

3.4.2 PID Corrections

We employ PID corrections to all simulated data to account for differences between the Monte Carlo simulation of the Belle II detector response and the real Belle II detector. These corrections were developed in [12].
The truth pion tracks are adjusted based on the efficiencies of the data $\epsilon_\pi^{\text{data}}$ and the efficiency of the MC data ϵ_π^{MC} as follows

$$w_\pi(p, \cos\theta) = \frac{\epsilon_\pi^{\text{data}}(p, \cos\theta)}{\epsilon_\pi^{\text{MC}}(p, \cos\theta)} \,. \tag{3.13}$$

For truth kaon tracks misidentified as pions, the correction weights are defined by the kaon-as-pion fake-rate ratio

$$w_K(p, \cos\theta) = \frac{F_K^{\text{data}}(p, \cos\theta)}{F_K^{\text{MC}}(p, \cos\theta)} \,, \tag{3.14}$$

where F_K denotes the probability that a kaon passes the pion PID selection. This adjustment is done for each of the three prongs, such that the event weight is

$$w_{\text{event}} = w_1\, w_2\, w_3 \,. \tag{3.15}$$

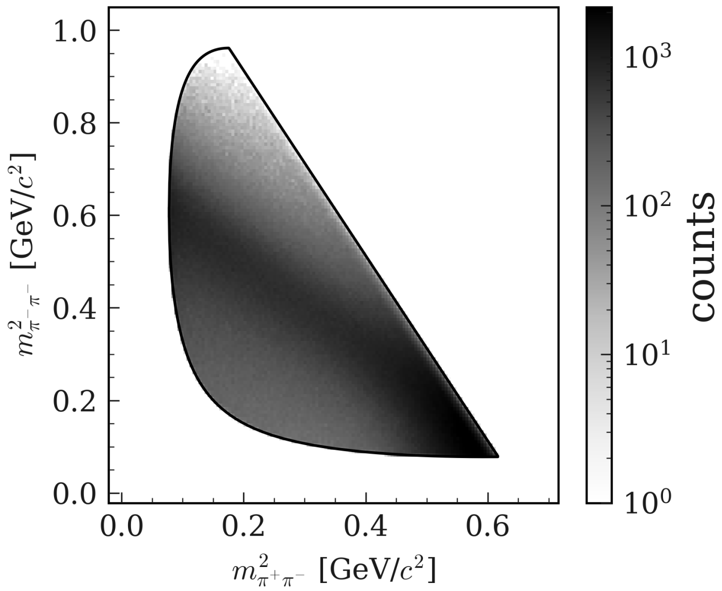


Figure 3.2: Phase-space coverage for $m_{3\pi} \in [1.10, 1.12]$ GeV. The kinematically allowed boundaries are shown in black.

The $(p, \cos(\theta))$ range of all three tracks is restricted to regions where the PID correction tables are valid by requiring the momentum to be within $0.1 < p < 5.0$ GeV/c. Additionally, the range corresponding to $(\cos\theta < -0.866 \ \wedge \ 3.5 < p < 4.5$ GeV/$c)$ is excluded.

3.4.3 Cuts on Phase Space

We cut the sample further on the kinematically allowed phase space using the 3 two-pion masses m^2_{12}, m^2_{13}, m^2_{23}. Two of them, m^2_{12} and m^2_{23} are cut based on a lower bound of the mass as $m^2_{\pi\pi} = (2m_\pi)^2 = (2 \times 0.13957)^2$ and an upper boundary, where the mass is constrained to be below $m^2_{\pi\pi} = (m_{3\pi} - m_\pi)^2$. The limits for the third squared two-pion mass, m^2_{13}, are calculated following the parametrization of ref. [3]. The maximum value for the same charged two-pion system is

$$m^2_{13(\max)} = (E_2 + E_3)^2 - \left(\sqrt{E_2^2 - m_2^2} - \sqrt{E_3^2 - m_3^2}\right)^2, \tag{3.16}$$

and the minimum value is

$$m^2_{13(\min)} = (E_2 + E_3)^2 - \left(\sqrt{E_2^2 - m_2^2} + \sqrt{E_3^2 - m_3^2}\right)^2. \tag{3.17}$$

The energies E_2 of particle 2 and E_3 of particle 3, are:

$$E_2 = \frac{m_{3\pi}^2 - m_1^2 + m_2^2}{2m_{12}}, \tag{3.18}$$

$$E_3 = \frac{m_{3\pi}^2 - m_{12}^2 - m_3^2}{2m_{12}}. \tag{3.19}$$

These cuts lead to a signal distribution in $m^2_{\pi^-\pi^-}$ and $m^2_{\pi^+\pi^-}$ as shown in Figure 3.2, with the phase-space borders shown in black.

3.5 Background Definitions

After the event selection, the final data sample contains 9.5% background contributions as estimated on simulated generic run-dependent MC15 data. The individual background contributions are listed in Table 3.3, and weighted by the luminosity to correspond to a dataset with an integrated luminosity of $\mathcal{L} = 361.654\,\mathrm{ab}^{-1}$.

Table 3.3: Number of expected events and corresponding fractions for each contribution. Fractions are computed with respect to the total expected number of events.

Contribution	Weighted Events	Fraction
Signal	17 028 687	0.905
$\tau \to 3\pi + \pi^0 + \nu_\tau$	1 109 758	0.059
$e^+e^- \to q\bar{q}$	308 627	0.016
other τ_{Bkg}	100 135	0.005
Rest	275 723	0.015
Total Background	1 794 243	0.095

The corresponding distributions in $m_{3\pi}$ of the $\tau^- \to \pi^+\pi^-\pi^-\nu_\tau$ signal (yellow) and total background (blue) events are shown in Figure 3.3, along with the four explicit categories of background contributions discussed in the following.
The dominant background contribution originates from the decay of $\tau^- \to \pi^+\pi^-\pi^-\pi^0\,\nu_\tau$ (green). Its relative contribution, however, depends strongly on the $m_{3\pi}$ region, as shown in Figure 3.3. While it is the leading background contribution for $m_{3\pi} \lesssim 1.35\mathrm{GeV}/c^2$, other background sources become more significant for larger $m_{3\pi}$ values.
This decay constitutes the dominant background contribution because the additional neutral pion is not always reconstructed correctly, i.e., the two photons from the decay of the neutral pion are not detected, so these events are wrongly reconstructed as signal and pass the event selection.
Similarly to the signal decay, the $\tau^- \to \pi^+\pi^-\pi^-\pi^0\,\nu_\tau$ decay proceeds via the intermediate resonance $a_1(1260)$ in the process of $\tau \to a_1\pi^0\nu_\tau$, where the a_1 subsequently decays to the three pion final state. Additionally, a large contribution is expected from the $\omega\pi$ intermediate state [30]. Thus, this background distribution exhibits resonance structures similar to the signal, making the two processes hard to distinguish. The distribution in

the three invariant masses (m_{12}, m_{23}, m_{13}) and the three angular variables ($\cos(\beta)$, γ, $\cos(\theta)$) for the $\tau^- \to \pi^+\pi^-\pi^-\pi^0\,\nu_\tau$ background contribution (green) are shown in Figure 3.4. They exhibit a ρ peak in the m_{13} mass spectrum at $\approx 770\,\mathrm{GeV}/c^2$, as well as a low-mass shoulder at $\approx 650\,\mathrm{GeV}/c^2$. Additionally, a comparison between signal and $\tau^- \to \pi^+\pi^-\pi^-\pi^0\,\nu_\tau$ data is given for the three invariant masses and three angles in Figure 3.5, showing some similarity in the structure of the variables making the two processes difficult to distinguish using these kinematic variables.
As shown in 3.3 above $m_{3\pi} \approx 1.35\,\mathrm{GeV}/c^2$ the background process of $e^+e^- \to q\bar{q}$ (purple), originating from continuum production of $u\bar{u}$, $d\bar{d}$, $s\bar{s}$, and $c\bar{c}$ pairs, which subsequently hadronize, becomes the most dominant contribution. As shown in Figure 3.4, the invariant two-pion masses do not exhibit dominant peaking structures, but we observe an increase in m_{13} roughly around the mass of the ρ resonance and a slight enhancement at the K^0 mass in m_{12}. Both of these mesons can be produced via hadronization of continuum $q\bar{q}$ pairs.
The background labeled as τ_{Bkg} (red) dominantly incorporates the decay of $\tau^- \to K^0\pi^-\nu_\tau$ where K^0 subsequently decays to $\pi^+\pi^-$. This background exhibits a dominant peaking structure in m_{12} and m_{13} at the K^0 mass, as shown in Figure 3.4, which corresponds to the decay of $\tau^- \to K^0\pi^-\nu_\tau$, which subsequently decays to $\pi^+\pi^-$.
The last category, summarized as "rest" (gray) are background contribution for $\tau^- \to K^-\pi^-\pi^+\nu_\tau$ and $\tau^- \to \pi^-\pi^0\nu_\tau$, as well as all remaining contributions. This contribution shows an increase in events in the low m_{12} region. This signature corresponds to events where a photon γ converts to an e^+e^- pair, which is wrongly reconstructed as $\pi^+\pi^-$. Thus, by combining these e^+e^- with a charged pion, e.g., from the decay of $\tau \to \pi^-\pi^0\nu_\tau$, a total of three charged tracks are observed. Since the e^+e^- is produced from a photon, the invariant mass m_{12} corresponds to 0. These tracks, however, are incorrectly reconstructed as pions, leading to the peak at $m_{12} \approx 2m_\pi$. The $\tau \to \pi^-\pi^0\nu_\tau$ contributions also exhibit the peaking structure at $\cos(\beta) \approx 0$, which means the CMS direction in the hadronic restframe lies approximately in the decay plane (see Equation (3.6)) for these events, indicating that this background has a characteristic phase-space distribution.
Additional minor background contributions originate from low-multiplicity processes, such as $e^+e^- \to \mu^-\mu^-$ or $e^+e^- \to e^+e^-X$, which are produced in two-photon scattering, and are also summarized in the rest.
The background events in this sample are weighted to account for the different luminosities at which they were generated. The signal events are generated using the TAUOLA generator [31], i.e., according to an implemented physics model. Since we want to be independent of this physics model, we introduce TAUOLA weights to deweight the signal events, effectively removing the generated signal-model dependence and leaving a distribution that follows only phase space and detector acceptance.

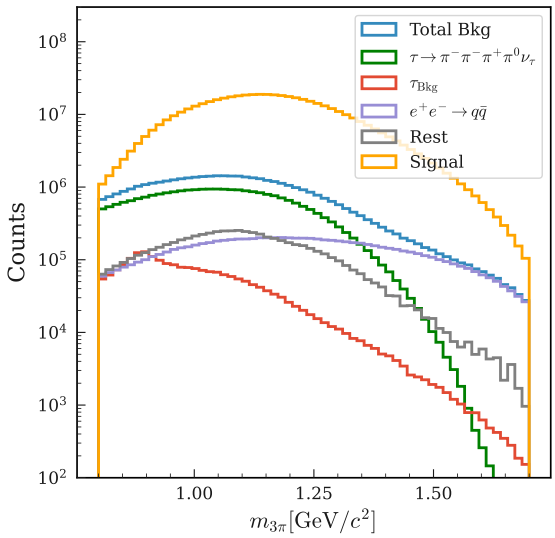


Figure 3.3: Signal and background contributions of the simulated MC15 data sample in $m_{3\pi}$.

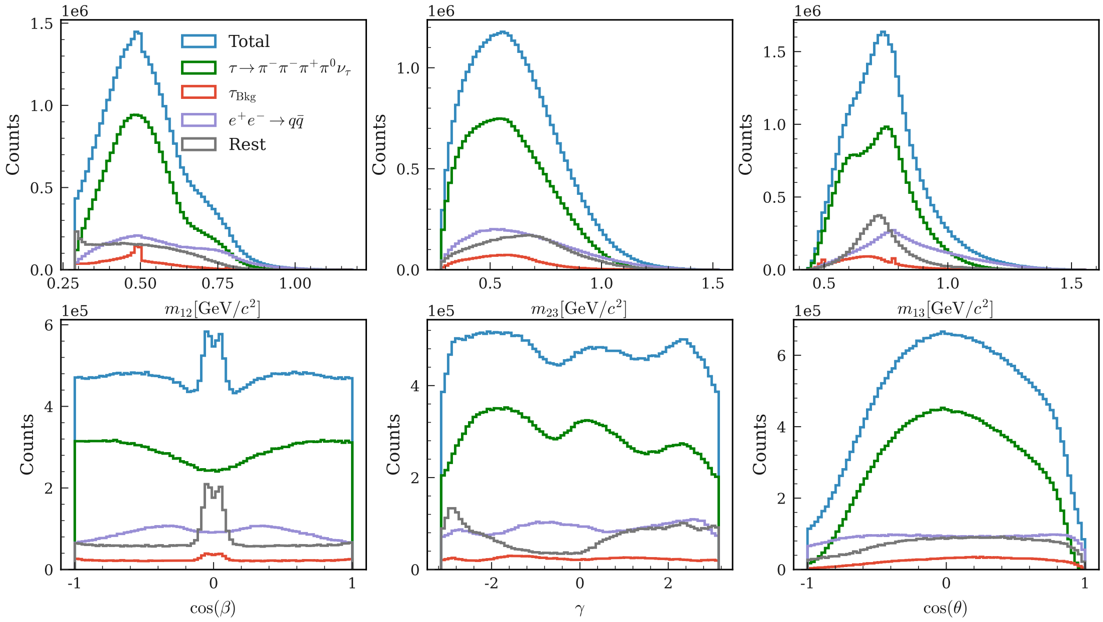


Figure 3.4: Kinematic distributions of the background contributions in the MC data sample in the three invariant masses (m_{12}, m_{23}, m_{13}) and three angular variables ($\cos(\beta)$, γ, $\cos(\theta)$).

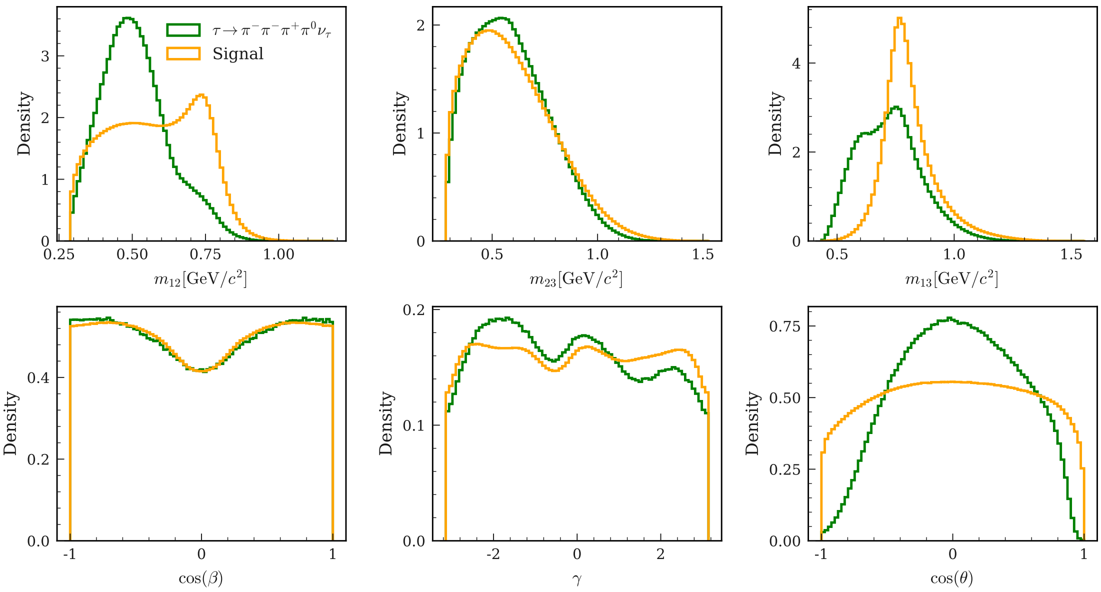


Figure 3.5: Kinematic distributions of $\tau^- \rightarrow \pi^+\pi^-\pi^-\nu_\tau$ and $\tau^- \rightarrow \pi^+\pi^-\pi^-\pi^0\nu_\tau$ decay in the three invariant masses (m_{12}, m_{23}, m_{13}) and three angular variables ($\cos(\beta)$, γ, $\cos(\theta)$).

4 Partial-Wave Analysis

The goal of partial wave analysis is to identify and separate intermediate meson resonances in the decay to determine their individual contributions to the decay and define their quantum numbers J^P. This is done in two steps: first, a partial-wave decomposition, followed by a resonance-model fit. In the first step, the input data is decomposed into contributions from different wave amplitudes representing certain intermediate resonances and decay paths. This is done using the isobar model, which is described in section 4.1 of this chapter. In order to be independent of the dependence of the partial wave amplitudes in the three-pion mass $m_{3\pi}$, i.e., not model it explicitly, we bin $m_{3\pi}$ into 45 bins with a bin-width of $0.02\,\mathrm{GeV/c^2}$ in the mass range from $m_{3\pi} = 0.8\,\mathrm{GeV/c^2}$ to $1.7\,\mathrm{GeV/c^2}$. The second step is a resonance model fit, which determines the mass and width of the resonances and which is based on the partial-wave decomposition. This thesis will focus on the first step. We discuss the description of the decay in Section 4.1, followed by detailing the partial-wave decomposition formalism in Section 4.2. We then discuss the Partial-wave decomposition fits in Section 4.3 and define the method to test the validity of the PWD results, along with the data and input model used, in section 4.4.

4.1 Isobar Model

The τ pairs produced in the e^+e^- collision at Belle II can subsequently decay to $\tau^- \to \pi^-\,\pi^-\,\pi^+\,\nu_\tau$. This decay is dominantly not immediate, but described by a sequence of two-body decays illustrated in Figure 4.1, (i) $\tau^- \to X^-\nu_\tau$, (ii) $X^- \to \pi^-\xi^0$, (iii) $\xi^0 \to \pi^+\pi^-$.

Initially, the τ^- decays to an intermediate resonance X, such as $a_1(1260)$, as well as a tau-neutrino (i). This is followed by the decay of the intermediate resonance into an isobar ξ^0 resonance, such as $\rho(770)$, and a bachelor pion (ii). Finally, the isobar decays into the final two pions (iii). This description of the decay as a sequence of two-body decays constitutes the isobar model.

4.2 Partial-Wave-Decomposition Formalism

In order to study the decay of $\tau^- \to \pi^-\,\pi^-\,\pi^+\,\nu_\tau$, which is cascading via intermediate resonances, the data is decomposed into contributions from individual so-called partial waves. A partial wave describes the decay of the τ lepton into an intermediate state with definite spin J and parity P, which subsequently decays into a specific isobar and a bachelor pion. Thus, each partial wave is fully defined by the quantum numbers J^P of

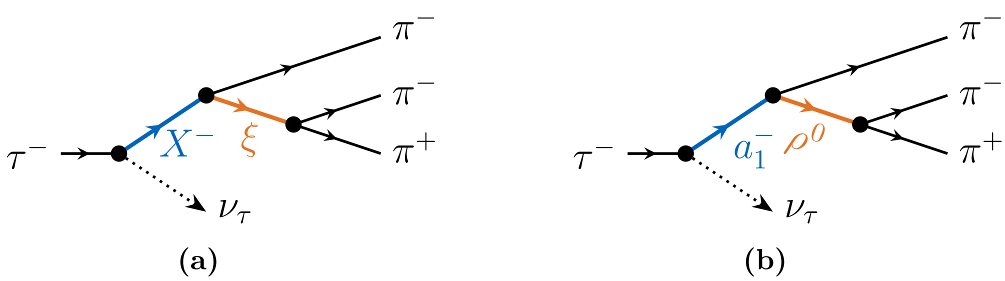


Figure 4.1: Isobar model representations of the decay $\tau^- \to \pi^-\pi^-\pi^+\nu_\tau$.

the intermediate state, the isobar resonance, and the relative orbital angular momentum L between the isobar resonance and the bachelor pion. These components together define the partial wave label as $J^P\xi^0\pi L$. The biggest contribution of this decay is expected from the intermediate resonance $a_1(1260)$ decaying to the isobar $\rho(770)$ with an S wave between the decay products, i.e. $L = 0$. This corresponds to a partial wave label of $1^+\rho(770)\pi S$. In addition, several smaller partial waves are expected to contribute, as observed in previous analyses of the $\tau^- \to \pi^-, \pi^-, \pi^+, \nu_\tau$ decay by Belle [11].

4.2.1 Decay Rate

The differential decay rate of $\tau^- \to \pi^+\,\pi^-\,\pi^-\,\nu_\tau$, is defined as

$$\frac{d\Gamma}{d\Phi_4} = |\mathcal{M}|^2, \tag{4.1}$$

where $\mathcal{M}$ is the matrix element encoding the dynamics of the decay. Following the parametrization of references [6, 25], the matrix element of this semi-leptonic decay is defined by the lepton current ℓ_μ and the hadron current J^μ, as well as the prefactor consisting of the Fermi constant G_F and the relevant CKM-Matrix element V_{ud} as

$$\mathcal{M} = \frac{G_F}{\sqrt{2}}\, V_{ud}\, \ell_\mu\, J^\mu\,. \tag{4.2}$$

The leptonic current coupling to the W^- Boson is a Vector - Axial (V-A) interaction of the form

$$\ell^\mu_\nu = \bar{u}_\nu \gamma^\mu (1 - \gamma^5) u_\tau\,, \tag{4.3}$$

using Dirac spinors u_τ and u_ν, as well as the Dirac matrices γ. The hadron current, discussed in equation 4.2, is governed by the strong interaction. Thus, we split the hadron current into individual hadronic currents j^μ_a, which represent the calculable kinematic and angular structure, and complex coefficient $\tilde{c}_a$, which defines the relative strengths and phases of an individual wave a that are not known a priori. The decomposed hadron current is:

$$J^\mu = \sum_a \tilde{c}_a\, j^\mu_a\,. \tag{4.4}$$

The individual hadron currents are described by the dynamic amplitude $B_X(m_X^2)$ of the intermediate resonance and the dynamic amplitude $B_\xi(m_\xi^2)$ of the isobar resonance, which can not be calculated from first principles and have to be parametrized. Along with the real-valued tensor structure t_a^μ corresponding to the angular dependence of the decay, the individual hadron currents are defined as

$$j_a^\mu = B_X(m_X^2)B_\xi(m_\xi^2)\, t_a^\mu \,. \tag{4.5}$$

The decay rate (equation 4.1) is proportional to the squared matrix element, which is ultimately defined by the lepton tensor $L_{\mu\nu}$ and the hadron tensor $H_{ab}^{\mu\nu} = j_a^\mu j_b^{\nu *}$, as follows

$$\begin{aligned} |\mathcal{M}'|^2 &= \sum_{a,b} \tilde{c}_a\, \tilde{c}_b^*\, (j_a^\mu\, \ell_\mu)\, (j_b^\nu\, \ell_\nu)^* \\ &= \sum_{a,b} \tilde{c}_a\, \tilde{c}_b^*\, (\ell_\mu\, \ell_\nu^*)\, (j_a^\mu\, j_b^{\nu *}) \\ &= \sum_{a,b} \tilde{c}_a\, \tilde{c}_b^*\, L_{\mu\nu}\, H_{ab}^{\mu\nu} \,. \end{aligned} \tag{4.6}$$

Generally, we need to consider the contribution of different τ helicities in the matrix element. However, the Belle II e^+e^- beam is unpolarized [32], meaning the τ leptons do not have a preferred helicity. Therefore, the squared matrix element is averaged over the τ helicities (λ_τ)

$$|\mathcal{M}|^2 = \frac{1}{2} \sum_{\lambda_\tau,\lambda_\nu} |\mathcal{M}'|^2 \tag{4.7}$$

Furthermore, as we cannot calculate the Euler angle α (see Section 3.1), we marginalize the matrix element $\overline{|\mathcal{M}|^2}$ over this unknown angle.
Only the lepton current depends on the Euler angle. We marginalize the lepton tensor $\overline{L}$:

$$\overline{L}_{\mu\nu} = \frac{1}{2\pi} \int d\alpha\, L_{\mu\nu} \tag{4.8}$$

To minimize the computational effort required for the calculation of the matrix element, we perform an eigenvalue decomposition of the lepton tensor. Using the eigenvalues λ_i and eigenvectors $\tilde{v}_\mu^i$ $(\tilde{v}_\nu^i)$ we construct v_μ^i (v_ν^i) as

$$v_\mu^i = \sqrt{\lambda_i}\, \tilde{v}_\mu^i \,. \tag{4.9}$$

The lepton tensor in equation 4.8 is therefore parametrized as:

$$\overline{L}_{\mu\nu} = \sum_i^4 v_\mu^i v_\nu^{i\,*} \,. \tag{4.10}$$

This leads to the final formulation of the matrix element, averaged over the helicities and marginalized over the angle α, using the parametrization of equation 4.10 for the lepton tensor as:

$$\begin{aligned}
\overline{|\mathcal{M}|^2} &= \frac{1}{2}\sum_{\lambda_\tau,\lambda_\nu}\sum_{a,b} c_a\, c_b^*\, \overline{L}_{\mu\nu}\, H_{ab}^{\mu\nu} \\
&= \sum_{a,b} c_a c_b^* \sum_{i}^{4} v_\mu^i\, v_\nu^{i\,*}\, j_a^\mu\, j_b^{\nu\,*} \\
&= \sum_{a,b}\sum_{i}^{4} c_a c_b^*\, \tilde{\psi}_a^i\, \tilde{\psi}_b^{i*}\,,
\end{aligned} \tag{4.11}$$

with the complex coefficients c_a and c_b^* as well as the decay amplitudes, denoted as:

$$\psi'^i_a = v_\mu^i\, j_a^\mu\,. \tag{4.12}$$

4.2.2 Symmetrization of Decay Amplitudes

The final state of $\tau^- \to \pi_1^+\pi_2^-\pi_3^-\nu_\tau$ contains two identical π^- mesons that are not differentiable. Thus, the decay amplitude ψ'^i_a in Equation (4.12) is symmetrized such that it obeys Bose symmetry, i.e., it is constrained to be symmetric under the exchange of the two pions as a coherent sum of two parts [6] as

$$\psi_a^i = \psi'^i_{a,(\pi_1,\pi_2,\pi_3)} + \psi'^i_{a,(\pi_1,\pi_3,\pi_2)}\,. \tag{4.13}$$

We further normalize this decay amplitudes ψ_a^i such that

$$\sum_i \int \left|\psi_a^i(\tau)\right|^2 \Phi(\tau)\mathrm{d}\tau = 1\,. \tag{4.14}$$

4.2.3 Observables

The distribution of produced events $\hat{N}$ predicted by our model in the phase space of the decay $\Phi(\tau)$ is proportional to the decay rate defined in equation 4.1. The proportionality factor is absorbed by the complex couplings, and we correlate

$$I(\tau) = \frac{d\hat{N}(\tau)}{d\Phi(\tau)} = \sum_i\sum_{a,b} c_a^i\, c_b^{i*}\, \psi_a^i(\tau)\, \psi_b^i(\tau)^*\,, \tag{4.15}$$

using the partial waves ψ_a^i, ψ_b^i defined in Equation (4.13) and the complex couplings c_a. The latter encodes the strength and the relative phase of the intermediate resonance. This is also called the modeled intensity $I(\tau)$.
We introduce the detector acceptance $\eta(\tau)$ to describe the distribution of measured events $\hat{\hat{N}}$ as predicted by our model, corresponding to[1]:

$$\frac{d\hat{\hat{N}}(\tau)}{d\Phi(\tau)} = \eta(\tau)I(\tau) \tag{4.16}$$

[1]Within this formalism, we neglect detector resolution

The model prediction of the total number of measured events $\hat{\hat{N}}$ is obtained by integrating the intensity in equation 4.16 over the phase-space density, i.e.,

$$\begin{aligned}\hat{\hat{N}} &= \int d\Phi(\tau)\,\eta(\tau)\,I(\tau)\\ &= \int d\Phi(\tau)\,\eta(\tau)\,\sum_i\sum_{a,b} c_a^i\,c_b^{i*}\,\psi_a^i(\tau)\,\psi_b^i(\tau)^*\\ &= \sum_{a,b} c_a\,\bar{I}_{ab}\,c_b^*.\end{aligned} \tag{4.17}$$

In the last step, the accepted integral matrix $\bar{I}_{ab}$ is introduced. Since we do not have an analytical form for the acceptance, and the computational effort to evaluate this integral is extensive, we approximate it using Monte Carlo integration. As defined in Chapter 3, the MC sample was generated with additional TAUOLA weights $w_g(\tau)$, introduced in Chapter 3, meaning we sample from a signal distribution that follows

$$\mathrm{PDF}(\tau) = \frac{\Phi(\tau)\,\eta(\tau)\,\omega(\tau)}{H}, \tag{4.18}$$

The normalization H of this signal PDF is evaluated in Appendix A using harmonic mean integration as

$$H = \int d\tau\,\Phi(\tau)\,\eta(\tau)\,\omega(\tau) \;= \bar{N}\,\frac{V_{\mathrm{PS}}}{\sum_e^N \frac{1}{\omega(\tau_e)}}. \tag{4.19}$$

The integral matrix is approximated using importance sampling [33, p. 80-83], effectively approximating the integral as a sum over uniformly drawn events from Equation (4.18). To account for the physics model in the Monte Carlo integration, we introduce $1 = \frac{H}{\omega(\tau)}\frac{\omega(\tau)}{H}$ in the calculation of the integral matrix and define

$$\bar{I}_{ab} = \int d\tau\,\frac{H}{\omega(\tau)}\frac{\omega(\tau)}{H}\Phi(\tau)\,\eta(\tau)\,\psi_a(\tau)\,\psi_b^*(\tau)\,, \tag{4.20}$$

and approximate the integral as

$$\bar{I}_{ab} \approx \frac{V_{PS}}{\sum_i^N \frac{1}{w_g(\tau_i)}}\sum_i^{\bar{N}}\frac{1}{w_g(\tau_i)}\,\psi_a(\tau)\,\psi_b^*(\tau)\,. \tag{4.21}$$

4.3 Partial-Wave Decomposition Fits

The partial-wave decomposition fits are done in 45 narrow $m_{3\pi}$ bins, since we want to measure the $m_{3\pi}$ dependence of the partial-wave amplitudes. These fits are done independently in each $m_{3\pi}$ bin using an unbinned extended maximum-likelihood fit with

the complex couplings c_a as free parameters.
We define the probability density function of an event e within a $m_{3\pi}$ bin as

$$\begin{aligned} P(\tau_e, m_{3\pi}) &= \frac{\Phi(\tau_e, m_{3\pi})\eta(\tau_e, m_{3\pi})\, I(\tau_e, m_{3\pi})}{\int d\Phi(\tau_e, m_{3\pi})\, \eta(\tau_e, m_{3\pi})\, I(\tau_e, m_{3\pi})} \\ &= \frac{\Phi(\tau_e, m_{3\pi})\eta(\tau_e, m_{3\pi})\, I(\tau_e, m_{3\pi})}{\hat{\bar{N}}}\,. \end{aligned} \tag{4.22}$$

The normalization corresponds to the expected number of total measured events, defined in equation 4.17.
The model intensity, defined in Equation (4.15), appears in both the numerator and the denominator of the PDF in Equation (4.22). Thus, a scaling factor of the complex couplings of the intensity does not impact the probability. To fix this scaling, we use an extended maximum likelihood (ML) approach.
The extended ML differs from the ML in that not only the shape of the distribution is fit, but also its size [34]. The total number of measured events is introduced by a Poisson term, corresponding to the Poisson probability to measure $\bar{N}$ given we expect to measure $\hat{\bar{N}}$ events, defined in 4.17. Thus, the extended ML is defined as:

$$\mathcal{L}_{\text{PWD}} = \frac{\hat{\bar{N}}^{\bar{N}} e^{-\hat{\bar{N}}}}{\bar{N}!} \prod_{e}^{\bar{N}} P(\tau_e, m_{3\pi}). \tag{4.23}$$

The complex couplings of the intensity are determined by minimizing the negative logarithm of this extended maximum likelihood with respect to these couplings, defined as [2]

$$\begin{aligned} -\log(\mathcal{L}_{\text{PWD}}) = &-\sum_{e}^{\bar{N}} \log\left(\sum_{i}\sum_{a,b} c_a^i c_b^{i*}\, \psi_a^i(\tau_e)\, \psi_b^{i*}(\tau_e)\right) \\ &- \sum_{e}^{\bar{N}} \log(\eta(\tau_e)) - \sum_{e}^{\bar{N}} \log(\Phi(\tau_e)) - \bar{N}\log(V_{PS}) \\ &- \bar{N}\log(\hat{\bar{N}}) + \log(\bar{N}!) + \hat{\bar{N}} + \bar{N}\log(\hat{\bar{N}}) \end{aligned} \tag{4.24}$$

Some terms in the log-likelihood do not depend on the couplings, which are the free parameters in the fit. Therefore, these terms do not change the location of the fit's minimum, and they can be disregarded during minimization. This leaves us with a simplified likelihood ($\mathcal{L}'$) used in the PWD fit, defined as

$$-\log({\mathcal{L}'}_{\text{PWD}}) = -\sum_{e}^{\bar{N}} \log\left(\sum_{i}\sum_{a,b} c_a^i c_b^{i*}\, \psi_a^i(\tau_e)\, \psi_b^{i*}(\tau_e)\right) + \hat{\bar{N}}. \tag{4.25}$$

The complex couplings always appear in pairs of $c_a\, c_b*$. Consequently, the model is ambiguous under a global phase rotation $c_a \to e^{i\varphi} c_a$. Therefore, the phase of one partial wave, called the reference wave, is fixed to be real-valued and positive, and the phases

[2] For readability, we drop the $m_{3\pi}$ dependency

of all other waves are measured with respect to the reference wave. We use the biggest partial wave, the $1^{+}[\rho(770)\pi]_S$ wave as the reference wave.
The start parameters of the couplings for the minimization are calculated from uniformly drawn coupling intensities and phases. For the non-reference waves, the intensity is drawn in the range of $0 - 10\%$ of the total intensity and the phases in the range of $0 - 2\pi$. We expect the biggest contribution from the reference wave, thus the intensity is drawn in the range of $0.9 - 1.1\%$ of the total intensity.

4.4 Input - Output Studies of Signal Events

We test the consistency of the PWD method by performing input-output studies. To this end, we use run-independent signal MC15 data, which incorporates the Belle II detector simulation, is passed through the event reconstruction, and we apply the event selection criteria. Therefore, we obtain a realistic comparison to real data. This MC sample is redistributed according to a predefined input model by using an accept-reject procedure [35], thus we generate a sample that incorporates resonant contributions with a known strength. This sample is referred to as reconstructed signal pseudo data, which incorporates all detector effects, i.e., acceptance and resolution.
We can also use the MC-truth-level information to obtain a dataset that excludes resolution effects, which we refer to as the MC-truth signal pseudo data.
In Chapter 7 and Chapter 8, we perform PWD fits to pseudo data and extract the complex partial-wave couplings in each of the 45 $m_{3\pi}$ bins. These resulting amplitudes are compared to the expectation from the input model, i.e., the model curve, to assess how well the results match the predefined input.
The model curve for each partial wave is calculated from the couplings of the intermediate resonances defined in the input model and the decay amplitudes of each $m_{3\pi}$ bin. The total number of events of the predicted model is fixed by setting the total number of events of the model equal to the total number of events in the fitted pseudo data.
In order to obtain a smooth model curve, the model predictions in each $m_{3\pi}$ bin are interpolated to a finer mass grid of 300 points.
Additionally, to quantify the agreement between the fit results f_i and the corresponding model expectation M_i, we calculate the pulls [36]. The pull in each $m_{3\pi}$ mass bin i is defined as the difference between the fit result f_i and the model prediction M_i, divided by the uncertainty σ_i of the fit result:

$$\text{pull}_i = \frac{f_i - M_i}{\sigma_i} \tag{4.26}$$

4.4.1 Input Model

The input model used in this thesis contains 9 partial waves, listed in Table 4.1, along with the corresponding intermediate resonances and coupling values used for the generation of the pseudo data. For the data generation, the dynamical amplitudes of the isobar

resonances ξ and intermediate resonances X are modeled using different parametrizations, which are detailed in the following subsections. The corresponding resonance parameters are summarized in Table 4.2.
In the PWD, the fit is performed independently in bins of $m_{3\pi}$ and no explicit parametrization of the intermediate 3π system is imposed. However, a parametrization of the isobar resonances is still needed, using the same dynamical amplitudes as in the generation model.

Table 4.1: Partial waves included in the input model and their contributing intermediate resonances with corresponding complex couplings, given in polar form.

Partial wave	**Intermediate resonance X**	**Coupling**
$1^+[\rho(770)\pi]S$	$a_1(1260)$	$7389\,e^{i\,0°}$
$1^+[\sigma\pi]P$	$a_1(1260)$	$2700\,e^{i\,121°}$
	$a_1(1640)$	$270\,e^{i\,91°}$
$1^+[\rho(770)\pi]D$	$a_1(1260)$	$507.6\,e^{i\,12°}$
	$a_1(1640)$	$907.2\,e^{i\,-168°}$
$1^+[\rho(1450)\pi]D$	$a_1(1640)$	$451.3\,e^{i\,154°}$
$1^+[f_2(1270)\pi]P$	$a_1(1260)$	$613.4\,e^{i\,-86°}$
$1^+[f_0(980)\pi]P$	$a_1(1420)$	$390\,e^{i\,-154°}$
$0^-[\rho(770)\pi]P$	$\pi(1300)$	$405\,e^{i\,-64°}$
$0^-[f_0(980)\pi]S$	$\pi(1300)$	$81\,e^{i\,-28°}$
$1^-[\omega(782)\pi]P$	$\rho(1450)$	$351.9\,e^{i\,145°}$

Relativistic Breit-Wigner

Most isobars are modeled using the relativistic Breit-Wigner amplitude of the form [3]:

$$A_{\mathrm{BW}}(m) = \frac{m_0\,\Gamma_0}{m_0^2 - m^2 - i\,m_0\,\Gamma(m)} \tag{4.27}$$

with the nominal mass m_0, the nominal width Γ_0 of the resonance and the mass-dependent width $\Gamma(m)$.
In the simplest case of the isobar $\omega(782)$, the width is approximated to be constant, i.e., $\Gamma(m) = \Gamma_0$. This may be used in the limit where the difference between the resonance mass and the threshold is large compared to the width of the resonance.

Table 4.2: Intermediate and isobar resonances used in the amplitude model together with the corresponding masses and widths.

Category	**Resonance**	**Mass** [GeV/c^2]	**Width** [GeV]
Three-body	$a_1(1260)$	1.23	0.425
	$\pi(1300)$	1.30	0.400
	$a_1(1420)$	1.41	0.150
	$a_1(1640)$	1.66	0.250
Isobar	$\rho(770)$	0.77526	0.1474
	$\rho(1450)$	1.465	0.400
	$f_0(980)$	0.99	0.055
	$f_2(1270)$	1.275	0.185
	$\omega(782)$	0.78266	0.00868

To incorporate the opening of the phase-space for different channels i, we use a mass-dependent width for the description of $\rho(1450)$ and $f_2(1270)$ of the form [3]:

$$\Gamma(m) = \sum_i \Gamma_i(m) = \sum_i \Gamma_{0,i} \left(\frac{q_i(m)}{q_i(m_0)}\right)^{2L} \frac{m_0}{m} \frac{F^2_{L,i}(q(m))}{F^2_{L,i}(q_0)}\,, \tag{4.28}$$

with the two-body break-up momentum q_i and the barrier damping factor $F_{L,i}$.
For generating our pseudo-data, all intermediate resonances are modeled by relativistic Breit-Wigner functions, defined in Equation (4.27).

Flatté Amplitude

The Breit-Wigner cannot be used to parametrize the isobar $f_0(980)$, since its mass is close to the $K\bar{K}$ threshold. Thus, we use the Flattè parametrization [37], using the parameters and formula defined in [38], which defines the amplitude as

$$A_{\text{Flatté}}(m) = \frac{1}{m_0^2 - m^2 - i\,(g_1\rho_{\pi\pi}(m) + g_2\,\rho_{K\bar{K}}(m))}\,, \tag{4.29}$$

with the Lorentz invariant two-body phase space factors $\rho(m)$ for the channels $\pi\pi$ and $K\bar{K}$ and the constant couplings g_1 and g_2 and the nomainal mass m_0

Gounari Sakurai Amplitude

Due to the large width of the $\rho(770)$ isobar, the lineshape is not well modeled by a relativistic Breit-Wigner function. Instead, we use a Gounari Sakurai amplitude, which was originally formulated in ref. [39]. We use the parametrization defined in ref. [40]:

$$A_{\text{GS}}(s) = \frac{m_0\Gamma_0}{m_0^2 - m^2 + f(m) - i\,m_0\Gamma(m)}\,, \tag{4.30}$$

with the mass-dependent width $\Gamma(m)$ from Equation (4.28) and the so-called f function, corresponding to the real part of the self-energy of the Gounari Sakurai amplitude, defined in equation 18 of ref. [40]. However, contrary to ref. [40], we change the normalization such that $A(m_0^2) = 1$ instead of $A(0) = 1$.

Au–Morgen–Pennington–Kachaev (AMPK) Parametrization

The $\pi\pi$ S-wave is parametrized using a modified version of the Au–Morgan–Pennington amplitude [41], as implemented by Kachaev [42]. In this modification, the coupling between the $\pi\pi$ and $K\bar{K}$ channels is removed and the contribution of the $f_0(980)$ pole is excluded.
The amplitude is constructed from the $\pi\pi \to \pi\pi$ element of the T matrix defined in Equation (13) of [42], i.e. the amplitude is defined as

$$A_{\pi\pi}(m) = T_{11}(m)\,. \tag{4.31}$$

5 Thresholding

As discussed in Section 4.4, we perform input-output studies on reconstructed signal MC data using the model defined in 4.4.1. Two representative examples of this PWD fit (blue), the $1^+ \, [\rho(770)\,\pi]_S$ and $1^+ \, [\rho(1450)\,\pi]_D$ waves, are shown in Figure 5.1 and compared to the model curve (red). Both waves show a bias introduces for $m_{3\pi} \lesssim 1.0\,\mathrm{GeV/c^2}$, the $1^+ \, [\rho(770)\,\pi]_S$ wave and the $1^+ \, [\rho(1450)\,\pi]_D$ wave is largely overestimated for a decreasing $m_{3\pi}$ with large uncertainties. We observe such deviations in the low $m_{3\pi}$ spectrum in almost all partial waves. They start for mass bins below $m_{3\pi} \approx 1.16\mathrm{GeV}/c^2$. The exact mass at which they occur varies between partial waves.

These deviations from the expected physical results can be understood by considering the overlap between the partial waves included in the model, i.e., how similar their kinematic distributions are. In general, partial waves with different quantum numbers (J^P) exhibit only small overlaps, whereas partial waves sharing the same quantum numbers can strongly overlap. Consequently, in the model considered here, significant overlaps are expected predominantly within the axial-vector ($J^P = 1^+$) and pseudoscalar ($J^P = 0^-$) currents.

These overlaps become particularly large for partial waves in which the nominal mass of the isobar lies well below the given $m_{3\pi}$. In that case, only the low-mass tail of the isobar lineshape is kinematically accessible, with the dominant contribution coming from the phase space [43]. Therefore, in the low $m_{3\pi}$ regions, two partial waves, for example, $1^+ \, [\rho(770)\,\pi]_D$ and $1^+ \, [\rho(1450)\,\pi]_D$ representing the same quantum numbers are almost indistinguishable.

In the $\pi^+\pi^-\pi^-$ final state, an additional challenge is introduced by the Bose symmetrization, defined in section 4.2.2. In this description, the amplitude is indifferent to the exchange of the same-charged pions, such that different assignments of pions to the isobar and the bachelor pion can not be distinguished. As a result, two partial waves that differ only in this assignment of pions are challenging to differentiate in the PWD fit. An example of this are the $1^+ \, [f_0(980)\,\pi]_P$ and $1^+ \, [\rho(770)\,\pi]_S$ waves.

To obtain unbiased results in the PWD fit, we have to exclude individual partial waves in $m_{3\pi}$ regions where we expect low intensity to resolve this ambiguity. This simplification is justified for partial waves in mass regions below the isobar resonance's nominal mass. Consequently, we define an individual waveset for each $m_{3\pi}$ bin by introducing lower thresholds on the $m_{3\pi}$ for each wave below which these waves are excluded from the PWD fit.

This is done by examining the integral matrix defined in Equation (4.20), which includes detector effects. The integral matrix can be viewed as the Gram matrix of the decay amplitudes, because it contains the outer products of all partial-wave decay amplitudes. As such, it encodes the degree of overlap between partial waves. We analyze this overlap

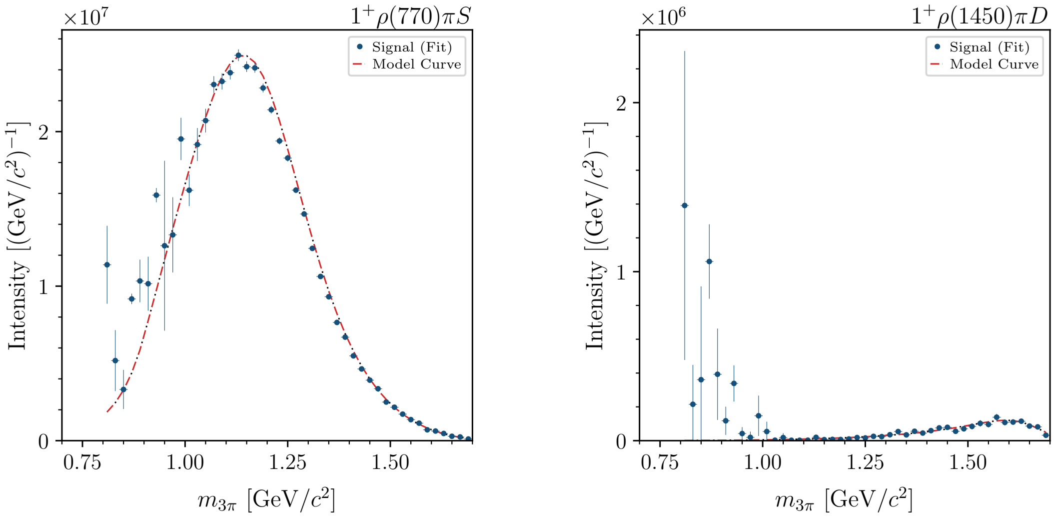


Figure 5.1: PWD fit results performed on reconstructed signal MC data using a non-threshold wave-set (blue) for the $1^+ [\rho(770)\,\pi]_D$ (left) and $1^+ [\rho(1450)\,\pi]_D$ (right) wave compared to the expected output of the model (model curve).

by performing an eigenvalue decomposition of the integral matrix, such that

$$I_{ab} = \sum_i \tilde{\lambda}_i \, v_{a,i} v^*_{b,i} \,, \tag{5.1}$$

where λ_i and $v_{a,i}$ denote the eigenvalues and eigenvectors of the integral matrix, respectively. Using this parametrization of the integral matrix, the total number of expected reconstructed events $\hat{\tilde{N}}$ can be expressed in the eigenvalue basis as

$$\hat{\tilde{N}} = c_a \, I_{ab} \, c_b = \sum_i \tilde{\lambda}_i \sum_{a,b} c_a v_{a,i} \, (c_b v^i_b)^* \,. \tag{5.2}$$

In this representation, each eigenvector entry corresponds to a partial wave, while the associated eigenvalue determines how strongly the eigenvector contributes to the $\hat{\tilde{N}}$.
Eigenvectors with large eigenvalues represent directions in parameter space that are well constrained by the data. In contrast, small eigenvalues indicate weakly constrained directions, where even large partial wave contributions have little impact on the description of the data. Large entries in the eigenvector of vanishing eigenvalues indicate partial waves that can not be distinguished in the PWD fit. In the fit procedure, larger couplings are favored, since we minimize the negative likelihood in section 4.3, unless properly constrained by $\hat{\tilde{N}}$, which explains the overestimation in the lower-mass bins in figure 5.1.
To summarize, we look for small eigenvalues together with large contributions from the partial wave in the corresponding eigenvectors, since they indicate weakly constrained directions in parameter space and lead to biased results in the PWD fit.

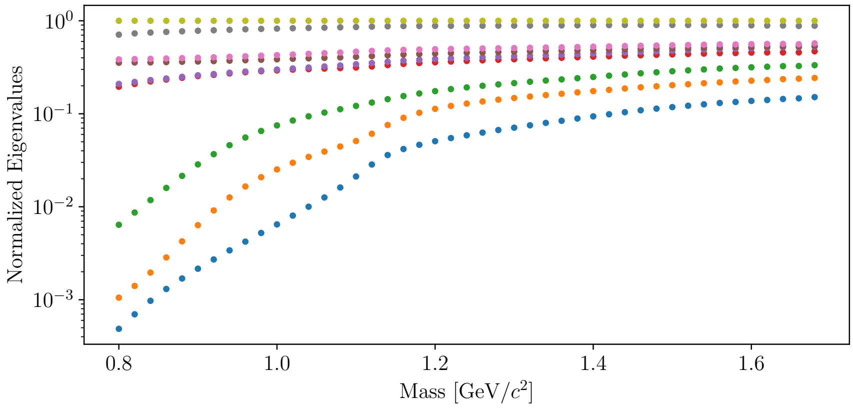


Figure 5.2: The 9 eigenvalues of the eigenvalue decomposition of the integral matrix (which includes acceptance effects) evaluated in each of the 45 $m_{3\pi}$ - bins and normalized to the biggest eigenvalue in each bin.

Figure 5.2 shows the eigenvalues normalized to the largest eigenvalue λ_{max}, corresponding to:

$$\lambda = \frac{\tilde{\lambda}_i}{\lambda_{max}} \,. \tag{5.3}$$

For $m_{3\pi} \lesssim 1.2\,\mathrm{GeV}/c^2$, the three lowest eigenvalues (blue, orange, green) strongly decrease as the mass reduces. These correspond to large overlaps in the $J^P = 1^+$ waves. By excluding waves one by one and repeating the eigenvalue decomposition, we can study the interference effects between different waves. All thresholds are listed in Table 5.1 and discussed in the following. We determine the exact thresholds from eigenvalue decomposition studies and input-output studies.

The smallest normalized eigenvalue (blue), which is within $\lambda \in [0.0005,\, 0.02]$ for $m_{3\pi} \lesssim 1.10$ GeV/c^2, corresponds to an overlap between the $1^+\,[f_0(980)\,\pi]_P$, the $1^+\,[\rho(770)\,\pi]_S$ and the $1^+\,[\sigma(500)\,\pi]_P$ partial waves according to its eigenvector. Based on these studies, we resolve this overlap by excluding the $1^+\,[f_0(980)\,\pi]_P$ wave in bins below $m_{3\pi} = 1.10\,\mathrm{GeV/c^2}$. This exact value is found in additional PWD studies, showing that this threshold effectively removes the bias.
The second lowest eigenvalue (orange) is in the range of $\lambda \in [0.001, 0.039]$ for $m_{3\pi} \lesssim 1.06\,\mathrm{GeV/c^2}$ arises from the $[\rho(1450)\,\pi]_D$, the $1^+\,[\rho(770)\,\pi]_D$ and the $1^+\,[f_2(1270)\,\pi]_P$ waves. Here we set the threshold for the $[\rho(1450)\,\pi]_D$ wave to $m_{3\pi} \approx 1.06\,\mathrm{GeV/c^2}$, since it is far away below the nominal mass of the $\rho(1450)$. Additionally, the $1^+\,[\rho(770)\,\pi]_D$ is also excluded from the waveset in the range $m_{3\pi} = [0.80, 0.86]$
The third vanishing eigenvalue (green) is below $\lambda = 0.048$ for $m_{3\pi} < 0.94\,\mathrm{GeV/c^2}$ and reflects an overlap between $1^+\,[\rho(770)\,\pi]_S$ and $1^+\,[\sigma(500)\,\pi]_P$. However, since the nominal masses of these partial waves lie within this kinematic range, no physically motivated constraint can be imposed to remove this overlap. Consequently, a bias between these

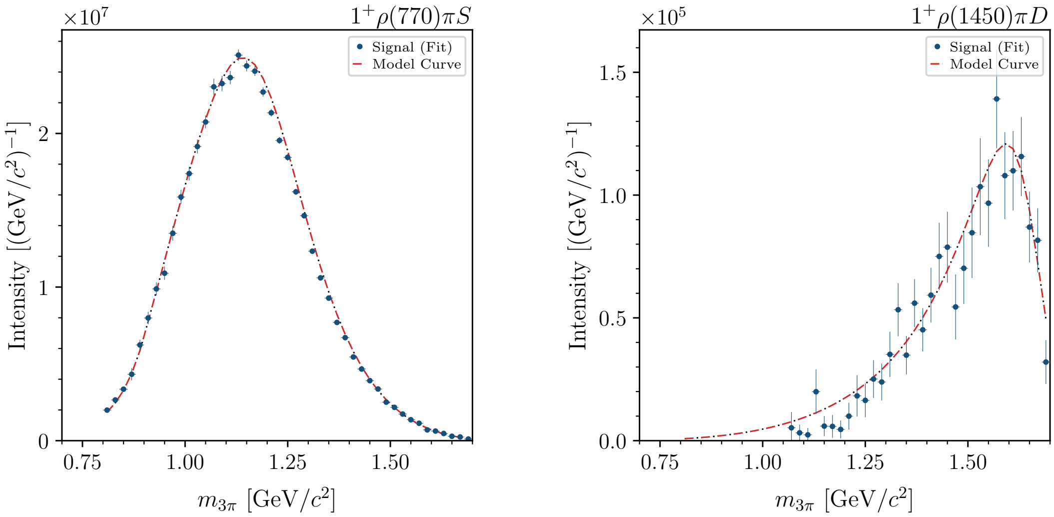


Figure 5.3: PWD fit results performed on reconstructed signal MC data wave-set including the determined thresholds (blue) for the $1^+\,[\rho(770)\,\pi]_D$ (left) and $1^+\,[\rho(1450)\,\pi]_D$ (right) wave compared to the expected output of the model (model curve).

two waves remains and cannot be resolved by the applied thresholds.
We additionally constrain the $0^-\,[f_0(980)\,\pi]_S$ and the $0^-[\rho(770)\pi]_P$ partial wave similarly to the $1^+[f_0(980)\pi]_P$ and the $1^+[\rho(770)\pi]_D$ partial wave, since we see in the PWD fit that we are not sensitive to the lower mass region in this wave.

Table 5.1: Partial waves included in the input-output studies and their lower thresholds in $m_{3\pi}$.

Partial wave	$m_{3\pi}^{\text{min}}$ [GeV/c^2]	**Partial wave**	$m_{3\pi}^{\text{min}}$ [GeV/c^2]
$1^+[\rho(770)\pi]_S$	0.80	$0^-[\rho(770)\pi]_P$	0.86
$1^+[\sigma(500)\pi]_P$	0.80	$0^-[f_0(980)\pi]_S$	1.20
$1^+[\rho(770)\pi]_D$	0.86	$1^-[\omega(782)\pi]_P$	0.80
$1^+[\rho(1450)\pi]_D$	1.06		
$1^+[f_2(1270)\pi]_P$	0.80		
$1^+[f_0(980)\pi]_P$	1.10		

Figure 5.3 shows the PWD fit (blue) for the $1^+\,[\rho(770)\,\pi]_S$ and $1^+\,[\rho(1450)\,\pi]_D$ waves with the developed thresholds employed. The large bias observed in these partial waves is mitigated, and the description of the $1^+\,[\rho(770)\,\pi]_S$ wave in particular is improved over the whole $m_{3\pi}$ range by thresholding the $1^+\,[\rho(1450)\,\pi]_D$ wave.

6 Background Modeling in Partial-Wave Decomposition Fits

As discussed in Section 3.5, we expect around 9.5 % background contamination in the real data sample. To quantify the impact of neglecting the background in the PWD, a PWD fit is done to pseudo data containing signal and 9.5 % background contributions without including any background components in the PWD model. This allows us to study how background contributions leak into the partial-wave intensities if not properly modeled. We test how much and in which $m_{3\pi}$ region background contributions project onto each partial wave.

In Figure 6.1, we compare the fit to signal and background pseudo data (green) to the fit to signal-only pseudo data (blue) and the input model curve (red) , corresponding to the expected output of the signal PWD model, for 4 partial waves: $1^+\,[\rho(770)\,\pi]_S$, $1^+\,[(\pi\pi)_S\,\pi]_P$, $1^+\,[f_2(1270)\,\pi]_P$ and $0^-\,[\rho(770)\,\pi]_P$.

We observe non-negligible deviations in the PWD fit results compared to the signal-only PWD fit results and signal model curve in the most dominant partial wave $1^+\,[\rho(770)\,\pi]_S$. For $m_{3\pi} \lesssim 1.14\,\mathrm{GeV/c^2}$, the background introduces a large bias, corresponding to vastly overestimated intensity compared to the signal model curve. Above $m_{3\pi} \approx 1.20\,\mathrm{GeV/c^2}$, the intensity is underestimated with only small uncertainties.

Above $m_{3\pi} \approx 1.5\,\mathrm{GeV/c^2}$, the impact of not modeling background contributions becomes negligible. This is equivalent for all partial waves.

The $1^+\,[(\pi\pi)_S\,\pi]_P$ wave is also strongly affected if background contributions are neglected in the PWD model. Especially in the mid range of $m_{3\pi} \approx [1.1, 1.35]$ GeV/c^2 the intensity exhibits a pronounced overestimation. Furthermore, the lower $m_{3\pi}$ range shows strong deviations as well.

Finally, a large enhancement of intensity is visible in $1^+\,[f_2(1270)\,\pi]_P$ and 0^-, $[\rho(770)\,\pi]_P$, similarly as in the $1^+\,[(\pi\pi)_S\,\pi]_P$ wave, specifically in the low $m_{3\pi}$ range of $1^+\,[f_2(1270)\,\pi]_P$ wave and in the full range of the $0^-\,[\rho(770)\,\pi]_P$ wave, shown in Figure 6.1.

These examples demonstrate that the background contributions introduce a significant bias if they are not accounted for in the PWD. The partial waves showing strong leakage effects are $1^+\,[\rho(770)\,\pi]_S$, $1^+\,[(\pi\pi)_S\,\pi]_P$, $1^+\,[f_2(1270)\,\pi]_P$ and $0^-\,[\rho(770)\,\pi]_P$. In the partial waves $1^-\,[\omega(782)\,\pi]_P$ and $1^+\,[f_0(980)\,\pi]_P$, the overall shape is similar, but the intensity of the peaks is biased. For $1^+\,[\rho(770)\,\pi]_D$, $1^+\,[\rho(1450)\,\pi]_D$ and $0^-\,[\rho(770)\,\pi]_P$ only fit results in $m_{3\pi}$ bins right above the thresholds set in Chapter 5 are affected. This indicates that partial waves in these poorly-constrained $m_{3\pi}$ regions are more sensitive to leakage of background contributions. Because of these biases introduced by the background, background contributions must be explicitly modeled and introduced in the PWD formalism.

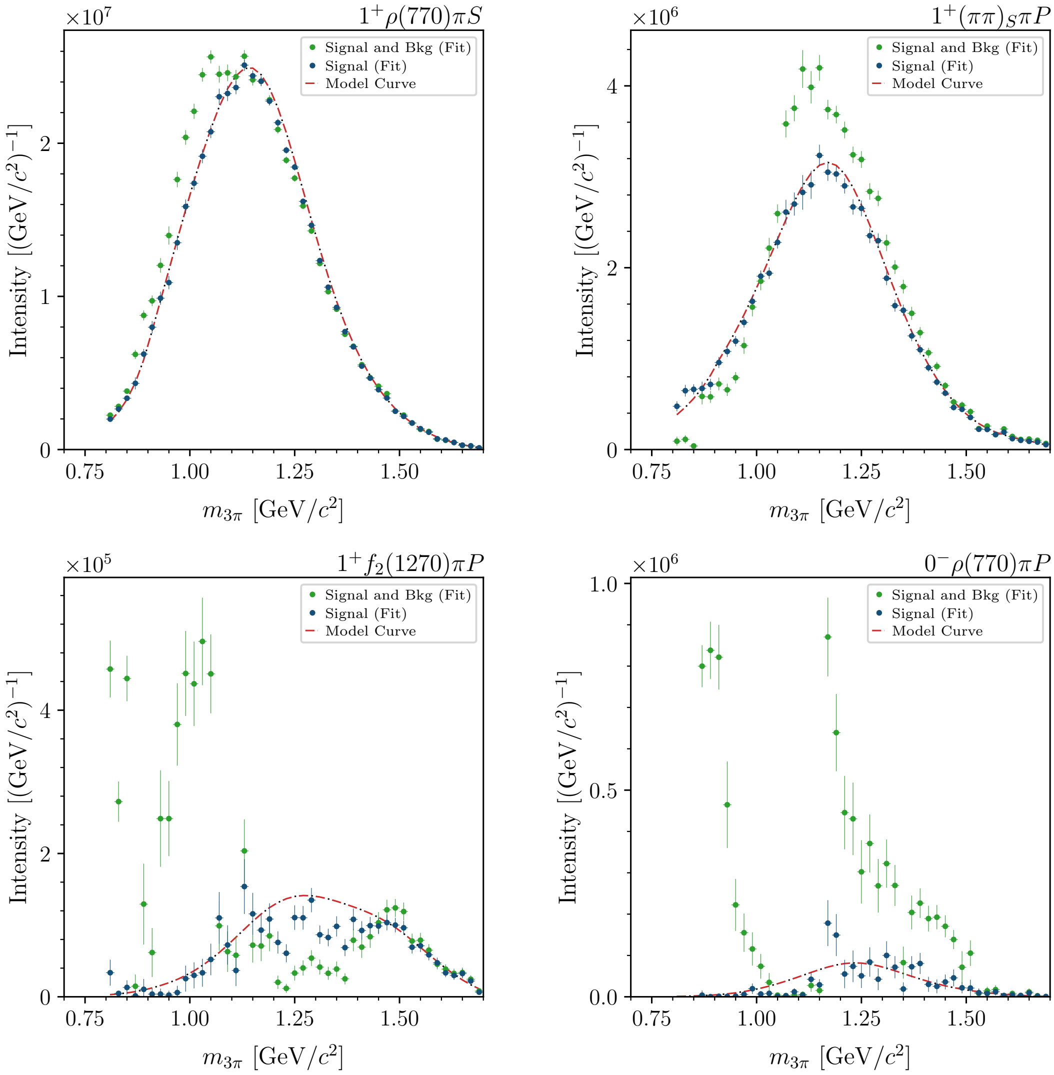


Figure 6.1: The intensity distribution in $m_{3\pi}$ is shown for a PWD fit using signal-only events (blue) and signal and background events (green) without correcting for the background in the PWD model for 4 partial waves: $1^+\,[\rho(770)\,\pi]_S$, $1^+\,[(\pi\pi)_S\,\pi]_P$, $1^+\,[f_2(1270)\,\pi]_P$ and $0^-\,[\rho(770)\,\pi]_P$. Additionally, the signal model curve is shown.

We discuss the inclusion of background contributions in the PWD formalism in Section 6.1. To model the background contributions, two background modeling methods are discussed, in Section 6.2 and in Section 6.4. Furthermore, an additional phase-space cut is detailed in Section 6.3 and the results of the background modeling and the background-subtracted PWD fits are shown in Chapter 7.

6.1 PWD with Background Components

In the partial wave decomposition formalism introduced in Section 4.2, we have considered signal contributions only. However, as already introduced in Section 3.5, the data sample is expected to contain several background contributions, with the biggest contribution coming from the decay $\tau \to \pi^-\pi^-\pi^+\pi^0\,\nu_\tau$. To account for these background events, the background distribution, which is related to the true reconstructed background yield $\hat{\hat{N}}_B$ times the background probability density function P_B^j for each considered background contribution, is incoherently added to the signal distribution. To get the total probability density function of an event, we normalize the sum over signal and background distributions, denoted by the normalization $\hat{\hat{N}}$, and get

$$P(\tau_e) = \frac{\eta(\tau_e)\,\Phi(\tau_e)\,I_S(\tau_e) + \sum_j \hat{\hat{N}}_B^j\,P_B^j(\tau_e)}{\hat{\hat{N}}}\,. \tag{6.1}$$

Since we consider multiple background channels, we sum over each individual component j. In the PWD fit using signal components, we have used the fact that the phase space density $\Phi(\tau)$ and the phase space volume V_{PS} are global factors in the likelihood and do not depend on any of the fit parameters, thus they can be neglected in optimization of the maximum likelihood fit in Section 4.3.

To similarly use this simplification in the maximum-likelihood fit including background components, we impose two modeling choices. First, the signal acceptance η and the signal differential phase space volume Φ are factored out, such that

$$= \frac{\eta(\tau_e)\,\Phi(\tau_e)\,\left[I_S(\tau_e) + \sum_j \hat{\hat{N}}_B^j\,B^j(\tau_e)\right]}{\hat{\hat{N}}}\,, \tag{6.2}$$

and we introduce the B^j function corresponding to

$$B^j(\tau) = \frac{P_B^j(\tau)}{\eta(\tau)\,\Phi(\tau)} = \frac{P_B^j(\tau)}{g_{\tilde{S}}(\tau)}\,. \tag{6.3}$$

Here, $g_{\tilde{S}}(\tau)$ is the reconstructed signal distribution of events distributed uniformly in phase space.

Using the signal intensity specified in Equation (4.15), written in terms of the decay amplitudes ψ (Equation (4.12)) and couplings c the probability density function of an event in bins of $m_{3\pi}$ is

$$
\begin{aligned}
P(\tau_e) = \frac{\eta(\tau_e)\,\Phi(\tau_e)}{\hat{\bar{N}}(m_{3\pi})} \Bigg(\sum_i \sum_{a,b} c_a^i(m_{3\pi}) c_b^{i*}(m_{3\pi}) \, \frac{\widetilde{\psi}_a^i(m_{3\pi},\tau_e)\,\widetilde{\psi}_b^{i*}(m_{3\pi},\tau_e)}{V_{PS}} \\
+ \frac{\sum_j \hat{\bar{N}}_B^j(m_{3\pi})\,\tilde{B}^j(m_{3\pi},\tau_e)}{V_{PS}} \Bigg),
\end{aligned}
\tag{6.4}
$$

using the $\tilde{B}^j$ function defined in Equation (6.18) and Equation (6.28)). The normalization of the probability density function factors into separate signal and background contributions as

$$
\begin{aligned}
\hat{\bar{N}}(m_{3\pi}) &= \int d\tau\; \eta(\tau)\,\Phi(\tau)\, I_S(\tau, m_{3\pi}) + \int d\tau\; \hat{\bar{N}}_B(m_{3\pi})\, P_B(\tau, m_{3\pi}) && (6.5) \\
&= \hat{\bar{N}}_S(m_{3\pi}) + \sum_j \hat{\bar{N}}_B^j(m_{3\pi})\,. && (6.6)
\end{aligned}
$$

The normalization of the signal is defined in Equation (4.17) and here denoted as $\hat{\bar{N}}_{S,MC}$, since we use signal MC data for its calculation. The background normalization for each contribution $\hat{\bar{N}}_B^{(j)}$ is simply the expected reconstructed background yield for each background contribution, which are the free parameters in the PWD fit.
For PWD fits considering signal and background contributions, the extended maximum likelihood fit (Section 4.3) is now performed using this adjusted probability density function, detailed in Equation (6.4). The simplified final logarithm of the likelihood with signal and background contributions is given by[1]:

$$
-\log(\mathcal{L}'_{\mathrm{PWD}}) = -\sum_e^{\bar{N}} \log\left(\sum_i \sum_{a,b} c_a^i c_b^{i*}\, \widetilde{\psi}_a^i(\tau_e)\,\widetilde{\psi}_b^{i*}(\tau_e) + \sum_j \hat{\bar{N}}_B^j(\tau_e)\tilde{B}^j(\tau_e) \right) + \hat{\bar{N}}. \tag{6.7}
$$

The fitted parameters here are the complex couplings c, and the background yields $\hat{\bar{N}}_B^{(j)}$. The constant prefactors are again disregarded in the minimization of the negative log-likelihood, since they do not depend on any free parameters in the fit, and they only change the overall value, not the location of the minimum.
The initial fit parameter values for the background yields are uniformly drawn between 0 and the total number of events in the data sample. As already discussed in Section 4.3, the start parameters for the couplings are generated from uniformly drawn intensities and phases, which are then used to compute the real and complex values of the couplings.
The high dimensionality and the complex structure of the background components in $\tau \to \pi^-\pi^-\pi^+\,\nu_\tau$ phase space make them difficult to model. The approach taken in this analysis is to model background contributions using multilayer feedforward networks, which can approximate any continuous function [13], and use these neural networks in the PWD fit to account for background events.

[1] For readability, the $m_{3\pi}$ dependency is dropped

The parametrization of the B function using neural networks will be discussed for two different approaches, the first one is detailed in the following section and the second one in Section 6.4.

6.2 Method 1: Density Estimation

The most straightforward idea is to train a neural network that learns the B function, i.e., the ratio of the background probability density function to the signal phase space and acceptance, denoted as

$$\mathrm{NN}(\tau) \propto \frac{P_B(\tau)}{g_S(\tau)} = B(\tau). \tag{6.8}$$

This follows the background modeling idea described in [44].
To determine the neural network parameters, we use a maximum-likelihood approach. The likelihood is given as the product over all reconstructed background events $\bar{N}_B$, weighted by the luminosity weights w_e, which account for differences in luminosity across the simulated sample, distributed according to the background PDF P_B, i.e.

$$\mathcal{L}_{\mathrm{bkg}} = \prod_{e}^{\bar{N}_B} \left(P_B(\tau_e)\right)^{w_e} \tag{6.9}$$

We can replace the background PDF with the description of the neural networks, defined in Equation (6.8), and write the background PDF as

$$P_B(\tau) = \frac{\mathrm{NN}(\tau)}{\mathcal{N}}\, g_S(\tau)\,. \tag{6.10}$$

In this parametrization, the neural network represents a function that effectively reweights signal data, uniformly distributed in phase space, to describe the background data. The function that best approximates the background PDF is found by optimizing the likelihood with respect to the neural network parameters θ.
By construction, the neural network must be normalized. We determine the normalization $\mathcal{N}$ from the condition

$$\begin{aligned} 1 &= \int P_B(\tau)\, d\tau \\ &= \int \eta(\tau)\Phi(\tau)\, \frac{\mathrm{NN}(\tau)}{\mathcal{N}} d\tau\,. \end{aligned} \tag{6.11}$$

This yields

$$\mathcal{N} = \int \eta(\tau)\Phi(\tau)\, \mathrm{NN}(\tau)\, d\tau\,. \tag{6.12}$$

However, the signal acceptance η is only known from Monte Carlo simulations, and evaluating the differential phase space volume Φ for each signal event is computationally expensive. Therefore, we use Monte Carlo integration [33], approximating the integral's value by sampling from a known distribution. The signal distribution we are sampling from is distributed according to

$$\mathrm{PDF}(\tau_i) \sim \frac{w_g(\tau_i)\,\eta(\tau_i)\,\Phi(\tau_i)}{H'}\,, \tag{6.13}$$

corresponding to signal data that was generated according to a model (weights w_g), signal acceptance η, and signal differential phase space volume Φ. The distribution is normalized by the normalization H'. We can approximate the normalization integral in Equation (6.15) by using importance sampling [33, p. 80-83], meaning we approximate the integral as a sum over uniformly drawn events $\bar{N}_{S,MC}$ from the distribution specified in Equation (6.13) as

$$\mathcal{N} = \frac{H'}{N_{S,MC}} \sum_{i}^{\bar{N}_{S,MC}} \frac{1}{w_g(\tau_i)}\,\mathrm{NN}(\tau_i)\,. \tag{6.14}$$

The signal normalization H' is defined in Appendix A, with which the normalization is fully defined as

$$\mathcal{N} = \frac{V_{PS}}{\sum_i^{N_{S,MC}} \frac{1}{w_g(\tau_i)}} \sum_{i}^{\bar{N}_{S,MC}} \frac{1}{w_g(\tau_i)}\,\mathrm{NN}(\tau_i) = \tilde{\mathcal{N}}\,V_{PS}\,. \tag{6.15}$$

With the final normalization, the likelihood used in the neural network training is now fully defined by

$$\mathcal{L}_{\mathrm{bkg}} = \prod_{e}^{\bar{N}_B} \left[\eta(\tau_e)\,\Phi(\tau_e)\frac{\mathrm{NN}(\tau_e)}{V_{PS}\,\tilde{\mathcal{N}}}\right]^{w_e}\,. \tag{6.16}$$

The negative log-likelihood is used during neural network training as a custom loss function L. We again use the fact that the acceptance η, the phase-space density $d\Phi$, and the phase-space volume V_{PS} do not depend on the parameters we optimize, since we optimize on the neural network parameters θ. Therefore, in the optimization, they do not change the location of the minimum, and the loss used in the training consists of only two terms:

$$L(\theta) = -\sum_{e}^{\bar{N}_B} w_e\,\log(\mathrm{NN}(\tau_e;\theta)) + \left(\sum_{e}^{\bar{N}_B} w_e\right) \log\left(\sum_{s}^{\bar{N}_S} \mathrm{NN}(\tau_s;\theta)\,\frac{1}{w_g(\tau)}\right)\,. \tag{6.17}$$

The first term is the sum over the logarithm of the network output for the background events e, weighted by the luminosity weights. In regions with background events, the first term becomes large, effectively corresponding to an increase in the background PDF. The second term normalizes the PDF and globally constrains increases in the neural network output. This means the neural network learns a function balanced by these two terms, enhanced in regions with background events but still consistent with the overall normalization.
With the neural network training concept discussed, we continue by detailing the exact B-function formulation using neural networks, which is required for the PWD fit in

Section 6.2.1. Further, the technical aspects of neural network training will be addressed in Section 6.2.2, including data handling and the neural network architecture. Finally, the results of the background modeling for this method are shown in Section 6.2.3.

6.2.1 Implementing Neural Networks in the PWD Fit

The B function is defined as the background PDF divided by the signal density in Equation (6.3). Using Equation (6.10) for the background PDF we can relate the B function to the neural network as

$$B(\tau) = \frac{\mathrm{NN}(\tau)}{\mathcal{N}} = \frac{\mathrm{NN}(\tau)}{\tilde{\mathcal{N}} V_{PS}}\,, \tag{6.18}$$

with the normalization $\mathcal{N}$ that is defined in Equation (6.15). In the last step, we use the fact that the normalization $\mathcal{N}$ is directly proportional to the phase-space volume V_{PS} to introduce the normalization $\tilde{\mathcal{N}}$. Therefore, the $\tilde{B}$ function used in the PWD formalism Equation (4.25) is defined as

$$\tilde{B} = \frac{\mathrm{NN}(\tau)}{\tilde{\mathcal{N}}} \tag{6.19}$$

6.2.2 Neural Network Training

We train the neural networks on the data sample described in Section 3.5. We require both signal and background data, since only background events enter the first term of the loss (Equation (6.17)), whereas the normalization is calculated from signal data.
The neural network architecture defined here serves as a baseline and is modified in the second method described in Section 6.4.2.
The entire data sample is split into 45 bins in $m_{3\pi}$, equivalent to the $m_{3\pi}$ bins used in the PWD fit, defined in Chapter 4. We train one neural network per bin and use all background contributions from the data sample. The $m_{3\pi}$ training ranges and the background contributions we train on will be changed in the second approach. For each of the 45 bins, the number of events and the distribution of signal and background vary significantly. Consequently, the optimal neural network architecture differs across the $m_{3\pi}$ bins, and we only specify the ranges of the hyperparameters. The exact neural network architecture for all bins is documented in [45].
The neural network is a fully connected feed-forward neural network, and training was optimal with a network architecture of 3-5 hidden layers with 20-300 hidden nodes.
We use the five phase space variables, defined in Section 3.1, as input features in the neural network. These are passed through successive linear transformations with Mish activation functions [46]. The final layer consists of a single sigmoid activation, representing an event-by-event weight that can be in the range $w_{\mathrm{NN}} \in (0, 1)$.
During training, we optimize using the Adam optimizer[47] with an exponentially decaying learning rate, and a custom scheduler that stops decreasing once the minimum learning rate, $\lambda_{min} = 5 \times 10^{-5}$, is reached. The initial value of the learning rate is between $\lambda_{start} = 0.003$ and $\lambda_{start} = 0.0008$.
To prevent overfitting, we employ two regularization approaches. We use dropout layers

[48] after each Mish activation with dropout rates between $d = 0.1$ and $d = 0.4$. This means that 10 % to 40 % of the neuron activations are not used during training, preventing overfitting.
An additional weight decay [49] of $\lambda_w = 0.003$ to $\lambda_w = 10^{-5}$ is used. Using weight decay means adding an additional term to the loss function that penalizes large neural network parameters, effectively favoring less complex models. These are less prone to overfitting and generalize better [49].
For the training, the background data is batched with a batch size of $\mathrm{s} = 2^9$, while the normalization is not batched, and all signal events are used to evaluate the normalization. The training was typically done for 700-2000 epochs, specified in more detail in [45]. The dataset is split into training, validation, and test samples. Out of 30 data chunks, 25 are used for training, 3 for validation, and the remaining 2 for testing.

6.2.3 Background Modeling Results

To represent the modeled background PDF and to assess how well the neural network approximates the background distribution, the outputs are used as reweighting factors for the signal data. Figure 6.2 shows an example neural network (red), as well as the background sample (blue) in the $m_{3\pi}$ range of $m_{3\pi} \in [1.10,\, 1.12]\,\mathrm{GeV/c^2}$.
The background is reasonably well described by the small deviations. This is similar across all neural network trainings in the mid $m_{3\pi}$ range, specifically in the $m_{3\pi} \in [0.98, 1.46]\,\mathrm{GeV/c^2}$.
However, especially for $m_{3\pi} \lesssim 0.98\,\mathrm{GeV/c^2}$ bins, the neural network training is unstable, which will be further discussed in the following section.

6.3 Outliers in Neural Network Outputs and Additional Phase Space Cut

We show the 2D distribution in $m_{\pi^+\pi^-}$ and $m_{\pi^-\pi^-}$ of the neural network weighted signal data in Figure 6.3 for a neural network trained in $m_{3\pi} = [0.84,\, 0.86]\,\mathrm{GeV/c^2}$. We observe individual neural network outputs that are significantly larger than the rest. An example for this is given in Figure 6.3, where we see an outlier at the squared two-pion masses of $m^2_{\pi^+\pi^-} \approx 0.1\,\mathrm{GeV^2/c^4}$ and $m^2_{\pi^-\pi^-} \approx 0.48\,\mathrm{GeV^2/c^4}$.
This individual neural network output is 2 orders of magnitude larger than the others. At the same time, the remaining distribution is described fairly well. This neural network training already incorporates regularization to prevent overfitting, specifically a relatively high dropout rate of $d = 0.3$ and a weight decay of $\lambda_w = 0.0003$. It was not possible to minimize the outliers further while still accurately describing the background distribution.
These irregular neural network outputs occur more frequently in regions near the kinematically allowed phase-space borders, where we do not have full coverage of signal events, but we do have background events. In addition, the difference between the outlier and the rest of the neural network outputs increases by several orders of magnitude

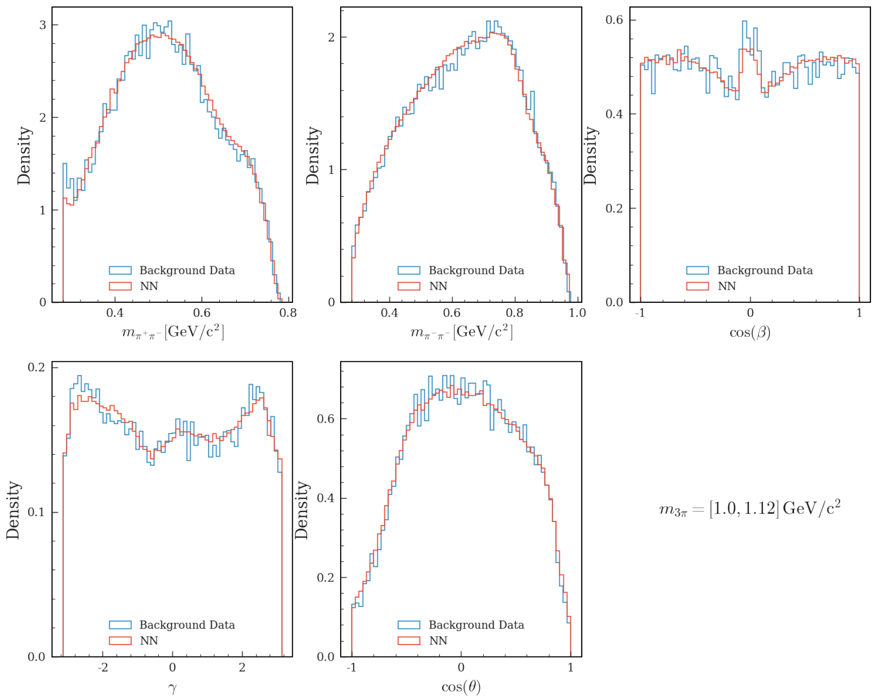


Figure 6.2: The signal data reweighted by the neural network outputs are shown in red for the 5 phase space variables ($m_{\pi^+\pi^-}$, $m_{\pi^-\pi^-}$, $\cos(\theta)$, $\cos(\beta)$, and γ) in the range $m_{3\pi}$ = [1.10, 1.12] GeV/c^2. These distributions are compared to the background data (blue).

as the neural network is trained for more epochs or with less regularisation. Thus, we conclude that the outliers are correlated to the phase-space distribution of the signal events and an unconstrained training procedure.

The outliers observed in the neural network results occur in regions where we have background data but are missing signal data. We have such regions because our signal data is not generated uniformly in phase space as required for the normalization, but rather using the TAUOLA generator [31], which generates events according to the implemented physics model. Consequently, some regions of the phase space are not fully populated and we can not fill the phase space by weighting if there are no events.

This becomes a problem because of the definition of the loss in the neural network approach described in the previous section in Equation (6.17). As described previously, the first term is evaluated on background data, where the neural network outputs increase when background events are observed, and the second term constrains the values and maintains the overall normalization calculated from signal data. That means during training, the neural network parameters are optimized through an interplay between background and signal data. In regions where we have background data but are missing

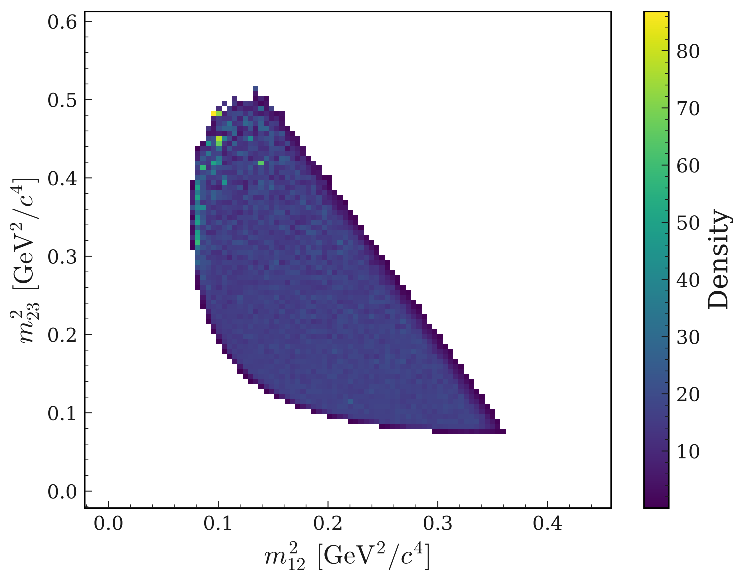


Figure 6.3: The signal events reweighted by the neural network outputs are plotted in $m_{12}^2 = m_{\pi^+\pi^-}^2$ and $m_{23}^2 = m_{\pi^-\pi^-}^2$ for $m_{3\pi} = [0.84,\ 0.86]\,\mathrm{GeV/c^2}$

signal data, the neural network can produce disproportionately large outputs that are unconstrained by the normalization. Since the final neural network output is a sigmoid activation function, its values are constrained to the interval $(0, 1)$ i.e., the neural netowrk output is limited to be below 1. However, the overall log-likelihood increases if large values are assigned to unconstrained regions, while the rest of the properly constrained outputs are pushed to very small values. As a result, individual outputs are several orders of magnitude bigger than all other outputs.

Following this discussion, the outliers observed in the neural network outputs, for example, shown in Figure 6.3, are the result of overfitting in regions where the signal sample does not fully cover the kinematically allowed phase space.

This poses a problem when using these neural networks in the partial-wave decomposition fit. In the fit, the neural networks are used twice when calculating the B function: once to evaluate them on the fitted data (signal and background data), and once in the normalization calculation, which sums over all neural network values evaluated on signal data. The first produces disproportionately large values for events in the poorly constrained region. At the same time, the normalization is dominated by individual neural network outputs that are significantly larger than the rest, corresponding to signal events in sparsely populated regions. Both of these lead to a bias in the PWD fit.

To preempt this behavior during neural network training, we employ a phase-space cut in the event selection, removing regions with missing signal data. Specifically, we introduce cuts on $m_{\pi^+\pi^-}^2 = m_{12}^2$ and $m_{\pi^+\pi^-}^2 = m_{13}^2$, supplementary to the cuts on the phase space borders discussed in Section 3.4.3.

Figure 6.4 shows the Dalitz plot for the three-pion final state in the invariant-mass range

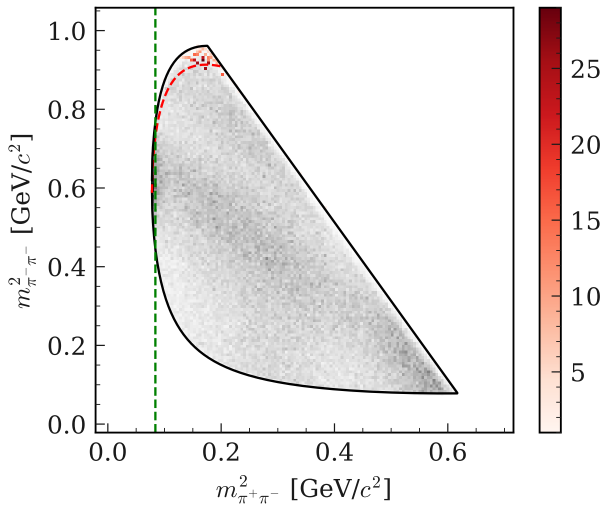


Figure 6.4: Overlay of background and signal data for $m_{3\pi} \in [1.10, 1.12]$ GeV. Regions with background data but without signal data are highlighted in red. The kinematically allowed phase space boundaries are shown in black. The additional cuts on m_{12} (green) and the event-dependent upper-bound cut on m_{13} are shown.

$m_{3\pi} \in [1.10, 1.12]$ GeV. It shows an overlay of signal and background events, with regions containing background but no signal data highlighted in red.
To exclude poorly populated regions near the phase-space boundaries, additional selection cuts are applied to the two-pion invariant masses. In particular, we restrict the upper boundary of the two-pion system m^2_{13} (red dotted curve in Figure 6.4) and the lower boundary of the m^2_{12} (green dotted line in Figure 6.4) by introducing small offsets δ_i

$$m^2_{13} \leq m^2_{13,\text{max}} - \delta_1, \qquad \delta_1 = 0.09\,, \tag{6.20}$$

$$m^2_{12} \geq (2m_\pi)^2 + \delta_2, \qquad \delta_2 = 0.005\,. \tag{6.21}$$

with $m_{13,\text{max}}$ calculated from Equation (3.16) for each event using the three invariant masses. Thus, the upper limit of m_{13} is effectively evaluated on an event-by-event basis and the red curve in Figure 6.4 does not represent a hard cut of all events above, but rather the lower envelope of the event-dependent kinematic boundary in the $m_{3\pi} \in [1.10, 1.12]$ GeV.

6.4 Method 2: Classifier

The unstable neural network outputs motivate the optimization of the training. Thus, we adopt a binary classifier approach, similarly used in [50, 51], to approximate probability

density ratios.
The binary classifier categorizes events into two labels, signal $\tilde{S}$ and background B. We use a neural network as a binary classifier that approximates the probability of an event being a signal event.

$$\mathrm{NN}(\tau) = P(\tilde{S}|\tau) = \frac{\nu_{\tilde{S}} P(\tau|\tilde{S})}{\nu_{\tilde{S}} P(\tau|\tilde{S}) + \nu_B P(\tau|B)} \,. \tag{6.22}$$

Here $\nu_{\tilde{S}}$ and ν_B denote the total number of reconstructed signal and total number of reconstructed background events. The signal events are reweighted with TAUOLA weights w_g, introduced in Section 3.5, to obtain a uniform phase-space distribution. Additionally, both samples are weighted by luminosity weights, so the distribution is corrected for events generated at different luminosities and further rebalanced such that $\nu_S = \nu_B$.

$$\nu_{\tilde{S}} = \sum_i^{\bar{N}_{\tilde{S}}} w_i \,, \qquad \nu_B = \sum_j^{\bar{N}_B} w_j \,. \tag{6.23}$$

The weighting simplifies Equation (6.22) to

$$\mathrm{NN}(\tau) = \frac{P_{\tilde{S}}(\tau)}{P_{\tilde{S}}(\tau) + P_B(\tau)} \,. \tag{6.24}$$

Here, $P_{\tilde{S}}(\tau)$ denotes the reweighted signal PDF (uniform in phase space), and $P_B(\tau)$ the luminosity-corrected background PDF. We find the neural network parameters that best approximate the probability ratio in Equation (6.22) by maximizing the likelihood of observing the true labels y_i given the input features τ_i. Thus, we are minimizing the cross-entropy function [52]

$$\begin{aligned} \mathrm{Loss}(\theta) &= -\sum_i^{\bar{N}} w_i \log(p(y_i|\tau_i)) \\ &= -\sum_i^{\bar{N}} w_i \left[\, y_i \log(\mathrm{NN}(\tau_\mathrm{i}, \theta)) \; + (1-y_i) \log(1 - \mathrm{NN}(\tau_\mathrm{i}, \theta))\right] . \end{aligned} \tag{6.25}$$

The weights w_i correspond to event weights, and the labels y_i can take the value 0 for background events and 1 for signal events.
We continue by discussing how neural networks are introduced in the PWD fit in Section 6.4.1, followed by a description of the neural network training in Section 6.4.2. The results will be shown in Chapter 7.

6.4.1 Implementing Neural Networks in the PWD Fit

The B function defined in Equation (6.3), i.e., the background PDF divided by the signal acceptance and phase-space, accounts for background components in the PWD fit. In order to relate the neural networks defined in Section 6.4.2 to the B function, we use Equation (6.24) and express the background PDF as

$$P_B(\tau) = P_{\tilde{S}}(\tau) \left(\frac{1 - \mathrm{NN}(\tau)}{\mathrm{NN}(\tau)} \right) . \tag{6.26}$$

The signal PDF $P_{\tilde{S}}$ is the uniformly distributed signal data i.e.,

$$P_{\tilde{S}} = \frac{g_{\tilde{S}}(\tau)}{\mathcal{N}} . \tag{6.27}$$

Using Equation (6.27) and Equation (6.26) the B function is defined as

$$B(\tau) = \frac{1 - \mathrm{NN}(\tau)}{\mathrm{NN}(\tau)} \frac{1}{\mathcal{N}} , \tag{6.28}$$

The normalization is fixed using the normalization condition of the background PDF

$$\begin{aligned} 1 &= \int d\tau \; P_B(\tau) \\ &= \int d\tau \; g_{\tilde{S}}(\tau) \frac{1}{\mathcal{N}} \left(\frac{1 - \mathrm{NN}(\tau)}{\mathrm{NN}(\tau)} \right) \end{aligned} \tag{6.29}$$

This yields a similar normalization as in the first method (Equation (6.12)),

$$\mathcal{N} = \int d\tau \; \eta(\tau) \, \Phi(\tau) \left(\frac{1 - \mathrm{NN}(\tau)}{\mathrm{NN}(\tau)} \right) . \tag{6.30}$$

We evaluate the integral using importance sampling [33], where we approximate the integral by randomly drawing $\bar{N}_{S,MC}$ from the signal distribution defined in Equation (6.13). Analogously to the first method, the normalization becomes

$$\mathcal{N} = \frac{V_{PS}}{\sum_i^{N_{S,MC}} \frac{1}{w_g(\tau_i)}} \sum_i^{\bar{N}_{S,MC}} \frac{1}{w_g(\tau_i)} \frac{1 - \mathrm{NN}(\tau_i)}{\mathrm{NN}(\tau_i)} = \tilde{\mathcal{N}} V_{PS} \tag{6.31}$$

Having calculated the normalization, the B function (Equation (6.3)) is fully defined. In Equation (6.31), the phase space volume V_{PS} is again directly proportional to the normalization. Therefore the $\tilde{B}$ function is defined as

$$\tilde{B}(\tau) = \frac{1 - \mathrm{NN}(\tau)}{\mathrm{NN}(\tau)} \frac{1}{\tilde{\mathcal{N}}} , \tag{6.32}$$

which is used in the PWD formalism in Equation (4.25).

The normalization in the B function used in this thesis is calculated with variables that neglect resolution.

6.4.2 Neural Network Training

Along with the neural network approach, we additionally improve the training procedure. This includes data handling and preprocessing, as well as neural network architectural choices.

Data Handling and Preprocessing

We apply an additional phase-space cut to the data, as detailed in Section 6.3, which prevents most spikes observed in the neural network outputs.
The training data range was previously chosen to be the same as the $m_{3\pi}$ mass binning used in the PWD fit, where the data is binned into intervals of $0.02\,\mathrm{GeV/c^2}$ in order to be independent of the $m_{3\pi}$-dependence of the partial wave amplitudes. We follow this fine binning approach again, training one neural network for each $m_{3\pi}$ bin used in the PWD fit, but we also separately train over wider $m_{3\pi}$ ranges in two additional binning approaches. This is justified because, during the neural network training, the signal data is de-weighted, making it uniformly distributed in phase space. Hence, we do not impose any modeling choice on $m_{3\pi}$ for the signal data during training, and it is not required to use the same binning. Therefore, we test the following two binning approaches: the first uses sliding-windows, and the second uses the full $m_3\pi$ range during training.
In the sliding-window approach, multiple neural networks are trained in windows of $m_{3\pi}$ with a size of $0.18\,\mathrm{GeV/c^2}$ with overlapping training regions. That means, for example, the first neural network is trained in the range $m_{3\pi} \in [0.71, 0.89]\,\mathrm{GeV/c^2}$, while the second is trained in the range $m_{3\pi} \in [0.80, 0.98]\,\mathrm{GeV/c^2}$. The full-$m_{3\pi}$-range approach avoids binning the data in $m_{3\pi}$ completely during neural network training, thus trains over the whole range of $m_{3\pi} \in [0.80, 1.70]\,\mathrm{GeV/c^2}$
The neural networks are trained either for all background contributions described in Section 3.5, or on samples split into individual background contributions, allowing for a fine-grained parametrization of each background type.
The overall data sample has a signal-to-background ratio of $\approx 11 : 1$ after all event selection cuts, meaning it is not equally represented, i.e., imbalanced. Because the training sample is split into different $m_{3\pi}$ ranges and background contributions, the imbalance is highly variable. Such imbalanced training sets tend to learn the majority class well, but model the minority class poorly [53], effectively ignoring the few events of the minority class. Thus, we preprocess the weights that are used in the loss, by normalizing them independently and reweighting the total event weight such that we have an effective 50:50 class balance.

Neural Network Architecture

The basic training structure is equivalent to that in Section 6.2.2, meaning we again use a fully-connected feed-forward neural network and optimize using the Adam optimizer. However, we adjusted the training architecture, specifically, we use around 12-16 hidden layers with 10-256 hidden nodes. The exact structure depends on the amount of available training data, with a tendency to use a more complex neural network as the

amount of data increases. The architecture for all trainings using different $m_{3\pi}$ binning and different background contributions is documented in [45].
Training performance can be improved by using different representations of the input data [54]. Thus, $\cos(\gamma)$ and $\sin(\gamma)$ are used complementary to the raw angle γ, incorporating the periodicity of the density of the kinematic angle [44]. Additionally, the helicity angle $\sin(\theta)$ and the Euler angle $\sin(\beta)$ are added, calculated from $\cos(\theta)$ and $\cos(\beta)$ respectively.
The invariant mass $m_{3\pi}$ is also included as an input feature. Thus, a total of 10 input features are used, which are passed through the neural network with Mish activation functions. The final layer is a sigmoid function, constraining the neural network outputs to $w_{\mathrm{NN}} \in (0, 1)$. A neural network output close to 0 corresponds to events that are classified as a background event, while an output close to 1 represents a classification as a signal event.
The training starts with a learning rate that decreases on a plateau in the validation loss, with an initial learning rate of $\lambda_{start} = 0.0008$ to $\lambda_{start} = 0.0006$. For some training runs, an optional finetuning step is applied, i.e., after $500 - 700$ epochs of training, the learning rate decays exponentially for $500 - 800$ epochs. Changing the learning rate ensures a more stable convergence during training and helps reduce fluctuations in the loss in later epochs [55].
Applying the phase-space cut (described in Section 6.3) prevents most spikes in the neural network outputs. However, to keep this cut to a minimum, the training is further adjusted to prevent overfitting in poorly constrained regions by applying batch normalization [56]. Additionally, a dropout layer is introduced after each Mish activation, with dropout rates between 0.1 and 0.3. The batch size s is determined dynamically and independently for training and validation, in order to achieve stable training with roughly 200 batches. This rule is violated by imposing a lower limit on the batch size of 2^8 and an upper limit of 2^{12} to keep it in a medium range. The batch size is further rounded to a multiple of 64 for computational efficiency.

6.5 Input-Output Studies including background contributions

We use generic run-dependent MC15 data as a background sample. These background events are only available on the reconstruction level, i.e., we do not have MCTruth-matched background variables. We combine this reconstructed background sample with the reconstructe signal pseudo data, defined in Section 4.4 using the expected fractions of individual background contributions defined in Section 3.5.

7 Results

The results of the background modeling using the classifier neural networks (Section 6.4) are split into two sections based on the background contributions used to train the neural networks. The first approach models all background contributions by a single neural network, i.e., all background contributions are combined during training (Section 7.1). In the second approach, multiple neural networks are trained within the same $m_{3\pi}$ range for different backgrounds by splitting the training data into separate contributions (see Section 7.2).
The different background modeling approaches are tested in input-output studies using reconstructed pseudo data, defined in Section 6.5. We define background modeling as working well if the modeled background distributions match the simulated background data, or if deviations between them have little effect on the PWD fit results.

7.1 Combined Background Modeling

As introduced in Section 6.4.2, we follow three different $m_{3\pi}$ binning approaches: the finely-binned approach (Section 7.1.1), the sliding-window approach (Section 7.1.2), and training on the full-$m_{3\pi}$-range (Section 7.1.3). The background modeling results will be discussed, followed by a comparison of the PWD fits using these neural networks in Section 7.1.4.

7.1.1 Neural Network Result: Finely-Binned Approach

In the finely-binning approach, training is performed in the same $m_{3\pi}$ bins as the PWD fit. Thus, we have 45 neural networks that are used to model the background contributions. We use the neural network classifiers to approximate the background probability density function, as defined in Equation (6.26). Figure 7.1 compares the resulting background PDF using the neural networks (red) to the background data (blue) in a central $m_{3\pi}$ bin of $m_{3\pi} \in [1.08,\, 1.10]\,\mathrm{GeV}/c^2$. The figure shows only the 6 phase-space variables, rather than the 10 input features, since the input features include 4 redundant angular variables. Based on this comparison, the neural network reproduces the overall background shape. However, a small discrepancy is observed in the angular variable $\cos\beta$, which exhibits a pronounced peak around $\cos\beta \approx 0$ in the background data that is somewhat smoothed in the neural-network model. The background description shows a comparable performance across the full mass range $m_{3\pi} \in [0.98,\, 1.36]\,\mathrm{GeV}/c^2$.
However, distinctive peaks are observed in the background data of all bins in the range of $m_{3\pi} \in [0.88, 0.96]\,\mathrm{GeV}/c^2$. One example is shown in Figure 7.2, the neural network (red) trained in $m_{3\pi} \in [0.90, 0.92]\,\mathrm{GeV}/c^2$ together with the background data (blue). The peak

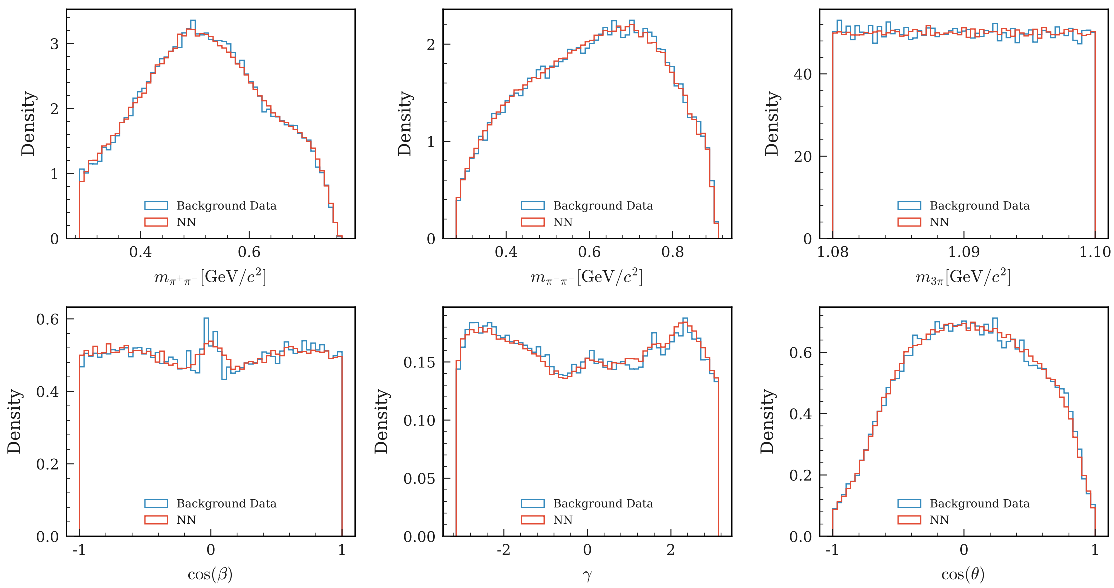


Figure 7.1: Background distributions as estimated by the neural network from the finely-binned combined-background modeling approach (red) compared to the background data distribution (blue) in the 6 phase-space variables for $m_{3\pi} \in [1.08,\ 1.10]\,\mathrm{GeV}/c^2$.

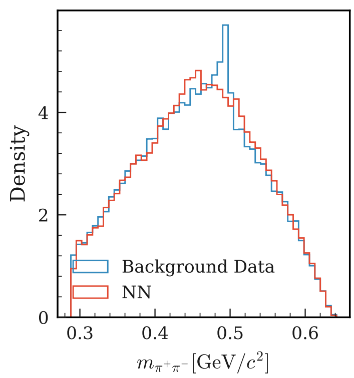


Figure 7.2: Background distributions as estimated by the neural network from the finely-binned combined-background modeling approach (red) compared to the background data distribution (blue) in $m_{\pi^+\pi^-}$ for $m_{3\pi} \in [0.90,\ 0.92]$ GeV/c^2.

in the background at an invariant mass $m_{\pi^+\pi^-} \approx 0.49\,\mathrm{GeV}/c^2$, is likely coming from the decay of $\tau^- \to \pi^- K^0 \nu_\tau$ where K^0 further decays as $K^0 \to \pi^+\pi^-$. An additional process corresponding to this signature in $m_{\pi^+\pi^-}$ is likely coming from $e^+e^- \to q\bar{q}$ events, producing K^0 mesons decaying to two pions. Both of these processes are described in Section 3.5.

It is not possible to model this specific peak with the neural networks used in this analysis, likely because they must learn the localized peaking structure while approximating the overall background distribution. This could not be resolved by increasing the neural network architecture's complexity, as this led to overfitting. Consequently, this specific K^0 signature is not modeled by this background modeling approach.

This limitation can be addressed by splitting the data during neural network training into distinct background contributions, as discussed in Section 7.2 We also study the impact of the K^0 peak on the PWD fit results for the combined background approach in Section 8.3.

Since the $m_{3\pi}$ binning is equivalent to the binning used in the PWD fit, the neural networks trained in each $m_{3\pi}$ bin are used in the corresponding $m_{3\pi}$ bin of the PWD.

7.1.2 Neural Network Results: Sliding-Window Approach

In the sliding-window approach, neural networks are trained using all expected background contributions in $m_{3\pi}$ windows of $0.18\,\mathrm{GeV/c^2}$ (see Section 6.4.2) totaling to 11 neural networks to be trained.

Figure 7.3 shows an example of this background modeling approach for the range of $m_{3\pi} \in [1.07,\, 1.25]\,\mathrm{GeV/c^2}$. The background PDF constructed using neural networks (red) and the background data (blue) agree well across all phase-space variables.

This agreement is similar across all neural networks, indicating a sufficient approximation of the background PDF over the full range of $m_{3\pi}$. However, as observed in the finely-binned approach discussed in Section 7.1.1 the neural networks cannot model the distinct K_S peak in $m_{3\pi} \in [0.80,\, 0.98]\,\mathrm{GeV/c^2}$ and $m_{3\pi} \in [0.89,\, 1.07]\,\mathrm{GeV/c^2}$.

As visualized in Figure 7.4, each neural network (each visualized with its own color) is trained on a wider $m_{3\pi}$ range (represented as the blue and orange squares) than the $m_{3\pi}$ range for which it is used in the PWD fit (filled squares of the same color). Except for the border regions, 4–5 adjacent $m_{3\pi}$ bins use the same neural network, which was trained on a range equivalent to 9 $m_{3\pi}$ bins. The first and last networks are only used for 2 bins.

7.1.3 Neural Network Results: Full-$m_{3\pi}$-Training-Range Approach

A single neural network is trained on the whole $m_{3\pi}$ range analyzed in this thesis, i.e., $m_{3\pi} \in [0.8, 1.7]\,\mathrm{GeV}/c^2$. The result of the background modeling is shown in Figure 7.5.

This yields similar results to those of the previous two binning approaches. While the overall shape of the background is well modeled using the neural network (red), it

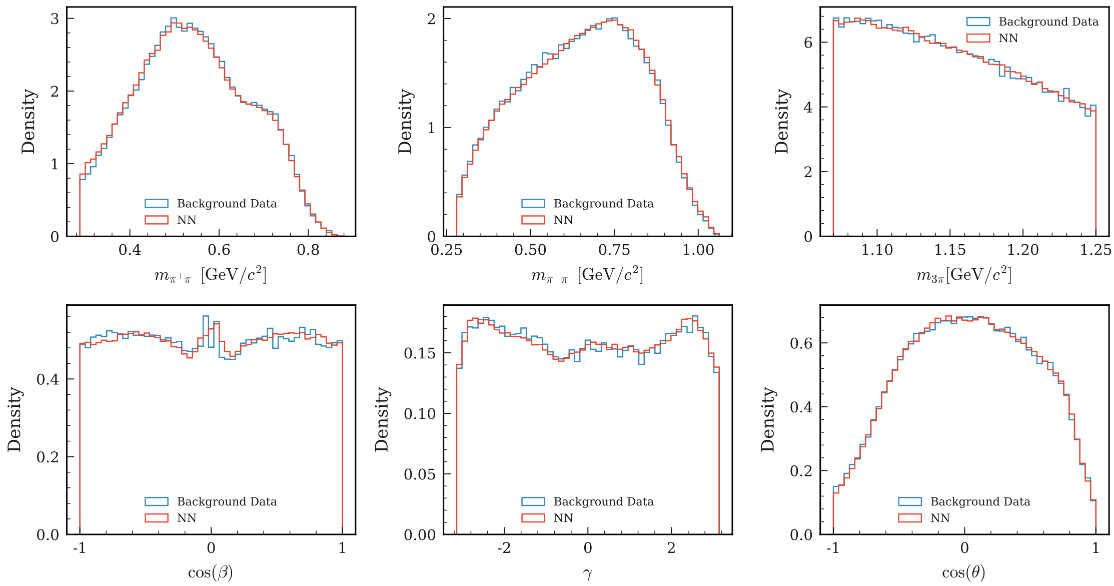


Figure 7.3: Background distributions as estimated by the neural network using the sliding-window combined-background modeling approach (red) compared to the background data distribution (blue) in the 6 phase-space variables for $m_{3\pi} \in [1.07,\ 1.25]\,\mathrm{GeV}/c^2$.

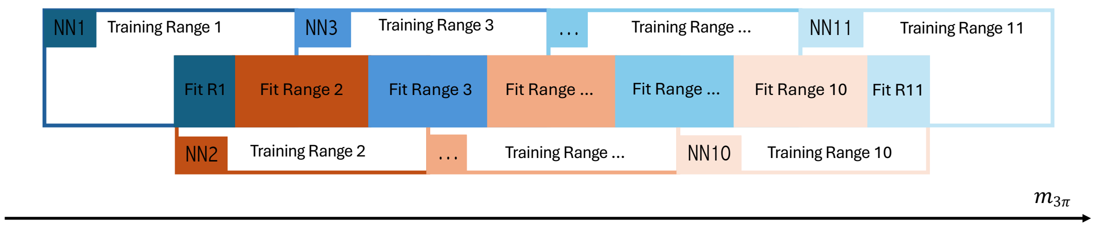


Figure 7.4: Visualized training ranges of the 11 neural networks (NN) of the sliding-window approach, represented as the transparent squares, along with the ranges in which neural networks are used in the PWD fit (filled squares) in $m_{3\pi}$. The color is chosen arbitrarily for visualizing the ranges of different neural networks.

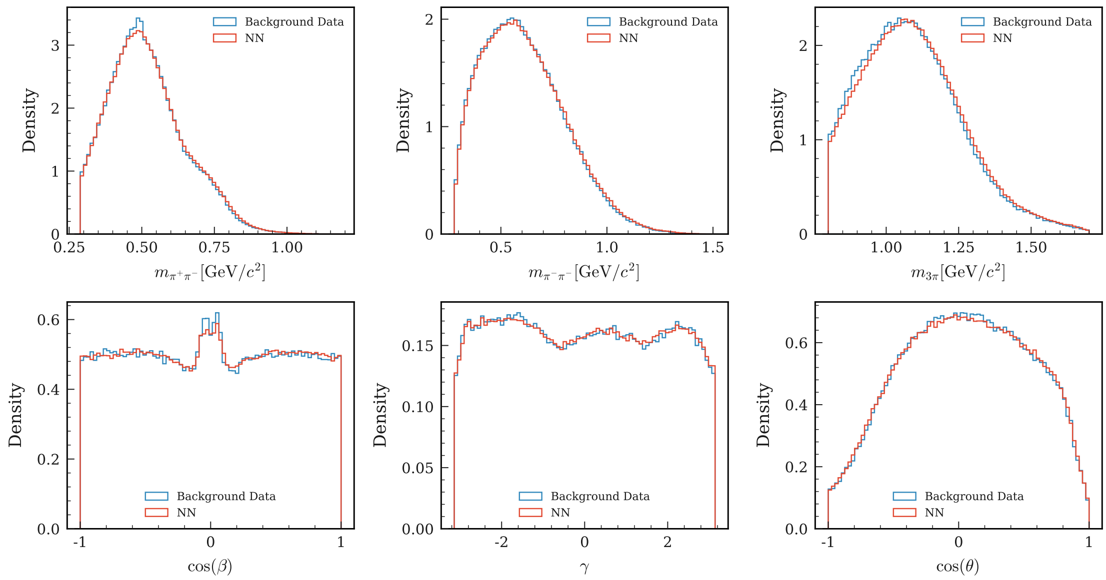


Figure 7.5: Background distributions as estimated by the neural network using the full-$m_{3\pi}$-training-range combined-background modeling approach (red) compared to the background data distribution (blue) in the 6 phase-space variables for $m_{3\pi} \in [0.80,\ 1.70]\,\mathrm{GeV}/c^2$.

imperfectly models the $m_{3\pi}$ distribution. It lightly underestimates the density for $m_{3\pi} \lesssim 1.05\,\mathrm{GeV}/c^2$, while is slightly overestimates the density for $m_{3\pi} \in [1.1, 1.4]\,\mathrm{GeV}/c^2$. The impact of this bias is assessed in input-output studies in Section 7.1.4. Additionally, the peak at $m_{\pi^+\pi^-} = 0.49\,\mathrm{GeV}/c^2$ is not accurately modeled, likely originating from K^0 decays. The bias introduced by not modeling the K^0 within the combined background modeling approach is discussed in Section 8.3.
The impact of this bias on the PWD fits will be discussed in Section 7.1.4. To model the background component in the PWD fits, the same neural network is used across all 45 $m_{3\pi}$ bins.

7.1.4 Input-Output Studies including Background Components

To assess the performance of the background modeling in the PWD fit, we conduct input-output studies which include signal and background contributions using the formalism established in Section 6.1. The background yield estimated by the PWD fit is compared to the true background yield in Figure 7.6 for the three different binning approaches. All of the yields are normalized by the bin width.

In all three cases, the neural network reproduces the overall shape of the background distribution in $m_{3\pi}$, characterized by a peak around $\approx [1.0, 1.1]\,\mathrm{GeV}/c^2$ and a falling tail extending towards higher masses of approximately $1.3\,\mathrm{GeV}/c^2$. All three approaches

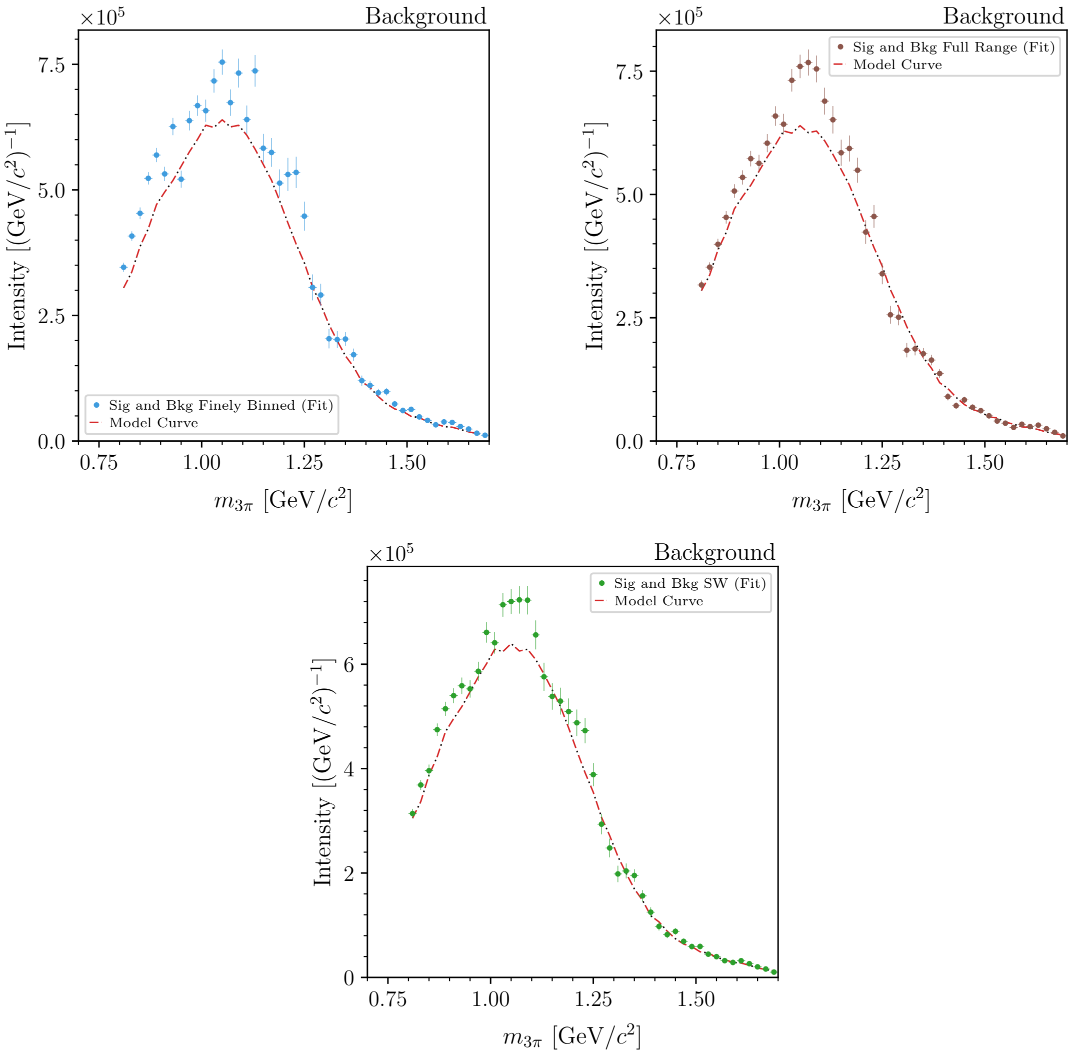


Figure 7.6: The background yields as estimated by the PWD for the three binning approaches of the combined background modeling: The Finely-binned approach (light blue, top left), the full-$m_{3\pi}$-training-range (brown, top right), and the sliding-window approach (green, bottom). Each yield is compared with the pseudo data's true background yield (red, same for all three subfigures).

slightly overestimate the background yield in the peak region compared to the true background contribution. This deviation suggests an interplay between signal and background components, which is investigated further in Section 8.1.
The background yields of the finely-binned training approach (light blue) are consistently overestimated, and we observe fluctuations within the range of $m_{3\pi} \approx [0.9, 1.25]\,\mathrm{GeV}/c^2$. Both of these are likely due to the overly fine binning during training. With fewer events per bin during training due to the small $m_{3\pi}$ region, the learned model becomes more sensitive to statistical noise, instead of accurately modeling the event signals within each bin, leading to a less stable background model by the neural network in the PWD.
From the comparison with the two binning approaches trained on a wider $m_{3\pi}$ range, it can be concluded that increasing the $m_{3\pi}$ training range reduces these fluctuations and provides a more stable approximation of the background contributions in the PWD fit. Thus, the finely-binned approach is disregarded.
Both the sliding-window approach (green) and training on the full-$m_{3\pi}$-range (brown) can approximate the background yields reasonably well. The sliding-window approach exhibits slightly smaller deviations from the model curve, most notably in the peak region of $m_{3\pi} \approx [1.0, 1.1]\,\mathrm{GeV}/c^2$. While this effect is minimal, it suggests that the complex background distribution is slightly more accurately described, compared to the full $m_{3\pi}$ training, which is slightly biased in the description of the background (see Section 7.1.3).

Examples of the background-subtracted partial wave results are shown in Figure 7.7 for the partial waves showing the largest leakage effects without accounting for background components (shown in Figure 6.1): $1^+\,[\rho(770)\,\pi]_S$, $1^+\,[(\pi\pi)_S\,\pi]_P$, $1^+\,[f_2(1270)\,\pi]_P$ and $0^-\,[\rho(770)\,\pi]_P$. The PWD fit results on signal-only data and the signal model curve are additionally shown. The intensity distribution of the waves in $m_{3\pi}$ is normalized by the bin width.
The fit results from the sliding-window approach (green) and the full range approach (brown) are consistent with the signal-only fit (blue), with no strong deviations visible in any of the partial waves. The dominant partial wave $1^+\,[\rho(770)\,\pi]_S$ is well reproduced, and the background no longer introduces the bias observed in the $m_{3\pi} \lesssim 1.14\,\mathrm{GeV}/c^2$ in the fit without accounting for background components (shown in Figure 6.1). Similarly, the PWD fit results for signal and background contributions match the signal model curve equally well as the signal-only-pseudodata fit for the $1^+\,[(\pi\pi)_S\,\pi]_P$ wave.
For the partial waves $1^+[f_2(1270)\,\pi]_P$ and $0^-\,[\rho(770)\,\pi]_P$, exhibiting strong structural deviations without accounting for background in the PWD fit, the agreement is substantially improved. This indicates that the neural networks detailed in Section 6.4 can be used to model background contributions in the pseudo data.
Further discussions on the impact of the background modeling on the PWD fit results will be given in Chapter 8.
Overall, the partial waves in the background-subtracted fits are equally well described compared to the signal-only fit.
The partial-wave results obtained using the sliding-window binning approach and the full $m_{3\pi}$-range approach are very similar. This suggests that the deviations in the background yields introduce a negligible bias on the description of the partial waves, and the

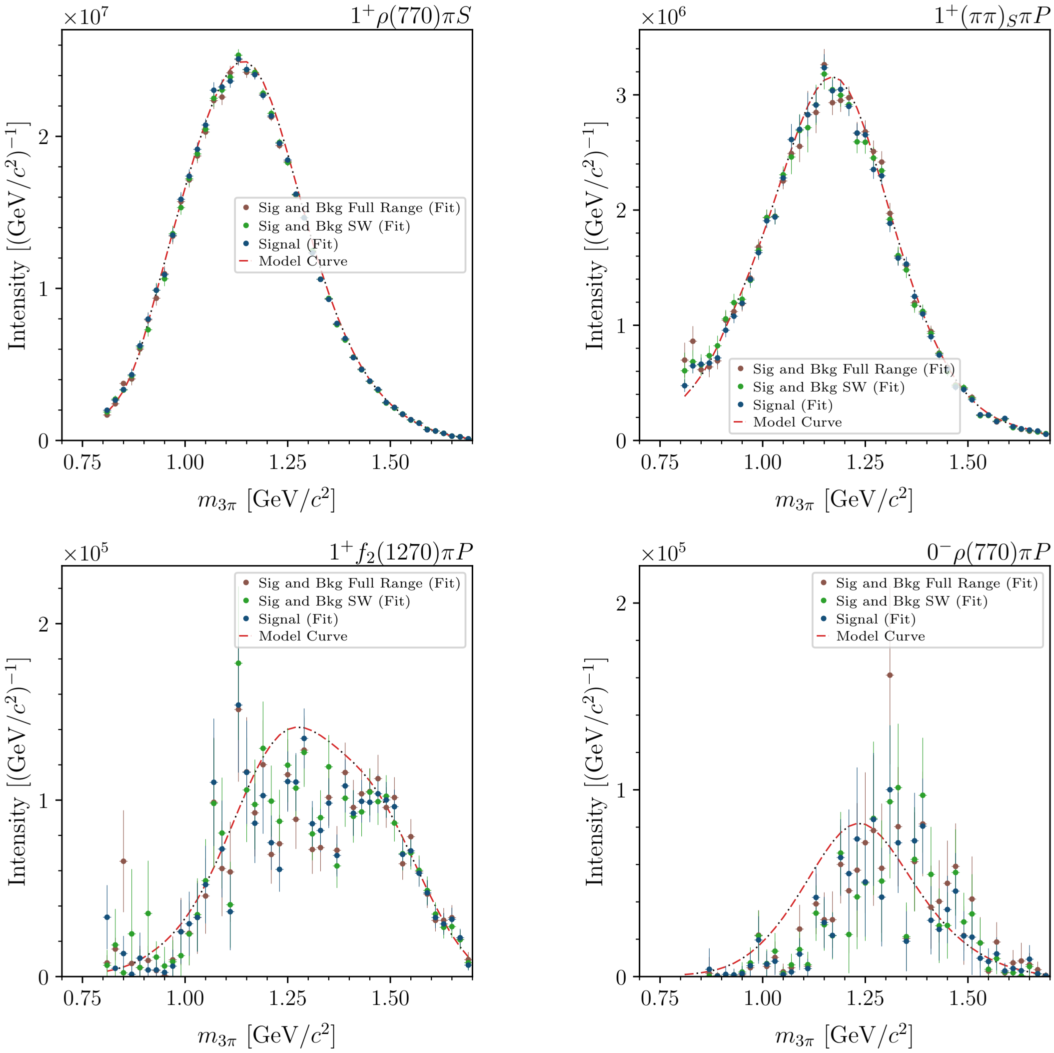


Figure 7.7: The background-subtracted PWD fits using the sliding-window approach (green) and the full $m_{3\pi}$ training approach (brown) compared to the signal-only PWD fits (blue) and the signal model curve (red).

two methods effectively serve as cross-checks of each other. Since both approaches describe the partial waves equally well, the sliding-window approach is used in the following chapter, as it provides a slightly more accurate estimate of the background yields.

7.2 Split Background Modeling

The motivation for modeling individual background components is to get a more fine-grained description of individual contributions. Moreover, we can avoid relying on MC simulations to determine the relative fractions of the individual background contributions, as their individual yields are directly extracted from the PWD fit.
The finely-binned approach, discussed previously in Section 7.1.1, is disregarded for the split background modeling, as it was shown that by training within small $m_{3\pi}$ bins, the neural network becomes unstable. Both the sliding-window approach and the full $m_{3\pi}$ training approach were tested, and the PWD results are similar in all partial waves. However, since the sliding-window approach showed slightly improved performance in the background yield approximation, we focus on it in the following discussion.

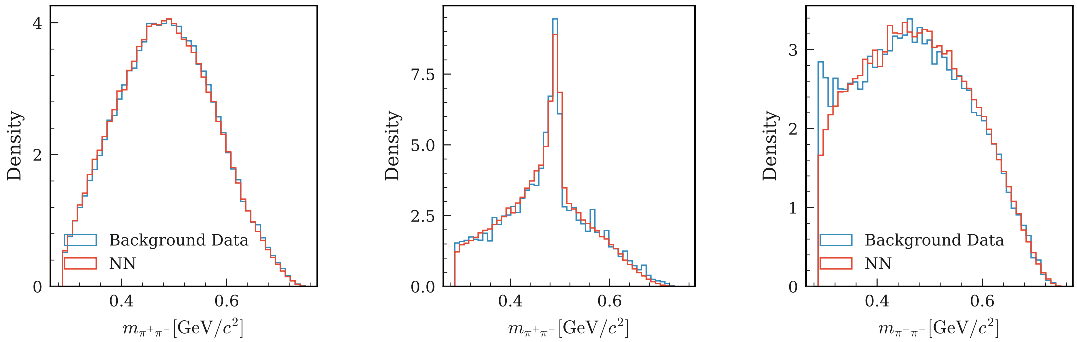
7.2.1 Neural Network Results: Sliding-Window Approach

The training data is split into separate contributions based on data availability. The distributions of individual background contributions in the phase-space variables are detailed in Figure 3.4. As defined in Table 7.1 the biggest background contribution $\tau \to \pi^-\pi^-\pi^+\pi^0\nu_\tau$ is individually modeled for $m_{3\pi} \in [0.8, 1.38]\,\mathrm{GeV}/c^2$. In the lower $m_{3\pi}$ region within $m_{3\pi} \in [0.8, 0.98]\,\mathrm{GeV}/c^2$, the τ_{Bkg} component is individually modeled, and in the intermediate $m_{3\pi}$ range, defined as $m_{3\pi} \in [0.98, 1.38]\,\mathrm{GeV}/c^2$ we include the individual contributions from $e^+e^- \to q\bar{q}$. In the upper $m_{3\pi}$ region, all background contributions are combined in a single component due to limited training sample size, preventing a reliable description of individual contributions.

Table 7.1: Training ranges for individual background-model contributions in the sliding-window split-background modeling approach.

$m_{3\pi}$ **range** [GeV/c^2]	**Background-Model Contribution**
0.80 – 0.98	$\tau \to \pi^-\pi^-\pi^+\pi^0\nu_\tau$ τ_{Bkg} all remaining contributions
0.98 – 1.38	$\tau \to \pi^-\pi^-\pi^+\pi^0\nu_\tau$ $e^+e^- \to q\bar{q}$ all remaining contributions
1.38 – 1.70	all combined

The invariant mass $m_{\pi^+\pi^-}$ of the background modeled using neural networks (red)



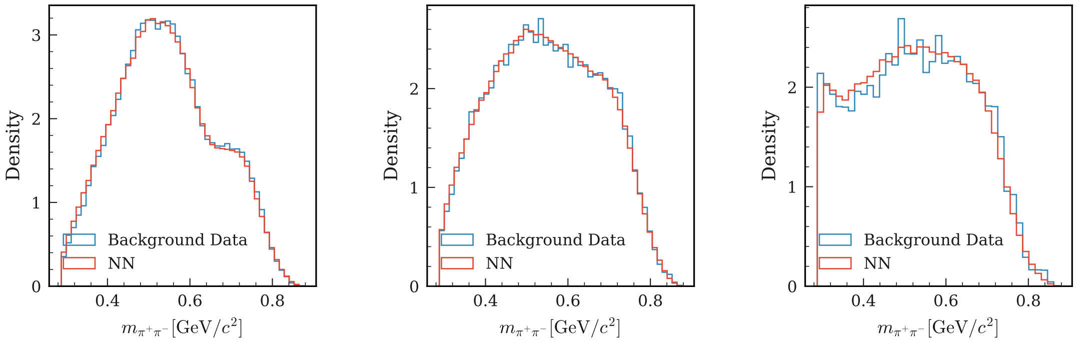
(a) $\tau \to \pi^-\pi^-\pi^+\pi^0\nu_\tau$ (b) τ_{Bkg} (c) Rest

Figure 7.8: Background distributions as estimated by the neural networks from the sliding-window split-background modeling approach (red) compared to the background distribution (blue) in $m_{\pi^+\pi^-}$ for $m_{3\pi} \in [0.80,\, 0.98]\,\text{GeV}/c^2$ for the explicitly modeled background $\tau \to \pi^-\pi^-\pi^+\pi^0\nu_\tau$ contributions and τ_{Bkg} and all remaining contributions denoted as rest.

trained in the range $m_{3\pi} \in [0.80, 0.98]\,\text{GeV}/c^2$ is shown for all three background contributions in Figure 7.8. By separating the background contributions, the previously discussed K^0 peak can be accurately modeled in the lower $m_{3\pi}$ range, unlike in the combined approach (Section 7.1). However, a peak in the neural network trained on everything except $\tau \to \pi^-\pi^-\pi^+\pi^0\nu_\tau$ and τ_{Bkg} (Figure 7.8, (c)) is not described at low $m_{\pi^+\pi^-}$ edge. These are likely coming from $\gamma \to e^+e^-$ conversion, where e^+e^- are misidentified as two pions and, when combined with an additional pion, show a similar signature as our signal decay of $\tau \to 3\pi\nu_\tau$, as described in Section 3.5.
The dominant $\tau \to \pi^-\pi^-\pi^+\pi^0\nu_\tau$ contribution is well described in both, the lower $m_{3\pi}$ range (Figure 7.8) and the intermediate $m_{3\pi}$ range. The latter is shown in Figure 7.9 for $m_{\pi^+\pi^-}$, using a neural network trained in $m_{3\pi} \in [1.07, 1.25]\,\text{GeV}/c^2$.
In the intermediate range, the neural network trained on $e^+e^- \to q\bar{q}$, as well as the neural network trained on all remaining contributions not explicitly modeled combined, are in good agreement with the background data, as shown in Figure 7.9. However, the latter neural network does not accurately model a small peak at the K_S mass, indicating that the distinct K_S signature is still not fully described in this approach.

7.2.2 Input-Output Studies including Background Components

The performance of the background modeling in the PWD fit is again tested by performing input-output studies on signal and background pseudo data, with background components included in the PWD likelihood in Equation (4.25).
In total, across the full-$m_{3\pi}$-range, we separate the background contributions 6 different ways: individually modeling the $\tau \to \pi^-\pi^-\pi^+\pi^0\nu_\tau$, the τ_{Bkg} and the $e^+e^- \to q\bar{q}$ con-



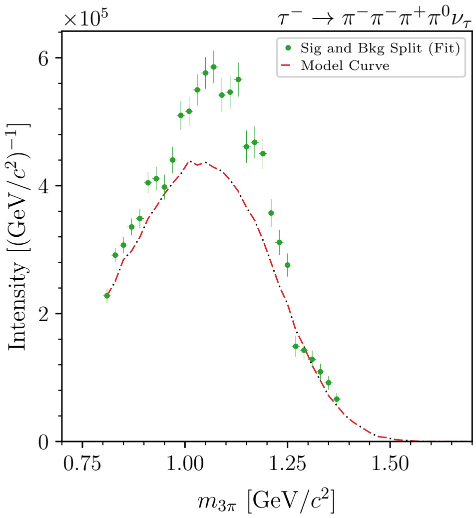
(a) $\tau \to \pi^-\pi^-\pi^+\pi^0\nu_\tau$ (b) $e^+e^- \to q\bar{q}$ (c) Rest

Figure 7.9: Background distributions as estimated by the neural networks from the sliding-window split-background modeling approach (red) compared to the background data distribution (blue) in $m_{\pi^+\pi^-}$ for the $m_{3\pi}$ training range of $m_{3\pi} \in [1.07,\ 1.25]\,\text{GeV}/c^2$ for the individually modeled background contributions of $\tau \to \pi^-\pi^-\pi^+\pi^0\nu_\tau$, $e^+e^- \to q\bar{q}$ and all remaining contributions, denoted as rest.

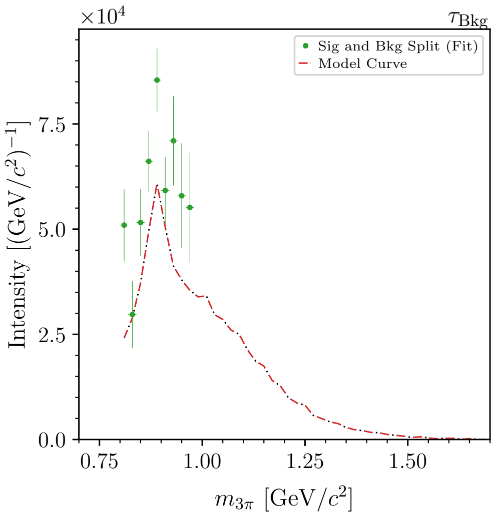
tributions and three different combined contributions. Their $m_{3\pi}$ ranges are defined in Table 7.1.

The fit results (green) for 4 of these background types are shown in Figure 7.10, together with the true background yield of the pseudo data (red). In the first $m_{3\pi}$ range, defined in Table 7.1, the τ_{Bkg}, as well as the $\tau \to \pi^-\pi^-\pi^+\pi^0\nu_\tau$ component, are slightly overestimated compared to the true background yield, as shown in Figure 7.10b. Within the same range, the fitted yield of the background component of the combined rest contributions is well described. The excess of intensity in the background yields within this $m_{3\pi}$ range indicates a small leakage of signal into background contributions.

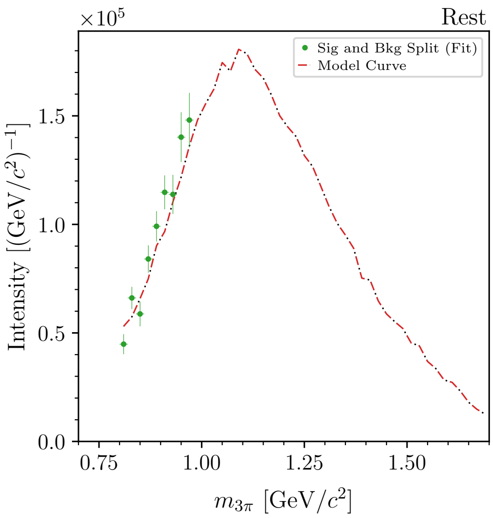
In the second $m_{3\pi}$ range defined in Table 7.1, the dominant $\tau \to \pi^-\pi^-\pi^+\pi^0\nu_\tau$ background is largely overestimated compared to the true background yield. The biggest discrepancy is in $m_{3\pi} \in [1.0, 1.16]\,\text{GeV}/c^2$, where the PWD yields a biased yield that is enhanced by $\approx 2 \cdot 10^5\,[\text{GeV}/c^2]^{-1}$, i.e., 1.5 times higher than the true yield. Within the same range, the background yield for the combined rest is approximated accurately compared to the true background yield of the pseudo data, with two outliers for the fit results in $m_{3\pi} \in [0.98, 1.02]\,\text{GeV}/c^2$, where the background yields are overestimated. Additionally, the background yields for the $q\bar{q}$ yields are within the expected range but exhibit large fluctuations between adjacent bins and large statistical uncertainties. The interplay between the $q\bar{q}$ component and the $\tau \to \pi^-\pi^-\pi^+\pi^0\nu_\tau$ component is further discussed in Section 8.1.2. The overall excess of intensity in the background contributions within this range indicates leakage of signal contributions into the background contributions, most dominantly in the $\tau \to \pi^-\pi^-\pi^+\pi^0\nu_\tau$, which will be further discussed in Section 8.1.



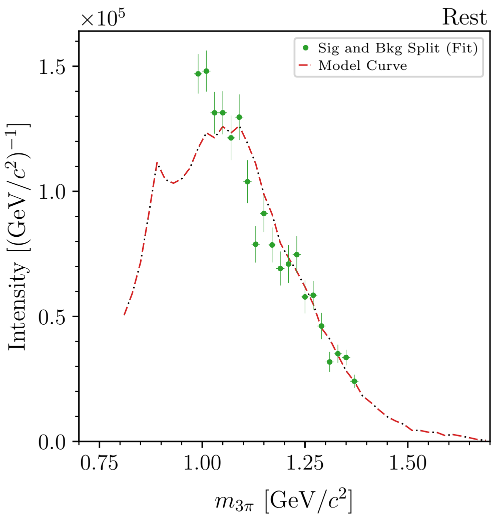
(a) Only $\tau \to \pi^-\pi^-\pi^+\pi^0\nu_\tau$



(b) Only τ_{Bkg}



(c) All expect $\tau \to \pi^-\pi^-\pi^+\pi^0\nu_\tau$ and τ_{Bkg}



(d) All except $\tau \to \pi^-\pi^-\pi^+\pi^0\nu_\tau$ and $q\bar{q}$

Figure 7.10: Fitted (green) and true (red) background yields for 4 different background components for the split background approach.

The last range defined in Table 7.1 uses the same neural networks as the combined background modeling approach. These results have already been discussed in Section 7.1.
Seven out of the nine partial waves of the input-output study are well described by the background-subtracted partial wave fit results using the split background modeling approach. Examples for the background-subtracted fit results (green) are shown in Figure 7.11. The $1^+\,[\rho(770)\,\pi]_S$ and the $1^+\,[\rho(770)\,\pi]_D$ agree well between the fit results (green) and the model curve (red). However, the $1^+\,[(\pi\pi)_S\,\pi]_P$ wave is consistently underestimated for $m_{3\pi} \in [1.04, 1.22]$ except for one result in $m_{3\pi} \in [1.14, 1.16]\,\mathrm{GeV}/c^2$. Furthermore, the background introduces a small bias for $m_{3\pi} \in [1.06, 1.12]$ in the $1^+\,[\rho(1450)\,\pi]_D$ wave. Thus, split background modeling introduces a bias in the description of partial-wave amplitudes.
We perform an additional background-subtracted PWD fit with a fixed $\tau \to \pi^-\pi^-\pi^+\pi^0\nu_\tau$ component to the true value of the pseudo data in order to assess the impact of the overestimation of the $\tau \to \pi^-\pi^-\pi^+\pi^0\nu_\tau$ component on the partial waves results. In Figure 7.11 the background-subtracted PWD fit results with the fixed $\tau \to \pi^-\pi^-\pi^+\pi^0\nu_\tau$ component (brown) for the $1^+\,[(\pi\pi)_S\,\pi]_P$ and $1^+\,[\rho(1450)\,\pi]_D$ waves are in good agreement with the signal-only PWD fits (blue). All other partial waves are equally well modeled. The background yields for the PWD fit with the $\tau \to \pi^-\pi^-\pi^+\pi^0\nu_\tau$ component fixed to the true background yield are identical to the PWD fit performed without the fixed $\tau \to \pi^-\pi^-\pi^+\pi^0\nu_\tau$ yield. This suggests an interplay between the $\tau \to \pi^-\pi^-\pi^+\pi^0\nu_\tau$ component and the $1^+\,[(\pi\pi)_S\,\pi]_P$ wave which is further discussed in Section 8.1 and Section 8.2.
These results indicate a better description of the partial waves for the sliding-window approach with the $\tau \to \pi^-\pi^-\pi^+\pi^0\nu_\tau$ yield fixed to the true pseudo data value.

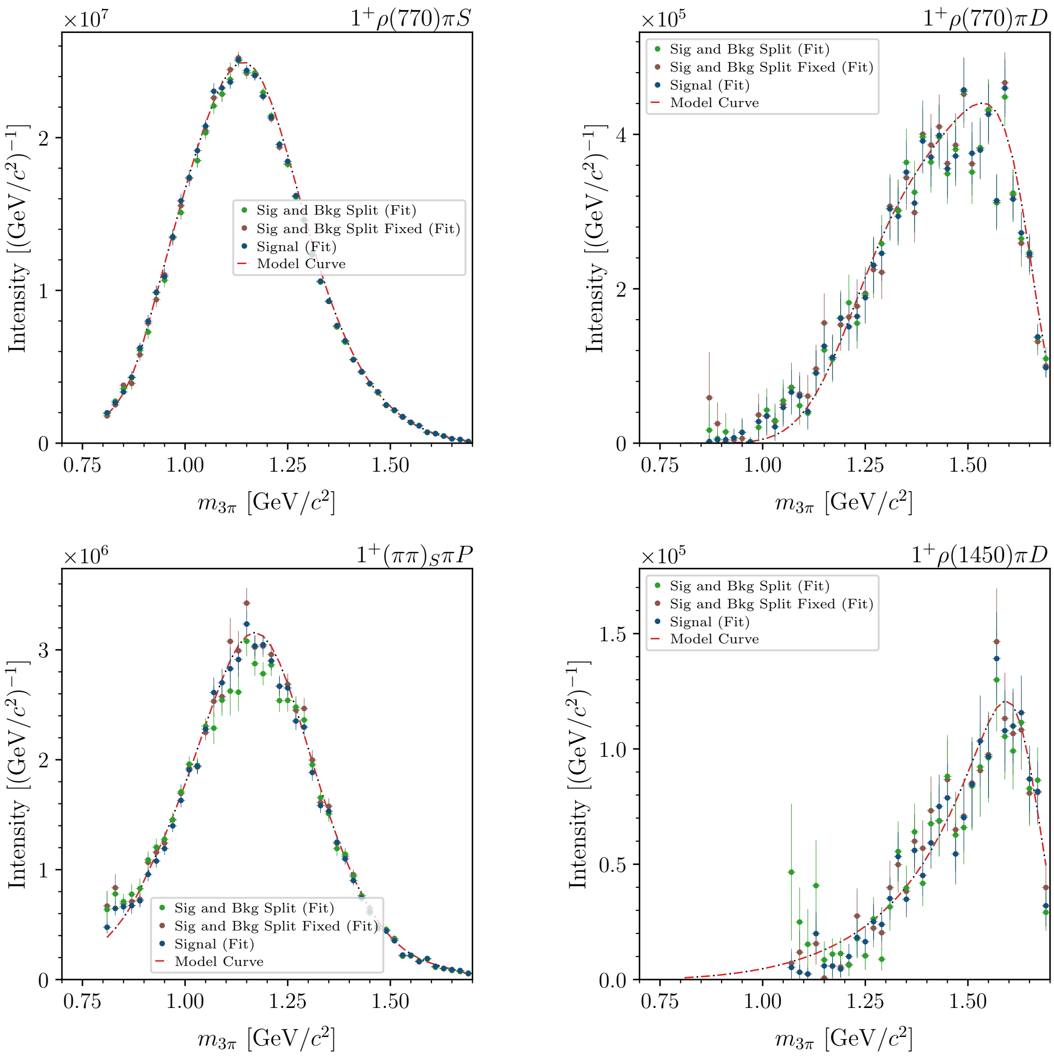


Figure 7.11: The background-subtracted PWD fits for the sliding-window approach (green), compared to PWD fits on signal-only pseudo data (blue) and the signal model curve (red).

8 PWD Studies

To understand how our modeling choices affect the PWD fit results, we conduct input-output studies by deliberately changing the input pseudo data or varying model components. We analyze the leakage of background into signal partial waves and vice versa in Section 8.1. Furthermore, we study resolution effects on signal-only (Section 8.2.1), as well as on signal and background (Section 8.2.2) pseudo data, followed by a discussion about the impact of contributions from K_S decays in Section 8.3. Finally, we explore the consequences of imperfectly modeling the background yields in Section 8.4.
The studies, including background contributions, are shown exclusively for the sliding window binning approach of the combined (see Section 7.1) and split (see Section 7.1.2) background modeling, not for the full $m_{3\pi}$-range approach.

8.1 Leakage Studies

The signal and background contributions may not always be perfectly separable in the PWD fit. We want to analyze whether, in which PWD component, and in which $m_{3\pi}$ regions, signal and background contributions might leak from the signal into the background component, or vice versa. We test this effect by fitting reconstructed pseudo data that includes either signal or background contributions only, while still including both signal and background components in the PWD fit. Leakage studies of signal into background contributions are discussed in Section 8.1.1 and the PWD results using background-only pseudo data are detailed in Section 8.1.2.

8.1.1 PWD Fits to Signal Only Data

To quantify whether signal data leaks into background components, we perform PWD fits to signal-only pseudo data while still including the background components as free parameters. Since the pseudo data does not contain any background events, any non-zero background yield corresponds to signal contributions that can be described as the specific background components, and thus quantify the leakage of signal contributions in the description of the background.
The results of this study are shown in Figure 8.1.

Combined Background Modeling

From the background yield results in Figure 8.1, we observe that including the background components in the PWD fit using the combined background modeling approach

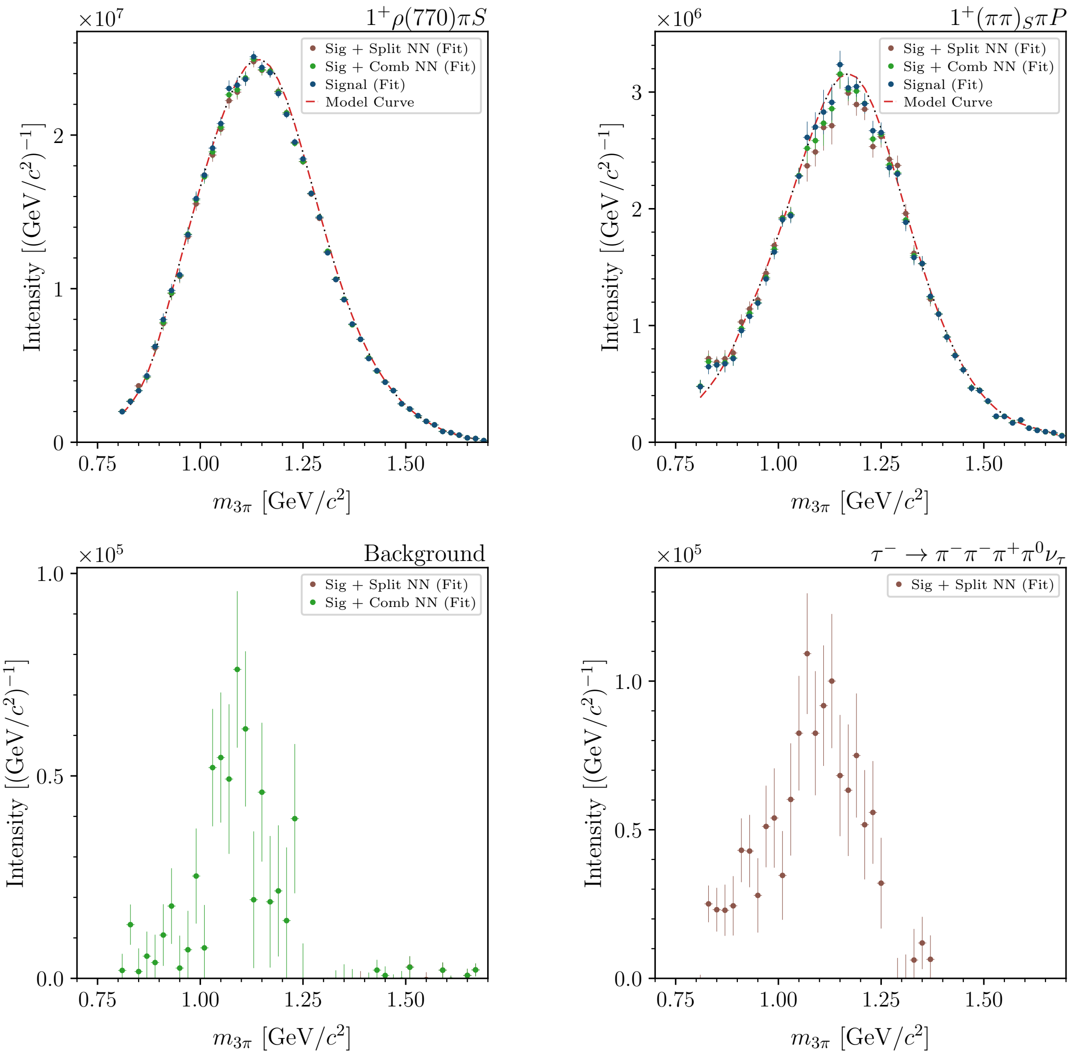


Figure 8.1: PWD fit results performed on signal-only pseudo data, including signal and combined (green) or split (brown) background components, compared to the PWD fit performed on signal-only reconstructed pseudo data using the signal component only (blue) and the input-model curve (red). The results are shown for the $1^+[\rho(770)\pi]_S$ wave, the $1^+[(\pi\pi)_S\pi]P$ wave, as well as the $\tau \to \pi^+\pi^-\pi^-\pi^0\,\nu_\tau$ and the combined background yield.

(gree), within $m_{3\pi} \in [1.04, 1.24]\,\mathrm{GeV}/c^2$ some signal contributions leak into the background component, corresponding to an intensity of $\approx 5 \cdot 10^4\,[\mathrm{GeV}/c^2]^{-1}$. Within the same $m_{3\pi}$ range, fit results for the $1^+[(\pi\pi)_S\pi]P$ wave minimally underestimate the truth signal. However, since the partial wave intensity is around one to two orders of magnitude larger than the observed leakage into the background contributions, this bias is considered to be negligible. All other partial waves show no deviations between the signal-only PWD fit, including only the signal component, and the PWD fit, including also the combined background component. An example is given in Figure 8.1 for the $1^+[\rho(770)\pi]_S$.

Split Background Modeling

The $\tau \to \pi^+\pi^-\pi^-\pi^0\,\nu_\tau$ background yield shown in Figure 8.1 shows non-negligible intensities in the PWD fit results on signal pseudo data (brown), which include signal and background components using the split background modeling approach within $m_{3\pi} \in [0.90, 1.24]\,\mathrm{GeV}/c^2$. In the same $m_{3\pi}$ range, the intensity of the $1^+[(\pi\pi)_S\pi]P$ wave is underestimated and deviates from the PWD fit results, including only the signal component (blue).
This means that within the $m_{3\pi} \in [0.90, 1.24]\,\mathrm{GeV}/c^2$ signal contributions potentially leak into the description of the $\tau \to \pi^+\pi^-\pi^-\pi^0\,\nu_\tau$ background contributions. This leakage in the background yield is up to twice as large as the leakage observed in the combined background modeling approach (green).
We have also seen an overestimation of $\tau \to \pi^+\pi^-\pi^-\pi^0\,\nu_\tau$ background component when fitting to signal and background pseudo data using the split background modeling approach (see Figure 7.10). This observed overestimation in the $\tau \to \pi^+\pi^-\pi^-\pi^0\,\nu_\tau$ background yield comes from the leakage effect of signal into the background contribution, observed in Figure 8.1.
The remaining background yields and all other partial waves in the split background modeling approach show only small leakage with observed intensities mostly below $\approx 2 \cdot 10^4\,[\mathrm{GeV}/c^2]^{-1}$ with individual outliers in the $e^+e^- \to q\bar{q}$ background yield of $\approx 5 \cdot 10^4\,[\mathrm{GeV}/c^2]^{-1}$.

8.1.2 PWD Fits to Background Only Data

To test leakage of background contributions into signal components and to assess how accurately we fit background contributions, we perform input-output studies on background-only pseudo data with signal and background components in the PWD fit. The intensity observed in the partial waves corresponds to leakage of background into the signal contributions.

Combined Background Modeling

The background yield for the PWD fit using the combined background modeling approach is shown in Figure 8.2 for background-only pseudo data. The background yield is accurately determined over the full $m_{3\pi}$ range, i.e., the yields estimated by the PWD

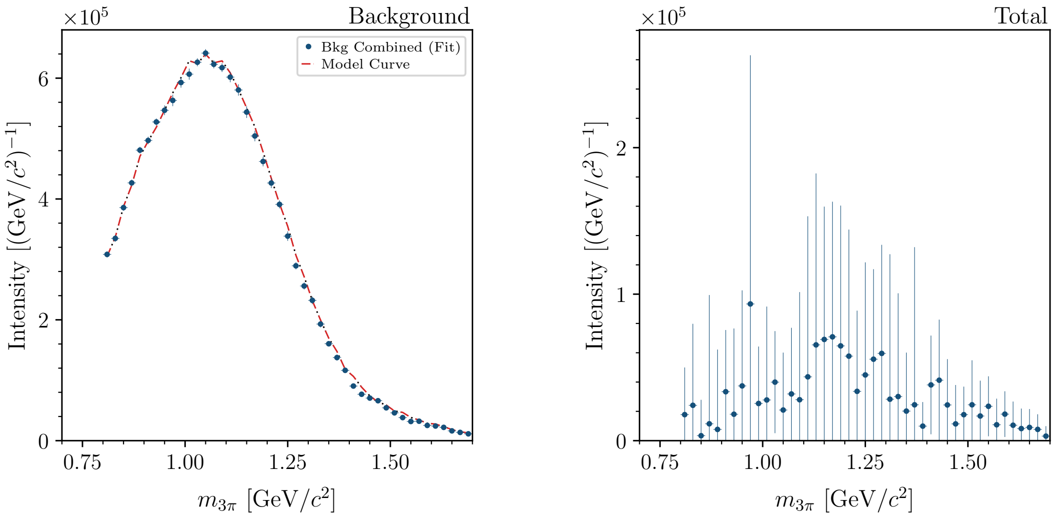


Figure 8.2: The PWD fit results performed on background-only pseudo data for the background yield as well as the total intensity of the combined background modeling approach.

fit (blue) and the true background yield from the pseudo data (red) match well.
The PWD fit shows some intensity in each partial wave, with the majority below $2 \times 10^4\,[\mathrm{GeV}/c^2]^{-1}$. Individual outliers can reach values up to $1 \times 10^5\,[\mathrm{GeV}/c^2]^{-1}$. The most sensitive partial waves are the $1^+[\rho(770)\pi]_D$ wave, the $1^+[f_2(1270)\pi]_P$ wave, the $1^+[\rho(1450)\pi]_D$ wave and the $1^+[(\pi\pi)_S\pi]P$. These fit results have large uncertainties, such that they agree with 0. This indicates that the leakage effect is strongly amplified by the lack of signal in the data, as we observe leakage especially in regions where the individual partial waves have low intensities and are not well constrained.
The total signal intensities of the PWD fit results are shown Figure 8.2 (right), which have a central value of $\approx 0.5 \cdot 10^5\,[\mathrm{GeV}/c^2]^{-1}$, and large uncertainties. These studies indicate that the leakage of background into signal components is fairly small for the combined background modeling approach.

Split Background Modeling

The same study is also done using the split background modeling approach, i.e., we perform a PWD on background-only pseudo data, which includes both signal and background components. We observe leakage in all partial waves. An example of this PWD fit (blue) in the partial wave of $1^+[\rho(1450)\pi]_D$ is shown in Figure 8.3 (bottom right), which shows intensities around $\approx 4 \times 10^4\,[\mathrm{GeV}/c^2]^{-1}$ to $\approx 9 \times 10^4\,[\mathrm{GeV}/c^2]^{-1}$ within $m_{3\pi} \in [0.98, 1.20]\,\mathrm{GeV}/c^2$. This reduces for larger $m_{3\pi}$. Additionally, three different background yields are shown for the same PWD fit, corresponding to the background contributions individually modeled as described in Table 7.1 within the middle $m_{3\pi}$ range.

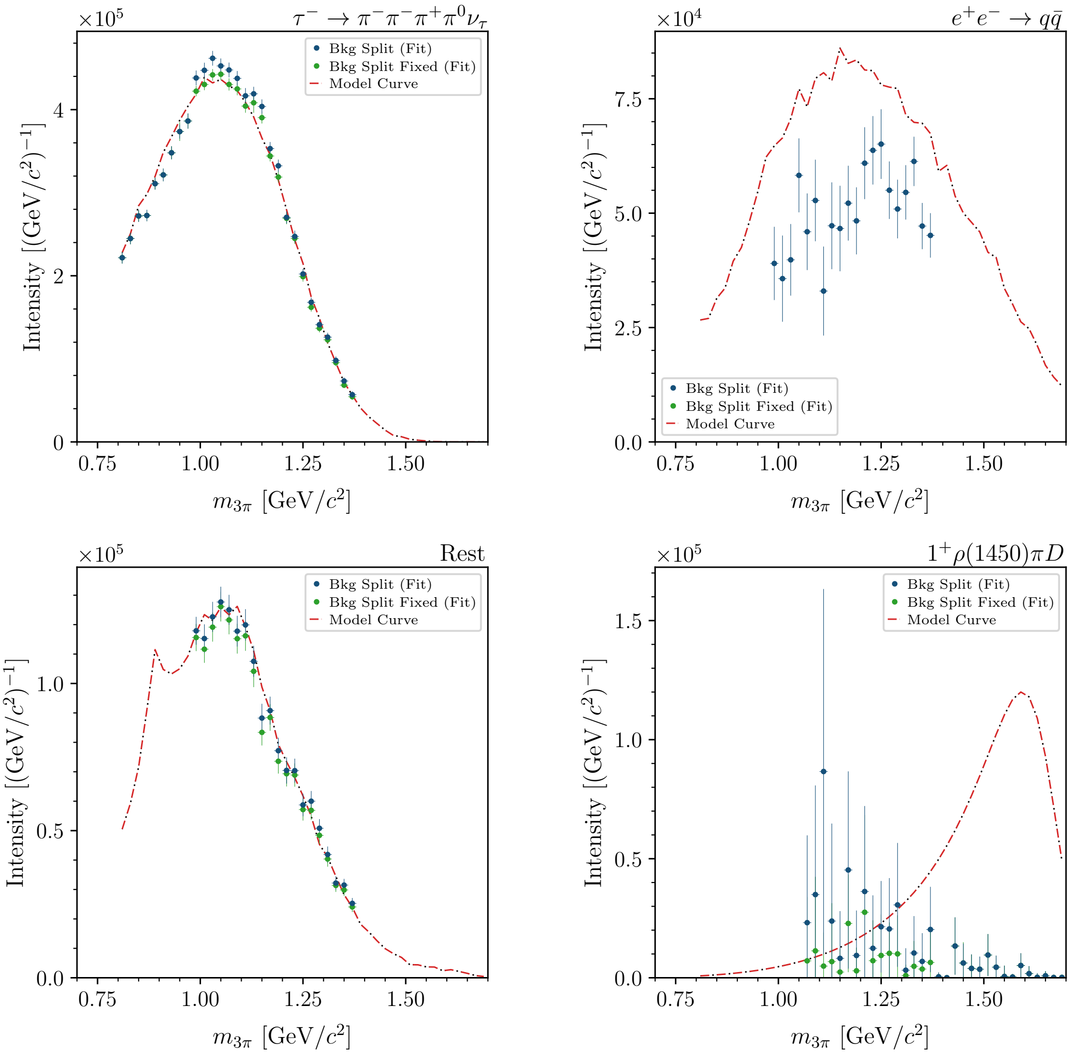


Figure 8.3: The PWD fit results (blue) performed on background-only pseudo data for the split background modeling approach for the background yields of the $\tau \to \pi^+\pi^-\pi^-\pi^0\,\nu_\tau$ (upper left), the $e^+e^- \to q\bar{q}$ (upper right), and the component denoted as *rest* (bottom left) within the middle $m_{3\pi}$ region of the split background modeling (see Table 7.1). Furthermore, the partial wave $1^+[\rho(1450)\pi]_D$ is shown on the bottom right. The PWD fit result on background-only pseudo data with a fixed $e^+e^- \to q\bar{q}$ component (green) is shown.

Within the same range of $m_{3\pi}$ the background yield of the $\tau \to \pi^+\pi^-\pi^-\pi^0\,\nu_\tau$ contribution is slightly overestimated. While the yield of the *rest* component matches the truth value of the pseudo data within that $m_{3\pi}$ range, the $e^+e^- \to q\bar{q}$ background yield is underestimated. To assess the correlation of the $\tau \to \pi^+\pi^-\pi^-\pi^0\,\nu_\tau$ and the $e^+e^- \to q\bar{q}$ components in the PWD fit, we perform PWD fits using the same configuration, with the e^+e^- background yields fixed to the data truth values.
When fixing the $e^+e^- \to q\bar{q}$ (green) component to its true MC value, the overestimation of the $\tau \to \pi^+\pi^-\pi^-\pi^0\,\nu_\tau$ background component is slightly reduced, leading to a much better approximation of the background yield compared to the true value (red). Furthermore, the leakage in all partial waves is also reduced, shown in Figure 8.3 for the $1^+[\rho(1450)\pi]_D$ wave.
The second background component, denoted as *rest*, i.e., all contributions excluding $e^+e^- \to q\bar{q}$ and $\tau \to \pi^+\pi^-\pi^-\pi^0\,\nu_\tau$, is not impacted by fixing the $e^+e^- \to q\bar{q}$ background yield.

8.2 Resolution Studies

In the partial-wave decomposition formalism described in Section 4.2, detector resolution effects, which lead to a smearing of the phase-space variables, are neglected, since the detector response is only available from Monte Carlo simulations, and we do not have a known analytical parametrization. However, the data we fit to are reconstructed (either reconstructed pseudo data or real data) and therefore biased by neglecting the detector resolution in the PWD formalism.
Thus, we analyze the effect of neglecting resolution in the PWD model for both signal-only (in Section 8.2.1) and signal and background contributions (in Section 8.2.2).

8.2.1 Resolution Studies on Signal-Only Pseudo Data

To quantify the impact of neglecting resolution in the PWD formalism, the fit is performed on signal pseudo data that uses MC-Truth variables, which do not include resolution effects. These results are compared to signal pseudo data using reconstructed variables, which have a realistic approximation of the resolution from MC simulation (see Section 4.4). Thus, the bias introduced by disregarding resolution in the PWD formalism is quantified by comparing these fit results. In most cases, the input models are equally well reproduced by the MC-Truth and the reconstructed variables affected by the resolution. However, we see two small biases in the waves $1^+\,[\rho(770)\,\pi]_S$ and $1^+\,[f_2(1270)\,\pi]_P$, which are shown in Figure 8.4, along with the pull (defined in Section 4.4) distribution of both fit results.
In the $1^+\,[\rho(770)\,\pi]_S$ wave, a small bias is introduced in $m_{3\pi} \in [1.16, 1.26]$ in the PWD fit, neglecting the resolution (green). The corresponding pulls in this $m_{3\pi}$ interval reach values of up to ≈ 5 standard deviations, whereas they are significantly smaller for the PWD fit performed using signal MC-truth variables (blue). In the remaining $m_{3\pi}$ range, the fit results are in good agreement with each other.
Additionally, a bias is introduced in the $1^+\,[f_2(1270)\,\pi]_P$ wave in $m_{3\pi} \in [1.16, 1.42]$ where

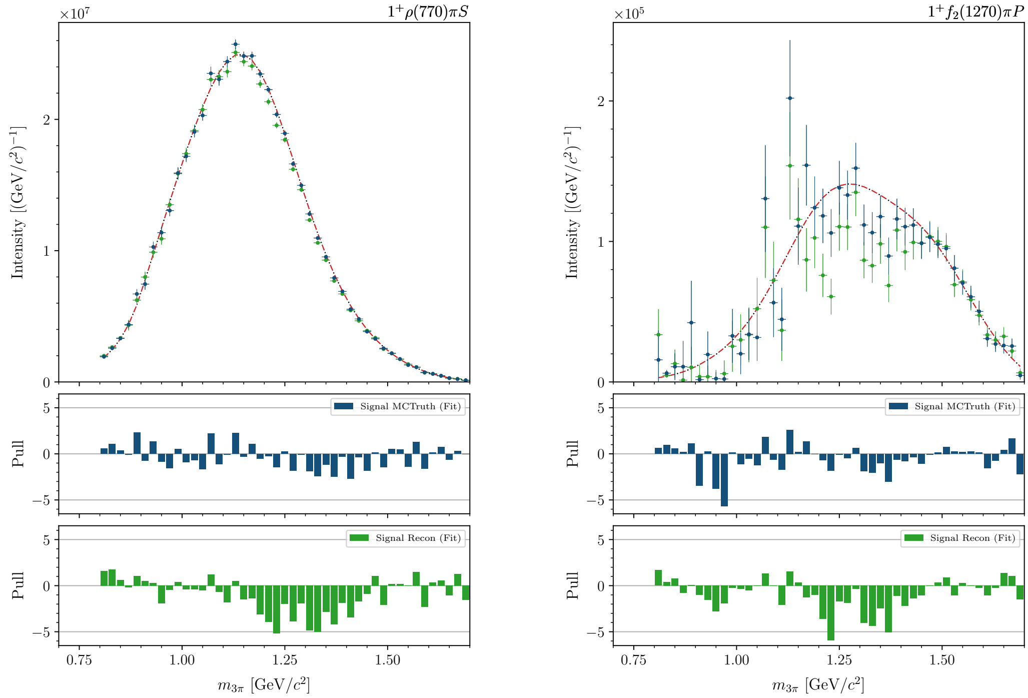


Figure 8.4: PWD fit results performed on signal MC-Truth pseudo data (blue) and reconstructed pseudo data (green) for the $1^+\,[\rho(770)\,\pi]_S$ wave and the $1^+\,[f_2(1270)\,\pi]_P$ wave, along with the pull distributions for both fit results.

the intensity is slightly underestimated compared to the input model. This is again visible in the pull distribution, which shows values of $\approx 4-5$ standard deviations for the reconstructed variables, as opposed to $\approx 2-3$ standard deviations using the MC-Truth variables.

These observations indicate that the impact of neglecting resolution effects is limited. Since only two partial waves exhibit small deviations while all other waves are negligible, we consider the approximation of neglecting resolution acceptable.

A small number of outliers are observed in the pull distributions of the PWD fits using the signal MC-truth variables for both partial-wave examples shown in Figure 8.4. These outliers are likely caused by an inaccurate estimation of the uncertainties in low-intensity regions, which results in artificially large pull values.

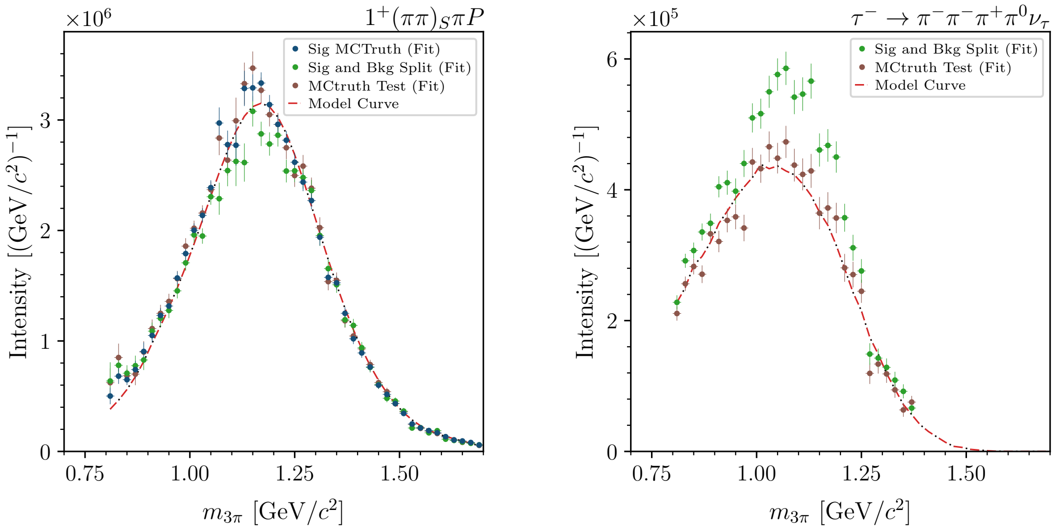


Figure 8.5: PWD fit results performed on signal and background MC-Truth variables (brown), compared to PWD fit on signal-only MC-Truth variables (blue), and the model curve (red), as well as on reconstructed variables (green) using the split background modeling approach. These results are shown in the $1^+[(\pi\pi)_S\pi]P$ wave and the $\tau \to \pi^+\pi^-\pi^-\pi^0\,\nu_\tau$ background component.

8.2.2 Resolution Studies on Signal and Background Pseudo Data

We also want to assess the impact of neglecting resolution in the PWD formalism when background contributions are included in the pseudo data. Thus, we perform PWD fits on the MC-Truth variables and compare them to fit results using reconstructed variables.

Split Background Modeling

Figure 8.5 shows the result for this study of the split background modeling approach for the $1^+[(\pi\pi)\pi]P$ wave and the dominant $\tau \to \pi^+\pi^-\pi^-\pi^0\,\nu_\tau$ background components. The remaining background components and partial waves show no impact and are thus not shown.

For a better assessment of the results, Figure 8.5 includes the PWD fit results of signal MCTruth pseudo data (blue), along with the corresponding model curve (red) and the split background modeling PWD fit on signal and background reconstructed pseudo data (green), which are also shown in Section 7.2.2. Since there are no significant deviations in the fit results across all other partial waves, only the $1^+[(\pi\pi)\pi]P$ wave is shown.

The results of the signal and background MC-Truth PWD fit (brown) and the PWD fit on MCTruth signal pseudo data are in good agreement in the $1^+(\pi\pi)\pi P$ wave. Additionally, the bias in the background yield of $\tau \to \pi^+\pi^-\pi^-\pi^0\,\nu_\tau$ observed for the split background modeling (discussed in Section 7.2.2) is vastly reduced for the signal and background MC-Truth PWD fit results within $m_{3\pi} \in [0.98, 1.2]\,\mathrm{GeV}/c^2$. This suggests

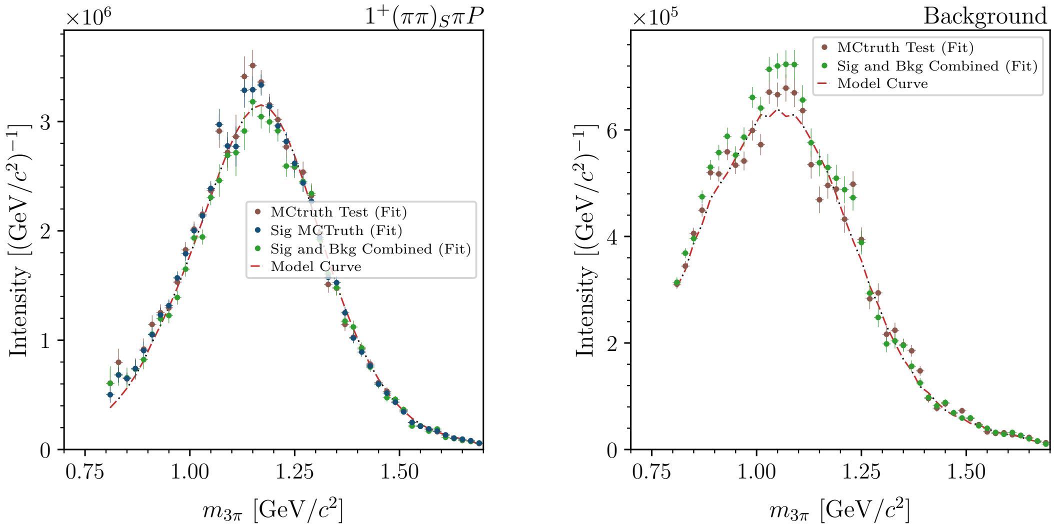


Figure 8.6: PWD fit results performed on signal and background MC-Truth variables (brown), compared to PWD fit on signal-only MC-Truth pseudo data (blue), and the model curve (dashed red line), as well as on reconstructed variables (green) using the combined background modeling approach. These results are shown in the $1^+[(\pi\pi)\pi]P$ wave and the combined background component.

that the previously observed overestimation in the $\tau \to \pi^+\pi^-\pi^-\pi^0\,\nu_\tau$ background yield is induced by the neglect of resolution effects in the PWD formalism. This indicates that the PWD fit absorbs resolution-induced imperfect modeling of the partial waves by absorbing contributions into the $\tau \to \pi^+\pi^-\pi^-\pi^0\nu_\tau$ background component.

Combined Background Modeling

Figure 8.6 shows equivalent studies (brown) on the impact of neglecting resolution for the $1^+[(\pi\pi)\pi]P$ and the combined background modeling approach (green), including the PWD fit on signal MC Truth variables (blue) and the corresponding model curve (red).

The results for the combined background modeling approach show a similar behavior. The results of the signal and background PWD fit using MCTruth variables and the signal-only PWD fit on MCTruth variables are in good agreement in the $1^+[(\pi\pi)\pi]P$ wave. The combined background modeling approach deviates slightly less from the model curve in Additionally, within $m_{3\pi} \in [1.04, 1.1]\,\mathrm{GeV}/c^2$, the description of the combined background component is improved for the signal and background MC-Truth PWD fit results compared to the PWD fit on reconstructed variables (green), yielding a background yield closer to the model curve (red). This suggests that the deviations observed in the reconstructed-variable fit are induced by the neglect of resolution effects in the PWD formalism. This suggests that the PWD fit absorbs resolution-induced mismodeling into the combined background component to improve the PWD likelihood.

8.3 K^0 studies

Both the combined and split background modeling results, discussed in Section 7.1.2 and Section 7.2.1, indicate that the distinct distribution in the two-pion invariant mass spectrum from $\tau^- \to K^0\pi^-\nu_\tau$ decays is not fully modeled by the neural networks. In order to estimate the impact of not modeling this specific background, we perform PWD fits on pseudo data, where we cut out the two two-pion masses $m_{\pi^+\pi^-}$ within $m_{\pi^+\pi^-} \in [0.485, 0.51]\,\mathrm{GeV}/c^2$ around the nominal mass of the K^0. This is denoted as the K^0 cut. We show this study only for the combined background modeling approach, as the split background modeling study shows similar results.
In this study, the PWD cut is consistently applied to all data used in the PWD fit, i.e., it is part of the event selection. However, the neural networks are trained on data including the K^0 background, since we want to include the full background information during training. Therefore, this study is only an estimation of the impact of the K^0 background and is not fully consistent.
The PWD fit results of the combined background modeling approach performed on reconstructed pseudo data with the K^0 cut applied are comparable to the PWD fit results without the additional cut, discussed in Section 7.1.4, i.e., the PWD fit with the K^0 cut shows no improvement. All partial waves are equally well described with and without the K^0 cut, and the results do not show any systematic deviations.
As an example, the fit results (blue) for $1^+[(\pi\pi)_S\pi]P$ wave are shown in Figure 8.7, and they are very consistent with the model curve (red).
The fitted background yield deviates from the expected yield in $m_{3\pi} \in [0.96, 1.2]\,\mathrm{GeV}/c^2$, which coincides with the $m_{3\pi}$ region that is sensitive to leakage between signal and background contributions (see Section 8.1). The leakage effect might be amplified in this study due to the inconsistency in the application of the cut, since the neural networks are trained on data where we do not apply the K^0 cut.
These observed partial wave results are consistent with the partial wave fit results obtained with pseudo data, including the $m_{\pi^+\pi^-}$ region of the K^0 cut. Thus, not modeling this specific background seems to have a small impact on the description of the partial waves.

8.4 Impact of ± 10 % Wrong Background Yields

In order to assess the impact on the partial wave description arising from uncertainties due to the potentially incorrect modeling of individual background yields in the PWD, we perform PWD fits with fixed background yields that are varied by $+10\%$ or -10% with respect to their true values. These values are chosen to be conservative compared to the uncertainties of the branching fraction [3] of the signal $\tau \to \pi^+\pi^-\pi^-\,\nu_\tau$ and the dominant $\tau \to \pi^+\pi^-\pi^-\pi^0\,\nu_\tau$ component. This is done for the combined and the split background modeling. In summary, we perform 4 PWD fits, once underestimating and

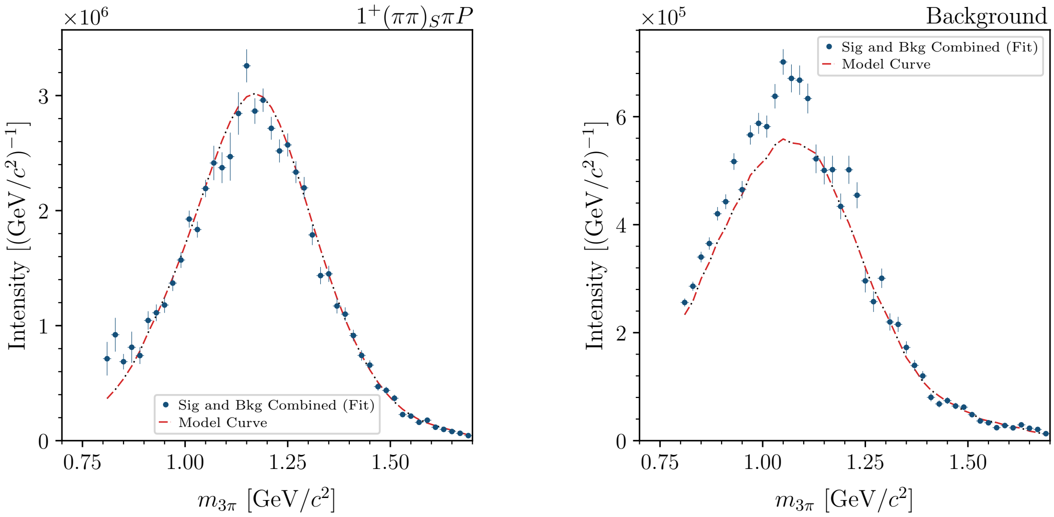


Figure 8.7: PWD fit results using the combined background modeling approach performed on reconstructed pseudo data with the K^0 cut applied (blue) compared to the model curve (red) in the $1^+[(\pi\pi)_S\pi]P$ wave and the combined background yield.

once overestimating the background yield for both background modeling approaches. For split background modeling, all individual background yields are set to be off by $\pm 10\%$. The results of overestimating (blue, green) and underestimating (brown, red) the background yield for the $1^+[(\pi\pi)_S\pi]P$, the $1^+[f_2(1270)\pi]P$, $0^-[\rho(770)\pi]P$, and the $1^-[\omega(782)\pi]P$ wave are shown in Figure 8.8. All other partial waves show no noticeable deviations between the fit results.

The intensity of the $1^+[(\pi\pi)_S\pi]P$ wave (top left in Figure 8.8) changes with varying background yield within $m_{3\pi} \in [1.06, 1.28]\,\mathrm{GeV}/c^2$ and in the low $m_{3\pi}$ tail of $m_{3\pi} \lesssim 0.96\,\mathrm{GeV}/c^2$. An underestimation of the background yield (blue, green) leads to an overestimation of $1^+[(\pi\pi)_S\pi]P$ wave intensity in the middle $m_{3\pi}$ range and a slight underestimation of the intensity at lower $m_{3\pi}$ compared to the model curve. This is exactly opposite in the case of underestimating the background yield (red, brown).

The $1^+[f_2(1270)\pi]P$ wave is slightly impacted by wrong background yields for most bins within $m_{3\pi} \in [098, 1.26]\,\mathrm{GeV}/c^2$. However, most biases are smaller than the statistical uncertainties, indicating that the impact of a wrongly estimated background yield of $\pm 10\%$ is small. Additionally, there are two outliers in the low $m_{3\pi}$ range. These are in the lower $m_{3\pi}$ range close to the kinematic limits of this wave, where the partial waves are generally not well constrained by data. An imperfect modeled background yield seems to enhance this effect, leading to the observed outliers.

The $0^-[\rho(770)\pi]P$ wave shows a systematic upward shift in the intensity if the background yield is underestimated for $m_{3\pi} \in [1.20, 1.44]\,\mathrm{GeV}/c^2$. Finally, the $1^-[\omega(782)\pi]P$ wave for $m_{3\pi} \lesssim 1.3\,\mathrm{GeV}/c^2$ is not affected by an imperfectly modeled background yield. Above that $m_{3\pi}$ until $m_{3\pi} \sim 1.52\,\mathrm{GeV}/c^2$, the PWD fit results overestimate its intensity

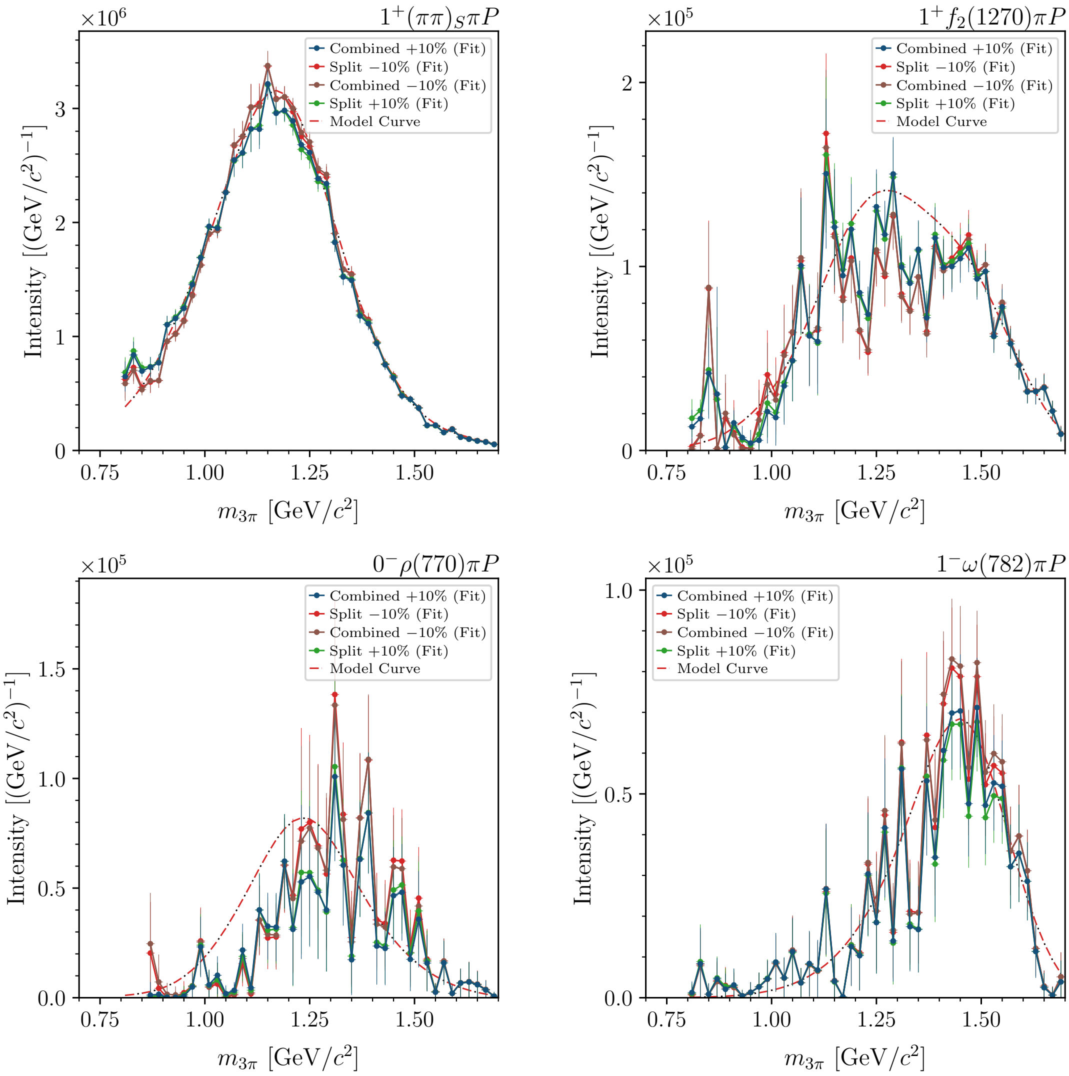


Figure 8.8: PWD fit results for the combined and split background modeling approach performed on reconstructed pseudo data with fixed background values overestimating the background yield by 10% (green, blue) and underestimating the background yield by 10% (brown, red) compared to the model curve (red) shown in the partial waves: $1^+[(\pi\pi)_S\pi]P$, $1^+[f_2(1270)\pi]P$, $0^-[\rho(770)\pi]P$, and $1^-[\omega(782)\pi]P$.

when the background yield is fixed to a too low value and vice versa.

However, this bias in the intensities of the $0^{-}[\rho(770)\pi]P$ and the $1^{-}[\omega(782)\pi]P$ wave are small compared to the uncertainties of the fit results. Thus, the effect of a wrongly modeled background yield is small.

9 Conclusion and Outlook

In this thesis, we investigate various neural-network-based approaches for the high-dimensional background modeling in the partial-wave analysis of the $\tau^- \to \pi^+\pi^-\pi^-\nu_\tau$ decay at Belle II. Furthermore, we integrate the neural networks into the partial-wave analysis formalism. To this end, we train feed-forward neural networks on simulated background and signal events and optimize the training procedure. We further test the ability of trained neural networks to account for background contributions in the partial-wave decomposition via input-output studies on MC-simulated events.

We find that poorly populated kinematic regions in signal MC samples lead to largely inflated neural network outputs (outliers), as the neural network overfits the unconstrained kinematic regions. These outliers dominate the PWD likelihood and introduce biases in the fitted partial-wave amplitudes. To mitigate these outliers, we implement regularization techniques and introduce a phase-space cut that excludes regions poorly populated by signal, removing only a negligible fraction of signal events.
We also explore the impact of different $m_{3\pi}$ ranges, on which each neural network is trained, on the PWD fit results. We find that broader training $m_{3\pi}$ regions of width $0.18\,\mathrm{GeV}/c^2$ (sliding-window approach) and $0.9\,\mathrm{GeV}/c^2$ (full $m_{3\pi}$ range) improve the stability of the neural network predictions compared to finely binned training (width $0.02\,\mathrm{GeV}/c^2$) by reducing statistical fluctuations and discontinuities between neighboring mass bins. While the training on the full $m_{3\pi}$ range eliminates these fluctuations, we find a slightly less accurate description of the background contributions. Thus, we recommend the sliding-window approach, which offers improved statistical stability of the background modeling while slightly improving the background descriptions. This represents an improvement over the finely binned training strategy used in the previous Belle analysis in ref. [11], where similar fluctuations in the background modeling were observed.

Two approaches to treat the individual contributions in background data are discussed and tested in this thesis: the combined background modeling approach, where we train one neural network on all background contributions, and the split background modeling approach, where we train multiple neural networks on individual background contributions. We investigate each of their performance in input-output studies and supporting signal and background studies.
The combined background modeling approach shows a stable estimation of the background yields. A small leakage of signal contributions into the background model components is observed. As this effect is small, we find that the PWD fit results on signal and background pseudo-data using the combined background modeling yield partial-

wave amplitudes similar to those on signal-only data and components, demonstrating that the bias introduced by the background modeling is small. Thus, we conclude that the combined background modeling approach can reliably describe background contributions within the partial wave decomposition framework.
The split modeling approach is motivated by the fact that we can determine the background yields of individual components from real data instead of relying on MC data. However, the observed leakage in the split background modeling approach is non-negligible. Signal is leaking into the dominant $\tau^- \to \pi^+\pi^-\pi^-\pi^0\nu_\tau$ background component. This component is particularly sensitive to leakage, because the background distribution shares similar decay dynamics with the signal decay of $\tau^- \to \pi^+\pi^-\pi^-\nu_\tau$, making these two components hard to separate in the PWD fit. Small imperfections in the PWD, such as the neglect of detector resolution effects, together with the similarity of these two components, explain the observed leakage. To account for these imperfections, the fit adjusts the background component and consequently biases the description of the partial waves, with the $1^+\ [(\pi\pi)_S\,\pi]_P$ wave being most sensitive to it. We can mitigate this effect by fixing the $\tau^- \to \pi^+\pi^-\pi^-\pi^0\nu_\tau$ yield to the true value of the pseudo-data. With this adjustment, the description of the partial waves is similar to the PWD fit on signal-only data and components. This result shows that the bias introduced by the background modeling is small, and the split background modeling approach with a fixed $\tau^- \to \pi^+\pi^-\pi^-\pi^0\nu_\tau$ yield can be reliably used to model background contributions.
Further studies on the bias of background yields fixed to a wrong value suggest that both background modeling approaches are robust to mismodeling of relative background fractions, which supports the use of these neural networks in the PWD.
In summary, both the combined background modeling approach and the split background modeling approach with a fixed $\tau^- \to \pi^+\pi^-\pi^-\pi^0\nu_\tau$ background, yield valid results of partial wave amplitudes with minimal bias. To reduce the dependence on MC simulations, the split background modeling approach is preferred.

Furthermore, we study potential biases introduced in the partial-wave decomposition using signal-only data. Resolution studies show that neglecting resolution in the data introduces only a small bias in the fitted partial-wave amplitudes, motivating this approximation in this analysis. In contrast, the introduced bias of partial waves strongly overlapping in kinematic and angular structures is considerable in poorly-constrained $m_{3\pi}$ regions. We determine thresholds for partial waves in $m_{3\pi}$ regions where their contribution is expected to be small, which considerably improves the description of the remaining partial waves within that region. Thus, we conclude that implementing thresholds is crucial and can reduce this bias. Identical thresholds can also be used on real data.

These results demonstrate that neural network-based background modeling can be used to model background contributions in the partial wave decomposition framework. With this thesis, we established and tested key components of the PWD framework. As a next step, the analysis can be extended to the larger MC16 data sample, which has recently become available. Repeating the developed analysis chain on these data and

subsequently on the corresponding real Belle II data will enable high-precision measurements of light meson resonances, including the $a_1(1260)$ and the $a_1(1420)$. The Belle II analysis will additionally benefit from better detector performance and reconstruction capabilities compared to those of Belle, improving the sensitivity for isospin-violating effects.
The extracted partial-wave amplitudes will serve as input to the subsequent resonance-model fit to determine the masses and widths of the resonances. Since the resonance-model-fit framework is currently being established in ref. [17] in collaboration with this work, this fit can easily be applied to the updated PWD results.

Appendix A

Evaluation of the Signal Normalization H

In the first step, we sample from a distribution that includes the TAUOLA weights but without acceptance, defined as

$$p(\tau) = \frac{\omega(\tau)\,\Phi(\tau)}{H'} \tag{A.1}$$

The integral defining the phase space volume is expanded by $1 = \frac{\omega(\tau)}{H'}\frac{H'}{\omega(\tau)}$, and rewritten in terms of Equation (A.1) as

$$\begin{aligned} V_{\text{PS}} = \int d\tau\,\Phi(\tau) &= \int d\tau\,\Phi(\tau)\frac{\omega(\tau)}{H'}\frac{H'}{\omega(\tau)} \\ &= \int d\tau\,p(x)\frac{H'}{\omega(\tau)} \\ &\approx \frac{1}{N}\sum_e \frac{H'}{\omega(\tau_e)} \end{aligned} \tag{A.2}$$

In the final step, we use harmonic mean integration to approximate the integral. Therefore, we can write the normalization H' as:

$$H' = \frac{V_{PS}}{\frac{1}{N}\sum_e \frac{1}{\omega(\tau)}}\,. \tag{A.3}$$

The actual sample is distributed, including the detector acceptance, meaning we sample from a distribution that follows

$$\tilde{p}(\tau) = \frac{\omega(\tau)\,\Phi(\tau)\,\eta(\tau)}{H} \tag{A.4}$$

We write the fraction of accepted to generated events in terms of the distribution defined in eq. (A.1) and the acceptance as:

$$\frac{\bar{N}}{N} = \int d\tau\,p(\tau)\,\eta(\tau) = \frac{1}{H'}\int d\tau\,\Phi(\tau)\,\omega(\tau)\,\eta(\tau) = \frac{H}{H'} \tag{A.5}$$

Therefore, we can use the fraction of the total number of events in Equation (A.5) to calculate the normalization H as

$$H = H'\frac{\bar{N}}{N} = \bar{N}\frac{V_{PS}}{\sum_e^N \frac{1}{\omega(\tau_e)}} \tag{A.6}$$

Appendix B

Partial Wave Fit Results on Signal and Background Components

The PWD fit results using the combined background modeling approach, and the split background modeling approach with a fixed $\tau^- \to \pi^+\pi^-\pi^-\nu_\tau$ component for the sliding window approach are shown in Figure B.1, Figure B.2, and Figure B.4.

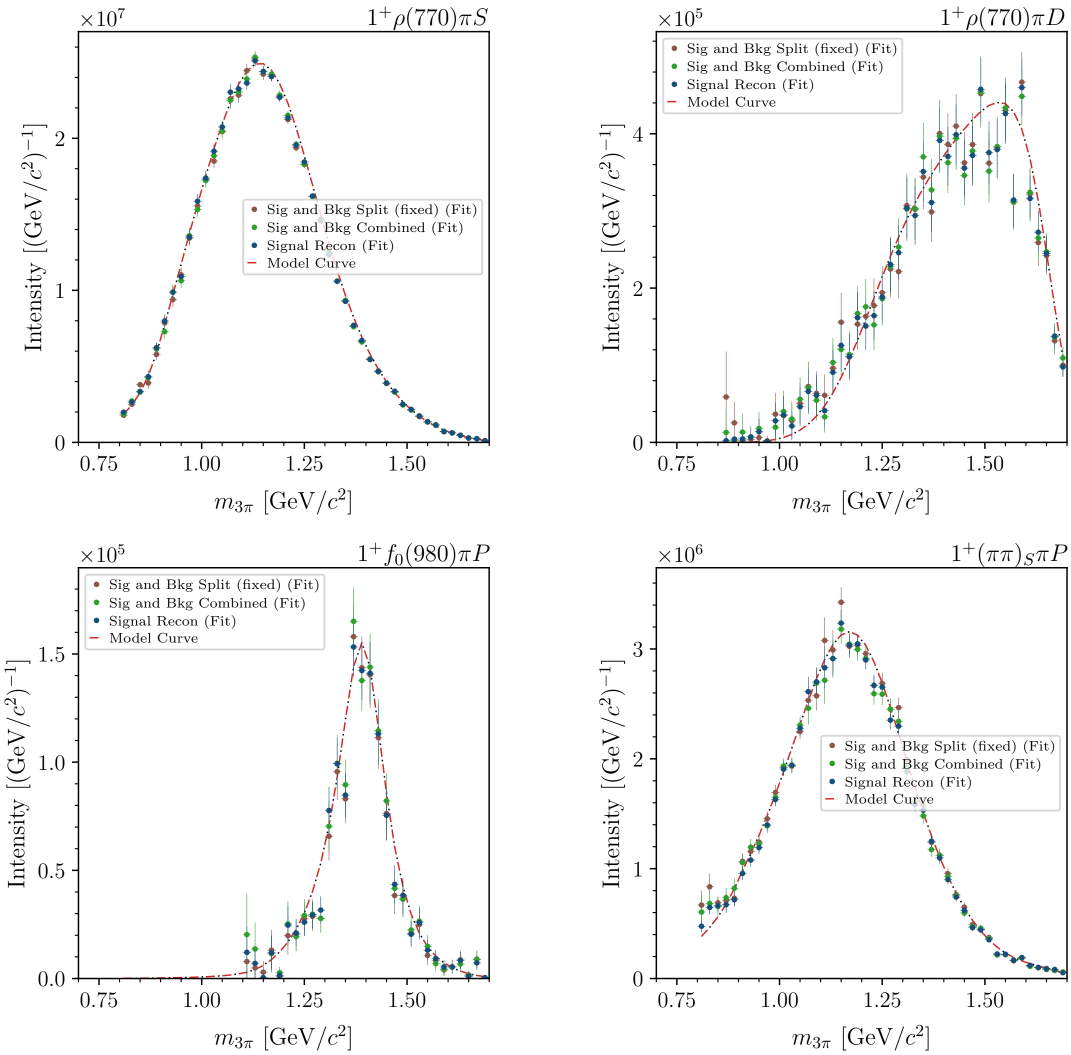


Figure B.1: Background-subtracted PWD fit results using the combined background modeling approach (green), and the PWD fit results using the split background modeling approach with a fixed $\tau^- \to \pi^+\pi^-\pi^-\nu_\tau$ component (brown) compared to the signal-only PWD fit results (blue).

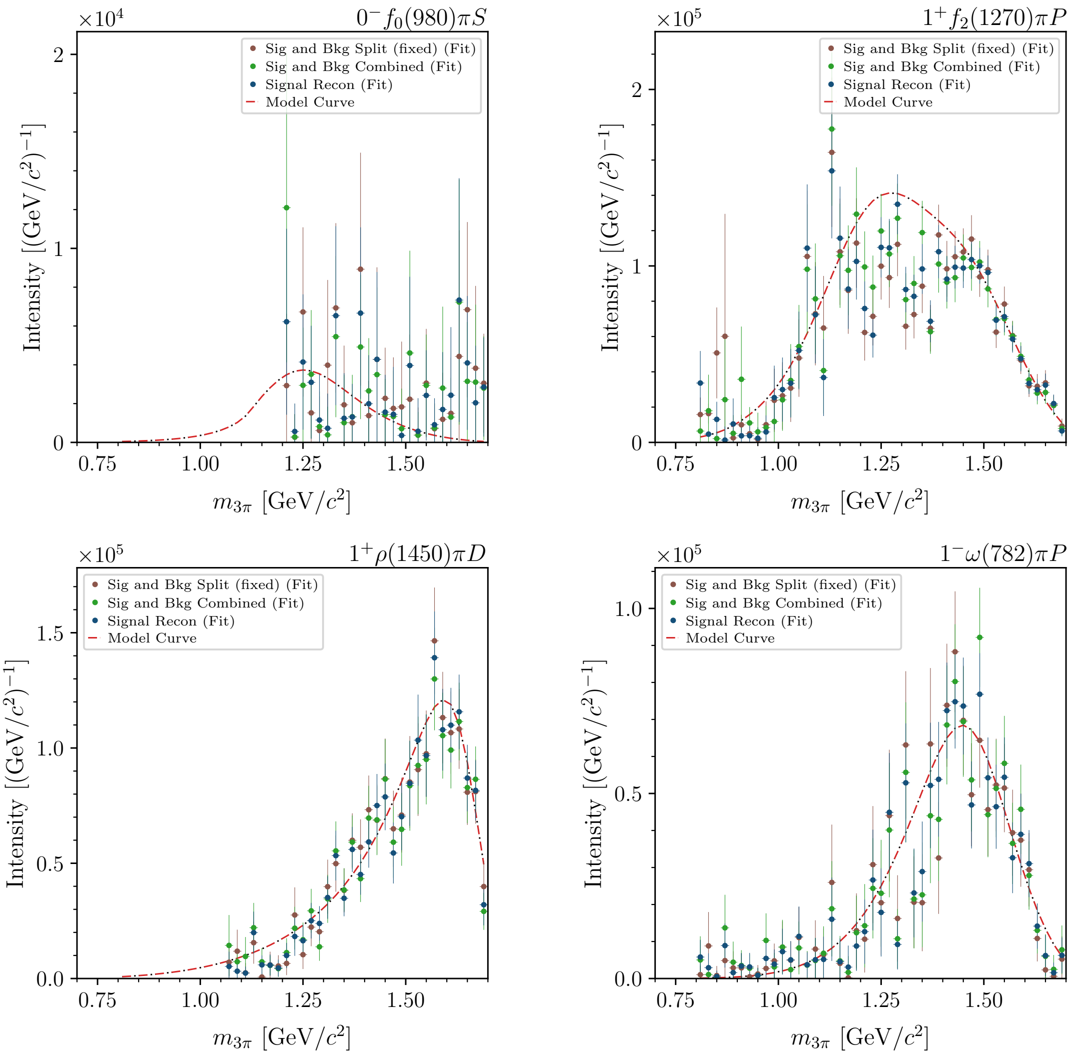


Figure B.2: Background-subtracted PWD fit results using the combined background modeling approach (green), and the PWD fit results using the split background modeling approach with a fixed $\tau^- \to \pi^+\pi^-\pi^-\nu_\tau$ component (brown) compared to the signal-only PWD fit results (blue).

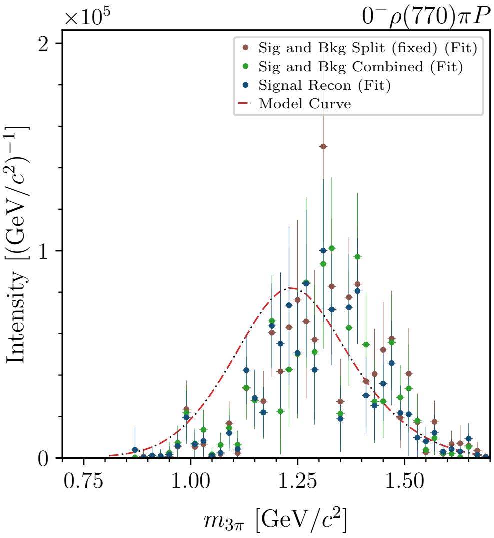


Figure B.3: Caption

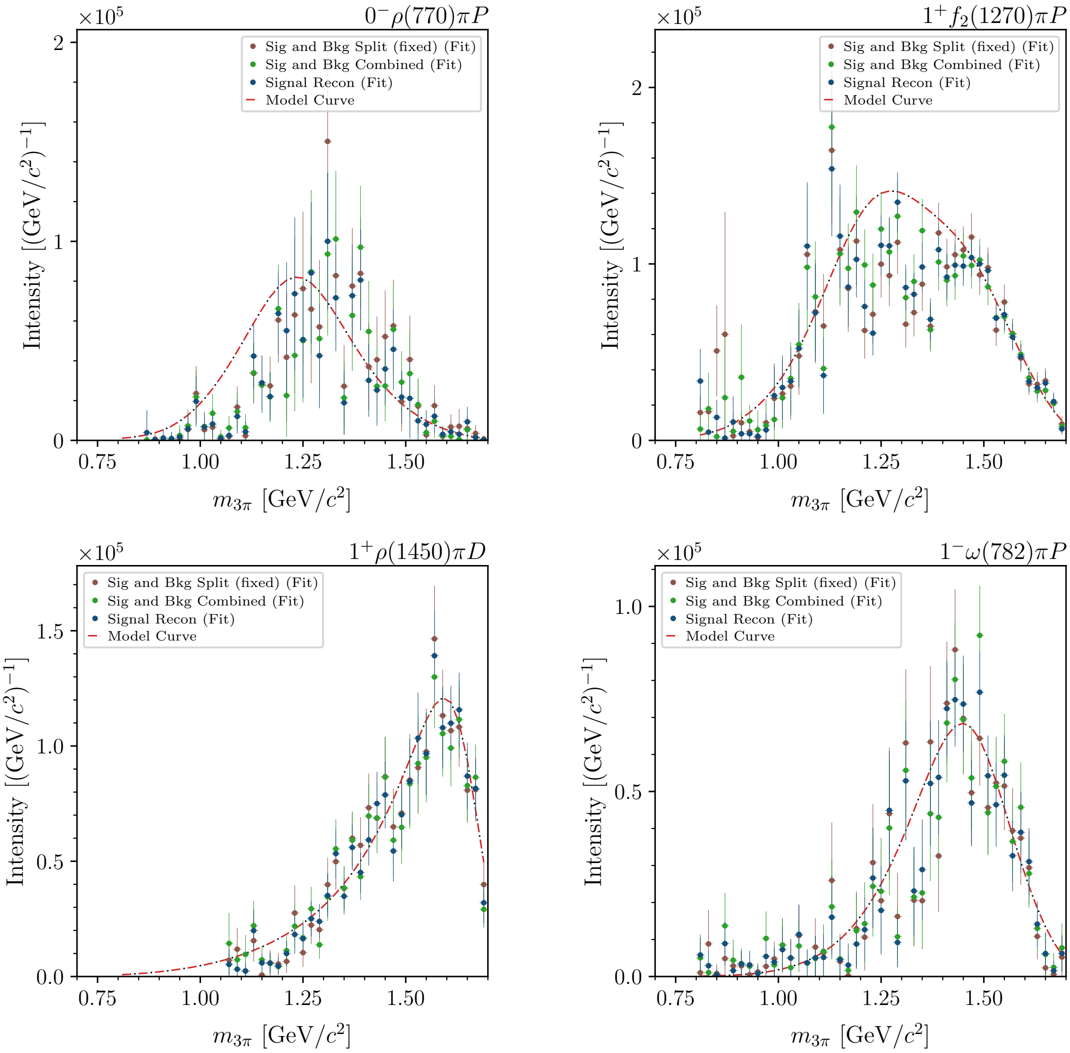


Figure B.4: Background-subtracted PWD fit results using the combined background modeling approach (green), and the PWD fit results using the split background modeling approach with a fixed $\tau^- \to \pi^+\pi^-\pi^-\nu_\tau$ component (brown) compared to the signal-only PWD fit results (blue).

Appendix C

Supporting Material for Chapter 8

C.1 Leakage Studies on Signal Only Data

Figure C.1 show the same studies as Section 8.1.1, including the neural networks trained on the full $m_{3\pi}$ range. Figure C.2 further shows the remaining individual background components of the split background modeling approach.

C.2 Leakage Studies on Background Only Data

All partial waves for the studies introduced in Section 8.1.2 are shown in Figure C.3, Figure C.4, and Figure C.5.

C.3 Resolution Studies on Signal data

The PWD fit results for the study introduced in Section 8.2.1 for the remaining partial waves are shown in Figure C.6, Figure C.7, Figure C.8, and Figure C.9.

C.4 Impact of $\pm 10\%$ Wrong Background Yield

The PWD fit results for the study introduced in Section 8.4 for the remaining partial waves are shown in Figure C.10 and Figure C.11

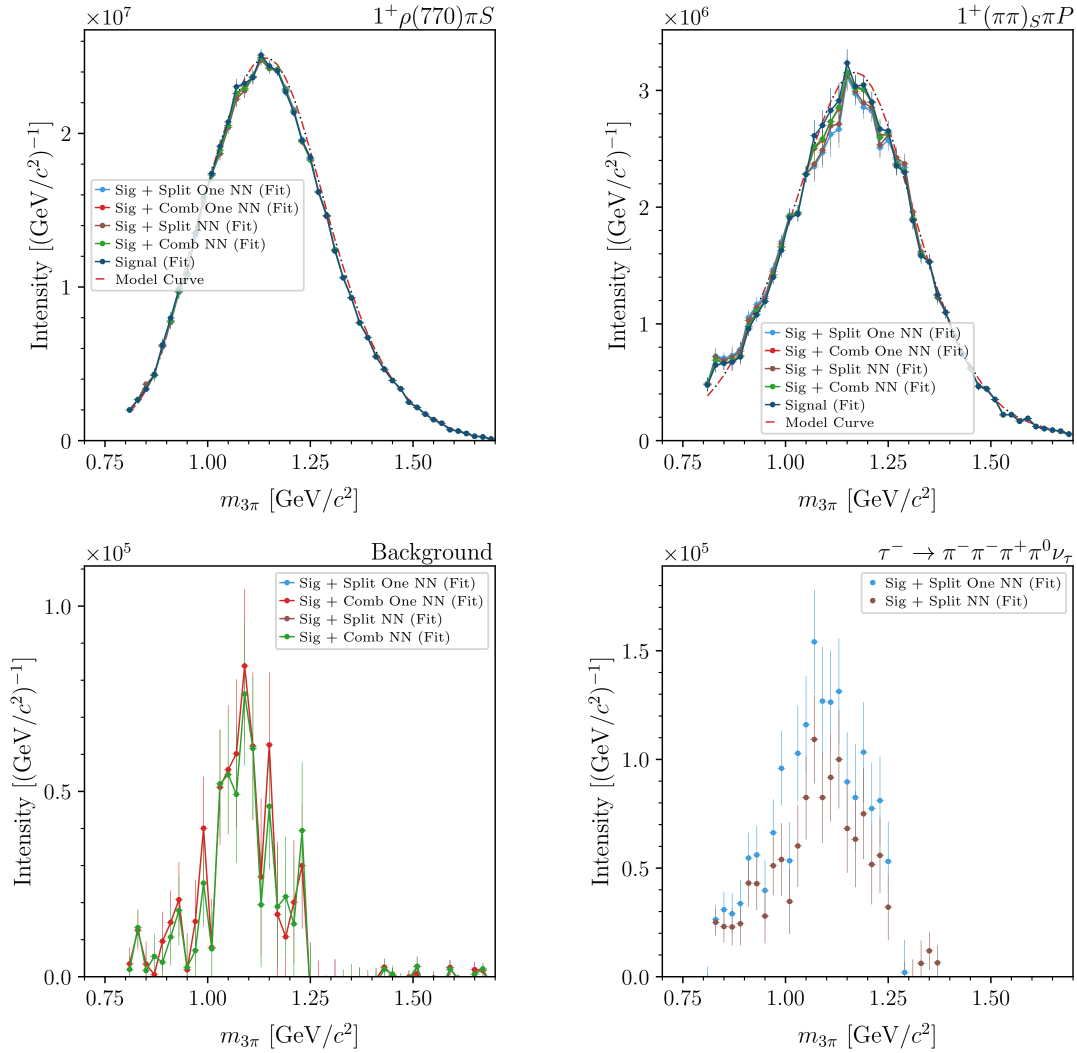


Figure C.1: PWD fit results performed on signal-only pseudo-data, including signal and either the combined sliding window (green), the combined full $m_{3\pi}$ (red), the split sliding window (brown) or the split full $m_{3\pi}$ (light blue) background component, background components, compared to the PWD fit performed on signal-only reconstructed pseudo data using the signal component only (blue) and the input-model curve (dashed red curve). The results are shown for the $1^+[\rho(770)\pi]_S$ wave, the $1^+[(\pi\pi)_S\pi]P$ wave, as well as the $\tau \to \pi^+\pi^-\pi^-\pi^0\,\nu_\tau$ and the combined background yield.

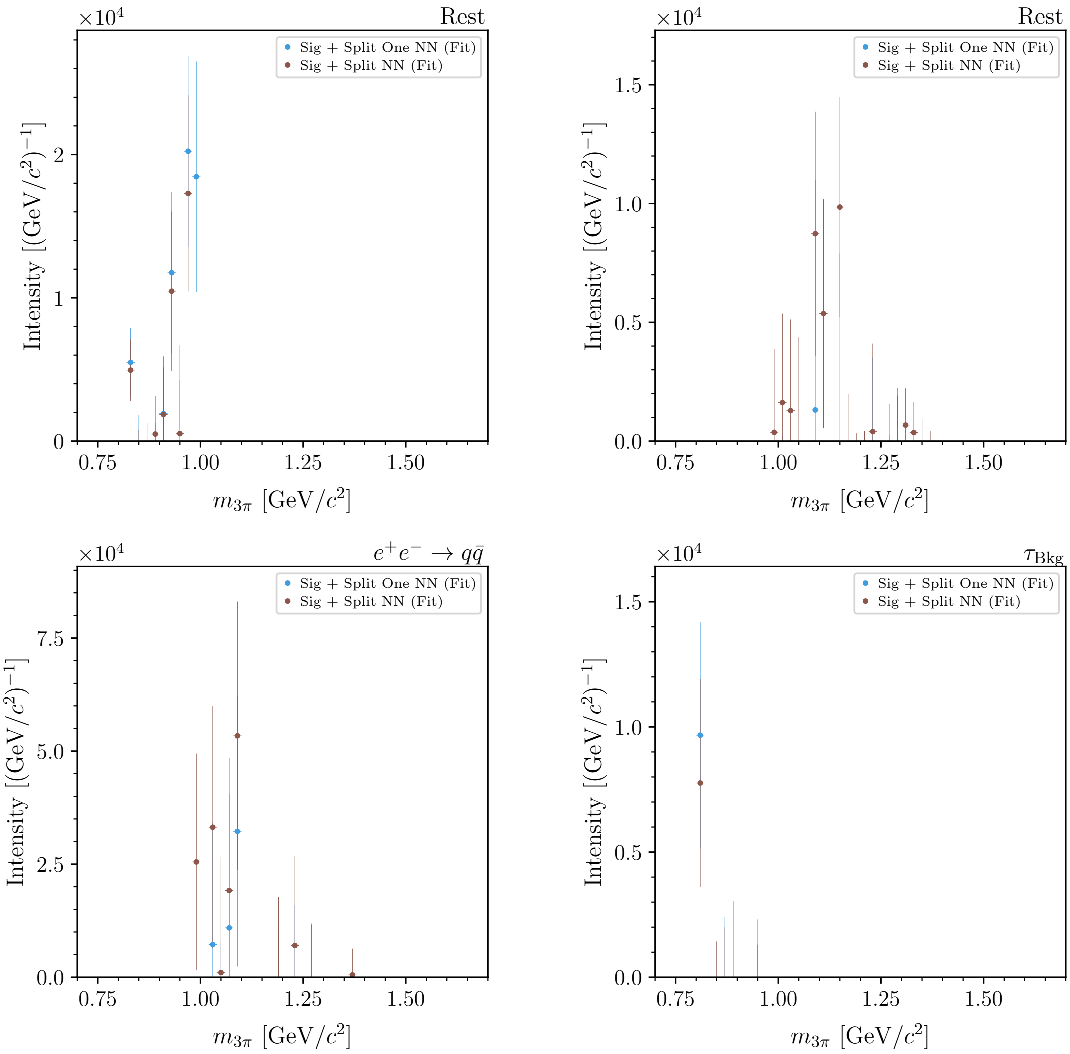


Figure C.2: PWD fit results performed on signal-only pseudo-data, including signal and split background components for the sliding window (brown) and the full $m_{3\pi}$ range training (green), compared to the PWD fit performed on signal-only reconstructed pseudo data using the signal component only (blue) and the input-model curve (dashed red curve). The results are shown for the

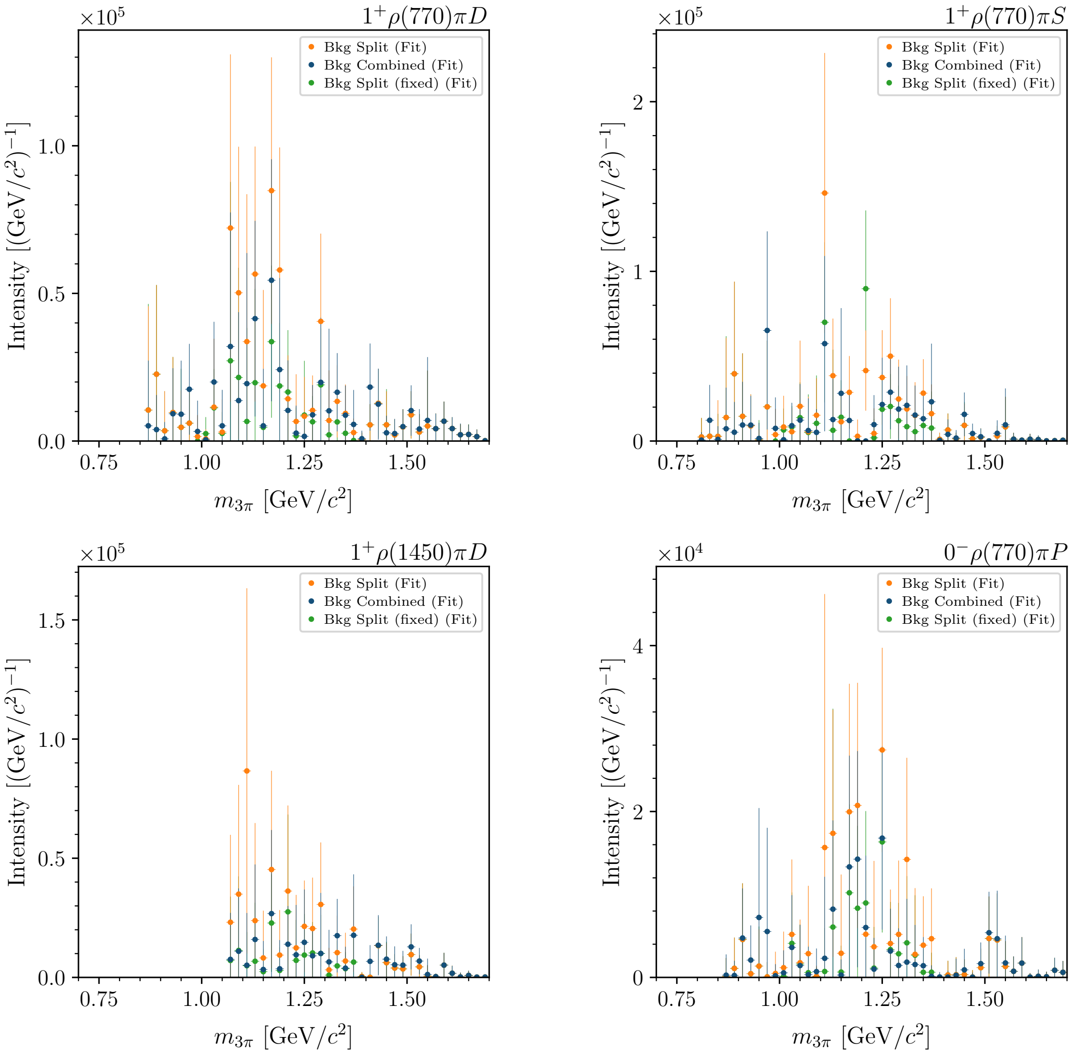


Figure C.3: PWD fit results performed on background-only pseudo data for the partial waves $1^+[\rho(770)\pi]D$, $1^+[\rho(770)\pi]S$, $1^+[\rho(1450)\pi]D$ and $0^-[\rho(770)\pi]P$ using the split background modeling (orange), the combined background modeling (blue) and the split background modeling with the dominant $\tau^- \to \pi^-\pi^-\pi^+ + \pi^0 + \nu_\tau$ fixed to the true MC value.

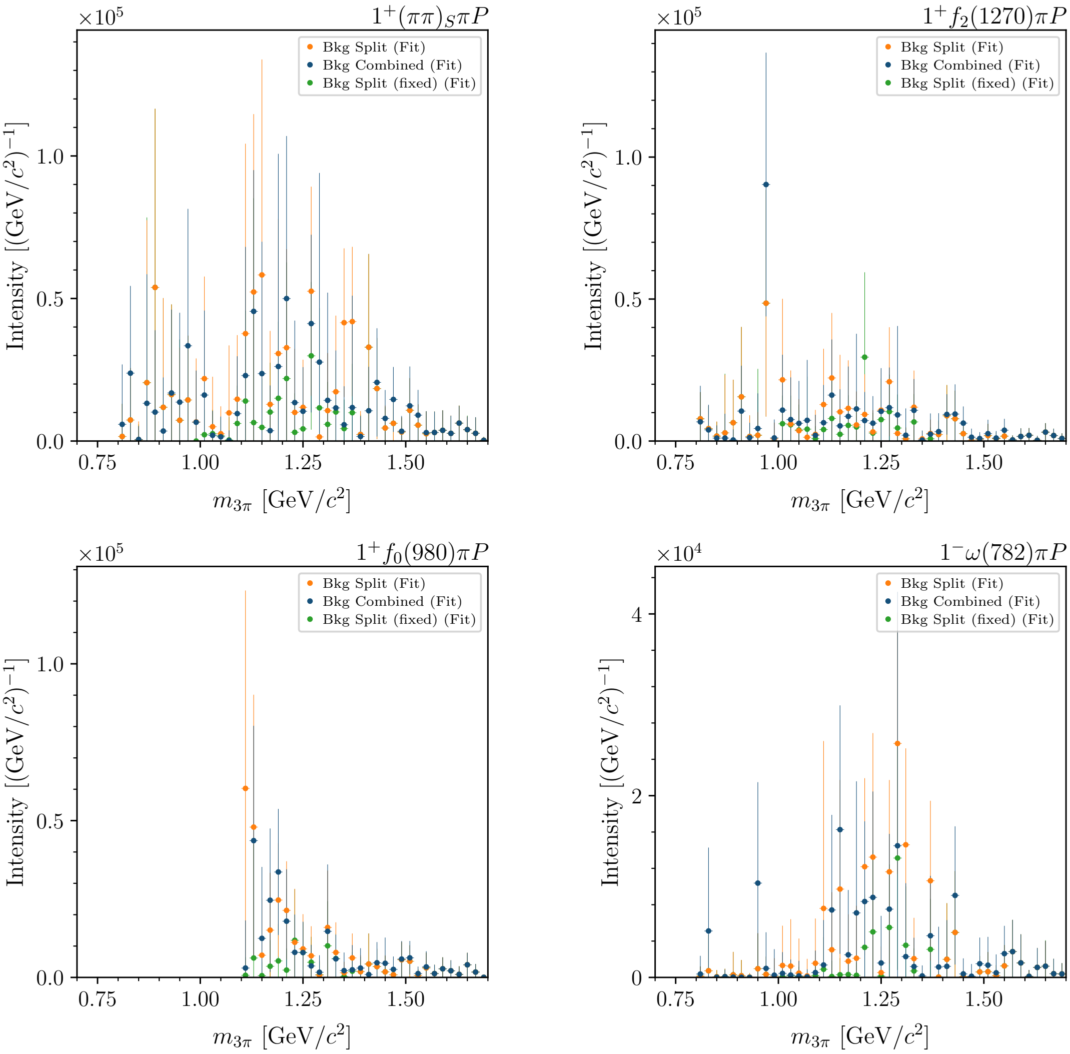


Figure C.4: PWD fit results performed on background-only pseudo data for the partial waves $1^+[(\pi\pi)_S\pi]P$, $1^+[f_2(1270)\pi]P$, $1^+[f_0(980)\pi]P$ and $1^-[\omega(782)\pi]P$ using the split background modeling (orange), the combined background modeling (blue) and the split background modeling with the dominant $\tau^- \rightarrow \pi^-\pi^-\pi^+ + \pi^0 + \nu_\tau$ fixed to the true MC value.

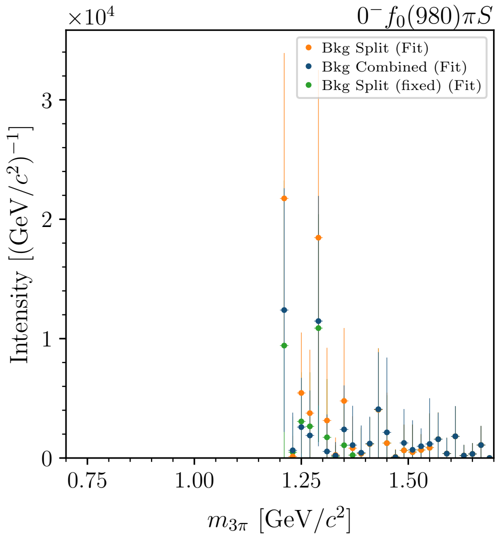


Figure C.5: PWD fit results performed on background-only pseudo data for the partial waves $0^-[f_0(980)\pi]S$ using the split background modeling (orange), the combined background modeling (blue) and the split background modeling with the dominant $\tau^- \to \pi^-\pi^-\pi^+ + \pi^0 + \nu_\tau$ fixed to the true MC value.

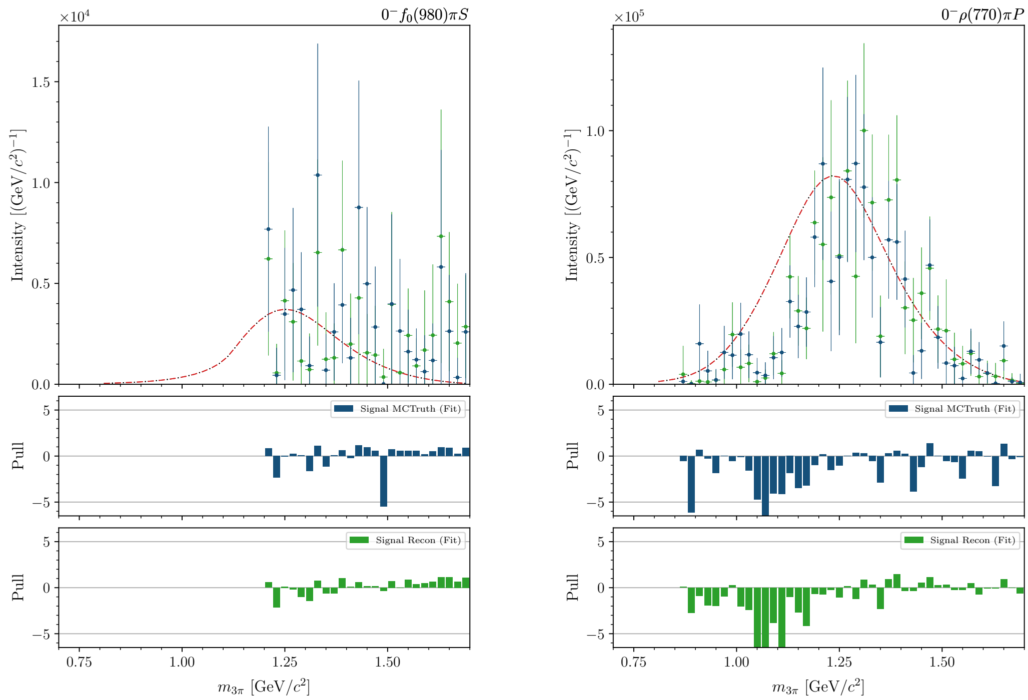


Figure C.6: PWD fit results performed on signal MC-Truth pseudo-data (blue) and reconstructed pseudo-data (green) for the $0^-[f_0(980)\pi]S$ wave and the $0^-[\rho(770)\pi]P$ wave, along with the pull distributions for both fit results.

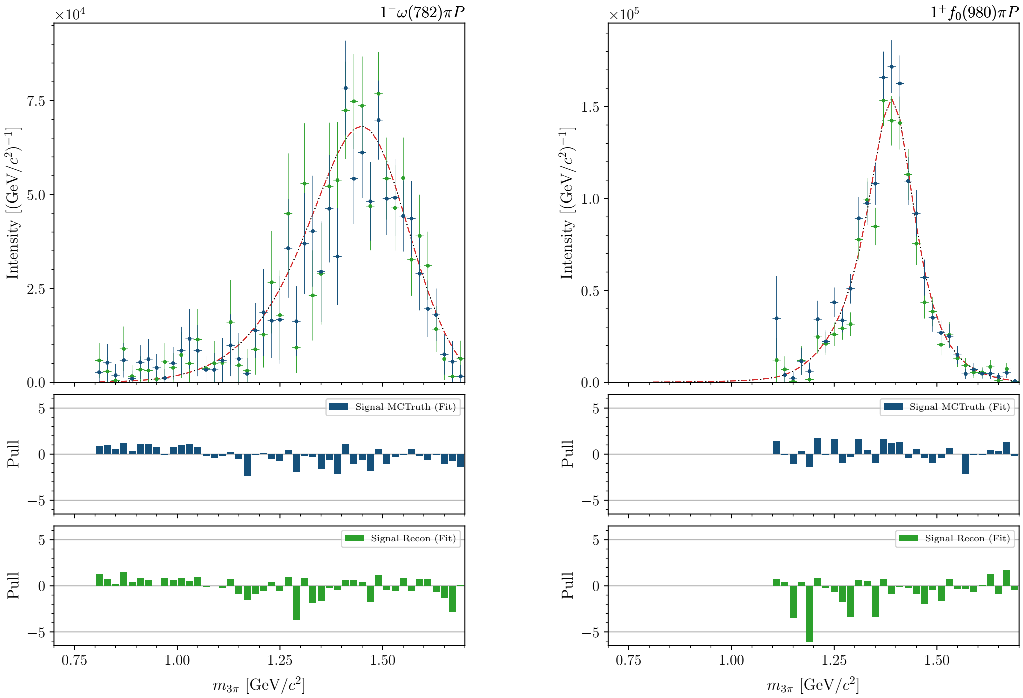


Figure C.7: PWD fit results performed on signal MC-Truth pseudo-data (blue) and reconstructed pseudo-data (green) for the $1^-[\omega(782)\pi]P$ wave and the $1^+[f_0(980)\pi]P$ wave, along with the pull distributions for both fit results.

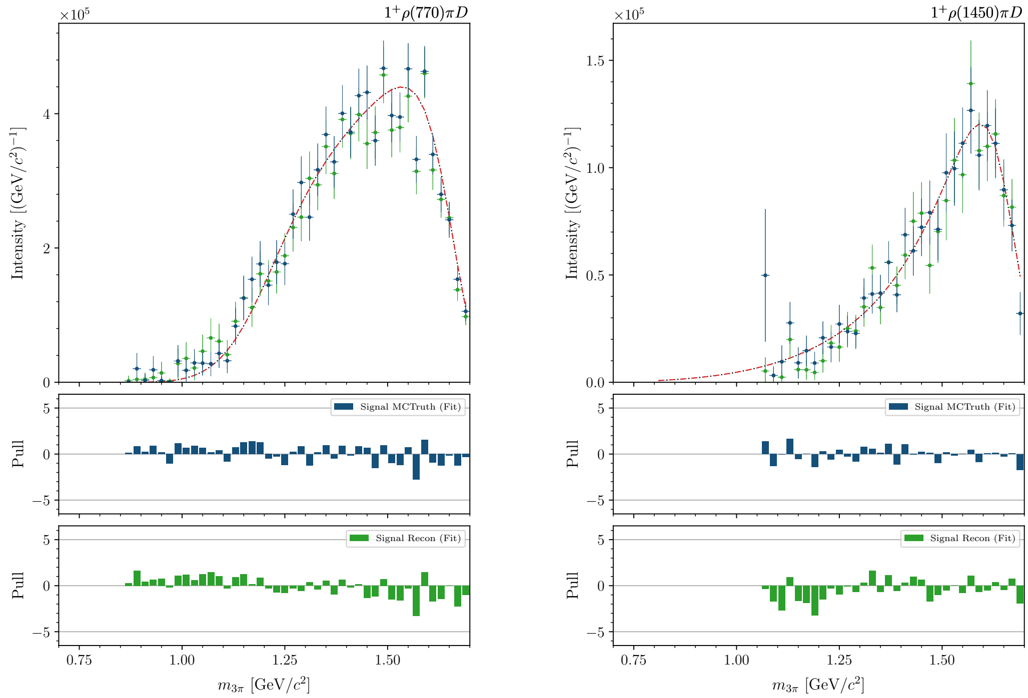


Figure C.8: PWD fit results performed on signal MC-Truth pseudo-data (blue) and reconstructed pseudo-data (green) for the $1^+[\rho(770)\pi]D$ wave and the $1^+[\rho(1450)\pi]D$ wave, along with the pull distributions for both fit results.

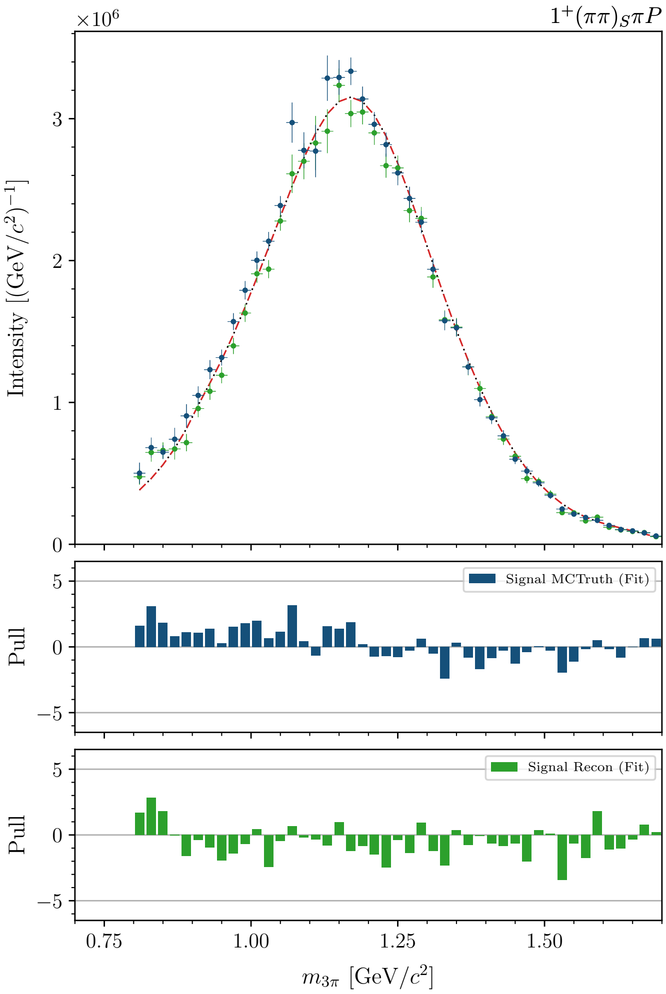


Figure C.9: PWD fit results performed on signal MC-Truth pseudo-data (blue) and reconstructed pseudo-data (green) for the $1^+[(\pi\pi)_S\pi]P$ wave, along with the pull distributions for both fit results.

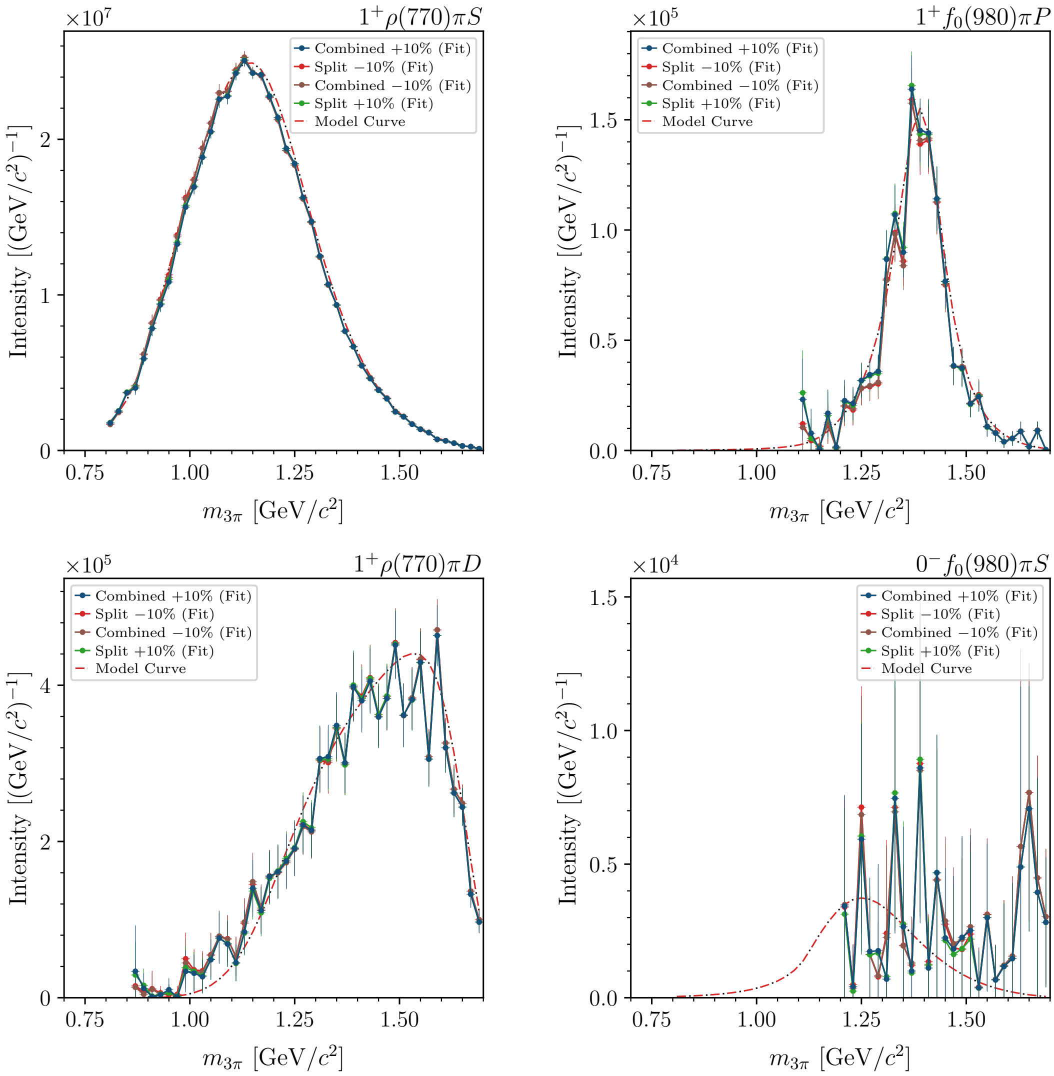


Figure C.10: PWD fit results for the combined and split background modeling approach performed on reconstructed pseudo-data with fixed background values overestimating the background yield by 10% (green, blue) and underestimating the background yield by 10% (brown, red) compared to the model curve (red) shown in the partial waves: $1^+[\rho(770)\pi]S$, $1^+[f_0(980)\pi]P$, $1^+[\rho(770)\pi]D$, and $0^-[f_0(980)\pi]S$.

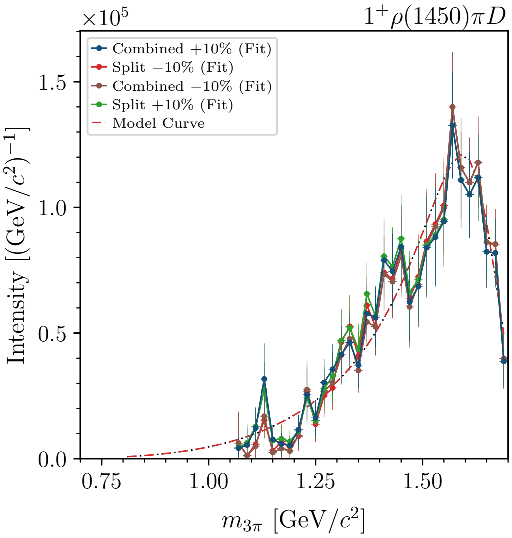


Figure C.11: PWD fit results for the combined and split background modeling approach performed on reconstructed pseudo-data with fixed background values overestimating the background yield by 10% (green, blue) and underestimating the background yield by 10% (brown, red) compared to the model curve (red) shown in the $1^+[\rho(1450)\pi]D$ wave.

Bibliography

[1] Mary K. Gaillard, Paul D. Grannis, and Frank J. Sciulli. The standard model of particle physics. *Reviews of Modern Physics*, 71(2):S96–S111, March 1999. ISSN 1539-0756. doi: 10.1103/revmodphys.71.s96. URL http://dx.doi.org/10.1103/RevModPhys.71.S96.

[2] R. D. Field. *Applications of Perturbative QCD*, volume 77. 1989.

[3] S. Navas et al. Review of particle physics. *Phys. Rev. D*, 110(3):030001, 2024. doi: 10.1103/PhysRevD.110.030001.

[4] Murray Gell-Mann. A Schematic Model of Baryons and Mesons. *Phys. Lett.*, 8: 214–215, 1964. doi: 10.1016/S0031-9163(64)92001-3.

[5] Kenneth G. Wilson. Confinement of quarks. *Phys. Rev. D*, 10:2445–2459, Oct 1974. doi: 10.1103/PhysRevD.10.2445. URL https://link.aps.org/doi/10.1103/PhysRevD.10.2445.

[6] Fabian Krinner and Stephan Paul. Hadronic currents and form factors in three-body semileptonic τ decays - the European Physical Journal C, Dec 2021. URL https://link.springer.com/article/10.1140/epjc/s10052-021-09876-1#citeas.

[7] M. Aghasyan et al. Light isovector resonances in $\pi^- p \to \pi^-\pi^-\pi^+ p$ at 190 GeV/c. *Phys. Rev. D*, 98(9):092003, 2018. doi: 10.1103/PhysRevD.98.092003.

[8] D. Gomez Dumm, A. Pich, and J. Portoles. tau —> pi pi pi nu(tau) decays in the resonance effective theory. *Phys. Rev. D*, 69:073002, 2004. doi: 10.1103/PhysRevD.69.073002.

[9] COMPASS Collaboration. Observation of a new narrow axial-vector meson $a_1(1420)$. *Phys. Rev. Lett.*, 115:082001, Aug 2015. doi: 10.1103/PhysRevLett.115.082001. URL https://link.aps.org/doi/10.1103/PhysRevLett.115.082001.

[10] M. Mikhasenko, B. Ketzer, and A. Sarantsev. Nature of the $a_1(1420)$. *Phys. Rev. D*, 91:094015, May 2015. doi: 10.1103/PhysRevD.91.094015. URL https://link.aps.org/doi/10.1103/PhysRevD.91.094015.

[11] Andrei Rabusov. *Partial-Wave Analysis of $\tau \to \pi^-\pi^-\pi^+\nu_\tau$ at Belle*. PhD thesis, Technische Universität München, 2023. URL https://mediatum.ub.tum.de/1715742.

[12] G. Cordes. Studying Detector Simulation Corrections for $\tau^- \to \pi^-\pi^-\pi^+\nu_\tau$ Decays at Belle II, May 2025. Bachelor Thesis.

[13] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989. ISSN 0893-6080. doi: https://doi.org/10.1016/0893-6080(89)90020-8. URL https://www.sciencedirect.com/science/article/pii/0893608089900208.

[14] S. Wallner. Personal communications.

[15] Arina Katscho. Selection of $\tau^- \to K^-\pi^-\pi^+\nu_\tau$ Decays for Data from the Belle II Experiment, March 2024. Bachelor Thesis, unpublished.

[16] C. Perez-Orive. Personal communications.

[17] G. Kurten. Partial-Wave Analysis of Hadronic Decays at Belle II, 2026. Master Thesis, unpublished.

[18] Belle II Collaboration. The Belle II Physics Book. *Progress of Theoretical and Experimental Physics*, 2019(12), December 2019. ISSN 2050-3911. doi: 10.1093/ptep/ptz106. URL http://dx.doi.org/10.1093/ptep/ptz106.

[19] Belle II Collaboration. Belle II Technical Design Report, 2010. URL https://arxiv.org/abs/1011.0352.

[20] Werner Herr and B Muratori. Concept of luminosity, 2006. URL https://cds.cern.ch/record/941318.

[21] Kazunori Akai, Kazuro Furukawa, and Haruyo Koiso. Superkekb collider. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 907:188–199, November 2018. ISSN 0168-9002. doi: 10.1016/j.nima.2018.08.017. URL http://dx.doi.org/10.1016/j.nima.2018.08.017.

[22] T. Hara, T. Kuhr, and Y. Ushiroda. Belle II coordinate system and guideline of Belle II numbering scheme. https://indico.mpp.mpg.de/event/2308/contributions/4092/attachments/3414/3799/Belle2NumberingScheme.pdf, 2011. Accessed: 2026-05-17.

[23] Paolo Branchini. The Belle II Experiment: Status and Prospects. *Universe*, 4(10), 2018. ISSN 2218-1997. doi: 10.3390/universe4100101. URL https://www.mdpi.com/2218-1997/4/10/101.

[24] Stefaan Tavernier. *Detectors Based on Ionisation in Semiconductor Materials*, pages 143–165. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010. ISBN 978-3-642-00829-0. doi: 10.1007/978-3-642-00829-0_5. URL https://doi.org/10.1007/978-3-642-00829-0_5.

[25] Johann H. Kuhn and E. Mirkes. Structure functions in tau decays. *Z. Phys. C*, 56: 661–672, 1992. doi: 10.1007/BF01474741. [Erratum: Z.Phys.C 67, 364 (1995)].

[26] J.H. Kühn and E. Mirkes. Angular distributions in semileptonic decays. *Physics Letters B*, 286(3):381–386, 1992. ISSN 0370-2693. doi: https://doi.org/10.1016/0370-2693(92)91791-7. URL https://www.sciencedirect.com/science/article/pii/0370269392917917.

[27] Stefan Wallner. Event selection of $\tau^{\pm} \to \pi^{\mp}\pi^{\mp}\pi^{\pm}\nu_{\tau}$, August 2023. Unpublished internal Belle II presentation, August 29, 2023.

[28] Belle II Collaboration. Reconstruction basf2. URL https://software.belle2.org/development/sphinx/mva/doc/index-01-mva.html. accessed on 09.05.2026.

[29] Belle II Collaboration. Variables Documentation basf2, 2026. URL https://software.belle2.org/development/sphinx/analysis/doc/Variables.html#variable-kaonIDNN. accessed on 2026-05-18.

[30] K. W. Edwards et al. Resonant structure of tau —> three pi pi0 neutrino(tau) and tau —> omega pi neutrino(tau) decays. *Phys. Rev. D*, 61:072003, 2000. doi: 10.1103/PhysRevD.61.072003.

[31] S. Jadach, Z. Was, R. Decker, and J.H. Kühn. The decay library tauola, version 2.4. *Computer Physics Communications*, 76(3):361–380, 1993. ISSN 0010-4655. doi: https://doi.org/10.1016/0010-4655(93)90061-G. URL https://www.sciencedirect.com/science/article/pii/001046559390061G.

[32] Belle II Collaboration. The Belle II Detector Upgrades Framework Conceptual Design Report, 2024. URL https://arxiv.org/abs/2406.19421.

[33] Christian P. Robert and George Casella. *Monte Carlo Integration*, pages 71–138. Springer New York, New York, NY, 1999. ISBN 978-1-4757-3071-5. doi: 10.1007/978-1-4757-3071-5_3. URL https://doi.org/10.1007/978-1-4757-3071-5_3.

[34] Roger Barlow. Extended maximum likelihood. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 297(3):496–506, 1990. ISSN 0168-9002. doi: https://doi.org/10.1016/0168-9002(90)91334-8. URL https://www.sciencedirect.com/science/article/pii/0168900290913348.

[35] Stephen A. Dupree and Stanley K. Fraley. *Monte Carlo Sampling Techniques*, pages 21–56. Springer US, Boston, MA, 2002. ISBN 978-1-4419-8491-3. doi: 10.1007/978-1-4419-8491-3_2. URL https://doi.org/10.1007/978-1-4419-8491-3_2.

[36] L. Demortier and L. Lyons. Everything you always wanted to know about pulls. URL https://api.semanticscholar.org/CorpusID:19545670.

[37] S.M. Flatté. On the nature of 0+ mesons. *Physics Letters B*, 63(2):228–230, 1976. ISSN 0370-2693. doi: https://doi.org/10.1016/0370-2693(76)90655-9. URL https://www.sciencedirect.com/science/article/pii/0370269376906559.

[38] et. al. M. Ablikim. Resonances in j/→+ and k+k. *Physics Letters B*, 607(3):243–253, 2005. ISSN 0370-2693. doi: https://doi.org/10.1016/j.physletb.2004.12.041. URL https://www.sciencedirect.com/science/article/pii/S0370269304017265.

[39] ALEPH Collaboration. Measurement of the spectral functions of vector current hadronic tau decays. *Zeitschrift für Physik C - Particles and Fields*, 76:15–33, 1997. doi: 10.1007/s002880050523.

[40] The ALEPH Collaboration. Measurement of the spectral functions of vector current hadronic tau decays. *Zeitschrift für Physik C Particles and Fields*, 76(1): 15–33, 1997. doi: 10.1007/s002880050523. URL https://doi.org/10.1007/s002880050523.

[41] K. L. Au, D. Morgan, and M. R. Pennington. Meson dynamics beyond the quark model: Study of final-state interactions. *Phys. Rev. D*, 35:1633–1664, Mar 1987. doi: 10.1103/PhysRevD.35.1633. URL https://link.aps.org/doi/10.1103/PhysRevD.35.1633.

[42] Igor Kachaev and IHEP. Modified s-wave $\pi\pi$ scattering amplitude for multiparticle pwa, 2023. URL https://arxiv.org/abs/2305.11711.

[43] B. Ketzer, B. Grube, and D. Ryabchikov. Light-meson spectroscopy with compass. *Progress in Particle and Nuclear Physics*, 113:103755, 2020. ISSN 0146-6410. doi: https://doi.org/10.1016/j.ppnp.2020.103755. URL https://www.sciencedirect.com/science/article/pii/S0146641020300028.

[44] A. Mathad, D. O'Hanlon, A. Poluektov, and R. Rabadan. Efficient description of experimental effects in amplitude analyses. *Journal of Instrumentation*, 16(06): P06016, June 2021. ISSN 1748-0221. doi: 10.1088/1748-0221/16/06/p06016. URL http://dx.doi.org/10.1088/1748-0221/16/06/P06016.

[45] M. Weiskopf. Background Modeling. URL https://gitlab.desy.de/weiskopf/backgroundmodelingpwd. GitLab repository, accessed on 2026-05-18.

[46] Diganta Misra. Mish: A self regularized non-monotonic activation function, 2020. URL https://arxiv.org/abs/1908.08681.

[47] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017. URL https://arxiv.org/abs/1412.6980.

[48] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15(1):1929–1958, January 2014. ISSN 1532-4435.

[49] Anders Krogh and John A. Hertz. A simple weight decay can improve generalization. In *Proceedings of the 5th International Conference on Neural Information Processing Systems*, NIPS'91, page 950–957, San Francisco, CA, USA, 1991. Morgan Kaufmann Publishers Inc. ISBN 1558602224.

[50] The ATLAS Collaboration. An implementation of neural simulation-based inference for parameter estimation in atlas. *Reports on Progress in Physics*, 88(6):067801, may 2025. doi: 10.1088/1361-6633/add370. URL https://doi.org/10.1088/1361-6633/add370.

[51] Kyle Cranmer, Juan Pavez, and Gilles Louppe. Approximating likelihood ratios with calibrated discriminative classifiers, 2016. URL https://arxiv.org/abs/1506.02169.

[52] Andrew Ng and Tengyu Ma, 06 2023. URL https://cs229.stanford.edu/main_notes.pdf.

[53] Haibo He and Edwardo A. Garcia. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1263–1284, 2009. doi: 10.1109/TKDE.2008.239.

[54] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives, 2014. URL https://arxiv.org/abs/1206.5538.

[55] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[56] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift, 2015. URL https://arxiv.org/abs/1502.03167.

List of Figures

List of Tables