IEKP-KA/2016-23

Karlsruher Institut für Technologie

Embedding - eine Datenbasierte Methode zur Abschätzung des $Z \to \tau\tau$ Untergrunds in $H \to \tau\tau$ Suchen

MASTERARBEIT

von

Artur Akhmetshin

Referent: Prof. Dr. Günter Quast
Korreferent: Prof. Dr. Roger Wolf

Institut für Experimentelle Kernphysik

Karlsruhe, den 09. November 2016

IEKP-KA/2016-23

Karlsruhe Institute of Technology

Embedding - a Data Driven Method to Estimate the $Z \to \tau\tau$ Background in the $H \to \tau\tau$ Searches

MASTER THESIS

by

Artur Akhmetshin

Referent: Prof. Dr. Günter Quast
Korreferent: Prof. Dr. Roger Wolf

Institute of Experimental Nuclear Physics

Karlsruhe, November 09, 2016

Contents

1. Introduction

The standard model of particle physics (SM) successfully describes the interactions of elementary particles at small scales. It has been tested by various experiments, including the discovery of a Higgs boson with a mass of $m_H = 125.09$ GeV and compatible with the SM expectations [1, 2], denoted as H. The Higgs boson is a particle predicted by the Brout-Englert-Higgs mechanism [3, 4], which is needed to provide a mass to the gauge bosons of the weak force, the W^+-boson, W^--boson and the Z-boson and for the fermions which constitute matter within the SM. This mechanism leads to couplings of the Higgs boson quadratic in the masses of the gauge bosons and linear in the masses of the fermions. In order to test the linearity to the fermion mass, it is essential to analyse the decay of the Higgs boson to fermions. The $H \to \tau\tau$ decay is the most promising decay for this purpose. To separate the $H \to \tau\tau$ signal from the background coming from other processes with similar kinematics, it is important to understand the expectation of these backgrounds. An alternative to a simulation based estimation of the main background resulting from the $Z \to \tau\tau$ decay is the embedding method, which is the topic of this thesis.
Within this chapter, the SM will be shortly introduced focusing on the Brout-Englert-Higgs mechanism, in the following referred as the Higgs mechanism, in the first section. The second section will describe the basics of the $H \to \tau\tau$ analysis. In the last section the topic and the goals of this master thesis will be described.

1.1. The standard model of particle physics

The standard model of particle physics is an application of the quantum field theory concept [5]. One of the key elements of a quantum field theory is the Lagrangian density $\mathcal{L}$ containing all information about the involved elementary particles and the possible interactions between them. After the building blocks of allowed physical processes, the Feyman rules, have been derived, it is possible to describe a transition from an initial state of elementary particles to a final state of elementary particles by a superposition of all possible contributing processes constructed from the Feynman rules. All these contributions can be grouped by the number of elementary interactions, called vertices, which are used to construct the contribution. If the coupling constant corresponding to the strengh of such an interaction is small enough, it's not neccessary to take all contributions into account to obtain a theoretical prediction with a precision comparable to the precision of the experimental results. An expansion in powers of the coupling constant can be made and only the leading terms of this expansion are important to achieve the desired precision.
The elementary particles described by fields in the SM can be subdivided in two main groups [6, p. 1-6]: the bosons and the fermions shown in Figure 1.1. The bosons have an integer spin and represent within the SM the particles responsible for interactions with spin equal to one, the gauge bosons, and the Higgs boson with spin equal to zero. The

gauge bosons correspond to three of the four fundamental forces of nature: the photon γ representing the electromagnetic force, the W-bosons and the Z-boson representing the weak force and eight gluons g representing the strong force. The fermions have a half-integer spin, are the constituents of matter within the SM and can be further subdivided into leptons and quarks, each group with three different generations corresponding to the columns within the boxes in Figure 1.1. All fermions are exposed to the weak force, whereas only the charged leptons - the electron e, the muon μ and the τ-lepton - and the quarks interact with the photon. The coupling strengths of the leptons to the gauge bosons of the electromagnetic and the weak force are independent from the generation of the leptons. This is referred to as the lepton universality. As a consequence, a Z-boson will decay into electrons as often as into muons or τ-leptons leading to similar decay widths: $\Gamma(Z \rightarrow ee) \approx \Gamma(Z \rightarrow \mu\mu) \approx \Gamma(Z \rightarrow \tau\tau)$. The quarks carry additionally color charge leading to interactions with gluons. As indicated in Figure 1.1 the W-bosons carry an electric charge and therefore interact with the photon. Furthermore, gluons couple with each other because of their color charge. Similarly, there are allowed couplings between the W-bosons and the Z-bosons. In contrast to various experimental observations [7], neutrinos ν are considered to be massless within the SM. All other fermions and the gauge bosons of the weak force receive their mass through the Higgs mechanism and the resulting coupling to the Higgs boson H. Because of a self-coupling, the Higgs boson is also massive.

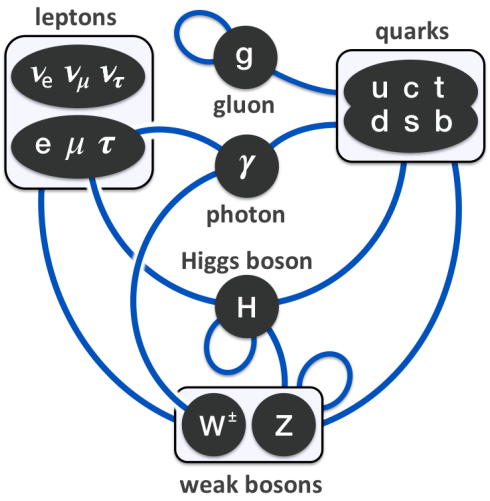

Figure 1.1.: The elementary particles and the interactions between them in the SM [8].

The SM is a gauge theory [6, p. 15-50]. This means that the terms in the Lagrangian density $\mathcal{L}$ obey local gauge symmetries: the unitary symmetry transformations applied on the fields contained in the Lagrangian density must leave it unchanged. This reflects the fact that only the absolute values of the fields are used in observables like the cross section that can be measured in an experiment. The transformations are dependent on the point in spacetime x_μ where the fields are evaluated. This means that kinetic terms of the Lagrangian density containing partial derivatives of the fields change after the transformations, unless covariant derivatives are introduced containing apart from the partial derivative an additional term with the gauge field. This leads to a certain transformation behaviour of the gauge field and the covariant derivative retaining the

local gauge symmetry of the theory. As a consequence, a field which should transform under this local gauge symmetry is forced to interact with the corresponding gauge field. The transformation behaviour of the gauge fields forbids the introduction of mass terms for all gauge bosons in contradiction to the experimental measurements in the weak sector [9, 10, 11]. One possible solution to this problem is a dynamical mass generation by introducing an additional Higgs sector. All terms in the Lagrangian density are still obeying all local gauge symmetries, but the vacuum ground state breaks the symmetry of the weak sector.

To describe all three fundamental forces, three gauge symmetries are imposed within the SM. For example, a field ψ corresponding to a fermion must transform under the $U(1)_Y$ symmetry, the $SU(2)_L$ symmetry and in the case of quarks the $SU(3)_C$ symmetry transformations, the latter corresponding to the strong force mediated by gluons. The first two symmetries build the electroweak sector of the SM, $SU(2)_L \times U(1)_Y$, spontaneously broken by the Higgs mechanism. The weak sector described by $SU(2)_L$ has the peculiarity that the corresponding charged bosons W^- and W^+ interact only with left-handed particles and right-handed antiparticles. Thus, a fermion field ψ must be decomposed into its left-handed and right-handed components: $\psi = \psi_L + \psi_R$. The component ψ_L is an isospin doublet in the $SU(2)_L$ space, the component ψ_R is an isospin singlet resembling the experimental observation that ψ_R does not take part in the weak interaction. For the electron and its neutrino the components read:

$$\psi_L = \begin{pmatrix} \nu_e \\ e_L \end{pmatrix} \qquad \psi_R = e_R$$

The gauge fields B_μ and W_μ^i; $i = 1, 2, 3$ corresponding to the $U(1)_Y$ and $SU(2)_L$ symmetries respectively don't reflect the physical behaviour of the interaction mediating particles observed in experiments. It was found, that the following linear combination of these fields delivers the fields A_μ, Z_μ, W_μ^+ and W_μ^- with the desired physical behaviour of photons, Z-bosons and W-bosons, respectively:

$$\begin{pmatrix} W_\mu^+ \\ W_\mu^- \\ Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & -i/\sqrt{2} & 0 & 0 \\ 1/\sqrt{2} & +i/\sqrt{2} & 0 & 0 \\ 0 & 0 & \cos\theta_W & -\sin\theta_W \\ 0 & 0 & \sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} W_\mu^1 \\ W_\mu^2 \\ W_\mu^3 \\ B_\mu \end{pmatrix}$$

The weak mixing angle θ_W is defined by the equation $\cos\theta_W = g/\sqrt{g^2 + g'^2}$, where g and g' correspond to the coupling constants of the $SU(2)_L$ and the $U(1)_Y$ interactions respectively. The quantum numbers which define a state in the $SU(2)_L \times U(1)_Y$ sector are the third component of the weak isospin I_3 and the hypercharge Y. These quantum numbers distinguish the strength of the corresponding coupling to different states of particles.

An example of these quantum numbers for the electron and its neutrino is shown in Table 1.1. The values of the hypercharge Y are chosen such that the electric charge q of a state matches the equation $q = I_3 + Y/2$. This charge can also be expressed in terms of the coupling constants g and g': $q = gg'/\sqrt{g^2 + g'^2}$. In that way, a description of the electromagnetic force and the weak force has been made by a single approach, known as the electroweak unification.

state	I_3	Y
e_L	-1/2	-1
e_R	0	-2
ν_e	+1/2	-1

Table 1.1.: Quantum numbers I_3 and Y for the states of the electron and of its neutrino. The electric charge q of a state fulfills the equation $q = I_3 + Y/2$.

As an example, a weak interaction converting a left-handed electron into an electron neutrino would correspond to the following piece of the Lagrangian density $\mathcal{L}(e_L \rightarrow \nu)$:

$$\mathcal{L}(e_L \rightarrow \nu) = \frac{g}{\sqrt{2}} W_\mu^+ \bar{\nu} \gamma^\mu e_L$$

There is no such term for the right-handed electron state e_R. This distinction of the left-handed and right-handed states by the weak force triggers the need for a dynamic mass generation also for the fermions which can be provided by the Higgs mechanism explained in the following.
To generate the masses of the gauge bosons dynamically, an additional Lagrangian density term $\mathcal{L}(Higgs)$ is introduced:

$$\mathcal{L}(Higgs) = (D_\mu \phi)^\dagger (D^\mu \phi) - \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \qquad \phi = \begin{pmatrix} \phi_+ \\ \phi_0 \end{pmatrix}$$

The Higgs field ϕ is a weak isospin doublet with the complex scalar components ϕ_+ and ϕ_0. The first term in $\mathcal{L}(Higgs)$ consists of the covariant derivative D_μ involving the B_μ and the W_μ^i fields, since ϕ transforms under the symmetry transformations of the $SU(2)_L \times U(1)_Y$ sector. The last two terms in $\mathcal{L}(Higgs)$ times -1 build the Higgs potential with a minimum at the non-vanishing point $v = \sqrt{\mu^2/(2\lambda)}$ when considered as a function of $|\phi|$. The value v is the vacuum expectation value. The direction of the minimum of ϕ in the weak isospin space is not unique but has to be chosen appropriately. The hypercharge Y of field ϕ is chosen such that the electric charge of the upper component ϕ^+ is one and the charge of the lower component ϕ^0 is zero. Since the vacuum is neutral, the minimum is chosen in the lower component of ϕ and the Higgs fields is expanded around this minimum of the Higgs potential:

$$\phi \rightarrow \begin{pmatrix} 0 \\ \sqrt{\mu^2/(2\lambda)} + H/\sqrt{2} \end{pmatrix}$$

The real scalar field H corresponds to the Higgs boson. Thus, if the expansion above is inserted into the first term of $\mathcal{L}(Higgs)$, the mass terms for the fields Z_μ, W_μ^+ and W_μ^- can be derived together with the interactions between the massive gauge bosons and the kinetic term of the Higgs boson. The field A_μ remains massless. The masses of the massive gauge bosons read:

$$m_W = gv/2 \qquad m_Z = \sqrt{g^2 + g'^2}\, v/2$$

For the fermions, the Yukawa coupling to the Higgs field is introduced. For an electron and its neutrino, the corresponding term reads:

$$\mathcal{L}(Yukawa) = -y_e \bar{e}_R \phi^\dagger \psi_L + y_e^* \bar{\psi}_L \phi e_R$$

The term is constructed such that after combining all quantum numbers for the electroweak sector, the term stays invariant under all required symmetry transformations. After expanding the Higgs field around the vacuum expectation value and inserting the expanded field into $\mathcal{L}(Yukawa)$, the following mass m_e can be identified for the electron field $e = e_R + e_L$:

$$m_e = y_e v$$

The interaction term of the Higgs boson to the electron in the Lagrangian density $\mathcal{L}(H \to ee)$ reads then:

$$\mathcal{L}(H \to ee) = -\frac{m_e}{\sqrt{2}v} H \bar{e} e$$

Aquivalently, this also holds for the τ-leptons with the advantage, that the mass m_τ and therefore the coupling y_τ are higher. The $H \to \tau\tau$ decay plays therefore a crucial role to test the concept of the Yukawa coupling.

1.2. $H \to \tau\tau$ analysis

During the Run I data taking period between the years 2011 and 2012, the search for a Higgs boson decaying into a pair of τ-leptons lead to an evidence for such a decay substantiating the Yukawa coupling of fermions to the Higgs boson [12]. The Run II data taking period with a centre-of-mass energy $\sqrt{s} = 13$ TeV started in 2015 and preparations for a rediscovery of the Higgs boson are ongoing.
As can be seen in the left plot of Figure 1.2, the main production processes of the Higgs boson of the SM for proton-proton collisions at the centre-of-mass energy $\sqrt{s} = 13$ TeV are the gluon-gluon fusion $pp \to H$, the vector boson fusion $pp \to qqH$ and the associated production with the massive vector bosons, $pp \to WH$ and $pp \to ZH$. The diagrams corresponding to these processes shown in Figure 1.3 indicate, that these processes can be distinguished by the particles accompaining the Higgs boson in the final state. This information can be used in the analysis to perform a categorization to enrich the signal yield.
From the right plot of Figure 1.2 it can be concluded that the decays $H \to b\bar{b}$ and $H \to \tau\tau$ are the most appropriate to test the Yukawa coupling structrure described in section 1.1. The $H \to \tau\tau$ analysis has the advantage, that it can profit from precise electron and muon reconstruction that can help to increase the signal to background ratio. These particles appear as possible final states within the analysis.
In Table 1.2 the branching ratios for the decay channels of the di-τ-system are summarized using the information of the τ-lepton decays [14, p. 15-19]. The names of the decay channels are constructed from the decay modes e, μ and τ_h of the individual τ-leptons. The decay mode e involves the leptonic decay of the τ-lepton into an electron and neutrinos. In the case of a negatively charged τ-lepton the decay reads $\tau^- \to e^- \nu_\tau \bar{\nu}_e$. Aquivalently,

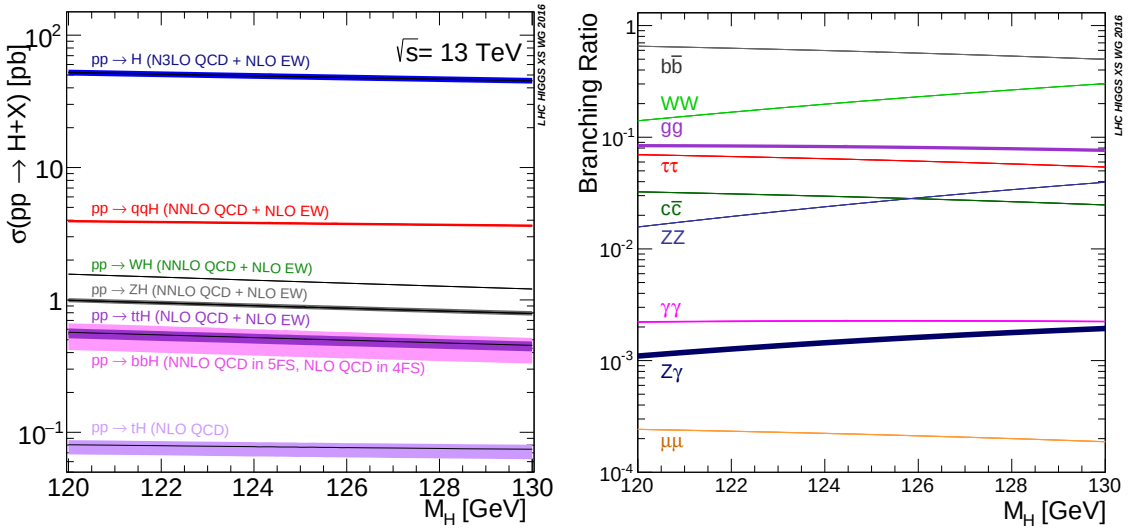

Figure 1.2.: Standard model production cross sections of the Higgs boson and the branching ratios of its decays in the left and the right plot respectively [13]. The quantities are shown in the mass window between 120 and 130 GeV.

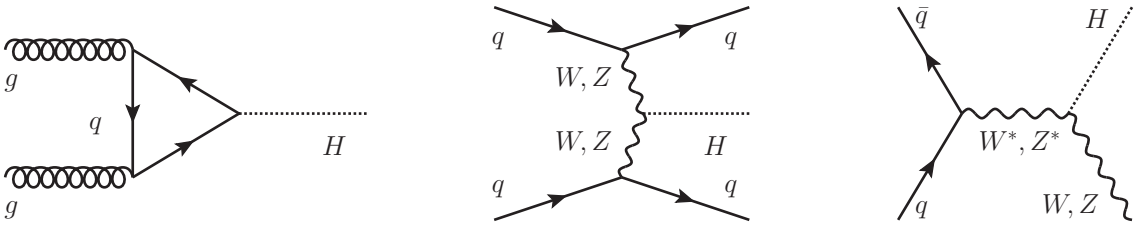

Figure 1.3.: Diagrams of the main production processes of the Higgs boson: gluon-gluon fusion in the diagram at the left, vector boson fusion in the diagram in the middle and associated production with a vector boson in the diagram at the right.

decay channel	branching ratio
$\tau_h\tau_h$	41.94%
$e\tau_h$	23.09%
$\mu\tau_h$	22.55%
$e\mu$	6.21%
ee	3.18%
$\mu\mu$	3.03%

Table 1.2.: The decay channels of the di-τ-system with the corresponding branching ratios.

the decay mode μ corresponds to the $\tau^- \rightarrow \mu^- \nu_\tau \bar{\nu}_\mu$ decay in the case of a negatively charged τ-lepton. All other decays are comprised in the τ_h decay mode containing all hadronic decays of the τ-lepton. In the case of a negatively charged τ-lepton the decays read: $\tau^- \rightarrow$ hadrons + ν_τ.
The aim of the analysis is to keep as much of the statistics provided by the recorded data as possible and at the same time to enlarge the signal to background ratio. This is achieved

by applying kinematic cuts on the reconstructed visible decay products of the τ-leptons and on the resonance candidate combined from these visible decay products. By imposing identification and isolation requirements the quality of the reconstructed particles can be improved. A further enrichement of the signal to background ratio can be achieved by a profound categorization used in the analysis. This can involve a further splitting of the decay channels by the number of jets, by the tags corresponding to the production process and by the transverse momentum of the visible decay products.
After the selection has been made the expected backgrounds must be compared to the recorded data. Therefore, a good understanding of the relevant backgrounds contributing to the decay channels is required [12, p. 13-18]. Most of the minor backgrounds are estimated from the simulation. These backgrounds comprise the processes with W-bosons accompanied by jets, the production of a top quark and the corresponding antiquark and processes with two vector bosons. The QCD multijet background is estimated by a data driven method using a control region enriched with these multijet events. The largest background contribution results from the Drell-Yan process where a pair of leptons is produced by a quark-antiquark annihilation into a Z-boson. The contribution of the $Z \rightarrow \tau\tau$ decay has been successfully estimated by the data driven embedding method during the Run I data taking period by the CMS [12, p. 13f] and ATLAS experiments [15, p. 15f] and subsequent studies were published to obtain a better understanding of the embedding and to improve this method [16, 17]. A data driven estimation of the events with a $Z \rightarrow \tau\tau$ decay is expected to provide a better description of pile-up and the underlying event compared to a completely simulated estimation. Therefore, it is essiential to prepare the embedding for the Run II $H \rightarrow \tau\tau$ search. A revision and a reimplementation of the embedding software is approached and studies are performed to increase the understanding of the individual steps of the embedding procedure within this thesis.
This thesis is organized in the following way: To understand the impact of the steps performed by the embedding method a description of the CMS experiment is made in chapter 2. It gives an overview of the detector design and the relevant reconstruction techniques. The embedding method and the qualitative studies related to it are described in chapter 3. The method is later validated quantitatively in chapter 4 and its performance is applied in chapter 5. Concluding remarks and an outlook to remaining questions to be answered in subsequent studies are given in chapter 6.

2. The CMS Experiment

The Compact Muon Solenoid (CMS) is a general purpose detector built for the Large Hadron Collider (LHC) at the Conseil Européen pour la Recherche Nucléaire (CERN) in Geneva. The collaboration of the CMS experiment involves more than 3500 members from 43 countries [18]. The data collected during the Run I data taking period from 2011 to 2012 allowed the discovery and the measuration of a Higgs boson compatible with the SM expectations [1, 2].
This chapter is devided into two sections: The first section will describe how the CMS detector is designed: its key components and their arrangement within the detector will be described. In the second section, the reconstruction techniques relevant for this thesis will be discussed.

2.1. Detector structure

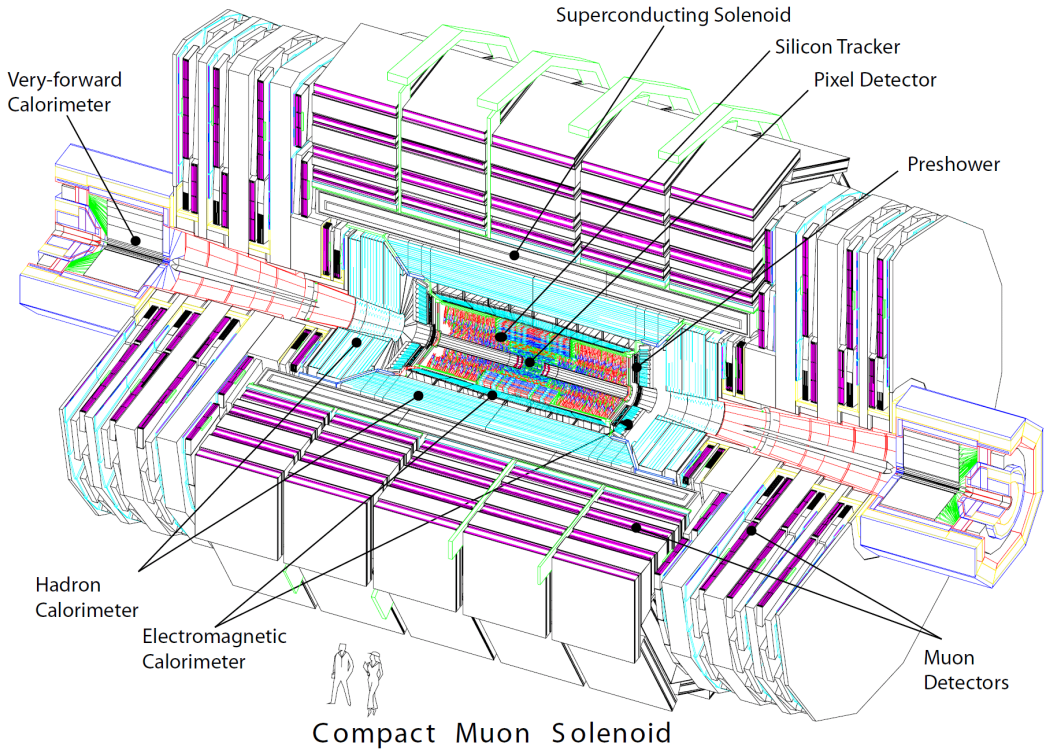

Figure 2.1.: An exploded view of the CMS detector [19, p. 8].

The CMS detector [19] can be devided into two main regions: the barrel region and the endcap region. The barrel region is built around the beam pipe to cover mainly the direc-

tions transverse to the beam axis. The endcap region closes the cylidnric barrel at both ends as shown in Figure 2.1. The barrel and the endcap region have a similar structure that will be explained in the following subsections, starting with the innermost and moving towards the outermost detector system.

2.1.1. Inner tracking system

The first detector system placed next to the beam pipe is the inner tracking system [19, p. 17-21] consisting of the pixel tracker and the strip tracker. The active detector material used for pixels and strips is silicon. The main purpose of this detector system is to measure precisely the tracks close to the interaction point in the beam pipe. This allows further vertex reconstruction which uses all information about the reconstructed tracks.

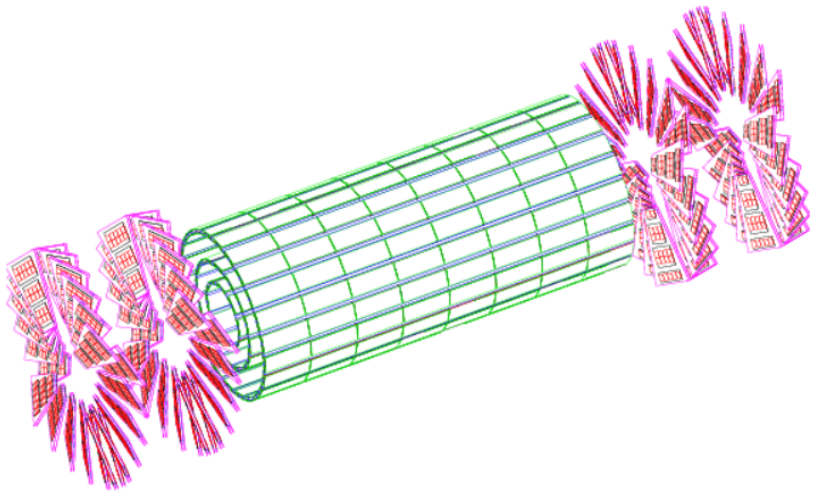

Figure 2.2.: A layout of the pixel tracker [19, p. 20].

The pixel tracker consists of three layers in the barrel and two disks for each side in the endcap region as shown in Figure 2.2. The barrel layers and the endcap disks consist of 1440 pixel modules in total. Each module contains several readout chips with 52x80 pixels. The size of each pixel is 100x150 μm^2. To improve the resolution of the pixel tracker, the Lorentz effect [20, p. 11] is exploited by sharing charge among nearby pixels. The modules in the barrel layers profit from the Lorentz effect because of the sensor plane oriented parallel to the magnetic field of the detector. To allow this also for the endcap disks, these are assembled in a turbine-like configuration with 7 modules per blade. The blades are rotated by 20° with respect to the plane transverse to the beam axis. The resulting resolution of the pixel tracker is about 10 µm for the (r, ϕ)-measurement and about 20 µm for the z-measurement.
The strip tracker surrounds the pixel tracker and is devided into four subregions as shown in Figure 2.3 in the (r, z)-view: the subregions placed in the barrel regions are the Tracker Inner Barrel (TIB) and the Tracker Outer Barrel (TOB). The subregions in the endcap region are the Tracker Inner Disks (TID) and the Tracker End Cap (TEC). The strip tracker

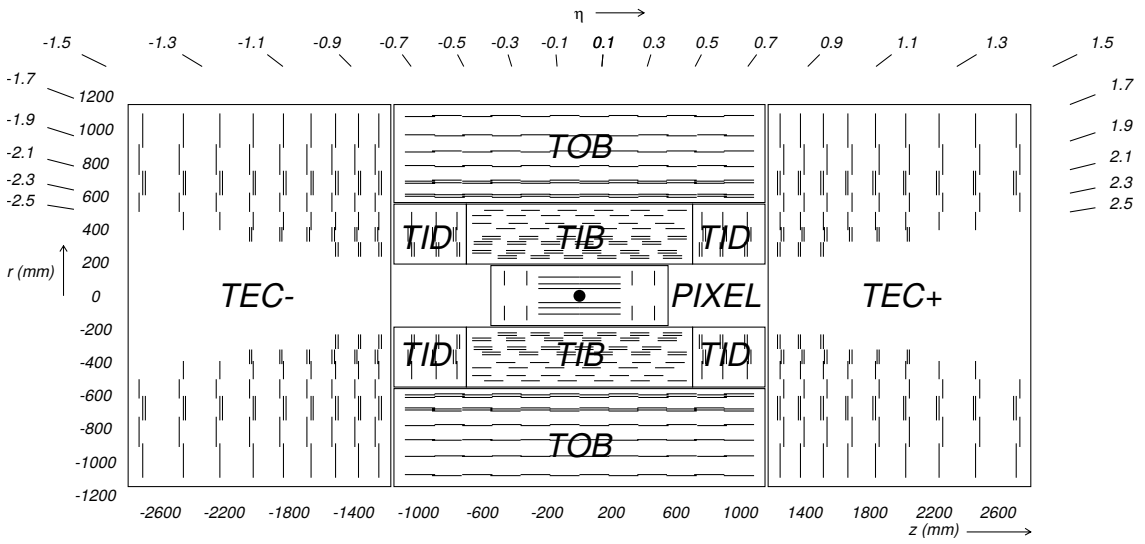

Figure 2.3.: A sketch of the strip tracker showing its different subregions in the (r, z)-view [21, p. 30].

consists of 15400 modules in total. The TIB and the TOB are made of 4 and 6 module layers respectively. In the first two layers of each barrel subregion double-sided modules are used to allow a stereo measurement in (r, ϕ) and (r, z) coordinates. One side of such a module is rotated against the other by a stereo angle of 100 mrad. The TID and the TEC are made of 3 and 9 module disks respectively. The first two disks in each endcap subregion and the fifth disk in the TEC contain double-sided modules to profit again from the stereo measurement. The same stereo angle of 100 mrad is used for the endcap region. The final resolution for the strip tracker is found to be between 20 an 50 µm in the (r, ϕ)-direction and between 230 and 530 µm in the z-direction.

2.1.2. Electromagnetic calorimeter

The electromagnetic calorimeter system is built from about 75000 homogeneous lead tungstate crystals and is placed next to the inner tracking system in the detector [19, p. 15, p. 146-150]. This material is chosen because of its short radiation length of 0.89 cm, fast light emission and raditation resistance. To compensate for its low light yield and for its operation within a magnetic field, suitable photodetectors with intrinsic gain are chosen. To guarantee a smooth operation temperature stability is required. As shown in Figure 2.4 the electromagetic calorimeter is devided into the barrel component (EB), the endcap component (EE) and the additional preshower component (ES) having a design different from the EE and EB. The EB and EE components cover a region in the pseudorapiditiy of $|\eta| < 1.479$ and $1.479 < |\eta|| < 3$ respectively to stop particles interacting electromagnetically, in particular electrons and photons. The ES component is a sampling calorimeter with two layers each formed from a lead radiator followed by a silicon strip sensor. The purpose of the ES is to identify neutral pions in the pseudorapidity range $1.653 < |\eta|| < 2.6$, to improve there the discrimination of electrons from ionizing particles and to imrove the position determination for electrons and photons.

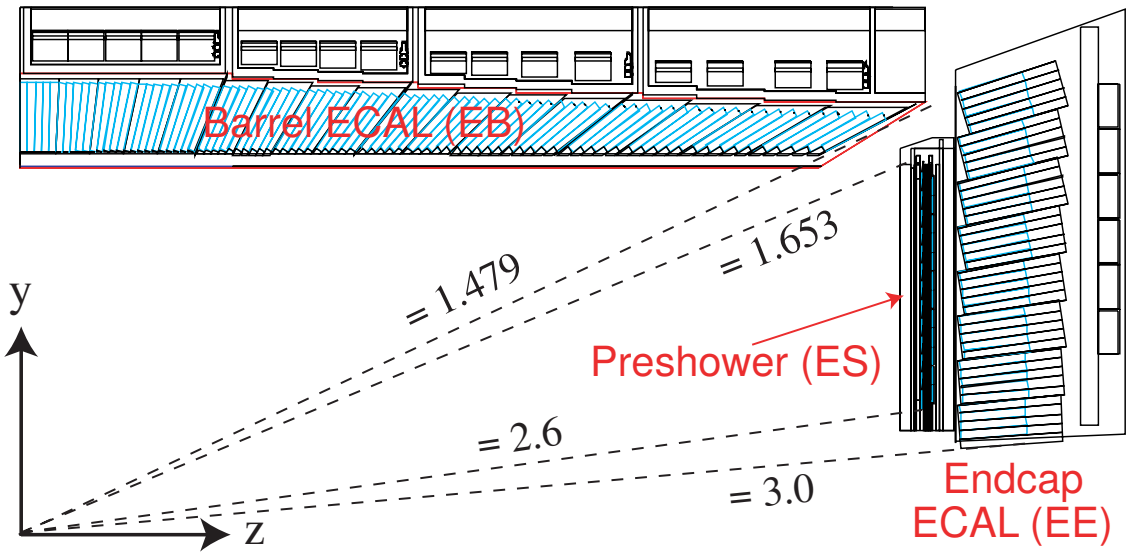

Figure 2.4.: A sketch of the electromagnetic calorimeter showing its different subregions in the spacial (y, z)-plane [19, p. 146].

2.1.3. Hadronic calorimeter

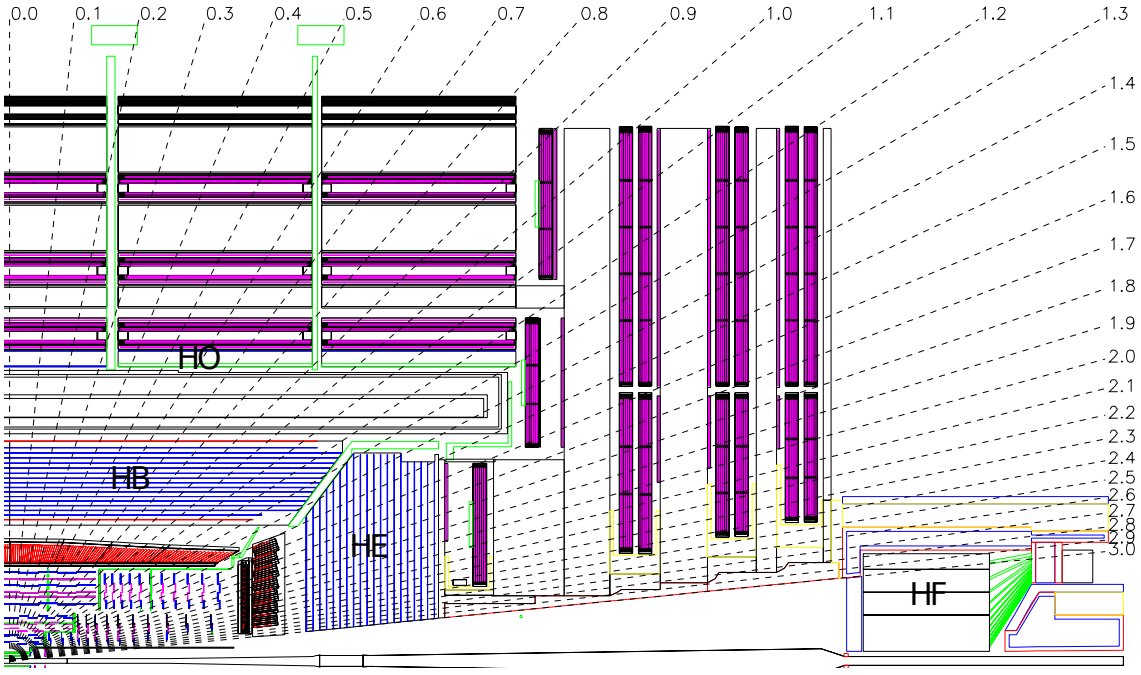

Figure 2.5.: A sketch of the hadronic calorimeter showing its different subregions in the spacial (y, z)-plane [21, p. 123]. Pseudorapidity directions are given as dashed lines.

The hadronic calorimeter is the next detector system placed after the electromagnetic calorimeter as shown in Figure 2.5. The purpose of the hadronic calorimeter is to stop strongly interacting particles, hadrons, and measure their energy [19, p. 15-17]. To achieve

this the amount of the absorber material is maximized in terms of the interaction length. The hadronic calorimeter is devided into the barrel region (HB), the endcap region (HE), the outer region (HO) and the forward region (HF) with an individual design for each region. In the HB and the HE brass is used as the absorber material since it is not magnetic and can be placed within the magnet solenoid. The tile/fibre technology is used to construct scintillators as active medium. The HB covers the $|\eta| < 1.4$ region and the HE covers the $1.3 < |\eta| < 3.0$ region. In contrast to the HB, the HO is located outside the magnet coil described in subsection 2.1.4. The HO consists of additional layers of scintillators arranged in 5 rings matching the geometry of the muon detector system described in subsection 2.1.5. The ring in the middle of the HO comprises two scintillator layers enclosing an iron absorber, the remaining rings have a single scintillator layer. The HO is designed to improve the energy resolution in the hadronic calorimeter by measuring hadrons passing the material of the magnet. The HF covers the forward pseudorapidity region of the detector, $3.0 < |\eta| < 5.0$. Since the HF is exposed mostly to proton remnants, it is designed as a radiation resistant steel/quarz fibre sampling calorimeter. The Cerenkov light emitted in the quarz fibres is led to photomultipliers to form a signal.

2.1.4. Magnet

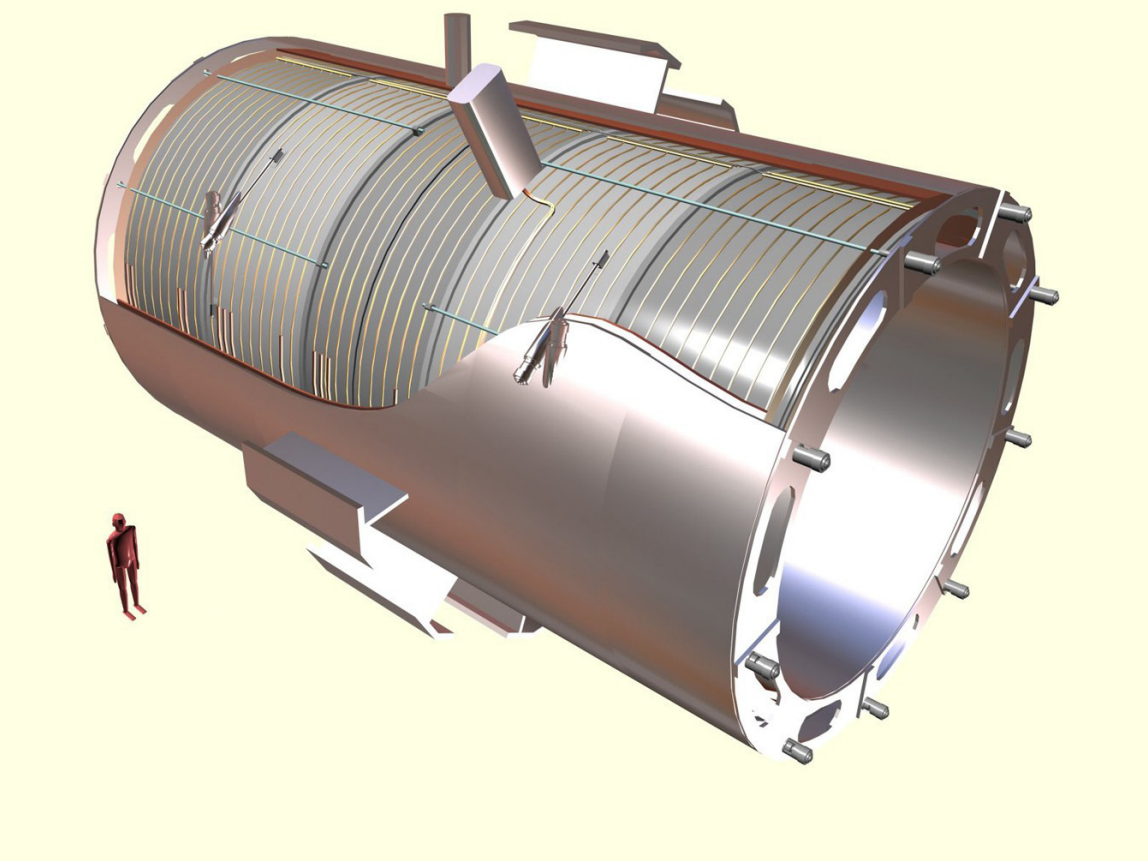

Figure 2.6.: An artistic view of the magnet used for the CMS experiment [21, p. 6].

A superconducting solenoid [19, p. 9-11] is used in the CMS experiment to provide a magnetic field of 3.8 T within the magnet coil shown in Figure 2.6. The magnetic field is

returned by the iron yoke built around the magnet coil. The direction of the field within the coil is parallel to the beam axis bending trajectories of the charged particles in the plane transverse to the beam axis abd in the opposite direction in the iron yoke.

2.1.5. Muon detector system

The muon detector system comprises three different types of gaseous detectors as shown in Figure 2.7: drift tubes (DT) in the barrel region of the detector, cathode strip chambers (CSC) in the endcap region and resistive plate chambers (RPC) in both detector regions [19, p. 11-13]. The DTs are chosen for the barrel region with $|\eta| < 1.2$ because of the small neutron induced background. In the endcap region with $|\eta| < 2.4$ both characteristics are high and therefore the CSCs were chosen. Both detector types are supported by RPCs with a complementary measurement: the RPCs provide a good time resolution, whereas the other detector components provide a good spacial resolution. The detector system components are placed within gaps in the iron yoke. The main purpose of the muon detector system is to determine precisely the tracks and the momentum of muons in the outermost region of the CMS detector. These measurements are used to complement the information on muons collected in the inner tracking system and the calorimeters.

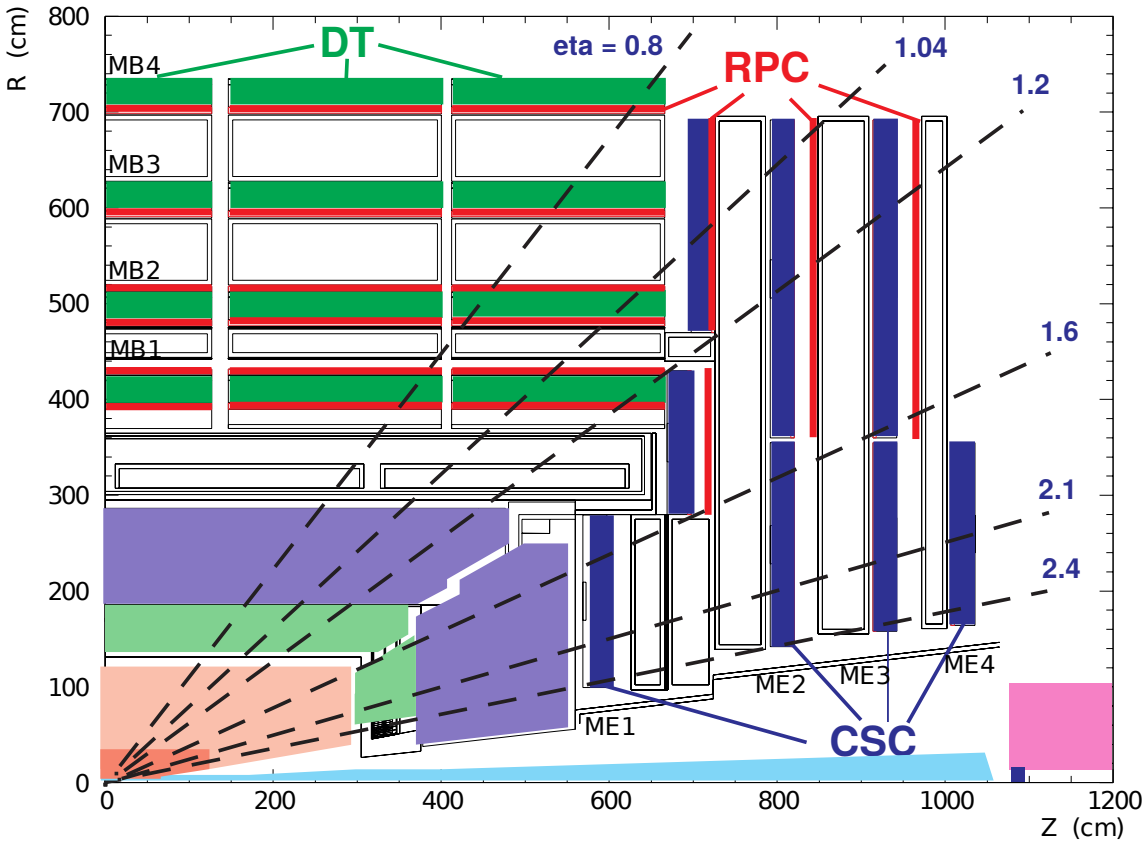

Figure 2.7.: A sketch of the muon detector system showing its different subregions in the spacial (R, z)-plane [19, p. 12].

2.1.6. Forward detectors

All detector systems mentioned so far cover the pseudorapidity region up to $|\eta| = 5.0$. Since a sizable amount of particles from the proton-proton collisions reach out to higher η values, additional forward detectors are installed [21, p. 156-161]. The Centauro And Strange Object Research (CASTOR) and the Zero Degree Calorimeter (ZDC) cover pseudorapidity regions of $5.2 < |\eta| < 6.6$ and $|\eta| \geq 8.3$ respectively. Both detectors measure electromagnetic and hadronic energy in corresponding sections.

2.2. Reconstruction

This section describes reconstruction techniques relevant for this thesis, beginning with the trigger system used in the online reconstruction during the data taking, alignment and calibration procedures and up to the offline reconstruction of several physics objects after the data has been recorded to tape.

2.2.1. Trigger

Proton-proton collisions at the LHC are provided at a high rate with proton beams crossed each 25 ns. For a deadtime free readout of interesting physics processes a trigger system is installed to filter events of interest [21, p. 247-260]. At the CMS detector, the trigger system is devided into a Level-1 Trigger (L1) and a High Level Trigger (HLT). The rate of 40 MHz provided by the detector is reduced to about 100 kHz by the L1 and then further to about 1 kHz by the HLT. The reduced output is then processed further to the data storage.

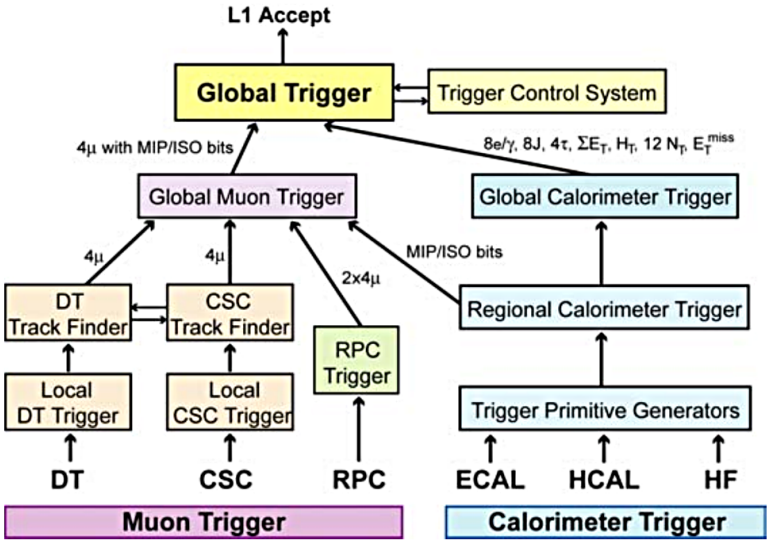

Figure 2.8.: Structure of the Level-1 Trigger in the CMS detector [21, p. 248].

The structure of the L1 is shown in Figure 2.8. The time allowed from reading a buffered input of a given bunch crossing to forwarding the trigger decision back to the detector front-end electronics in order to transfer the full detector output further in case of a positive decision is 3.2 μs. The resulting time to compute the trigger answer is below 1 μs.

This motivates the design of the L1, where only the information from the calorimeters and from the muon detector system is used. The L1 is subdivided into a Muon Trigger and Calorimeter Trigger. Both have local, regional and global components.
In the Calorimeter Trigger, trigger towers in the (η, ϕ)-plane are computed for the each calorimeter by summing up the energy within the trigger tower range. This range is usually larger than the range used for the offline analysis to save time. After this local stage, patterns are used to identify particle candidates for electrons, photons and minimum ionizing particles and regional transverse energy sums are computed within the regional stage. The full information is then processed further to the last global stage of the trigger, which reconstructs objects from the entire detector information like the missing transverse energy, jets and the electrons and photons. Furthermore, the information on minimum ionizing particles and on the regional transverse energy sums is transferred to the Global Muon Trigger.
The Muon Trigger provides hit patterns and track segments in the drift tubes and the cathode strip chambers as local information. After that, the local information from DTs and CSCs is used to construct regional tracks. Additionally, tracks are constructed also from resistive plate chambers using regional hit patterns. The information from all three components of the muon detector system and the information on minimum ionizing particles and transverse energy sums from the Calorimeter Trigger is combined at the global stage to provide muon candidates with improved momentum resolution and efficiency.
If an event from a bunch crossing is accepted, the full data on the event is transferred further to the Data Aquisition (DAQ) system which is used by the HLT. The HLT is a software based trigger which applies selection code similar to the offline selection on the L1 trigger objects. These objects serve as L1 seeds to construct HLT objects. The selection used by the HLT can be usually devided into two main steps. The first step, referred as Level-2, is, like the L1, based only on the information from the calorimeters and the muon detector system but using a higher resolution. After that, intermediate steps can be performed using only a subregion of the inner tracking system, which is determined from parameters providTed by the Level-2. For the final step of the HLT, referred as Level-3, the full information from the pixel and strip trackers is used to construct a Level-3 trigger object. Several trigger objects passing predefined cuts on their transerve momentum, pseudorapidity and isolation are combined to HLT trigger paths used in the offline analysis to select appropriate events.

2.2.2. Alignment and calibration

To make a proper interpretation of the recorded data from the CMS detector, the exact conditions must be known, in particular the calibration parameters and the global position of all detector components within the CMS detector [19, p. 70-76].
There are three different workflows to measure these conditions [22]: Online, Prompt Calibration Loop (PCL) and Offline workflows. The Online worflow is needed for the HLT and the first reconstruction of the data and must be done within one minute. The PCL workflow is performed within 48 hours and provides conditions for the reprocessing of the first data reconstruction. The Offline workflow contains state-of-the-art measurements of the detector conditions and is used for reprocessing campaigns for already available data to improve the accuracy.

The calibration parameters of detector components like gain and pedestal are measured for corresponding detector components under controlled conditions and saved for certain periods of time called Intervals of Validity (IOV). To measure the alignment, laser systems are available in case of the inner tracking system and the muon detector system. Additionally, a track-based alignment measurement is performed using $Z \rightarrow \mu\mu$ events. For each detector unit, six parameters must be saved corresponding to the three rotational and three translational degrees of freedom. The meaurement of the alignment is challenging and is required to be precise in particular for the pixel tracker because of its high resolution of up to 10 µm and high sensitivity to displacements of the same order. Like for the calibration, the alignment parameters are also stored for certain IOVs.

2.2.3. Track and vertex reconstruction

For each bunch crossing in the CMS detector multiple proton-proton collisions take place. Therefore, it is very important to reconstruct precisely the tracks from charged particles traversing the detector and subsequently determine the primary vertices of these collisions. In that way, the main collision in a bunch crossing can be distinguished from other collisions sorting the reconstructed primary vertices by the transverse momentum contribution of the tracks to a vertex. The first primary vertex in this sorted list corresponds to the main collision and the remaining collisions are subsumed to pile-up. The separation of particles coming from the first primary vertex from particles coming from pile-up plays a crucial role in the reconstruction of physics objects.
The track reconstruction is devided in two main steps [23, p. 5-26]: first, the recorded signals in the pixel tracker and in the strip tracker must be clustered to reconstructed hits. To achieve this, the calibration parameters for the complete inner tracking system are used to determine the position of the charge barycenter of a cluster in the local coordinate system of a sensor. This step is referred as local reconstruction.
After the local reconstruction is performed, reconstructed hits are used to reconstruct tracks in the inner tracking system by an iterative approach using the Combinatorical Track Finder (CTF) algorithm within each iteration. To prepare for the track reconstruction, the local coordinates of the reconstructed hits are translated into the global coordinate system taking into account the differences between the ideal and real detector geomentry stored in the alignment records.
With each new iteration, tracks are reconstructed, which are more difficult to find compared with the previous iteration. The difficulty is characterised by the transverse momentum of the track and the distance to the interaction point. At the beginning of an iteration, seeding trajectories are constructed from two or three reconstructed hits. The CTF algorithm based on the Kalman filter [24] is used for track finding by extrapolating the seeding trajectories further into the inner tracking system and searching for additional hits that fit to the track. After all fitting reconstructed hits are found, a final fit is performed to provide the best possible estimate of the track. Finally, quality criteria are assigned to the track and the hits of the tracks passing the criteria of the current iteration are removed for the next iteration to reduce the combinatorical complexity.
After the track reconstruction is finished, the vertex reconstruction is performed done in three steps. At first, appropriate tracks are selected fulfilling the requirements on the number of assigned reconstructed hits in the pixel tracker and the strip tracker, on the significance of the transverse impact parameter to the beam spot and the χ^2 value from

the trajectory fit. Next, the tracks are clustered to groups originating from the same vertex by an algorithm based on statistical mechanics. Finally a vertex fit is performed for each track cluster. Several variations of the usual track reconstruction are used for electrons, muons and High Level Trigger objects.

2.2.4. Muon reconstruction

The first step in the muon reconstruction [25, p. 5-7] is to determine tracks in the inner tracking system and the muon detector system independently. Next, two basic approaches are performed: the Global Muon reconstruction and the Tracker Muon reconstruction.
The Global Muon reconstruction starts in the muon detector system propagating a track in the muon detector system towards the inner tracking system. If a set of tracks in the inner tracking system is found to be near to the propagated track, the best matching inner track is taken by comparing the paramaters of each inner track with the corresponding parameters of the track from the muon detector system, all tracks propagated to a common surface. After the best pair of an inner and outer track is found, the Kalman filter [24] is applied on the reconstructed hits of both tracks to deliver a track traversing the full CMS detector. The Global Muon reconstruction has its highest efficiency for muons with multiple hits in the muon detector system.
The Tracker Muon starts in the inner tracking system considering all reconstructed tracks passing certain momentum thresholds as muon candidates. The tracks are extrapolated to the muon detector system, taking the magnetic field, energy losses and multiple sccattering into account. If a single hit in a muon detector component is found to be near the track extrapolation this muon detector component, the candidate is identified as a valid Tracker Muon. The Tracker Muon reconstruction has its highest purity for low energetic muons. During the further reconstruction process, additional characteristics are added to the muons, like their isolation and identification criteria.

2.2.5. Photon and electron reconstruction

Before the photon and the electron reconstruction can be performed, the energy in the electromagnetic calorimeter must be clustered [19, p. 365-368]. At first, 3x3 or 5x5 arrays of crystals in the electromagnetic calorimeter are built. Since the photons and the electrons must traverse the detector material before reaching the electromagnetic calorimeter and since they are under the influence of a the magnetic field, the energy of a photon or electron is spread in the (η, ϕ)-plane in the electromagnetic calorimeter. The 3x3 or 5x5 cluster may not cover this effect and therefore superclustering algorithms are applied additionally to build a supercluster from a certain amount of 3x3 or 5x5 clusters.
To reconstruct photons [19, p. 375-390], discriminating techniques must be used to distinguish prompt photons with high transverse momentum from converted photons or photons from a π^0 decay. To measure the spread of the energy of a photon, the energy within a 3x3 cluster, centered at the crytal with the highest energy deposit, is compared with the energy of the corresponding supercluster. The ratio between these two energy values allows to distinguish between the prompt photons with the ratio near one and the converted photons with a convertion vertex far from the electromagnetic calorimeter and therefore a ratio closer to zero. Isolation requirements can additionally be used to separate

prompt photons from photons originating from secondary decays.
During the electron reconstruction several difficulties have to be solved [19, p. 390-403]. Before reaching the electromagnetic calorimeter, electrons traverse detector material and may radiate off a sizable amount of their initial energy in form of photons. This effect introduces a spread of the electron energy in the electromagnetic calorimeter. Some amount of the initial energy may be lost completely because the radiated photons may convert into e^+e^- pairs. These electrons can be forced by the magnetic field to stay within the region enclosed by the electromagnetic calorimeter and deposit most of their energy for example in the material of the inner tracking system. To account for these effects, the clustering in the electromagnetic calorimeter and the track reconstruction are adapted. The seeds used in the track reconstruction are found by matching a supercluster in the electromagnetic calorimeter to them. The default track finding algorithm, the Kalman filter, which assumes Gaussian fluctuations for the measured energy, has a poor accuracy because of the energy loss of the electrons. The Kalman filter can account for fluctuations due to multiple scattering but not for fluctuations due to radiation in the detector material. A generalization of this filter, the Gaussian Sum filter (GSF), is found to be the appropriate track finding algorithm for electrons.

2.2.6. Particle flow algorithm

For each recent analysis it is important to have a high efficient reconstruction of particles in a wide energy range. The solution at CMS is the particle flow algorithm, that provides collections of all measureable particles by combining all information from the detector [26, 3-9].
First, an adapted particle-flow clustering in all seven calorimeter components described in subsection 2.1.2 and in subsection 2.1.3 is applied in three steps. At the beginning of the clustering, seeds are determined in each calorimeter component by choosing cells in the calorimeter component with an energy deposit above a given threshold. Next, topological clusters are grown sharing at least one cell side with the already existing cluster and having an energy deposit significantly above a given noise threshold. These topological clusters are then separated into particle-flow clusters of the same amount as the number of seeds within the tolopogical cluster.
After the calorimeter clustering, a linking between the detector components is performed to provide all reasonable possibilities to combine tracks in the inner tracking system, the energy deposits in the calorimeters and the tracks in the muon detector system to particles. Tracks in the inner tracking system reconstructed as described in subsection 2.2.3 are first propagated into the components of the calorimeter and particle-flow clusters crossed by the propagated track are linked to it. By computing the distance within a linked pair the quality of the link is defined. To account for photons radiated by electrons traversing the inner tracking system, tangents to the reconstructed tracks are extrapolated at the interception points of the tracks and the layers of the pixel tracker and strip tracker. Calorimeter clusters crossed by the tangents are also linked to the corresponding tracks. Similarly, clusters in the electromagnetic calorimeter and the hadronic calorimeter are linked. The last possible link is between the tracks in the inner tracking system and the muon detector system. A link is established when the global fit of a pair of an inner and an outer track is sufficiently good. At the end of the linking procedure, blocks of linked objects are created.

Finally, a particle reconstruction is performed for each block. First muons are reconstructed, followed by the reconstruction of electrons. By comparing the momentum of the tracks with the energies of the calorimeter clusters, the amount of charged hadrons, neutral hadrons and photons is determined in each block. To perform this step, the energy in the electromagnetic and hadronic calorimeter is recalibrated to allow for energy deposits of hadrons in the electromagnetic calorimeter and of photons in the hadronic calorimeter. This algorithm provides a unique assignment of reconstructed energy deposits in the detector components to physics particles and avoids double counting.

2.2.7. Jets and missing transverse energy

For hadron colliders, a profound reconstruction of high-level objects like jets and the missing trasverse energy is necessary to perform a decent analysis [19, p. 404-436].
Jets are reconstructed by iterative jet algorithms, which combine the energy deposits in the calorimeters. Since the electromagnetic and the hadronic calorimeters have different granularity, the energy of the crystals in the electromagnetic calorimeters are summed to towers in the (η, ϕ)-plane, which correspond to individual cells of the hadronic calorimeter. After supression of noise in the calorimeters, the jet algorithms combine the towers to jets starting at towers with highest amount of energy. Usually, the anti-k_T-algorithm [27] is used to reconstruct jets. In contrast to other jet algorithms, the anti-k_T-algorithm provides highly desired properties like infrared and collinear safety. It reconstructs reliably jets with high transverse momentum originating usually from the hard process and jets with low transverse momentum coming from the underlying event. The term underlying event summarizes all interactions excluding the hard process, which come from the same proton-proton collision. Usually, such interactions are low energetic.
Assuming that right before a proton-proton collision both protons have a vanishing transverse momentum, the transverse momentum of particles which rarely interact with the detector material like neutrinos can be determined by summing up all transverse energy measured by the detector. Because of the momentum conservation, this sum is expected to be zero in the case that all particles coming from the proton-proton collision are measured. This sum, the negative of the missing transverse energy, gives a handle on the reconstruction efficiciency for an event because it is sensitive to all particles reconstructed in the detector. Different techniques are used to compute the missing transverse energy, including also multivariate regression methods.

3. Embedding

Embedding describes a method to estimate the $Z \to \tau\tau$ background in $H \to \tau\tau$ searches. The aim of this method is to benefit both from Monte Carlo simulations and data. Possible discrepancies in the description of pile-up and the underlying event in the simulation are avoided by taking the event from data. Due to lepton universality one can select the well understood Z-boson decay into two muons in data, $Z \to \mu\mu$, and replace the di-muon system with the di-τ-system of a simulated $Z \to \tau\tau$ decay obeying conservation laws and using the same kinematics as given by the muons. In that way, only the decay of two τ-leptons must be simulated and therefore, uncertainties related to the simulation are reduced, especially those related to the remaining event. A graphical description of embedding is shown in Figure 3.1.

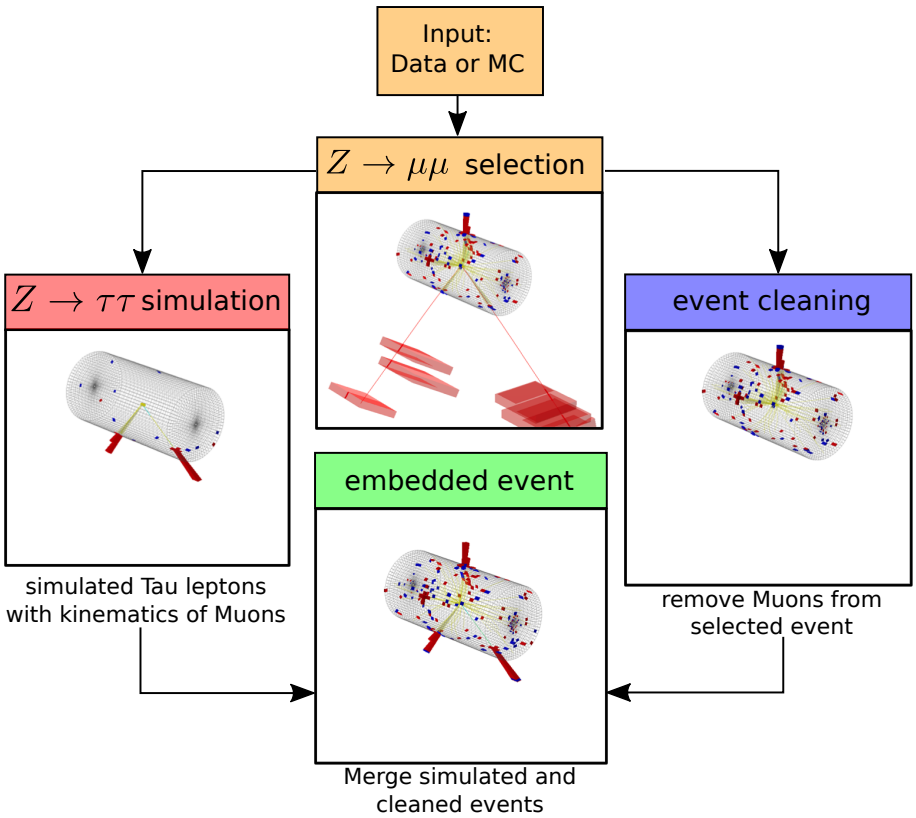

Figure 3.1.: Graphical description of embedding. First, appropriate events with two muons are selected from data or simulation (MC). Next, the two selected muons from the $Z \to \mu\mu$ decay are used to simulate a $Z \to \tau\tau$ decay. Besides that, their footprint is removed from the selected data. These steps are called $Z \to \tau\tau$ *simulation* and *event cleaning* respectively. In the last step, the simulated Z-boson decay and the data after the cleaning step are merged to a hybrid event containing both data and simulation, called *embedded* event in the following.

This chapter is devided into four sections. In section 3.1, the $Z \to \mu\mu$ selection is described. The contribution of several processes to the event selection with two reconstructed muons is quantified explicitly with the $Z \to \mu\mu$ decay found to be the main contributing process. A method to avoid biases due to the selection of the two reconstructed muons is introduced in subsection 3.1.1.
In section 3.2, the simulation of the Z-boson decay is described. For validation of the method itself, it is possible to simulate $Z \to \mu\mu$ events instead of $Z \to \tau\tau$ events and to compare these with usual $Z \to \mu\mu$ simulated events. Several adjustments to the simulated Z-boson decay are described like the correction of its vertex and the 4-momenta of the leptons. A procedure is described that allows to retain all selected events by forcing the simulated decay to obey kinematic thresholds. Because the masses of the muons and τ-leptons are different, $m(\mu) = 0.106$ GeV and $m(\tau) = 1.777$ GeV [14, p. 14, 15], respectively, effects due to final state radiation (FSR) of the leptons are examined in subsection 3.2.1.
In section 3.3, the *event cleaning* is described. During this step, the footprints of the two selected muons are removed from the data on an event-by-event basis. It is discussed, at which object level in the reconstruction workflow this removal can be performed and which effects may be introduced by this.
In section 3.4 the *merging* step of the simulated and the cleaned events is described. Also for this step, the object level will be discussed, at which the merging should be performed. This choice eventually depends on the alignment and the calibration of the detector described in subsection 2.2.2, which may be different between data and simulation.

3.1. $Z \to \mu\mu$ selection

There are two main goals, that must be satisfied by the $Z \to \mu\mu$ selection used for embedding: on one hand, the selected muons should have a good quality. On the other hand, it must be avoided to introduce biases to the data. Selection criteria chosen for the embedding are described in the following.
At first, the event is required to have passed one of the two high level triggers (HLT) constructed for muons [28]. For these triggers, two muons with transverse momentum $p_T > 17$ GeV and $p_T > 8$ GeV respectively must be reconstructed online during the data taking. The muon with the higher p_T is required for both triggers to be a muon with a successfull L3 Muon reconstruction. The muon with the lower p_T must be either a L3 Muon for the first HLT or a Tracker Muon for the second HLT path. For each muon loose track isolation requirements are imposed to reduce the rate of each HLT path.
Next, kinematic requirements on the transverse momentum p_T and the pseudorapidity η of the muons are applied: $p_T > 8$ GeV and $|\eta| < 2.5$. Additionally, the muons are required to be reconstructed as global as explained in chapter 2.
The last step is an attempt to combine the two muons with opposite charge to a neutral Z-boson candidate required to have an invariant mass $m(\mu\mu)$ larger than 20 GeV. The event is taken into account, if at least one Z-boson candidate is found. In case more than one candidate is found, the one closer to the measured Z-boson mass $m(Z) = 91.188$ GeV [14, p. 9] is selected. The two muons of this Z-boson candidate are chosen for further embedding steps.
The comparison between data and the expected background contributions to the distribution of the invariant mass of the di-muon-system, $m(\mu\mu)$, after applying the $Z \to \mu\mu$

selection discussed is shown in Figure 3.2.
The largest contribution to the selection is made up from the Drell-Yan process, where a Z-boson is produced through quark-antiquark annihilation and decays into a pair of muons or τ-leptons. A final state with two muons can be achieved directly through the $Z \rightarrow \mu\mu$ decay or a decay chain $Z \rightarrow \tau\tau \rightarrow \mu\mu + 4\nu$, where the τ-leptons decay into muons and neutrinos through the weak interaction. In Figure 3.2 the Drell-Yan process is summarized as $Z \rightarrow ll$ in the histogram.
Next contribution results from a pair of two tops quarks, $t\bar{t}$, decaying in two muons, two bottom quarks and two muon neutrinos through the weak interaction, $t\bar{t} \rightarrow \mu\mu + b\bar{b} + 2\nu$. This contribution is summarized as $t\bar{t}$ + jets in the histogram in Figure 3.2.
Processes, which involve a combination of two vector bosons also contribute to the $Z \rightarrow \mu\mu$ selection. The combinations of W and Z bosons, WW, WZ and ZZ, can lead to a presence of at least two muons with opposite charge in the final state. Events with more than two reconstructed muons can be filtered if necessary but in case the reconstruction of the surplus muons fails, the event would pass the filter. These processes are summarized as Di-boson in the histogram.
The last considered contribution, for which a W-boson with an additional jet passes the selection can contribute in cases where a missidentified muon. This contribution is summarized as W + jets in the histogram.
So far, only contributions were considered for which simulated datasets are used. Another possible contribution resulting from QCD processes is usually estimated from data, but neglected here.

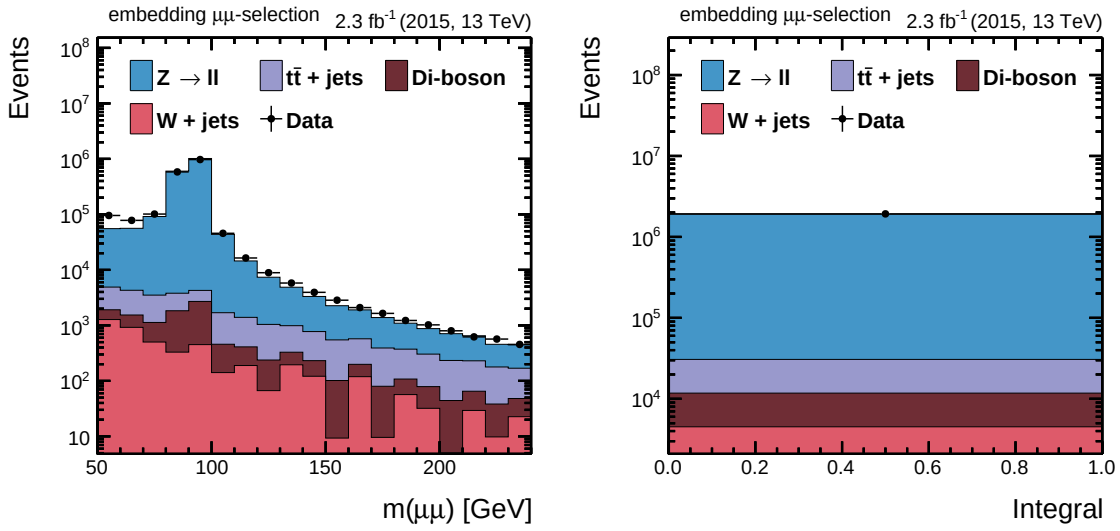

Figure 3.2.: Distribution of the visible mass $m(\mu\mu)$ and the integral for selection described in this section. An additional selection cut $m(\mu\mu) > 50$ GeV is imposed for both plots to focus on the region near the Z peak.

The yields of the expected backgrounds that constitute the integral in the right plot of Figure 3.2 and their fractions of the total yield are shown in Table 3.1. The selection used for the embedding provides as expected a sufficiently pure sample with about 98% of the total yield resulting from the Drell-Yan process and a decent agreement with the data for mass values of interest that are above 50 GeV.

background	yield	fraction of the total yield
$Z \to ll$	1879340.0	98.39%
$t\bar{t}$ + jets	18919.4	0.99%
Di-boson	7256.1	0.38%
W + jets	4510.7	0.24%
total background	1910026.2	100.00%

Table 3.1.: The yields of the expected backgrounds of the $Z \to \mu\mu$ selection used for embedding.

3.1.1. Mirroring

Through the selection requirements for embedding a bias is introduced when stricter isolation requirements are imposed on the selected muons. An environment around these muons is chosen, that obeys these requirements. In turn this leads to a τ-lepton reconstruction efficiency different from the one, that is expected for an untouched environment. This effect is referred as selection bias.
The main idea of this subsection is to introduce a method, which changes the phase space of the simulated τ-leptons with respect to the selected muons to reduce the effect. At the same time, the method should leave the distributions of quantities derived from 4-momenta and the polarization of the Z-boson candidate invariant. From these two restrictions, an appropriate transformation will be derived, called *mirroring*.
First the restriction on the Z-boson candidate polarization will be discussed. Since the Z-boson is a massive vector boson, it has three different polarization states, two transverse and one longitudinal [29, p. 16,17]. These polarizations correspond to eigenstates of the spin operator in direction of the 3-momentum of the Z-boson. The two transverse polarizations have eigenvalues ± 1, whereas the longitudinal polarization has the eigenvalue 0. This is illustrated in Figure 3.3.

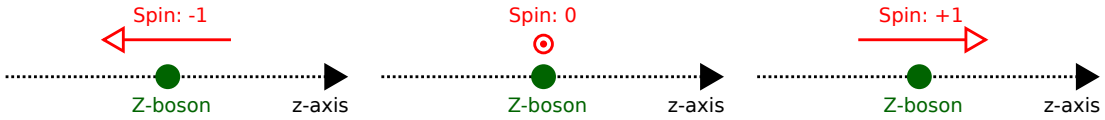

Figure 3.3.: Spin eigenstates of the Z-boson, that can be associated with its polarisation states. The left and right plots show the spin states according to the transverse polarizations, the middle plot shows a spin state according to the longitudinal polarization.

In Figure 3.4 the spin configuration of the $Z \to \mu\mu$ decay is shown. The distribution of the angle θ between the axis of the Z-boson 3-momentum in the laboratory frame and the 3-momentum of the muon in the Z-boson rest frame is non-trivial and can change with the Z-boson p_T. This means, the $Z \to \mu\mu$ decay retains only a cylindrical symmetry and only rotations of the muons in the plane perpendicular to the axis of the Z-boson 3-momentum are allowed to guarantee the conservation of the Z-boson polarization and the spin. Since the 3-momentum components perpendicular to the boost direction remain unchanged after applying boosts in the Z restframe along the axis of the Z-boson 3-momentum, these rotations can be applied in the laboratory frame instead of the Z-boson rest frame.

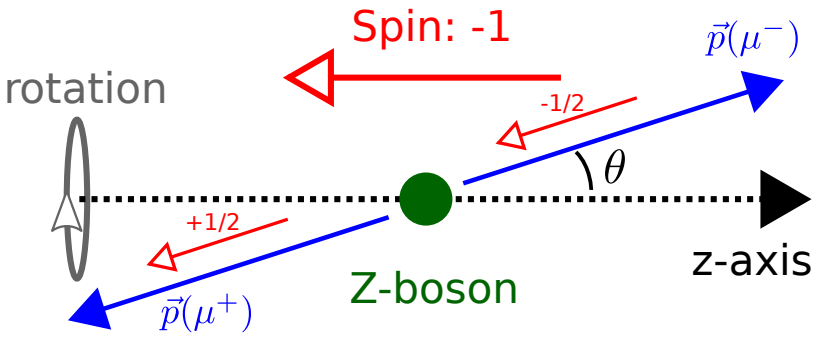

Figure 3.4.: Spin configuration of a Z-boson decay. In this case, a transverse polarized Z-boson is shown with spin eigenvalue -1. A rotational symmetry remains for the decay, marked as a gray arrow. This is also true for other spin eigenstates of the Z-boson. All spins are shown as red arrows, the momenta of the muons as blue arrows. The flight direction of the Z-boson corresponds to the black z-axis.

Next the restriction on the invariance of distributions for the 4-momenta of the selected muons will be considered. The importance of this restriction will be discussed on an example of an orthogonal transformation, which may come first to mind: a rotation by 180°.
Important quantities derived from the 4-momentum are for example the transverse momentum p_T and the pseudorapidity η. These quantities are computed with respect to the beam axis and therefore depend directly on the laboratory frame they are defined in. This means, the distributions stay invariant only for orthogonal transformations like rotations around the beam axis. Transformations with respect to other directions like the axis of the Z-boson 3-momentum may change the distributions of p_T and η.
To examine the changes of the p_T value, one can define a plane spanned by the axis of the Z-boson 3-momentum, $\vec{p}(Z)$, and the direction of the beam, called *Z-beam plane* in the following. The transverse momentum vector, $\vec{p_T}$, can then be decomposed into components perpendicular and parallel to the Z-beam plane, $p_T^{\perp}$ and $p_T^{||}$ respectively. The $Z \to \mu\mu$ decay itself spans another plane, which is rotated around the axis of the Z-boson 3-momentum with respect to the Z-beam plane.
Figure 3.5 shows the impact of the rotation by 180° around the axis of the Z-boson 3-momentum, $\vec{p}(Z)$, on the value of the transverse momentum of a muon from the $Z \to \mu\mu$ decay, $p_T(\mu)$. In the left plot of Figure 3.5 the perpendicular component $p_T(\mu)^{\perp}$ is examined. After the rotation by 180°, $p_T(\mu)^{\perp}$ only changes the sign, thus its absolute value remains unchanged. For any other rotation this wouldn't be the case.
In contrast to that as shown in the right plot of Figure 3.5, a rotation by 180° around $\vec{p}(Z)$ affects the absolute value of the parallel component, $p_T(\mu)^{||}$. Moreover, the component of the muon 3-momentum parallel to the beam direction, $p_z(\mu)$, is also changed. The full value of the transverse momentum of the muon reads:

$$p_T(\mu) = \sqrt{p_T(\mu)^{\perp} \cdot p_T(\mu)^{\perp} + p_T(\mu)^{||} \cdot p_T(\mu)^{||}}$$

Thus, $p_T(\mu)$ changes after a rotation by 180° around $\vec{p}(Z)$. The pseudorapidity of the muon, $\eta(\mu)$, also changes as a function of $p_T(\mu)$ and $p_z(\mu)$.
To avoid these changes in p_T and η, the following transformation is proposed: the per-

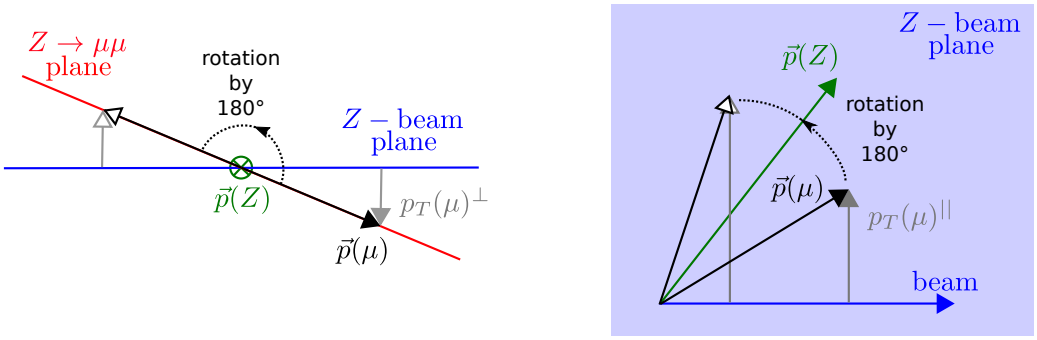

Figure 3.5.: The effects of a rotation by 180° around the axis of the Z-boson 3-momentum. The left plot shows the rotation for the perpendicular component of p_T with respect to the Z-beam plane. The view is towards the flight direction of the Z-boson such that the Z-beam plane appears as a line. The right plot shows the rotation for the parallel component of p_T with respect to the Z-beam plane. The view is from the top on the Z-beam plane. Initial vectors have filled arrowheads, transformed vectors have outlined arrowheads.

pendicular component of the transverse momentum, $p_T(\mu)^{\perp}$, is mirrored on the Z-beam plane, whereas the parallel component $p_T(\mu)^{||}$ and $p_z(\mu)$ are left unchanged. In that way, $p_T(\mu)$ and $\eta(\mu)$ stay invariant under the transformation. The only quantity of the muon 4-momentum which changes after the transformation eventwise is the azimutal angle of the muon, $\phi(\mu)$. This is not a problem, since the distribution of $\phi(\mu)$ is flat. In Figure 3.6 the mirroring transformation is summarized.

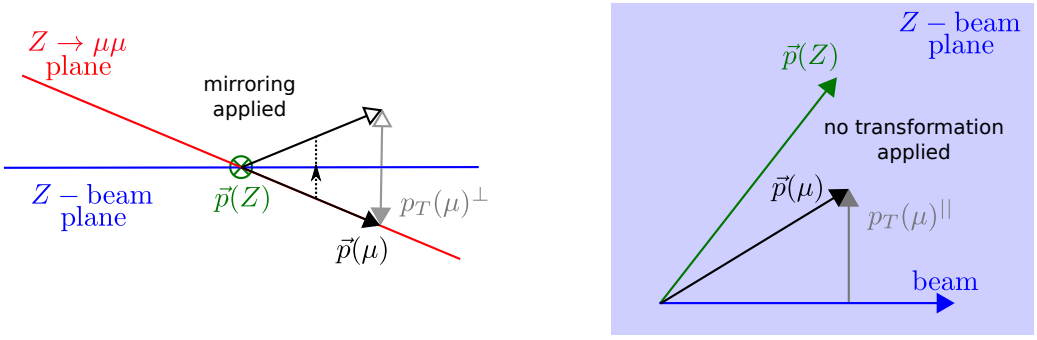

Figure 3.6.: The mirroring transformation. The left plot shows the impact of mirroring on the perpendicular component of p_T with respect to the Z-beam plane. The view is towards the flight direction of the Z-boson such that the Z-beam plane appears as a line. The right plot shows the impact of the mirroring on the parallel component of p_T with respect to the Z-beam plane. The view is from the top on the Z-beam plane. Initial vectors have filled arrowheads, transformed vectors have outlined arrowheads.

3.2. Simulation of the Z-boson decay

At the beginning of the simulation step described in this section there are two ways to proceed. For the purpose to validate the method itself a $Z \to \mu\mu$ decay can be simulated, instead of the application of the embedding method, for which a simlated $Z \to \tau\tau$ decay is needed. For the first decision, one simply saves the 4-momenta of the two selected muons and the Z-boson candidate built from these two muons in a format suitable for the event simulation software, chosen to be the Les Houches Event (LHE) format [30]. For the second decision, additional adjustments and calculations must be made:
At first one has to correct the 4-momenta of the two τ-leptons for the difference between $m(\mu)$ and $m(\tau)$. This correction has to obey conservation laws.
The first restriction is that the invariant mass of the di-lepton-systems, $m(\mu\mu)$ from selected data and $m(\tau\tau)$ to be simulated, must be equal, since the leptons form a Z-boson candidate:

$$m(Z) = m(\mu\mu) = m(\tau\tau) \tag{3.1}$$

From this equation the energy of each τ-lepton in the rest frame of the Z-boson candidate, $E(\tau)$, can be derived. Keeping in mind, that the two τ-leptons are back to back in this frame, $E(\tau) = m(\mu\mu)/2$ follows. With this information, the magnitude of the τ-lepton 3-momentum, $\vec{p}(\tau)$, can be further constrained. Then, a scalefactor s can be derived for the known muon 3-momentum in the Z-boson candidate rest frame, $\vec{p}(\mu)$, to compute $\vec{p}(\tau)$:

$$\begin{aligned} \vec{p}^{\,2}(\tau) &= \frac{1}{4}m^2(\mu\mu) - m^2(\tau) \\ s &:= \frac{|\vec{p}(\tau)|}{|\vec{p}(\mu)|} \end{aligned} \tag{3.2}$$

The τ-lepton 4-momentum in the Z-boson candidate rest frame reads:

$$p^\alpha(\tau) := (\frac{1}{2}m(\mu\mu), s \cdot \vec{p}(\mu)) \tag{3.3}$$

Since the 4-momentum of the di-muon system, $p^\alpha(\mu\mu) = p^\alpha(Z)$, is known in the laboratory frame, A boost into the Z-boson candidate rest frame and back can be constructed. Therefore, $p^\alpha(\tau)$ can be transformed into the laboratory frame. After this correction, the derived 4-momenta of the two τ-leptons are saved in LHE format.
In the case of the $Z \to \tau\tau$ simulation, the embedded dataset is used as background estimate and passed through selections done on analysis level. These selections depend on the particular decay channel considered in the analysis. In case of the $H \to \tau\tau$ search, there are six channels in total depending on the decay modes of each τ-lepton. Four of these channels are relevant for the analysis as discussed in section 1.2: $e\mu$, $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$.
For each event, the 4-momenta of the simulated τ-leptons are given by the two muons selected from data as described in section 3.1. But it is not guaranteed, that subsequently simulated decays of the two τ-leptons pass the selection criteria used in the $H \to \tau\tau$ search and such that a significant amoint of events might be lost. To reduce this loss of events, kinematic requirements on the visible decay products compatible with the selection of the $H \to \tau\tau$ analysis are imposed already at the simulation level. These requirements include in particular cuts on the transverse momentum p_T and the absolute value of the pseudorapidity $|\eta|$. As a consequence, a bias is introduced explained on an example in the

following:
The $e\tau_h$ decay channel has to be simulated with the two τ-leptons having initial transverse momenta with $p_T(\tau^-) = 25$ GeV and $p_T(\tau^+) = 20$ GeV. The transverse momentum of the electron is required to be larger than 23 GeV. The transverse momentum of the visible decay product of the hadronically decayed τ-lepton, τ_h, is required to be larger than 18 GeV. Thus, τ^- is forced to decay into e^- and two neutrinos, τ^+ is forced to decay hadronically. Moreover, both visible decay products are restricted to carry most of the transverse momentum of the τ-leptons. The free phase space for this decay is small and therefore, it has a small probability. This means, if the $e\tau_h$ decay is simulated inclusively for the same initial τ-leptons, it will more likely fail the imposed selection criteria. Thus, by forcing the p_T requirements on the visible decay products, the contribution of this event is overestimated. For a complete dataset, this results in a shape changing effect and has to be corrected.
Therefore, selection efficiencies are calculated to obtain a sample which is unbiased. This can be achieved by simulating a $Z \to \tau\tau$ decay multiple ($N = 1000$) times and checking, how often the requirements for one specific decay channel are passed (P times). One of the passing decays is taken for further processing and additionally the efficiency $\epsilon = P/N$ is saved on an event-by-event basis. This efficiency estimates the weight with which an event contributes to the full dataset. The efficiency ϵ will be discussed further as a function of the transverse momenta of the two simulated τ-leptons in chapter 4 for decay channels $e\mu$, $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ with the corresponding kinematic requirements on p_T and η of the visible decay products for each channel.
After the complete $Z \to \mu\mu$ or $Z \to \tau\tau$ decay process is simulated up to the hadronization level, the primary vertex of the process has to be corrected to be at the same position as in the selected $Z \to \mu\mu$ event. For the simulation, the primary vertex is usually calculated from the beamspot applying an additional smearing, which follows a beta function distribution for the transverse plane and a Gaussian in the z-direction. Both distributions are centered at the beamspot. The beamspot in turn is set to a reasonable single value for the complete dataset. Since the simulated Z-boson decay has to be merged into the data, the smearing procedure is replaced with a shift to the reconstructed primary vertex of the selected muons from data. Additionally, the beamspot in the simulation of the Z-boson decay is set to the one measured in data. A quantitative discussion of the performance of the vertex correction will be given in chapter 4.
All subsequent procedures required to arrive at a fully reconstructed simulated event are equal to those used in the usual simulation chain. The simulated Z-boson decay must be processed through detector simulation, which determines the hits and energy deposits of the simulated particles in the active detector material. These hits and energy deposits are then digitalized and translated into a format as returned by detector electronics. This is followed by a reconstruction to high-level physics objects like electrons, muons, particle flow candidates and jets. Several objects built during reconstruction are saved to be used further for the merging step described in section 3.4.
In the following subsection, an effect introduced by the replacement of muons with τ-leptons will be discussed in more detail.

3.2.1. Study of FSR in embedding

Final state radiation (FSR) is dependent on the mass of the radiating lepton and changes its kinematics. Since a muon has a smaller mass than a τ-lepton, it radiates photons more often. If these photons are not captured and consequently removed with the muon footprint during the reconstruction, a bias may be introduced by the replacement of muons with τ-leptons. First, the impact of FSR on the reconstructed invariant mass of the di-muon-system, $m(\mu\mu)$, will be studied. Then, the consequences for the $\mu \rightarrow \tau$-replacement will be discussed.

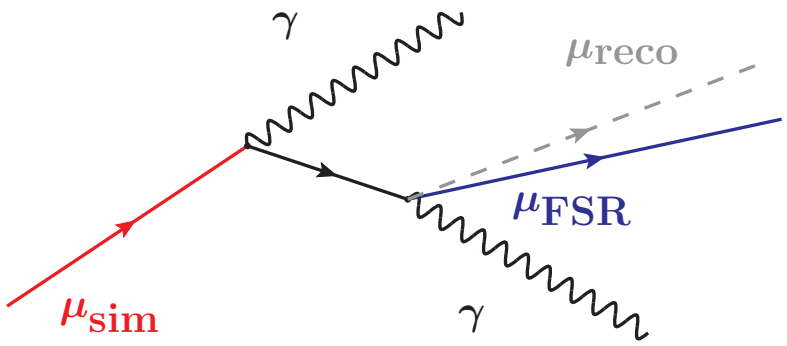

Figure 3.7.: Three different states for a muon accesible in simulation are shown: the reconstructed state, μ_{reco}, the simulated state before final state raditation, μ_{sim} and the simulated state after final state raditation μ_{FSR}. The reconstructed state is expected to be similar to the simulated state after FSR, since photons from FSR aren't recovered during the standard reconstruction algorithms.

As shown in Figure 3.7, three states of the simulated muon are of relevance for this study: the reconstructed state, μ_{reco}, the simulated state before FSR, μ_{sim}, and the simulated state after FSR, μ_{FSR}. In the illustrated case, the FSR consists of two photons, γ, radiated off the muon. Whereas the energy loss due to ionization is covered in the muon reconstruction, since the ionization happens along the track curve of the muon, FSR photons aren't recovered by the standard reconstruction algorithms. An attempt of FSR recovery has been made only in the $H \rightarrow ZZ$ search using adjusted lepton reconstruction algorithms with an efficiency not more than 50% [31].
Due to radiation, the reconstructed muons lose energy, so a shift in the invariant mass $m(\mu\mu)$ to lower values is expected with respect to the simulated state before FSR. In that way, the tail of the $m(\mu\mu)$ distribution below the Z-boson mass is enriched because of events migrating from the peak. This is shown in Figure 3.8 achieved by processing the full simulation chain of embedding up to the reconstructed event for each of the three muon states μ_{reco}, μ_{sim} and μ_{FSR} as initial input. These states are accessed from the simulated Drell-Yan datasets.
If the true muon state before FSR, μ_{sim}, could be accessed in reconstruction and then replaced by a simulated τ-lepton, no effects due to FSR would appear. This can be achieved by proper selection steps. Since the software and settings for the simulation of the τ-lepton decays and the subsequent reconstruction are chosen to be the same as used for the standard CMS simulation chain, the systematic effect on the distribution of the invariant mass of the di-τ-system, $m(\tau\tau)$, will arise only from the unrecovered photons from muon

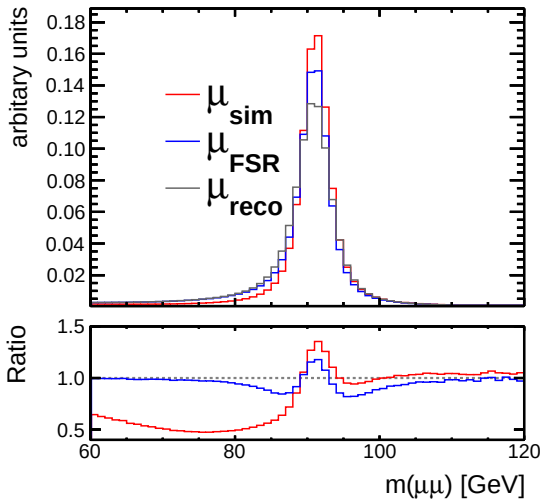

Figure 3.8.: A shape comparison in the invariant mass of the di-muon-system, $m(\mu\mu)$, for the three different states of the muons, μ_{reco}, μ_{sim} and μ_{FSR} is shown. The ratio is built with respect to the distribution of μ_{reco}. The distribution for μ_{reco} is smeared with respect to the distribution for μ_{FSR} because of the repeated reconstruction performed during the embedding procedure as can be seen in the ratio. The distribution of μ_{sim} has a decreased amount of events in the tail below the Z-boson mass due to FSR of the selected muons. As a consequence, the μ_{reco} distribution is shifted to lower mass values with respect to the μ_{sim} distribution.

FSR. A comparison of the embedded dataset and the $Z \rightarrow \tau\tau$ simulation is made in chapter 5.

3.3. Cleaning

This section describes the procedure to remove as much of the footprint of the muons from a Z-boson decay as possible from the selected data. This procedure is called *cleaning*. The data content provided by the detector is at first digitalized. This digitalized information is used to build hit clusters in the pixel and strip trackers and reconstructed hits in the electromagnetic and hadronic calorimeters and the muon chambers. This collection of basic reconstructed objects is used for all following steps in the reconstruction process: track and vertex reconstruction, reconstruction of physics objects like muons and electrons and the particle flow algorithm. A final high level reconstruction step is performed to compute quantities and objects dependent on all so far reconstructed particles in the detector, for example jets, missing transverse energy and isolation sums of the particles. The footprint of the muons is removed at the level of clusters in the pixel and strip trackers and reconstructed hits in the calorimeters and muon detector systems. This is done for the following reasons: at first, performing the cleaning step as early as possible in the reconstruction chain has the advantage, that it is stable against software changes in all subsequent reconstruction steps, which remain untouched by the cleaning step. On the other hand, the amount of effort which is needed to perform the removal of the selected

muons should be taken into account. Cleaning right after the digitalization step is possible, but involves additional work on the association of the digitalized information with the muons to be removed.
Two conceptually different algorithms are used to perform the cleaning. Clusters in the inner tracking system and reconstructed hits in the muon detector system are directly used to reconstruct tracks and therefore referenced in the track object within the software. This means, the clusters and hits used for muon tracks can be directly removed.
Reconstructed hits to be cleaned in the calorimeter system first need to be associated with the muon track. This is done by propagating the tracks of the selected muons through the calorimeter systems identifying those reconstructed hits which are crossed by the trajectories of the muons. The energy deposited by the muon in the crossed hit is then subtracted from each hit. This is illustrated in Figure 3.9 for the electromagnetic calorimeter as an example. The plots show the (x, y) plane of the CMS detector coordinates in cm units. Yellow, green and violet squares show the energy deposited in a certain segment of the detector. The size of the square with the darker shade within the square of the brighter shade of a given color represents the amount of energy that has been deposited.

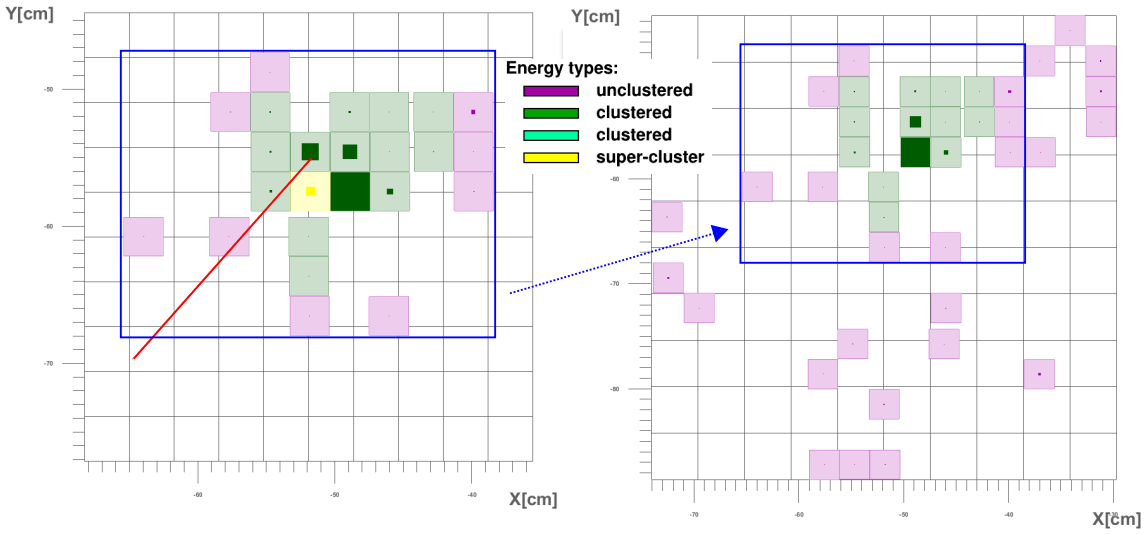

Figure 3.9.: An example for the cleaning of reconstructed hits in the electromagnetic calorimeter associated with the muon track. The trajectory of the muon is shown as a red line, the calorimeter deposits are shown as yellow, green and violet squares in the (x, y) plane. The darker shade of the color within the squares with a brighter shade represents the amount of the energy deposited at the place enclosed by the bright square. The red trajectory of the muon in the left plot is first associated with a yellow and a green energy deposit, which are crossed by the trajectory. In this case, the complete energy of these deposits is removed from the calorimeter. The result of the cleaning is shown in the right plot: the red track and the assosiated energy deposits are missing. The event displays are produced with the Fireworks software [32].

The left plot of Figure 3.9 shows the region in the electromagnetic calorimeter, where a selected muon passed through, marked as a red line. First, two energy deposits are associated with the red track of the muon crossing the calorimeter. Then, the track and the

two calorimeter deposits are removed. The complete energy of these deposits is associated with the track and substracted. The result is shown in the right plot of Figure 3.9: compared to the right plot, the red trajectory of the muon and the green and the yellow energy deposits at the beginning of this trajectory are missing.
After the cleaning and the subsequent reconstruction have been performed for the complete dataset, the following expectation must be checked: Only the muons selected for embedding have been removed from the data while the remaining part of the event is left unchanged. For this reason, the number of muons, electrons and all particle flow candidates are compared between the selected $Z \rightarrow \mu\mu$ data and the data after the cleaning step. For this comparison only global mouns, which are matched with a particle flow muon, are considered, since only these are selected for the embedding procedure. For electrons, loose cut-based identification requirements are imposed [33].
To understand the influence of the cleaning better, this procedure is done in two separated steps. First only the track footprint in the inner tracking system and in the muon detector system is removed. Then, the calorimeter cleaning is applied. All in all, three different datasets are compared with each other: (i) reference data with the selected $Z \rightarrow \mu\mu$ events in black, (ii) data after track cleaning in dark green and (iii) data after full cleaning in bright green.
In Figure 3.10 the number of particle flow candidates is shown. The upper left plot shows the number of all reconstructed particle flow candidates, whereas the remaining upper right, lower left and lower right plots show the number of charged hadrons, photons and neutral hadrons respectively. Remaining particle flow candidates are electrons, muons and electromagnetic and hadronic particles in the forward hadronic calorimeter. For completeness, the comparison plots for the particle flow electrons and muons are shown in Figure A.1. The particles in the forward hadronic calorimeter aren't considered, since this detector component is beyond the relevant pseudorapidity range $|\eta| < 2.5$.
Since the footprint of the two muons is removed from the data, it is expected, that the number of all particle flow candidates is smaller for the cleaned datasets. This is only true for the full cleaning as can be seen in the upper left plot of Figure 3.10. Thus, removing the energy deposits in the calorimeter material not removed in data after track cleaning has a sizeable influence on the reconstruction, which can be examined further by looking at the other three plots. The binning of these plots is chosen equally with a bin width $\Delta N = 10$. For charged hadrons only fluctuations are visible for all datasets after cleaning. In contrast to that, a trend towards higher values for the number of particles is seen after the track cleaning for photons and neutral hadrons. This trend is reduced in both distributions after the full cleaning is performed. Because of the remaining difference for the photons and the neutral hadrons it is expected, that this has an effect on isolation quantities used for reconstructed leptons. For this reason, the p_T sum of particle flow candidates within a $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$-cone around the embedded muon will be discussed later in chapter 4 in the context of the $\mu \rightarrow \mu$-embedding closure test and the mirroring method described in subsection 3.1.1.
The main reason for the surplus of neutral hadrons after the full cleaning has been applied is that not all energy deposits of the selected muons are removed from the hadronic calorimeter. As a consequence, these energy deposits are reconstructed as additional neutral hadrons. A possible solution not studied within this thesis would be to remove not only the assigned energy in the crossed detector segment, but also a certain amount of energy in the surrounding area.

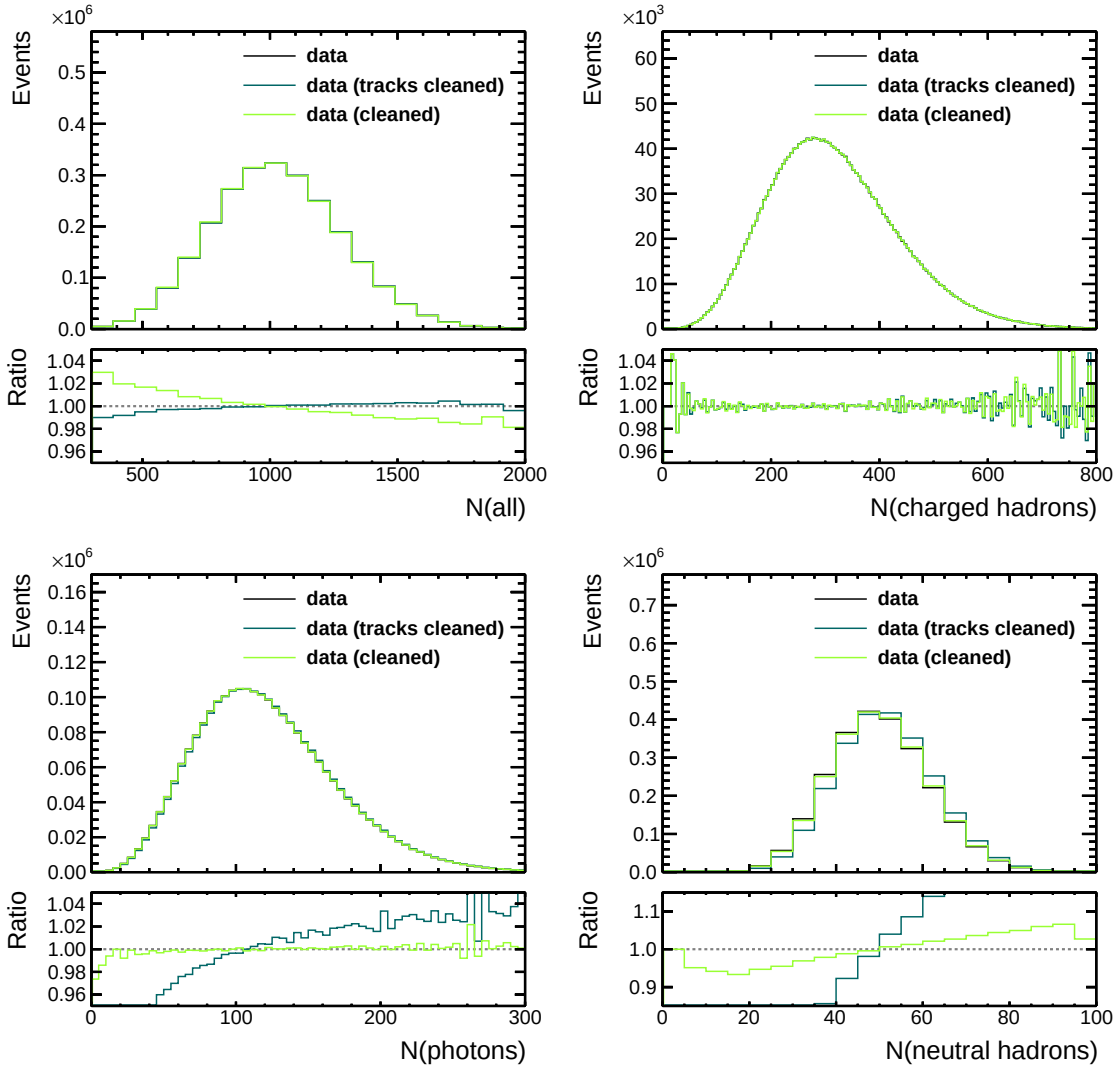

Figure 3.10.: Comparison of the number of particle flow candidates for three datasets: (i) reference data in black, (ii) data after track cleaning in dark green and (iii) data after full cleaning in bright green. The upper left plot shows the number of all particle flow candidates, the upper right, lower left and lower right plot show the number of charged hadrons, photons and neutral hadrons respectively. Trends introduced by the track cleaning for photons and neutral hadrons are reduced, after the calorimeters are cleaned in addition. As a consequence, only the full cleaning shows the expected shift to lower values in the upper left plot, when compared to the reference data.

Next, the number of global muons matched to a particle flow muon is studied, shown in Figure 3.11. For most of the events, the cleaned datasets have two muons less than the reference data, thus showing an expected behaviour. A closer look at the number of muons is taken in chapter 4.
For the number of loose electrons shown in Figure 3.12, the behaviour is unexpected: deviations grow, if additionally to the track cleaning calorimeters are cleaned as can be seen in the left plot of Figure 3.12. The expected ideal behaviour would be, that the reconstructed

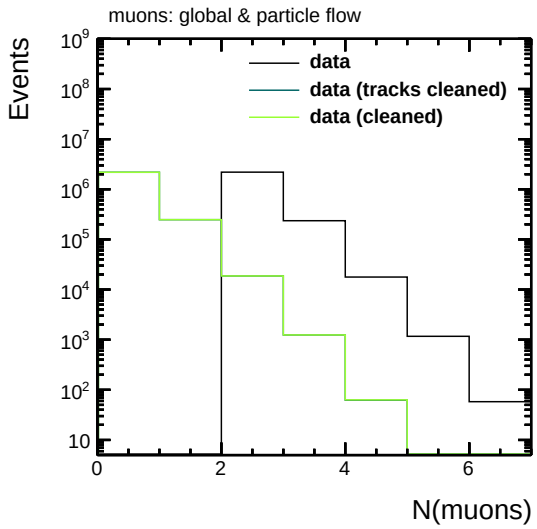

Figure 3.11.: Comparison of the number of global muons matched to a particle flow muon for three datasets: (i) reference data in black, (ii) data after track cleaning in dark green and (iii) data after full cleaning in bright green. As expected, this distribution is shifted by two for all cleaned datasets.

electrons are not affected by the cleaning step, so a closer look at the electrons has to be made and will be presented in the following.
For electron identification, different approaches can be used. A cut-based approach with cuts on several discriminating variables has been chosen in this study. It should be kept in mind, that there are quantities among these discriminating variables, that are influenced by the cleaning: the impact parameters of the electron track with respect to the first primary vertex, denoted as Δr and Δz for the component in the transverse detector plane and the component in the direction of the beam axis respectively and the ratio between the hadronic and the electromagnetic energy deposited in a ΔR-cone around the electron flight direction, $E_{\mathrm{h}}/E_{\mathrm{em}}$. First the influence on the impact parameters will be discussed. Then, the influcence on the energy ratio. The relevant requirements on these variables used in this study are summarized in Table 3.2.

variable	barrel region: $\lvert\eta\rvert <= 1.479$	endcap region: $1.479 < \lvert\eta\rvert < 2.5$
$E_{\mathrm{h}}/E_{\mathrm{em}}$	0.104	0.0897
Δr	0.0261	0.118
Δz	0.41	0.822

Table 3.2.: The requirements on the variables used in the electron identification, which are influenced by the cleaning procedure.

During the vertex reconstruction, the primary vertices are ordered by the total amount of energy of the tracks used in the fit for each vertex. By removing the footprint of the muons and in this way, the main part of the hard process, a big amount of the contribution to the energy assigned to the first primary vertex is removed. As a consequence, the ranking of the vertices may change and a vertex originating from pile-up in the reference data may

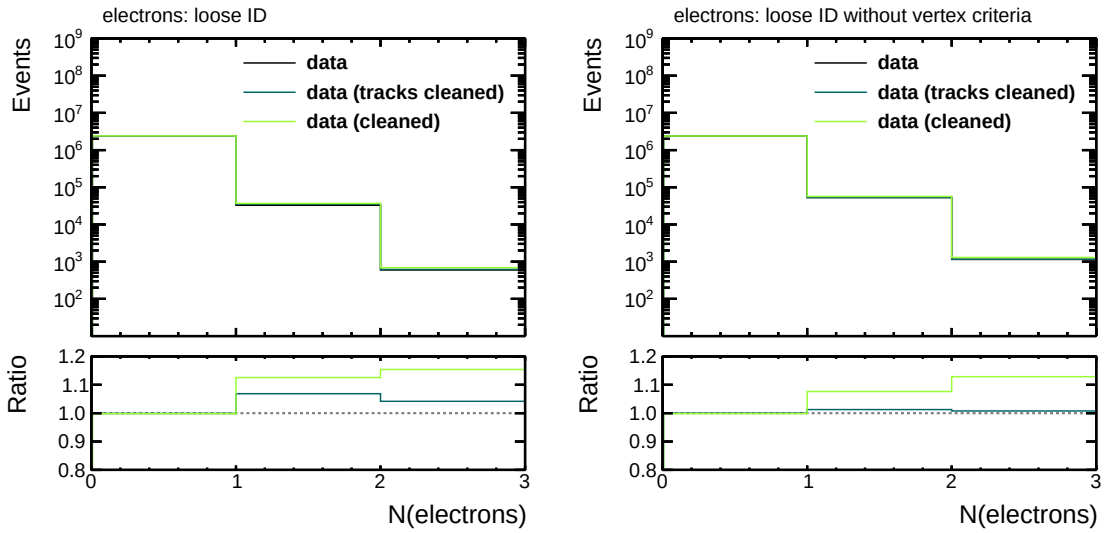

Figure 3.12.: Comparison of the number of electrons for three datasets: (i) reference data in black, (ii) data after track cleaning in dark green and (iii) data after full cleaning in bright green. For the left plot, all loose identification requirements are applied on the electrons, for the right plot, the vertex dependent criteria of these requirements are omitted. Events with a number of electrons higher or equal to 3 are not shown in the plots, since they are populated with less than 100 events. The bias common to track cleaning and full cleaning is reduced after the vertex dependent criteria are released as can be seen in the ratio subplots. This effect is caused by the changed ranking of vertices after the cleaning is applied as explained in the text.

become the first primary vertex in the cleaned data. This influences directly the impact parameters Δr and Δz of all electron tracks. Thus, a pile-up electron in the reference data may become an electron associated to the newly determined first primary vertex as shown in Figure 3.13.
Due to this effect introduced by the cleaning procedure a difference between the reference data and data after the cleaning step is expected, when the identification requirements are applied. This difference would decrease, if the vertex dependent requirements on the electrons, Δr and Δz, are released as confirmed in Figure 3.12. The effect on the vertex ranking is expected to disappear, when the simulated part of the hard process, $Z \to \mu\mu$ or $Z \to \tau\tau$, is inserted back into the cleaned data.
Another effect caused by the cleaning procedure is shown in Figure 3.14. In upper left plot of Figure 3.14, the area surrounding a selected muon in the (η, ϕ)-plane is shown. The muon marked as a red cross is also reconstructed as a jet indicated by a purple circle. The muon 3-momentum is surrounded by electromagnetic and hadronic energy deposits, shown as red and blue towers respectively. The largest electromagnetic energy deposit with the hadronic energy deposit on top is assigned to the muon.
In the right plot of Figure 3.14 the same area is shown after only the track of the muon has been removed. As a consequence, the remaining energy deposit is reconstructed as an electron, which may happen, if there are reconstructed tracks close to the removed muon track. In this particular case, the ratio $E_{\mathrm{h}}/E_{\mathrm{em}}$ is too high to let this electron pass

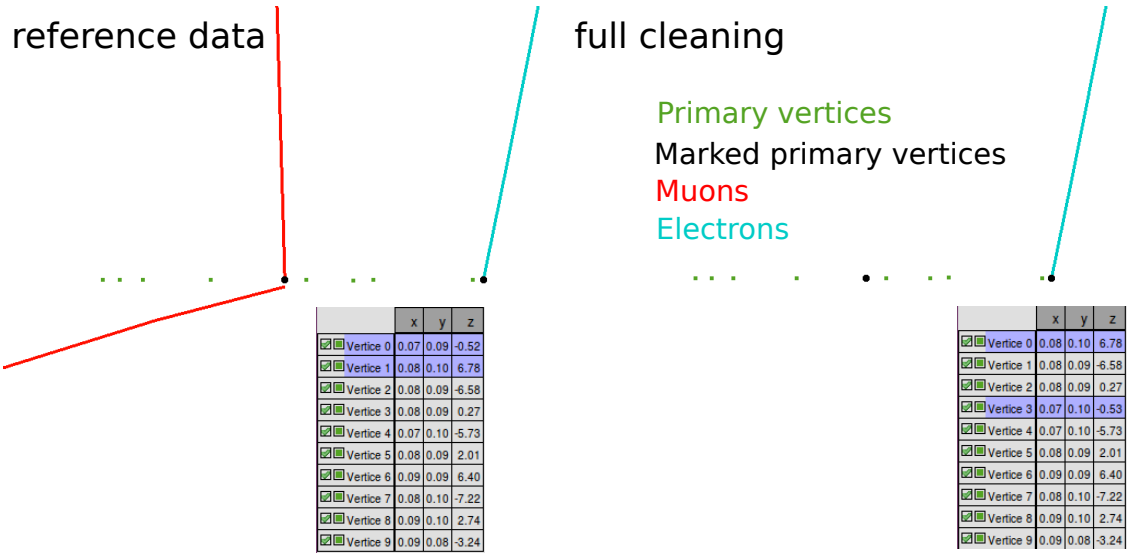

Figure 3.13.: The change of the vertex ranking due to the cleaning step is illustrated. In the left plot, an event from the reference data is shown, in the right plot the corresponding event in the cleaned data. The primary vertices are shown as green dots, the muon and electron tracks as red and blue lines respectively. The the two primary vertices of interest belonging to the muon and the electron tracks respectively are marked in black. After the two muons have been removed from data, the order of the marked vertices has changed as can be seen in the vertex listing in the lower right corner of each plot, where the marked vertices are highlighted. The event displays are produced with the Fireworks software [32].

the identification requirements. In the lower left plot of Figure 3.14 the reconstruction result is shown, after the full cleaning is applied on the muon. The energy tower formely assigned to the muon and now reconstructed as an electron has changed: The hadronic energy has disappeared completely and the electromagnetic energy deposit has become smaller. As a consequence, the ratio $E_{\mathrm{h}}/E_{\mathrm{em}}$ has decreased and the same electron now has passed the identification requirements.
To sum up, the observed effect results from the cleaning of the calorimeter deposits assigned to the muon. It affects the ratio of the electromagnetic energy to the hadronic energy in the area around the muon. Since this value is used for electron identification, an already existing electron near the muon can be invalidated by the cleaning procedure or a new electron candidate can be created, which may pass the identification.
Depending on the level, where the merging of the simulated Z-boson decay and the cleaned dataset happens, the effects introduced by the cleaning, for example the surplus electrons and neutral hadrons, may remain. This is to be checked during the validation in chapter 4 and is qualitatively discussed in the next section.

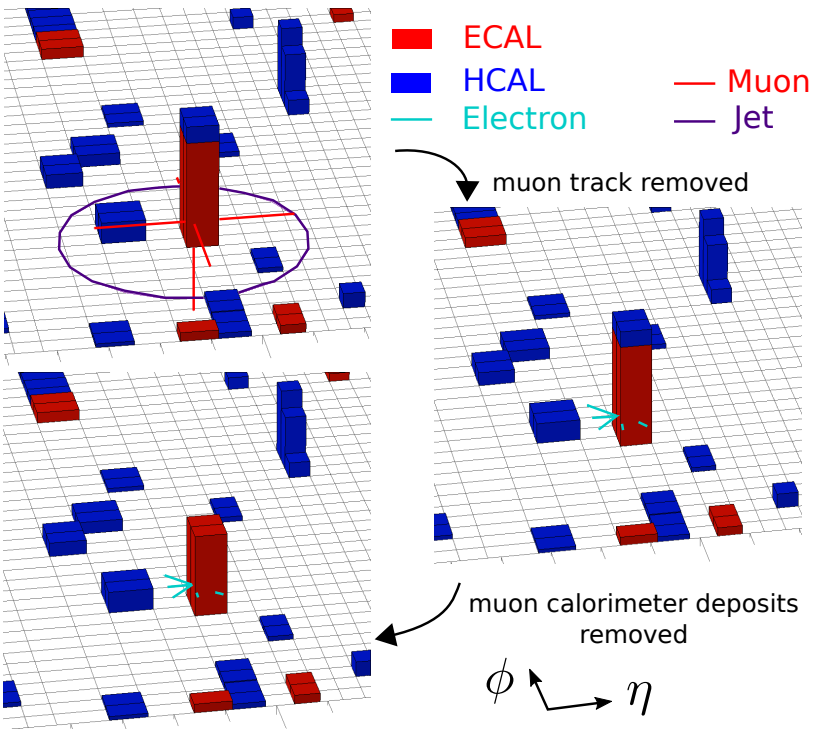

Figure 3.14.: The area around the selected muon in the (η, ϕ)-plane is shown. The plots correspond to the different stages of the embedding procedure: the selection of the muon in the upper left plot, the track cleaning in the right plot and the calorimeter cleaning in the lower left plot. The muon marked as a red cross is also reconstructed as a jet, shown as a purple circle. The reconstructed electron is shown as a light blue cross. The electromagnetic (ECAL) and hadronic (HCAL) energy deposits are shown as red and blue towers respectively. The effects that appear after the cleaning procedure are explained in the text. The event displays are produced with the Fireworks software [32].

3.4. Merging

It is desired that the merging of the simulated Z-boson decay with the data after the cleaning step is performed as early as possible for the same reason as described in section 3.3: the procedure would be more stable against changes in the software. So the first attempt would be to do it at the same level as the cleaning procedure: the merging should happen for clusters in the pixel and strip trackers and for reconstructed hits in the calorimeters and the muon detector system. The difficulties, that must be faced and solved on that level will be described in the following.
It must be taken into account that the detector geometry, alignment and calibration for simulation and data is different, influencing already the reconstructed clusters and hits. To save resources, this information isn't stored on an event-by-event basis, but for periods of time denoted as Intervals of Validity discussed in subsection 2.2.2. This means, to gather the geometry, alignment and calibration information for an event, the run, luminosity section and event numbers must be associated with the corresponding time Interval of Validity. In turn, all detector components don't have any information about their global position and the absolute amount of charge recorded for a signal. Every detector component has it's

own local coordinates and it's own units for measuring a signal.

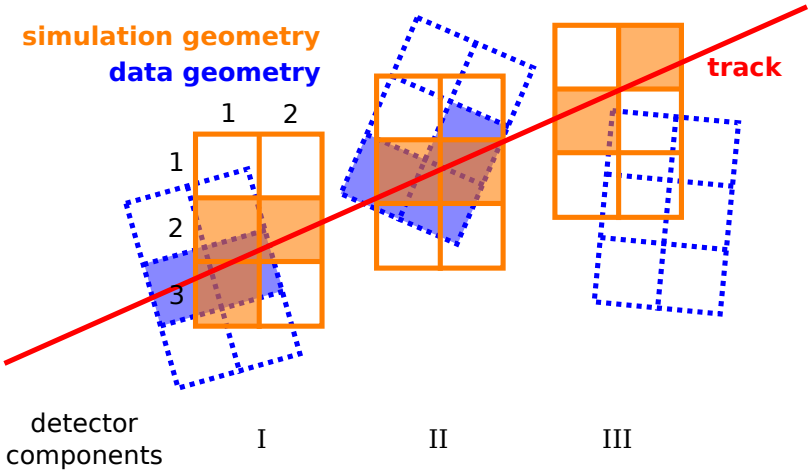

Figure 3.15.: Difference in geometry between simulation and data. The three detector components I, II and III are rotated and displaced against each other between data and simulation shown in blue and orange respectively. Each detector component has six segments in total placed in three rows and two columns. A segment crossed by the red track is filled with the corresponding color to indicate, that it has recorded a signal.

Consequences from the translation from local into global coordinates are sketched in Figure 3.15. Three detector components for the track measurement are considered. Each component in the data is displaced and rotated against the corresponding component in the simulation, causing an overal difference of the detector geometry between data and simulation, shown in blue and orange respectively in Figure 3.15. Every detector component has a local coordinate system, in this case with three rows and two columns for each component. This means, each detector component has six segments, where a signal can be recorded in case of a charged particle crossing. In Figure 3.15, the segments crossed by a charged particle are filled with the corresponding color. An example for a track of a charged particle crossing the detector is shown in red.
The records of the segments with signal are summarized in Table 3.3 as points (row, column). The numbering begins at the top left segment. The interpretation of these records can only be made under the proper geometry assumption: either data or simulation geometry. Since the merged event contains clusters and hits from both data and simulation, subsequent reconstruction steps, which can be only performed either under data or simulation geometry assumption, will fail. For the outer detector systems like the muon detector and the calorimeters the geometry difference between data and simulation is small compared to the size of the detector systems and therefore not problematic. But for the inner tracking system, it has a significant effect, such that in the end, a track wouldn't be reconstructed properly.
Another difference is in the calibration of the detectors. For each segment of a detector component, where a signal is recorded, a certain amount of charged particles is captured. This number is dependent on how the corresponding detector component is calibrated at a certain point of time. The difference in this calibration can influence for example the position of the charge barycenter of a cluster composed from several detector segments with

detector component	simulation	data
I	(2,1), (2,2), (3,1)	(2,1), (2,2)
II	(2,1), (2,2)	(2,2), (3,1), (3,2)
III	(1,2), (2,1)	

Table 3.3.: The records in the detector components in Figure 3.15 for simulation and data.

a signal. In Figure 3.15, the segments (2,1), (2,2) and (3,1) of the first detector component may form a cluster for the simulation.
Possible solutions would be to use the geometry information from data already at detector simulation or to correct for the difference between data and simulation at the level of signals in detector segments. Currently, this is not possible.
All this leads to the decision, that the merging should take place, after the objects dependent on geometry, alignment and calibration have already been reconstructed. In particular, it is performed on the level of particle collections: muons, electrons, photons and particle flow candidates. Additional merging of tracks must be performed, since they are referenced by the particle objects.
Furthermore, the primary vertices must be refitted, because their position in the simulated and cleaned datasets is reconstructed by using the tracks only from the corresponding dataset. Then, further high level reconstruction can be run to compute quantities dependent on the complete event. Since the merging step happens at a late stage of reconstruction, it is expected that features introduced by the cleaning step mentioned in section 3.3 remain.
Results of the performance validation of the embedding method are discussed in the next chapter.

4. Validation of Embedding

This chapter is devoted to the validation of the embedding method by using $\mu \rightarrow \mu$-embedding, within which the selected $Z \rightarrow \mu\mu$ decay is replaced by the simulated $Z \rightarrow \mu\mu$ decay. The naive expectation is that the $\mu \rightarrow \mu$-embedded event is the same as the corresponding selected $Z \rightarrow \mu\mu$ event in data. As already mentioned within chapter 3, differences are expected because of the finite precision of the reconstruction algorithms, which must be applied during the embedding and because of the way the embedding is implemented. Relevant effects are checked and if necessary uncertainties are quantified: the performance of the vertex refitting is checked and compared to the vertex resolution for Run II [34]. Additionally, a physics object validation is performed similarly to the one in section 3.3. Furthermore, the kinematics of the embedded muons and the effect of the mirroring introduced in subsection 3.1.1 on the isolation variables are checked.

4.1. Vertex refitting

After the vertex of the simulated Z-boson decay is set to the position of the reconstructed vertex in the selected $Z \rightarrow \mu\mu$ event as described in section 3.2 and after the merging of the simulated Z-boson decay and the event from data after the cleaning step as described in section 3.4, a vertex refitting is performed, where the merged track collections are used as input for the fitting algorithm. As will be discussed in section 4.4, the embedding with mirroring will be chosen to be the default method. Therefore, the vertex comparision is done between the selected $Z \rightarrow \mu\mu$ data and the $\mu \rightarrow \mu$-embedding with mirroring.
Because of the finite precision of the vertex reconstruction, which leads to a finite vertex resolution, it is expected that the refitted position of the first primary vertex in the $\mu \rightarrow \mu$-embedded event can be different from the vertex position in the origina event from data. Additional differences could appear, because the tracks of the two simulated muons are slightly different from the corresponding tracks of the selected muons in data. As long as the net difference is within the boundaries given by the vertex resolution, all such effects can be accepted.
The differences Δx, Δy and Δz between the position coordinates of the first primary vertices of the hard interaction process in the $\mu \rightarrow \mu$-embedded dataset and the selected $Z \rightarrow \mu\mu$ data are computed eventwise. Corresponding histograms are shown in Figure 4.1 and fitted within the range of ± 0.01 cm with a Gaussian. All values beyond this range are neglected for the fit and filled into the overflow bins of the histograms. The overflow bins cover only 0.6% of all considered events. From the fit the following parameters can be obtained:

$$\sigma_x = 9.67\ \mu\text{m} \qquad \sigma_y = 9.63\ \mu\text{m} \qquad \sigma_z = 14.70\ \mu\text{m} \tag{4.1}$$

These values are smaller than the given vertex resolution values for 2016 data [34] shown in Figure B.1. Thus in the case of the vertex reconstruction, all effects that may arise due

to embedding can be neglected.

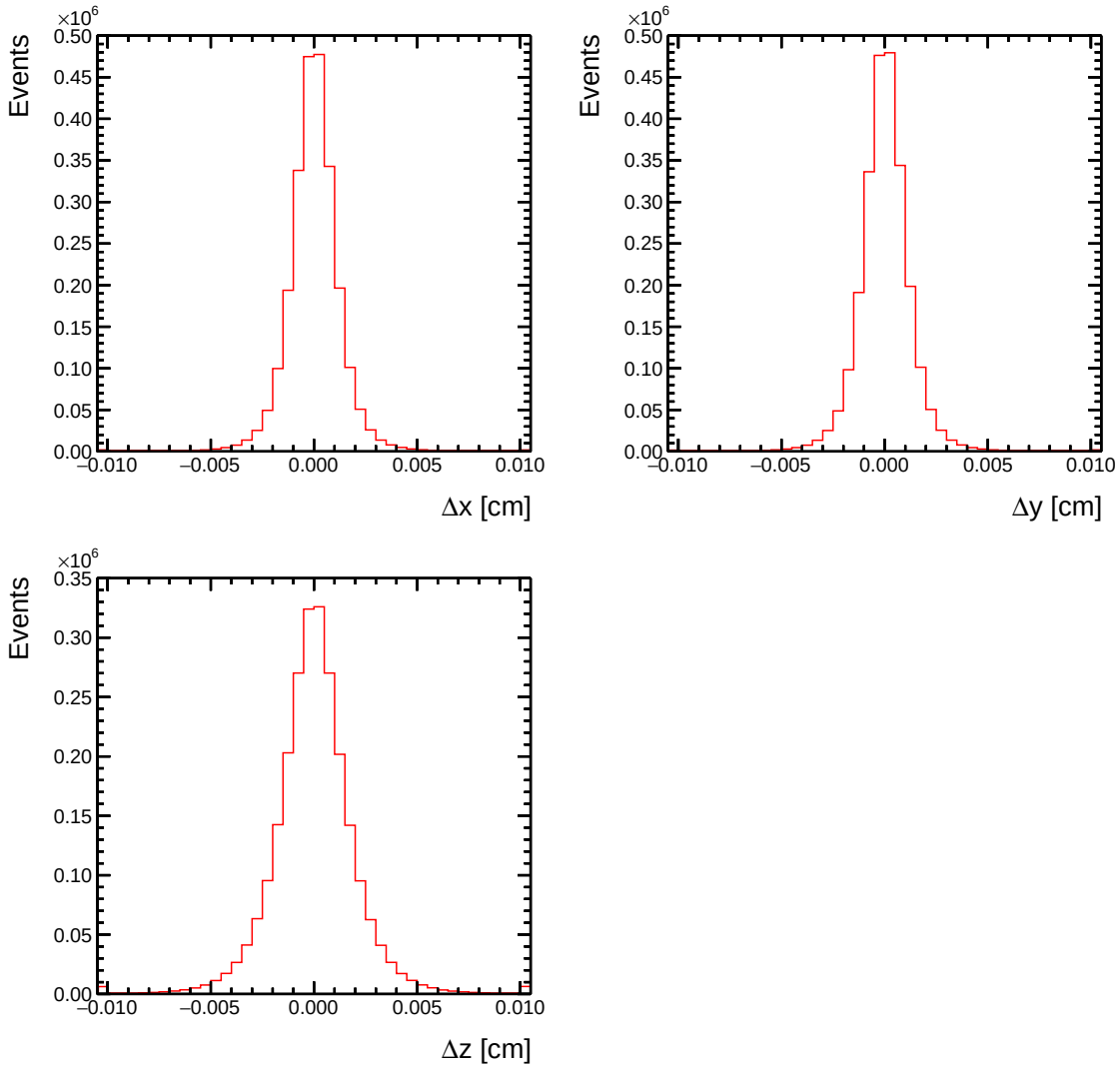

Figure 4.1.: Performance of vertex refitting in $\mu \rightarrow \mu$-embedding shown for the differences Δx, Δy and Δz of the vertex coordinates in the selected $Z \rightarrow \mu\mu$ data and the mirrored $\mu \rightarrow \mu$-embedded dataset. In the range of ±0.01 cm a Gaussian is fitted to the histograms. Bins beyond this range are overflow bins. The parameters of the fit in Equation 4.1 are smaller than the resolution values for 2016 data [34].

So far, only $\mu \rightarrow \mu$-embedding was considered. Additional changes of the vertex position may appear by performing $\mu \rightarrow \tau$-embedding, because the tracks of the decay products of the τ-leptons will look different from the tracks of the selected muons. But this is a wanted feature, since these effects would also appear in real $Z \rightarrow \tau\tau$ events. Therefore, discussing vertex refitting within $\mu \rightarrow \mu$-embedding only is sufficient.

4.2. Physics object validation

The topic of this section is to validate, whether the event apart from the embedded muons is affected by the method. Similarly to the study done in section 3.3 numbers of muons, electrons and particle flow candidates are compared between the reference data with selected $Z \to \mu\mu$ events and the $\mu \to \mu$-embedded datasets. Additionally, data after the cleaning step is added for comparison where appropriate.

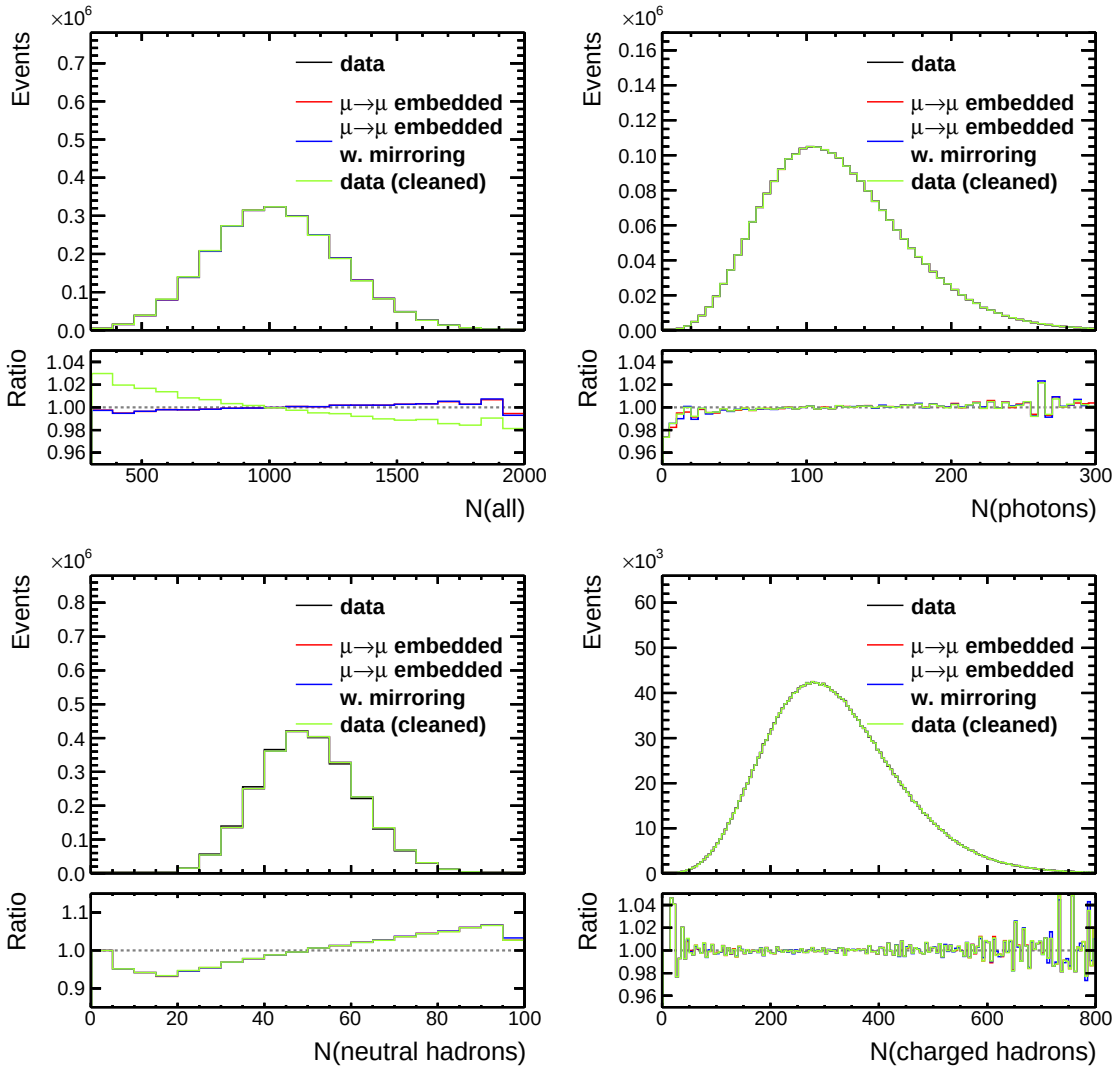

Figure 4.2.: Comparison of the number of particle flow candidates for four datasets: (i) reference data in black, $\mu \to \mu$-embedded datasets (ii) without mirroring in red and (iii) with mirroring in blue and (iv) data after the cleaning step in bright green. The upper left plot shows the number of all particle flow candidates, the upper right, lower left and lower right plots show the number of photons, neutral and charged hadrons respectively. The features introduced by the cleaning remain, no new effects are introduced by the merging.

First, the number of particle flow candidates is discussed, shown in Figure 4.2. As expected in section 3.4 all features of the cleaning remain after the merging is performed, as can be

seen in the upper right and the lower plots in Figure 4.2. The same applies for the number of loose electrons and the number of global muons matched to a particle flow muon in Figure 4.3. Thus, a small trend to higher values is visible for the number of all particle flow candidates comparing the reference data and the embedded dataset in the upper left plot of Figure 4.2. For the muons shown in the right plot of Figure 4.3 a migration of events into the first and second bin is visible. The reason is, that due to the finite precision of the muon reconstruction, muons initialy passing kinematic thresholds can fail these thresholds after the reconstruction is applied repeatedly during the embedding procedure.

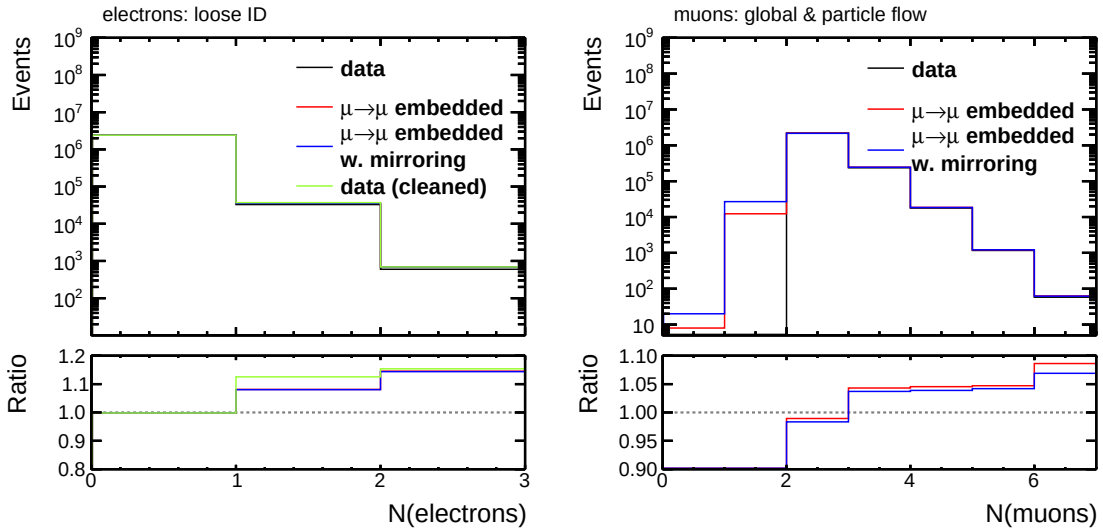

Figure 4.3.: Comparison of the number of electrons and the number of muons for three datasets respectively: (i) reference data in black, $\mu \rightarrow \mu$-embedded datasets (ii) without mirroring in red and (iii) with mirroring in blue. For the plot on the left the data after the cleaning step is also shown in bright green. Electrons shown in the left plot are chosen to pass loose identification requirements, muons shown in the right plot are required to be global and matched to a particle flow muon. The features introduced by the cleaning remain, no new effects are introduced by the merging. Due to the finite precision of the muon reconstruction the first and the second bin of the number of muons are populated.

This leads to the conclusion, that the merging step works as expected without any new systematic effects introduced to the collections of particles.
The next question to be answered is whether the remaining differences influence quanitities sensitive to all particles reconstructed for an event. One class of such quantities is the scalar and vectorial sum of transverse momenta of all particles, denoted as H_T and E_T respectively. This means that E_T contains the information about the direction of the energy deposit in the plane transverse to the beam axis and H_T doesn't have this information. The sum E_T is directly related to the missing transverse energy $\not{E}_T$. Distributions of these quantities are shown in Figure 4.4 for the reference data and the embedded datasets after the selection described in section 3.1 applied. All quantities are described well by the embedding within a wide range. As expected, the distribution of $\not{E}_T$ is similar to the distribution of E_T and

therefore only the latter quantity is discussed further. The distribution of E_T is slightly underestimated by the embedding for values below 10 GeV and overestimated in the tail region above 50 GeV. This points to a slight shift in the distribution to higher values.

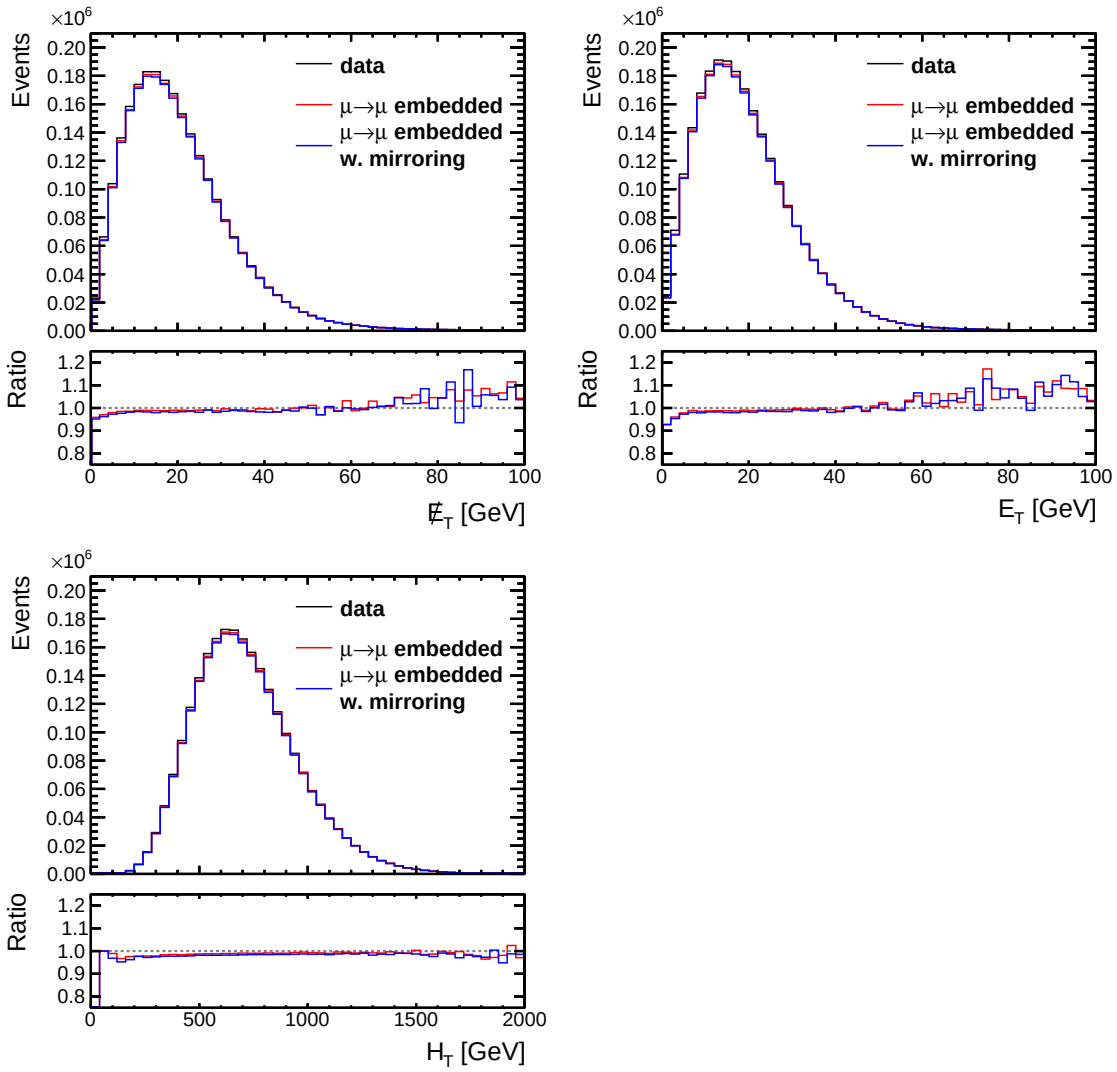

Figure 4.4.: Distributions of $\not{E}_T$, E_T and H_T for three datasets: (i) reference data in black and $\mu \rightarrow \mu$-embedded datasets (ii) without mirroring in red and (iii) with mirroring in blue. All three quantities are described well in a wide range. A slight shift to higher values is visible for $\not{E}_T$ and E_T.

In order o check whether this is an artefact resulting from the additional reconstruction of the two selected muons for embedding, both sums E_T and H_T are computed without the contribution of these muons as shown in Figure 4.5. The shift in the distribution of E_T has disappeared. This means, the reason for the shift is the finite precision of the muon reconstruction, which smears the p_T distribution of muons as discussed in section 4.3. This change is then propagated to E_T and H_T.
To sum up, differences between the reference data and the embedded dataset concerning the number of reconstructed particles are small, resulting mainly from the cleaning step. Individual features have been investigated in section 3.3. Quantities dependent on the

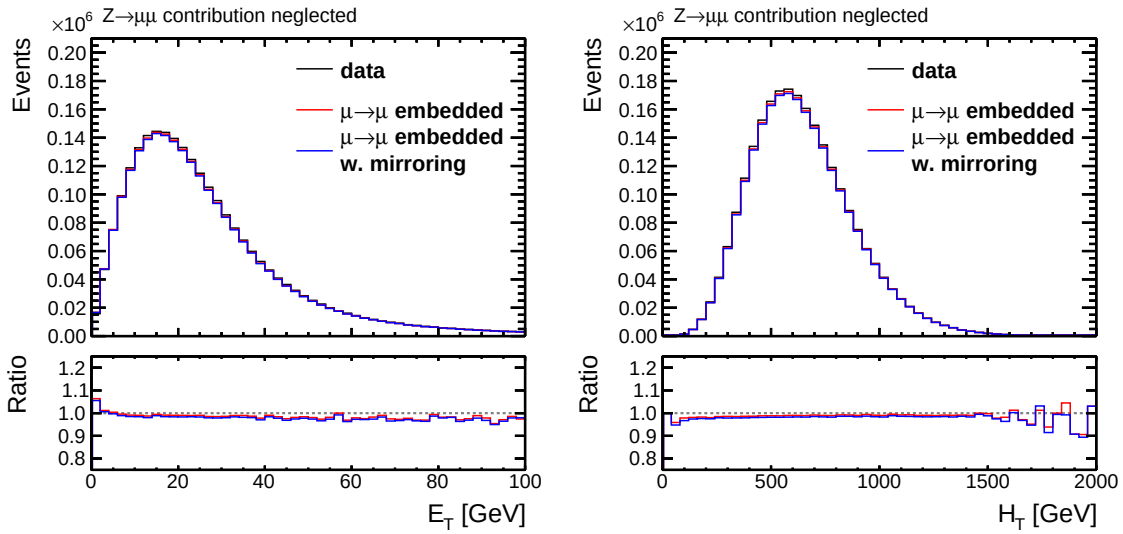

Figure 4.5.: Distribution of E_T and H_T without the contribution of the muons used for embedding for three datasets: (i) reference data in black and $\mu \rightarrow \mu$-embedded datasets (ii) without mirroring in red and (iii) with mirroring in blue. The shift for E_T as visible in Figure 4.4 has disappeared.

complete event are also well described. It can be concluded, that the part of the event without the inserted $Z \rightarrow \mu\mu$ decay is influenced only slightly by the embedding method.

4.3. Muon kinematics

Next a closer look is taken at the selected muons in the data and their simulated counterpart in the $\mu \rightarrow \mu$-embedded dataset. Within this context the performance of mirroring is tested. A general selection described in section 3.1 is applied to study the resulting effects on the absolute distributions of the muons.

dataset	events passing the selection	fraction of reference
reference data	2450930	100.00%
$\mu \rightarrow \mu$-embedding	2424583	98.93%
$\mu \rightarrow \mu$-embedding with mirroring	2408535	98.27%

Table 4.1.: The number of events passing the selection for three datasets: (i) reference data and $\mu \rightarrow \mu$-embedded datasets (ii) without mirroring and (iii) with mirroring.

The yields after applying the selection on the selected reference data and the $\mu \rightarrow \mu$-embedded datasets with and without mirroring are summarized in Table 4.1. The relative event loss in the embedded datasets of about 2% is investigated in the following.
In the upper plots and the lower left plot of Figure 4.6 the distributions of η, ϕ and p_T of the leading muon building the Z-boson candidate are shown. Similar distributions can be seen for the trailing muon in Figure B.3. In the lower right plot the distribution of the invariant mass of the Z-boson candidate is shown.
Regarding the distribution of η in the upper left plot of Figure 4.6 a general loss of events

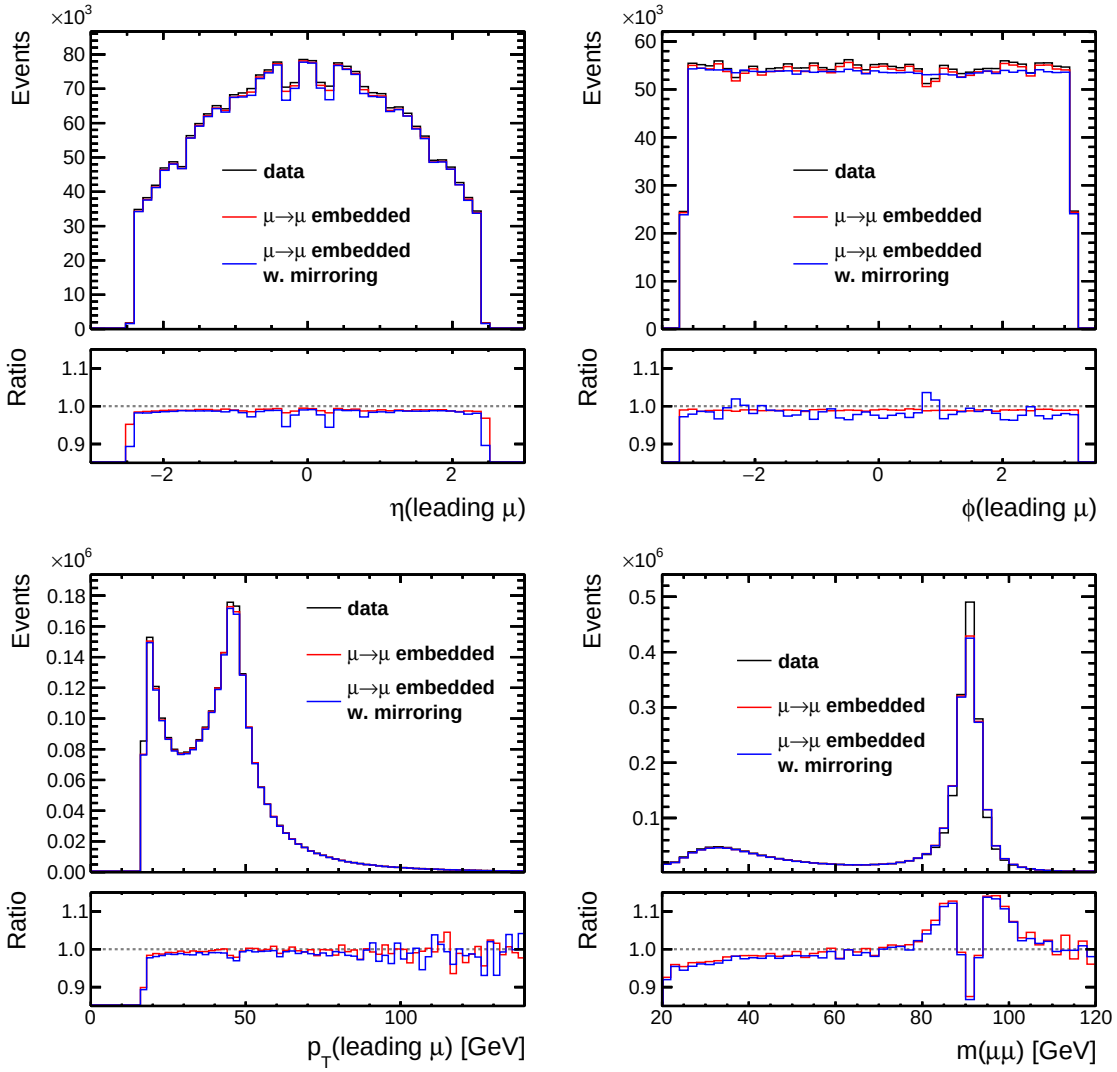

Figure 4.6.: Kinematics of the leading muon and the invariant mass of the Z-boson candidate.

can be seen by comparing the reference data with the embedded datasets as shown in the ratio. This can be explained by the finite precision of the muon reconstruction: Muons passing the selection criteria in data can have a slightly different momentum in the case of the embedded datasets. Thus, they may fail kinematic selection criteria in the case that their momentum components were originally near the selection thresholds. For the $\mu \rightarrow \mu$-embedded dataset with mirroring additional loss of events is seen for certain η bins: $|\eta| \approx 2.5$ and $|\eta| \approx 0.3$. The mirroring changes only the ϕ component of the muon momentum. Therefore, a possible explanation would be that by performing the mirroring, the momenta of the muons are mapped into regions of the (η, ϕ)-plane of the detector, which have a worse reconstruction efficiency. The dips at $|\eta| \approx 0.3$ seen already in the distribution for the reference data point to that explanation. An example is the geometry dependent trigger efficiency of the muon detector system as shown in Figure B.2.
For the ϕ distribution of the leading muon in the upper right plot of Figure 4.6 an effect due to the mirroring can be observed: For the $\mu \rightarrow \mu$-embedding without mirroring, the

same fluctuations as in the reference data are seen, whereas for $\mu \rightarrow \mu$-embedding with mirroring the distribution is flattened, since ϕ is changed eventwise by the mirroring.
For the p_T distribution of the leading muon in the lower left plot of Figure 4.6 an effect is introduced by $\mu \rightarrow \mu$-embedding due to the finite precision of the muon reconstruction: a smearing of all peak structures in the distribution, in this case, the two peaks at ≈ 20 and ≈ 45 GeV can be seen for the embedded datasets. Since the lower peak is near the selection threshold and no migration of events from below this threshold to values above is apparent, a more pronounced event loss is seen at this threshold.
The smearing of the momenta of the leading and the trailing muon directly affects the 4-momentum of the Z-boson candidate and in particular its distribution of the invariant mass shown in the lower right plot of Figure 4.6. First of all, a very pronounced smearing is seen for the peak at ≈ 90 GeV around the mass value of Z-boson resonance. This sizable effect reaching differences in the ratio above 10% is dedicated to the $\mu \rightarrow \mu$-embedding. In the case of $\mu \rightarrow \tau$-embedding, this effect would be covered by the resolution. The second effect can be seen for mass values below 50 GeV. Muons contributing to these mass values have mostly p_T values located close to the lower p_T peak. Therefore, the event loss due to the p_T selection threshold appears and is propagated to the invariant mass distribution.
Reaching the end of this section it can be concluded, that most of the differences regarding the muon kinematics appear due to the finite precision of the muon reconstruction, which is repeatedly performed during the embedding procedure. All resulting effects can be accepted, since the invariant mass resolution of the di-τ-system is expected to be worse and the discussed smearing effects would be covered.

4.4. Muon isolation sums

The aim of this section is to ensure, that the area around the selected muons, which are eventually replaced by τ-leptons in the application of the embedding procedure is not affected by the selection of the muons itself. Therefore, distributions of the p_T-flow are discussed for different particles. These distributions are a measure for the energy surrounding the muon. The amount of this energy is used to compute the isolation of the particle and quantities used for its identification. Again, the general selection described in section 3.1 is applied.
Distributions shown in Figure 4.7 are received as follows: for each event the leading muon building the Z-boson candidate is taken. Next, for each particle flow candidate within a distance of $\Delta R = 0.4$ in the (η,ϕ)-plane to the muon its transverse momentum p_T is filled into a corresponding ΔR bin. Since the charged hadrons have a track, they can be split into a contribution from the first primary vertex of the hard interaction (PV) and a contribution from pile-up (PU). In order to obtain the average p_T-flow for a single leading muon, the yields in the bins of the p_T-flow histograms are devided by the number of events in Table 4.1 for each dataset.
In the upper plots of Figure 4.7, histograms for the charged hadrons are shown, the pile-up contribution in the left and first primary vertex contribution in the right histogram. The ower left plot shows the contribution of neutral hadrons, the lower right plot the contribution of the photons.
An unbiased environment around the muon would imply that the p_T-flow is linearly rising

with ΔR because the linear ΔR-dependence of the ring areas in the (η,ϕ)-plane corresponding to a certain ΔR bin. An additional p_T-flow dataset is computed from the reference data by taking a random direction instead of a muon. The direction is chosen from the cylidrical coordinates θ and ϕ, which are distributed uniformly in the intervals $[0, \pi]$ and $[-\pi, \pi]$ respectively. The resulting p_T-flow would correspond to an unbiased environment if the averaged energy of the collisions would be distributed equaly over the spacial directions.

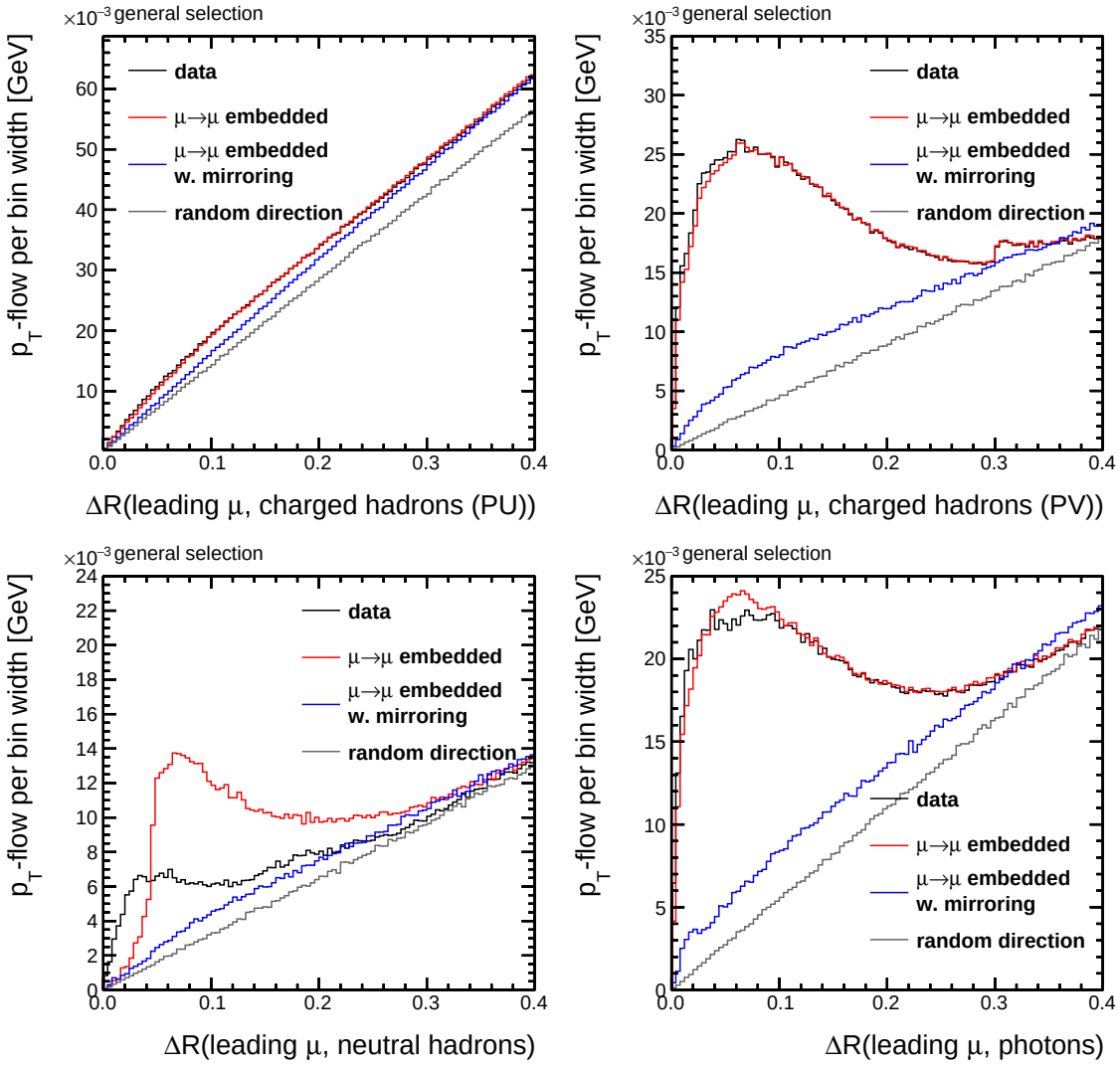

Figure 4.7.: The p_T-flow as a function of ΔR for different particle flow candidates computed for the leading muon. The comparison is done for four datasets: (i) reference data in black, $\mu \rightarrow \mu$-embedded datasets (ii) without mirroring in red and (iii) with mirroring in blue and (iv) a dataset with a random direction chosen instead of the direction of the leading muon in grey.

From the plots in Figure 4.7 it can be concluded, that the muons are placed by the mirroring into a nearly unbiased environment because of its linearity, whereas the reference data and the embedding without mirroring show an excess of the particle flow candidates surrounding the leading muon. For the reference data, this excess results from the reconstruction and the selection of the muons caousing a selection bias. This bias is reproduced by the

embedding without mirroring only for the charged hadrons compared to the reference data in the upper plots of Figure 4.7. The spike at $\Delta R \approx 0.3$ in the upper right plot of Figure 4.7 is a result of the track isolation applied within the used muon HLT. This affects the charged hadrons coming from the same hard interaction vertex as the selected muons. As expected in section 3.3 and section 3.4, the differences between the reference data and the embedding without mirroring appear for photons in the lower left plot of Figure 4.7 and for neutral hadrons in the lower right plot of Figure 4.7. This supports the need for a study to tune the performance of the calorimeter cleaning. In all plots of Figure 4.7 the expected linear dependence of the p_T-flow on ΔR for a randomly chosen direction is clearly visible and the p_T-flow of the mirroring is nearest to it. This leads to the conclusion, that an unbiased environment for a $Z \rightarrow \mu\mu$ decay can be approximately described by a flat distribution of the energy over the spacial directions.
In order to separate the selection bias from the effects introduced by the embedding procedure, the contribution to the p_T-flow is split by regions defined for the invariant mass distribition $m(\mu\mu)$. For the peak region $m(\mu\mu)$ is defined to be between 80 and 100 GeV, otherwise it belongs to the sideband region. In that way, the peak region contains mostly pure $Z \rightarrow \mu\mu$ events that will always fulfill the imposed $Z \rightarrow \mu\mu$ selection, whereas the sideband region is more contaminated by muons originating from other processes.
As expected, the selection bias is reduced for the peak region and originates mostly from the sideband region shown in Figure B.4. The p_T-flow from the peak region shown in Figure 4.8 is discussed in the following for charged hadrons from pile-up, charged hadrons from the first primary vertex, neutral hadrons and photons respectively. The p_T-flow of the charged hadrons from pile-up is unbiased for all four datasets as shown in the upper left plot of Figure 4.8. In the upper right plot of Figure 4.8 surplus charged hadrons from the first primary vertex are visible for the reference data and the $\mu \rightarrow \mu$-embedded dataset without mirroring. Since this is not apparent if the mirroring is applied, it must be an effect specific to the muons. In the lower left plot of Figure 4.8 there is a significant difference in the p_T-flow between the reference data and the $\mu \rightarrow \mu$-embedded dataset without mirroring. This points to the expected inefficiency in the embedding method, where too few energy deposits are cleaned in the hadronic calorimeter causing the observed excess for the neutral hadrons. For the reference data there are less neutral hadrons compared to the unbiased environment of the mirroring method. This feature must be again specific to the muon selection but th question of its origin remains open. The p_T-flow for the photons shown in the lower right plot of Figure 4.8 has a common excess at ΔR values between 0 and 0.04. One possible explanation could be a shortcoming in the determination of the energy deposited by the muon in the electromagnetic calorimeter. The broader excess for ΔR values between 0 and 0.15 visible only for the reference data and $\mu \rightarrow \mu$-embedded dataset without mirroring might be caused by FSR, since it is reduced by the mirroring. Another possible solution would be to selection muons without radiated photons.
Summing up this chapter, it can be concluded that the embedding does not affect global quantities like the missing transverse energy. The kinematics of the muons in the $\mu \rightarrow \mu$-embedded datasets obtain only slight changes due to the repeatedly performed reconstruction during the embedding step. Local effects in the area around the muons selected for the embedding are visible that are caused by the selection of the muons or by the cleaning step. To reduce these effects, the mirroring is tested and is found to map the embedded muons into a nearly unbiased environment. Further studies of the muon isolation can be performed to study the origins of the observed differences by using $Z \rightarrow \mu\mu$ simulation as

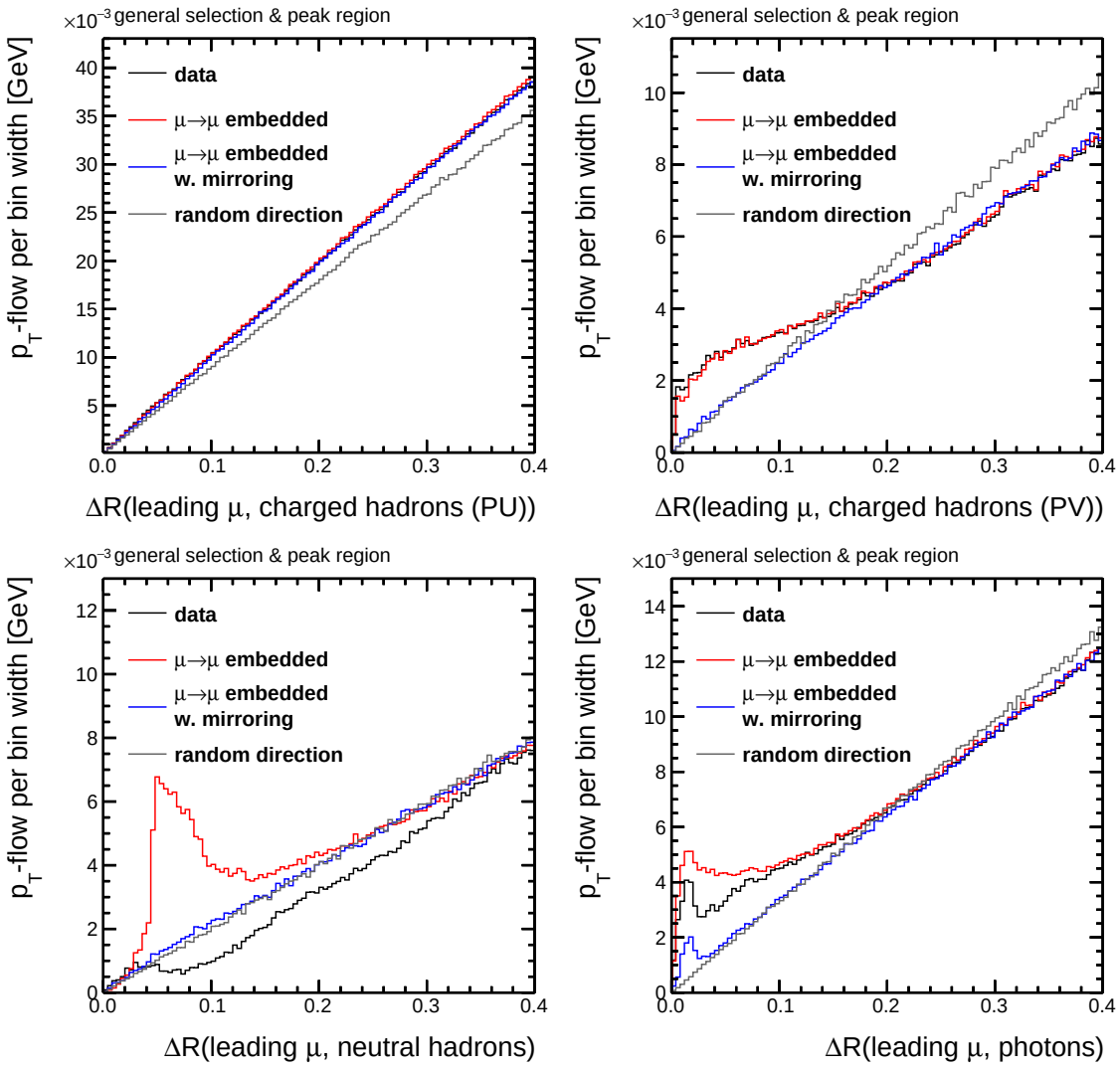

Figure 4.8.: The p_T-flow as a function of ΔR for different particle flow candidates computed for the leading muon selected from the peak region of the invariant mass $m(\mu\mu)$. The comparison is done for four datasets: (i) reference data in black, $\mu \rightarrow \mu$-embedded datasets (ii) without mirroring in red and (iii) with mirroring in blue and (iv) a dataset with a random direction chosen instead of the direction of the leading muon in grey.

input.

5. Application of Embedding

Within this section the application of the $\mu \to \tau$-embedding will be discussed, for which the selected $Z \to \mu\mu$ decay is replaced by the simulated $Z \to \tau\tau$ decay. The method will be tested in all four selection channels used for the $H \to \tau\tau$ searches explained in section 1.2: $e\mu$, $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$. First, the acceptance requirements on the simulation level compatible with the selection of the $H \to \tau\tau$ analysis will be described for each selection channel and the resulting efficiencies will be calculated. Afterwards, the migrations of the simulated decay channels into reconstructed selection channels due to a misidentification of τ-lepton decays will be quantified. In the last section, a comparison between the $Z \to \tau\tau$ simulation and the $\mu \to \tau$-embedded dataset will be discussed.

5.1. Acceptance efficiencies

decay channel	kinematic cuts on visible decay products			
$e\mu$	$p_T(e) > 16$ GeV		$p_T(\mu) > 8$ GeV	
	or			
	$p_T(e) > 11$ GeV		$p_T(\mu) > 16$ GeV	
$e\tau_h$	$p_T(e) > 23$ GeV	$\lvert\eta(e)\rvert < 2.2$	$p_T(\tau_h) > 18$ GeV	$\lvert\eta(\tau_h)\rvert < 2.4$
$\mu\tau_h$	$p_T(\mu) > 18$ GeV	$\lvert\eta(\mu)\rvert < 2.2$	$p_T(\tau_h) > 18$ GeV	$\lvert\eta(\tau_h)\rvert < 2.4$
$\tau_h\tau_h$	$p_T(\tau_h) > 38$ GeV	$\lvert\eta(\tau_h)\rvert < 2.2$		

Table 5.1.: Kinematic acceptance requirements for relevant decay channels in the $H \to \tau\tau$ search on simulation level. The selection is performed on the transverse momentum, p_T, and the absolute value of the pseudorapidity, $|\eta|$, of the visible decay products. In the case of a leptonic decay of a τ-lepton, the visible decay product is an electron or a muon. For a hadronic decay of a τ-lepton, τ_h corresponds to the sum over the 4-momenta of all detectable particles, in particular all particles simulated in the decay except the neutrinos. Each line corresponds to a selection required to be passed. If more than one line is available for a channel, then only one of these selections must be satisfied.

The acceptance requirements mentioned in section 3.2 and compatible with the selection of the $H \to \tau\tau$ analysis are summarized in Table 5.1 for each selection channel. These requirements are chosen to be looser than the requirements used on the reconstrution level to account for differences between 4-momenta in simulation and in reconstruction. The acceptance efficiency ϵ is computed on an event-by-event basis. Each decay channel is obtained separately by imposing only the corresponding selection in Table 5.1 on the

inclusively simulated τ-lepton decays. In that way, branching ratios are accounted for the efficiencies. The number of simulation attempts N is chosen to be 1000. After the number of passed events P is counted, the efficiency $\epsilon = P/N$ is calculated and the last attempt passing is taken as the simulated Z-boson decay for the subsequent embedding steps.

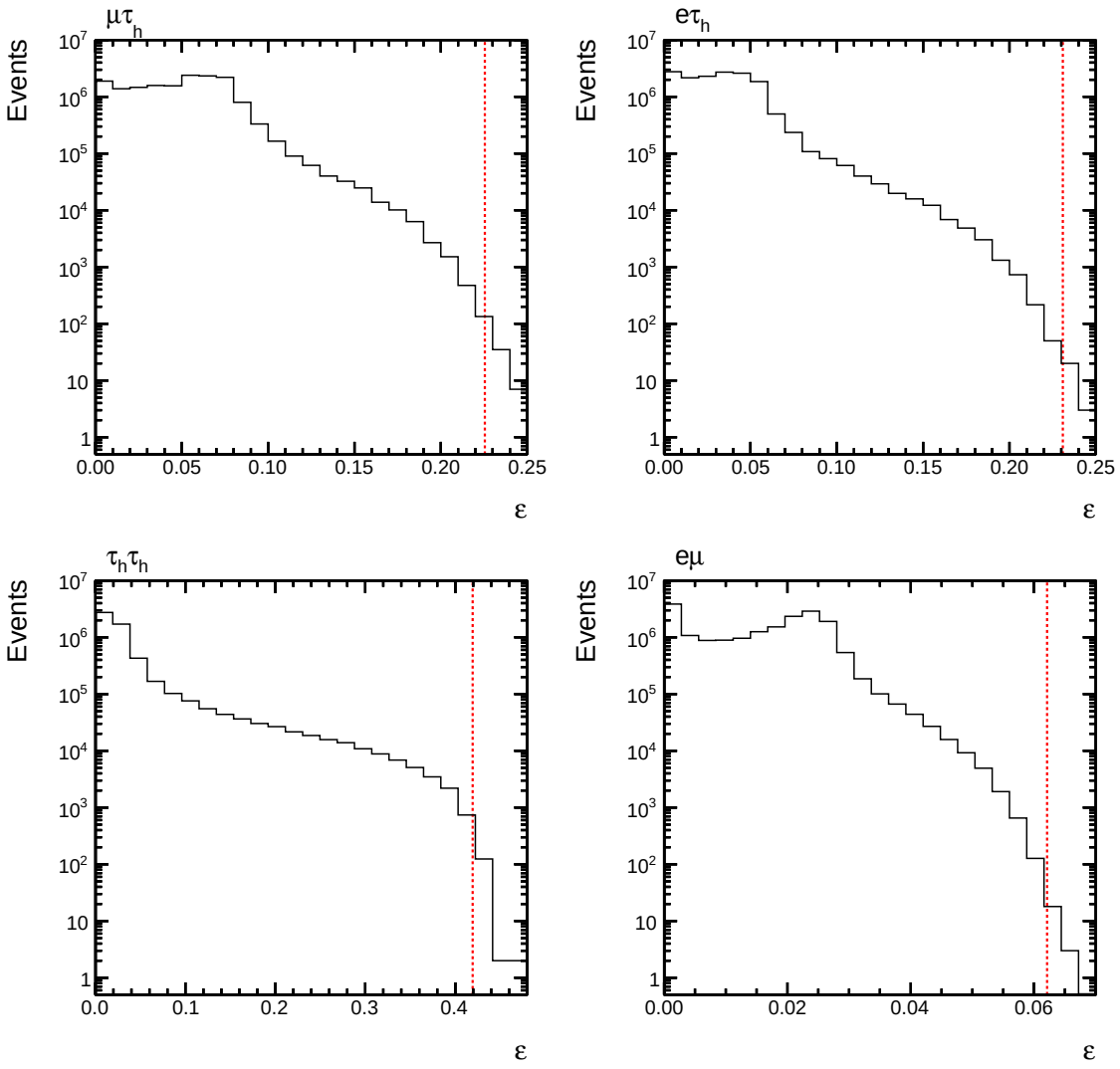

Figure 5.1.: Distributions of the efficiency ϵ for the decay channels $\mu\tau_h$ in the upper left plot, $e\tau_h$ in the upper right plot, $\tau_h\tau_h$ in the lower left plot and $e\mu$ in the lower right plot. The values for the branching ratios of each decay channel are shown in form of vertical red dashed lines.

In Figure 5.1 the distribution of the efficiency ϵ is shown for all four decay channels. For each decay channel the value of the efficiency varies within its full range from 0 to a value corresponding approximately to the branching ratio of the decay channel given in Table 1.2. The highly populated region with small ϵ values with respect to the branching ratios looks differently for each decay channel and comprises several millions of events. For the decay channels $\mu\tau_h$ and $e\tau_h$ shown in the upper plots of Figure 5.1 there is a plateau for the low values of ϵ from zero up to approximately 0.07. Comparing the decay channel $\tau_h\tau_h$ shown in the lower left plot of Figure 5.1 with $\mu\tau_h$ and $e\tau_h$, the peak at the low ϵ values

is narrow. For the $e\mu$ decay channel in the lower right plot of Figure 5.1 a double peak structure is visible with peaks at $\epsilon \approx 0$ and $\epsilon \approx 0.03$. Beyond the highly polulated region the number of events decreases approaching values near 100 at ϵ equal to the branching ratio for each channel.

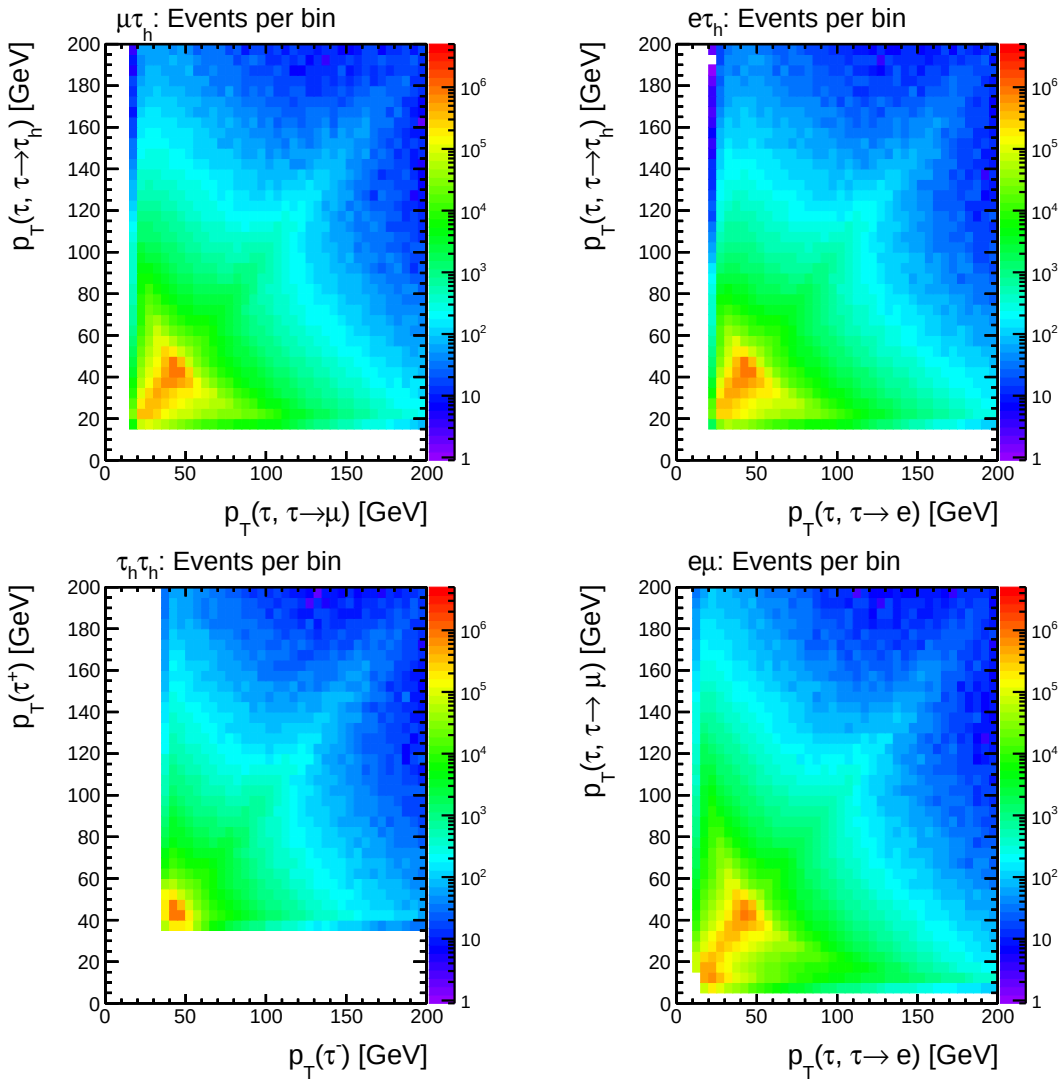

Figure 5.2.: The distribution of the events passing the requirements in Table 5.1 as a function of the transverse momenta for the decay channels $\mu\tau_h$, $e\tau_h$, $\tau_h\tau_h$ and $e\mu$ in the upper left, upper right, lower left and lower right plot respectively. For $\mu\tau_h$, $e\tau_h$ and $e\mu$, the horizontal axis is used for the τ-lepton decaying into the first visible decay product, the vertical axis for the τ-lepton decaying into the second visible decay product. In the case of $\tau_h\tau_h$, the horizontal axis is used for negatively charged τ-lepton, the vertical axis is used for the positively charged τ-lepton.

In order to study the features specific for each decay channel, the distribution of the events passing the requirements in Table 5.1 as a function of the transverse momenta of the simulated τ-leptons and the efficiency ϵ as a function of these momenta are shown in Figure 5.2 and Figure 5.3 respectively for all considered decay channels.

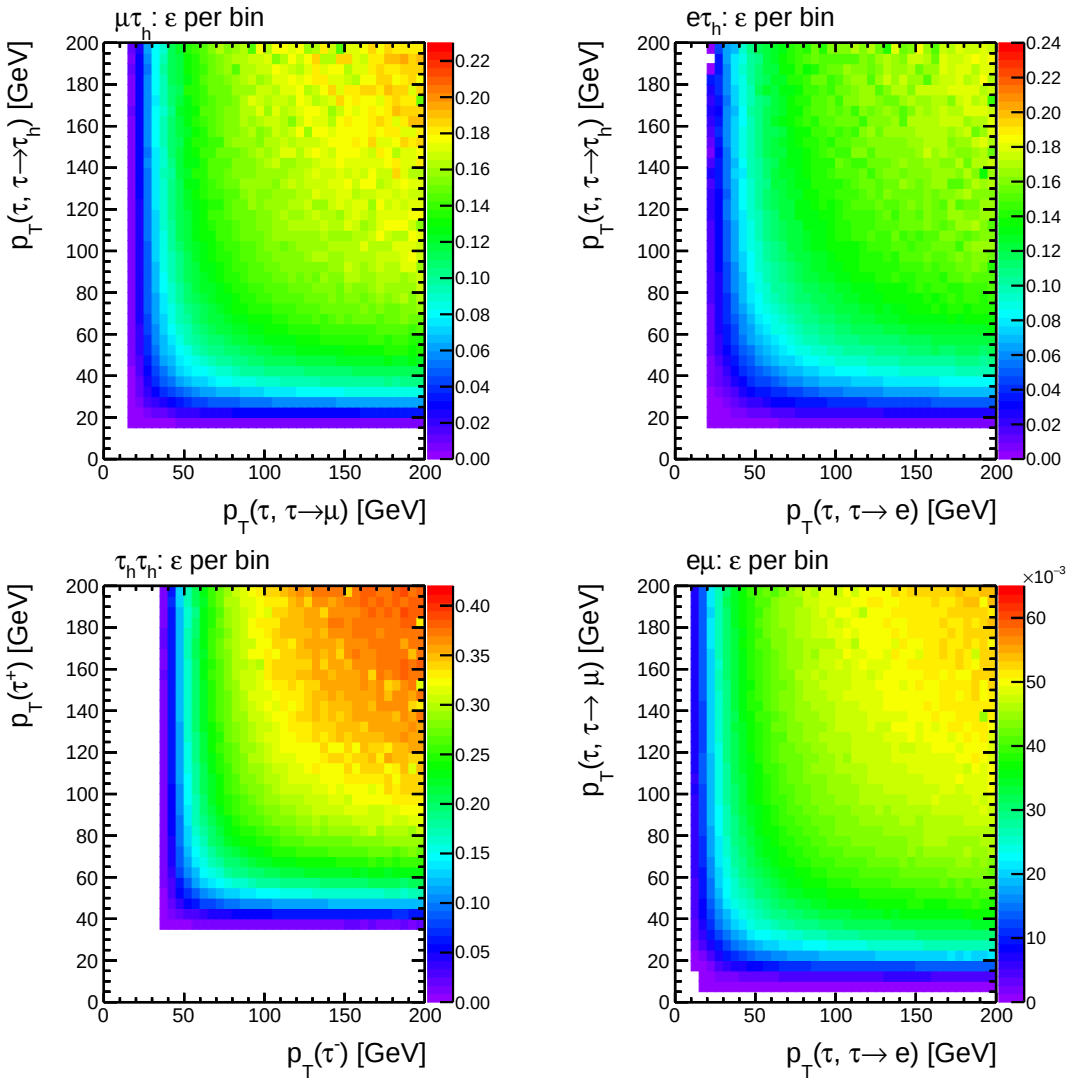

Figure 5.3.: The efficiency ϵ as a function of the transverse momenta of the simulated τ-leptons for the decay channels $\mu\tau_h$, $e\tau_h$, $\tau_h\tau_h$ and $e\mu$ in the upper left, upper right, lower left and lower right plot respectively. The upper bound on ϵ is chosen to be near the branching ratio of each decay channel. For $\mu\tau_h$, $e\tau_h$ and $e\mu$, the horizontal axis is used for the τ-lepton decaying into the first visible decay product, the vertical axis for the τ-lepton decaying into the second visible decay product. In the case of $\tau_h\tau_h$, the horizontal axis is used for the negatively charged τ-lepton, the vertical axis is used for the positively charged τ-lepton.

The distributions of the events in Figure 5.2 look similar for all four decay channels except for the white areas. The white area of the $\tau_h\tau_h$ decay channel is the biggest because of the high p_T thresholds on the transverse momenta of the visible decay products which have a smaller transverse momentum than the corresponding simulated τ-leptons. The majority of the events populates the region with p_T values roughly between 10 and 60 GeV for both simulated τ-leptons. Most of these events correspond to a Z-boson resonance with an invariant mass of 90 GeV. Furthermore, the events are mostly located at the diagonal line

from the lower left to the upper right corner reflecting the fact, that the Z-boson is mostly produced at the rest. The significant decrease of events for p_T values above 50 GeV on the diagonal correspond to an off-shell Z-boson having an invariant mass larger than 90 GeV. Using this information the peak structures which are specific to each decay channel and are shown in Figure 5.1 can be explained by Figure 5.3 where the efficiency ϵ is shown as a function of the transverse momenta of the simulated τ-leptons. For the decay channels $\mu\tau_h$ and $e\tau_h$ shown in the upper right and the upper left plot of Figure 5.3, the efficiency ϵ picks up the values between 0 to 0.08 leading to the plateaus seen in Figure 5.1. The narrow peak for the $\tau_h\tau_h$ decay channel results from the high p_T thresholds imposed for this channel, since only a small part of the resonance region is left for ϵ to pick up small values from 0 to 0.05 in the lower left plot of Figure 5.3. The peak at zero of the double peak structure that can seen for the decay channel $e\mu$ in the lower right plot in Figure 5.1 corresponds to the events for which the simulated τ-leptons have transverse momenta between 15 and 25 GeV.
Concluding this section, the general features of the efficiency ϵ are summarized in the following. Its distribution is dominated by the $Z \to \tau\tau$ events with a Z-boson produced on the mass shell. These events obtain mostly low efficiency values compared to the branching ratio of the corresponding decay channel, because the neutrinos involved in the τ-lepton decays carry a part of the energy of the Z-boson such that the visible decay products may fail to pass the thresholds in Table 5.1. The shape of the ϵ distribution specific to each decay channel depends on the thresholds imposed on the visible decay products of each corresponding decay channel. The events with τ-leptons with transverse momenta slightly above the imposed threshold allways receive low efficiency values since the phase space for the corresponding visible decay product is restricted. The events with off-shell Z-bosons with a high invariant mass are rare and achieve a high ϵ value approaching the branching ratio of the corresponding decay channel. These events build the tail of the efficiency distribution. The intermediate region of the efficiency distribution is filled by events with boosted Z-bosons.
All in all, the attempt to save the statistics which is explained in section 3.2 leads as expected to shape changing effects as can be concluded from the distributions of the efficiency ϵ. In order to correct for these effects, the calculated efficiencies have to be applied as weights on an event-by-event basis.

5.2. Misidentification of τ-lepton decays

In this section, the migration of the simulated decays of the di-τ-system into the reconstructed decay channels is studied. This can happen due to a misidentification of τ-lepton decays. For example, a τ-lepton decaying into an electron can be misidentified as a hadronically decaying τ-lepton. The $Z \to \tau\tau$ simulation is produced inclusively and the reconstruction is applied on the complete sample. As a consequence, a reconstructed decay channel $\tau\tau$ can be contaminated by other simulated decay channels. In contrast to that, the $\mu \to \tau$-embedding provides a pure sample corresponding to the chosen simulated decay channel. This is illustrated in Figure 5.4.
The results presented in Table 5.2 quantify the differences that may appear between the $Z \to \tau\tau$ simulation and the the $\mu \to \tau$-embedding after the full selection of the $H \to \tau\tau$ analysis discussed in section 1.2 is applied. The study is done for the pure $\mu \to \tau$-embedded

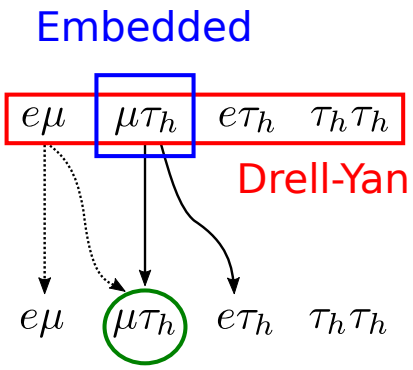

Figure 5.4.: The difference is illustrated that appears for the Drell-Yan simulation marked red, where all decay channels are simulated and the embedded dataset marked blue, where only the $\mu\tau_h$ decay channel is simulated. The migrations shown as solid arrows will happen for both datasets. In contrast to that, the migrations shown as dashed arrows only appear for the Drell-Yan simulation. As a consequence, the reconstructed $\mu\tau_h$ decay channel marked green is contaminated by the $e\mu$ decay channel only for the simulation and not for the embedded sample.

samples with each of the four decay channels corresponding to a row in Table 5.2. Each pure embedded sample is run through the reconstruction and the $H \to \tau\tau$ analysis providing the reconstructed decay channels corresponding to the columns in Table 5.2 as a result. The sum of the events weighted by the acceptance efficiencies discussed in section 5.1 which pass the selection of one of the four reconstructed decay channels is taken as the reference yield for each given simulated decay channel. The individual yields in the reconstructed decay channels are then devided by the reference yield. In that way the migration of a simulated decay channels into reconstructed decay channels can be quantified.

	$\mu\tau_h$	$e\tau_h$	$\tau_h\tau_h$	$e\mu$
$\mu\tau_h$	99.81%	0.00%	0.00%	0.18%
$e\tau_h$	0.00%	99.79%	0.05%	0.16%
$\tau_h\tau_h$	0.33%	1.02%	98.60%	0.04%
$e\mu$	1.68%	0.00%	0.07%	98.25%

Table 5.2.: Influence of the misidentification of the τ-lepton decays. The percentages of the events migrating from the simulated decay channel in the row to the reconstructed decay channel in the column are shown.

The largest migrations above 1% correspond to the events with a τ-lepton decaying into an electron misidentified as a hadronically decaying τ-lepton and vice versa. These effects are discussed in the measurements of the τ-lepton reconstruction performance [35] giving a handle on the reconstruction efficiencies and the misidentification rates.

5.3. Comparison between simulation and embedding

Within this section a comparison between the simulated $Z \to \tau\tau$ events and the $\mu \to \tau$-embedded events is presented for the visible mass of the di-τ-system, $m_{vis}(\tau\tau)$ in order to check, whether potential differences between the simulation and the embedding are covered by shape changing uncertainties. As an example for such an uncertainty the variation of the energy scale of the τ-leptons and electrons is chosen. In the case of the $e\mu$, the energy scale of the reconstructed electron is varied. For all other channels the energy scale of the visible decay products of the hadronically decaying τ-lepton, τ_h, is varied. The upward and downward variation of the energy scale is performed on the 4-momentum of the considered particle as follows:

$$p_\alpha^{\text{up/down}}(\tau_h) = (1 \pm 0.03) \cdot p_\alpha^{\text{nominal}}(\tau_h)$$
$$p_\alpha^{\text{up/down}}(e) = (1 \pm 0.02) \cdot p_\alpha^{\text{nominal}}(e)$$

In Figure 5.5 a shape comparison between the $\mu \to \tau$-embedded events with the energy scale variation, the simulated $Z \to \tau\tau$ events and the simulated $H \to \tau\tau$ events is shown after the full selection of the $H \to \tau\tau$ analysis has been applied. The distributions are normalized such that the integrated yield after the full selection is equal to the one for the $\mu \to \tau$-embedded dataset, the $Z \to \tau\tau$ simulation and the $H \to \tau\tau$ simulation. The energy scale variations of the $\mu \to \tau$-embedded dataset are normalized with the same scalefactor as the nominal $\mu \to \tau$-embedded dataset to preserve the relative differences between the nominal datase and its variations.
The distribution of the $H \to \tau\tau$ events hown in Figure 5.5 determines the region of interest for the visible mass of the di-τ-system. For the decay channels $\mu\tau_h$, $e\tau_h$ and $e\mu$ this range comprises the interval between 20 and 120 GeV. In the case of the $\tau_h\tau_h$ decay channel, the region of interest is between 30 and 170 GeV. Comparing the $\mu \to \tau$-embedded dataset and the $Z \to \tau\tau$ simulation, a slight shift of the $Z \to \tau\tau$ simulation to higher values similar to the one observed in Figure 3.8 is visible for all decay channels, most pronounced for the $e\mu$ decay channel. Furthermore, a trend for the event distributions above 80 GeV for the decay channels $\mu\tau_h$, $e\tau_h$ and $e\mu$ and above 110 GeV for the $\tau_h\tau_h$ decay channel is visible: there are less events available for the $Z \to \tau\tau$ simulation than for the $\mu \to \tau$-embedded dataset. This trend is pronounced for the $e\mu$ and $\tau_h\tau_h$ decay channels.
These differences between the $\mu \to \tau$-embedded dataset and the $Z \to \tau\tau$ simulation are covered by the energy scale variations in the peak region of the $H \to \tau\tau$ distribution, but exceed the variation bands in the tails of the $H \to \tau\tau$ distribution. The best agreement can be seen for the $\mu\tau_h$ decay channel in the upper left plot of Figure 5.5 and the $e\tau_h$ decay channel in the upper right plot of Figure 5.5. These channels are followed by the $\tau_h\tau_h$ decay channel in the lower left plot of Figure 5.5. The largest and most pronounced deviations can be seen for the $e\mu$ decay channel in the lower right plot of Figure 5.5.
It remains to be studied, whether these differences are caused solely by the effects from FSR discussed in subsection 3.2.1 or by additional effects caused by the reconstruction. In particular the slight shift discussed above could be explained by effects due to FSR.
All in all, the embedding method provides a result consistent with the one obtained from the $Z \to \tau\tau$ simulation in the region of interest defined by the $H \to \tau\tau$ distribution. To follow the improvements of the energy scale measurements and to have a better control of the high mass regions important for searches beyond the SM as well as for high precision

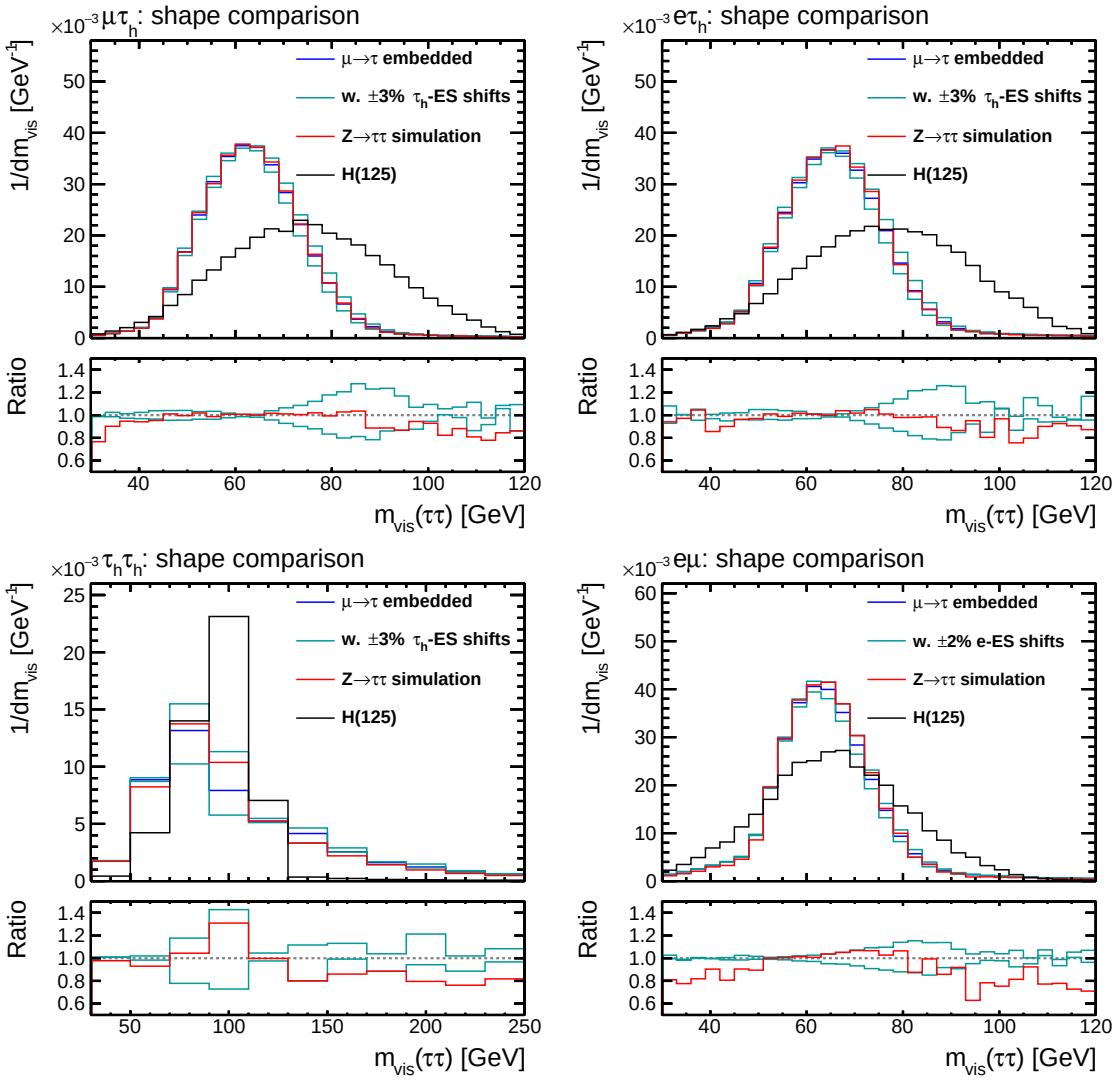

Figure 5.5.: The distribution of the visible mass of the di-τ-system, $m_{vis}(\tau\tau)$, for the decay channels $\mu\tau_h$, $e\tau_h$, $\tau_h\tau_h$ and $e\mu$ in the upper left, upper right, lower left and lower right plot respectively. A shape comparison for the following datasets is performed: (i) the $\mu \rightarrow \tau$-embedded dataset in blue with the upward and downward variations of the energy scale (ES) in light blue, (ii) $Z \rightarrow \tau\tau$ simulation in red and (iii) the $H \rightarrow \tau\tau$ simulation in black. In the case of the decay channel $e\mu$, the energy scale of the electron is varied by 2%. For all other decay channels, the energy scale of the visible decay product τ_h is varied by 3%.

measurements, improvements are needed to better understand and potentially reduce the discrepancies observed in the tails of the obtained $Z \rightarrow \tau\tau$ distributions.

6. Conclusions and Outlook

The embedding method is now in a shape to provide an important consistency check for the $H \to \tau\tau$ analysis as indicated in Figure 6.1. The comparison of the two methods for the estimation of the $Z \to \tau\tau$ background, the $\mu \to \tau$-embedding in the left plot and the $Z \to \tau\tau$ simulation in the right plot, allows to gain a better understanding of the $H \to \tau\tau$ analysis and provides the possibility to cross-check the results. The revision of the software is completed for the Run II data taking period. The well understood steps of the embedding method described in chapter 3 are implemented in a simplified and maintainable way that is aimed to become part of the standard reconstruction workflow in CMS. The validation and the application of the embedding method discussed in chapter 4 and chapter 5 were performed on the data recorded by the CMS detector in 2015 and 2016. The $\mu \to \tau$-embedded dataset provides results consistent with the $Z \to \tau\tau$ simulation and can be used for the standard model $H \to \tau\tau$ search.

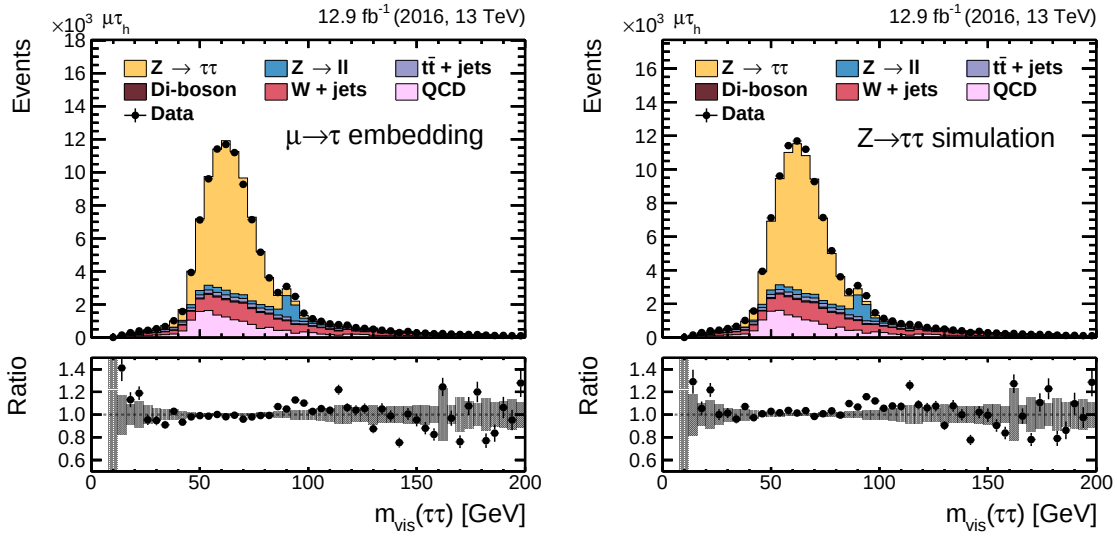

Figure 6.1.: The $H \to \tau\tau$ analysis chain is performed for the $\mu\tau_h$ decay channel with the $Z \to \tau\tau$ estimation by the embedding method in the left plot and the simulation in the right plot. The normalization of the embedded sample is scaled to the intergrated yield. The distribution of the visible di-τ-mass, $m_{vis}(\tau\tau)$, is compared.

Nevertheless, further improvements of the embedding method can be developed. As stated in chapter 4, the effects caused by the final state radation of the muons that isn't recovered during the muon reconstruction can be reduced by proper selection steps choosing muons without radiated phothons. Furthermore, there are shortcomings in the cleaning of the energy deposits in the calorimeters, in particular in the hadronic calorimeter, as discussed in section 3.3. A study could be done where the optimal amount of energy

is determined that should be removed from the energy clusters surrounding the cluster directly crossed by the muon track. figure of merit for this study could be provided by the isolation sums discussed in section 4.4. The merging step described in section 3.4 depends on the realization of the alignment and calibration of the CMS detector. A worthwhile improvement would be to use the information of the alignment and calibration measured in data within the detector simulation. Another approach to remove the difference in the detector geometry and calibration between data and simulation would be to correct the local information stored for each detector component. In order to distinguish the effects coming from a Z-boson selection both for a $Z \to \mu\mu$ and a $Z \to \tau\tau$ decay from the effects specific to the $Z \to \mu\mu$ decay, simulated $Z \to \mu\mu$ events can be used as input for the embedding procedure instead of data. This would help to tune the $Z \to \mu\mu$ selection used for embedding and to understand better the features discussed in section 4.4.
Further improvements for a proper understanding of the embedding method within the $H \to \tau\tau$ analysis include the calculation of the normalization constant of the $\mu \to \tau$-embedded dataset from the cross section of the selected $Z \to \mu\mu$ data. A proper quantification of the systematic uncertainties dedicated to the embedding method is also needed for the statistical inferrence.
Concluding this thesis, the embedding method is an interesting and relevant topic rich with possibilities to learn about physics and software and to gain insights into a modern particle physics experiment.

Bibliography

[1] Serguei Chatrchyan, Vardan Khachatryan, Albert M Sirunyan, et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B*, 716(arXiv:1207.7235. CMS-HIG-12-028. CERN-PH-EP-2012-220):30–61. 59 p, Jul 2012.

[2] Georges Aad, Brad Abbott, Jalal Abdallah, et al. Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV. *J. High Energy Phys.*, 08(arXiv:1606.02266. CERN-EP-2016-100. ATLAS-HIGG-2015-07. CMS-HIG-15-002):045. 70 p, Jun 2016.

[3] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.

[4] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.

[5] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. 1995.

[6] Roger Wolf. *The Higgs Boson Discovery at the Large Hadron Collider*. Springer International Publishing Switzerland, 2015.

[7] G. Bellini, L. Ludhova, G. Ranucci, and F. L. Villante. Neutrino oscillations. *Adv. High Energy Phys.*, 2014:191960, 2014.

[8] Wikipedia. Standard model. `https://en.wikipedia.org/wiki/Standard_Model`, 2016.

[9] Measurement of the mass of the z boson and the energy calibration of lep. *Physics Letters B*, 307(1):187 – 193, 1993.

[10] Victor Mukhamedovich Abazov et al. Measurement of the W Boson Mass with the D0 Detector. *Phys. Rev. Lett.*, 108:151804, 2012.

[11] T. Aaltonen et al. Precise measurement of the W-boson mass with the CDF II detector. *Phys. Rev. Lett.*, 108:151803, 2012.

[12] Serguei Chatrchyan et al. Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons. *JHEP*, 05:104, 2014.

[13] Marco Pieri. Lhc higgs cross section working group. `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG`, 2016.

[14] K. A. Olive et al. Review of Particle Physics. *Chin. Phys.*, C38:090001, 2014.

[15] Evidence for Higgs boson Yukawa couplings in the $H \to \tau\tau$ decay mode with the ATLAS detector. Technical Report ATLAS-CONF-2014-061, CERN, Geneva, Oct 2014.

[16] Georges Aad et al. Modelling $Z \to \tau\tau$ processes in ATLAS with τ-embedded $Z \to \mu\mu$ data. *JINST*, 10(09):P09018, 2015.

[17] Benjamin Treiber. Estimation of the background from z $\longrightarrow \tau\tau$ in h $\longrightarrow \tau\tau$ analyses. Ms, Karlsruher Institut für Technologie (KIT), 2015. Karlsruher Institut für Technologie (KIT), Masterarbeit, 2015.

[18] Eleanor Rusack. CMS collaboration. `http://cms.web.cern.ch/content/cms-collaboration`, 2015.

[19] G. L. Bayatian et al. CMS physics: Technical design report. 2006.

[20] S Chatrchyan, K Abadjiev, D Abbaneo, G Abbiendi, et al. Commissioning and Performance of the CMS Pixel Tracker with Cosmic Ray Muons. *J. Instrum.*, 5(arXiv:0911.5434. CMS-CFT-09-001):T03007. 37 p, Nov 2009.

[21] S. Chatrchyan et al. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[22] Salvatore Di Guida. CMS Alignment and Calibration Workflows: Lessons Learned and Future Plans. `https://indico.cern.ch/event/369822/contributions/873710/attachments/735540/1009090/IEEE2014-AlCa_v1.pdf`, 2015.

[23] Serguei Chatrchyan et al. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009, 2014.

[24] R. Fruhwirth. Application of Kalman filtering to track and vertex fitting. *Nucl. Instrum. Meth.*, A262:444–450, 1987.

[25] Serguei Chatrchyan et al. Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV. *JINST*, 7:P10002, 2012.

[26] Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. 2009.

[27] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Anti-k(t) jet clustering algorithm. *JHEP*, 04:063, 2008.

[28] Piet Verwilligen. Muons in the CMS high level trigger system. *Nuclear and Particle Physics Proceedings*, 273–275:2509 – 2511, 2016. 37th International Conference on High Energy Physics (ICHEP).

[29] Michał Szleper. The Higgs boson and the physics of WW scattering before and after Higgs discovery. 2014.

[30] Johan Alwall et al. A Standard format for Les Houches event files. *Comput. Phys. Commun.*, 176:300–304, 2007.

[31] Studies of Higgs boson production in the four-lepton final state at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-HIG-15-004, CERN, Geneva, 2016.

[32] D. Kovalskyi, M. Tadel, A. Mrak-Tadel, B. Bellenot, V. Kuznetsov, C. D. Jones, L. Bauerdick, M. Case, J. Mulmenstadt, and A. Yagil. Fireworks: A physics event display for CMS. *J. Phys. Conf. Ser.*, 219:032014, 2010.

[33] Electron and Photon performance using data collected by CMS at $\sqrt{s}$ = 13 TeV and 25ns. Dec 2015.

[34] Primary vertex resolution in 2016. Technical Report CMS-DP-2016-041, Jul 2016.

[35] Performance of reconstruction and identification of tau leptons in their decays to hadrons and tau neutrino in LHC Run-2. Technical Report CMS-PAS-TAU-16-002, CERN, Geneva, 2016.

[36] S Chatrchyan et al. Performance of the CMS Level-1 Trigger during Commissioning with Cosmic Ray Muons. *JINST*, 5:T03002, 2010.

A. Appendix for chapter 3

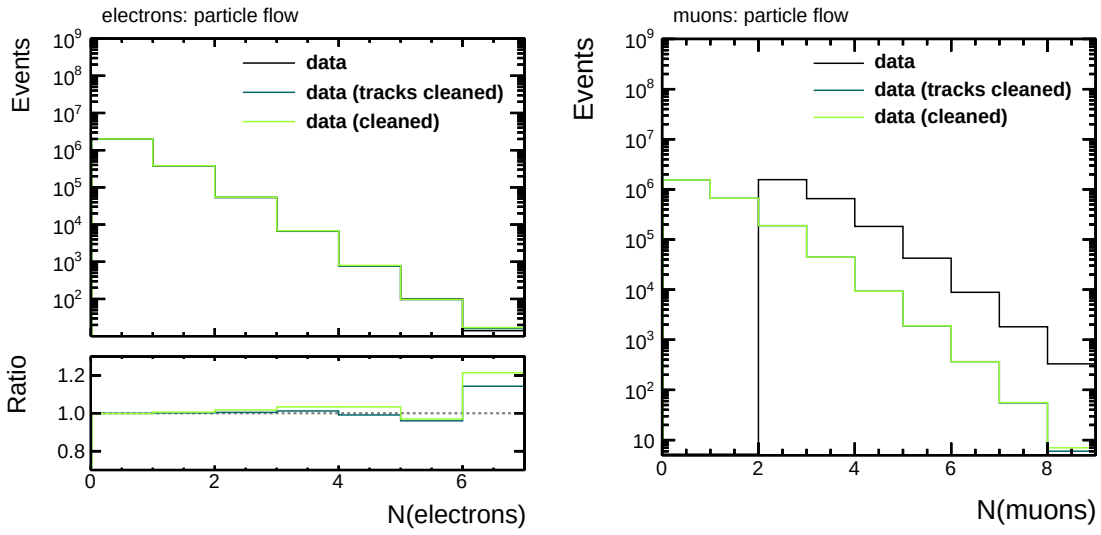

Figure A.1.: Comparison of the number of particle flow electrons in the left plot and particle flow muons in the right plot for three datasets: (i) reference data in black, (ii) data after track cleaning in dark green and (iii) data after sideband cleaning in bright green.

B. Appendix for chapter 4

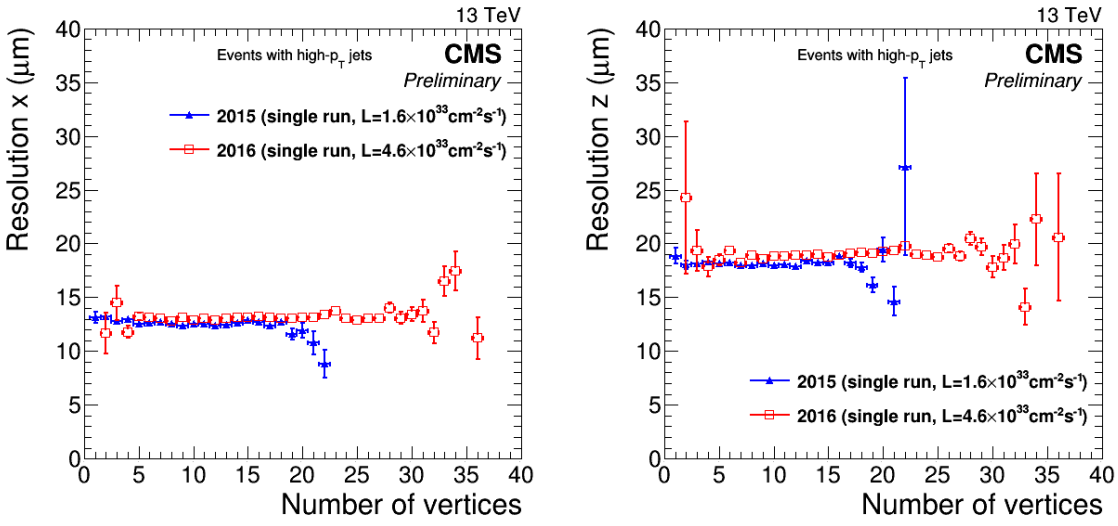

Figure B.1.: The vertex resolution for 2016 data as a function of the number of reconstructed primary vertices [34, p. 6].

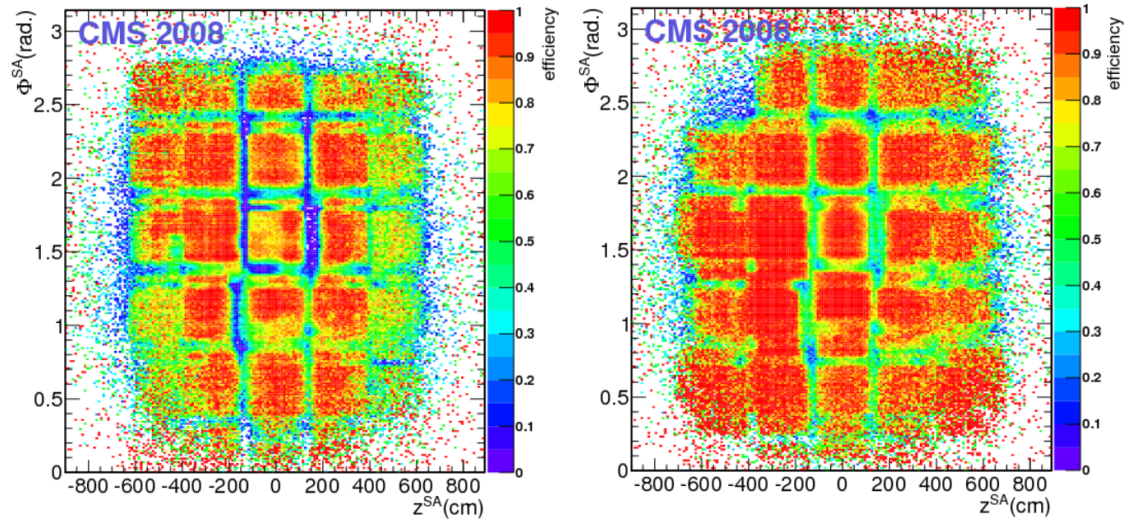

Figure B.2.: The trigger efficiency as a function of the (ϕ, z)-plane for the resistive plate chambers (RPC) and the drift tubes (DT) in the left plot and in the right plot respectively [36]. Low efficient regions corresponding to the gaps in the active detector material are clearly visible.

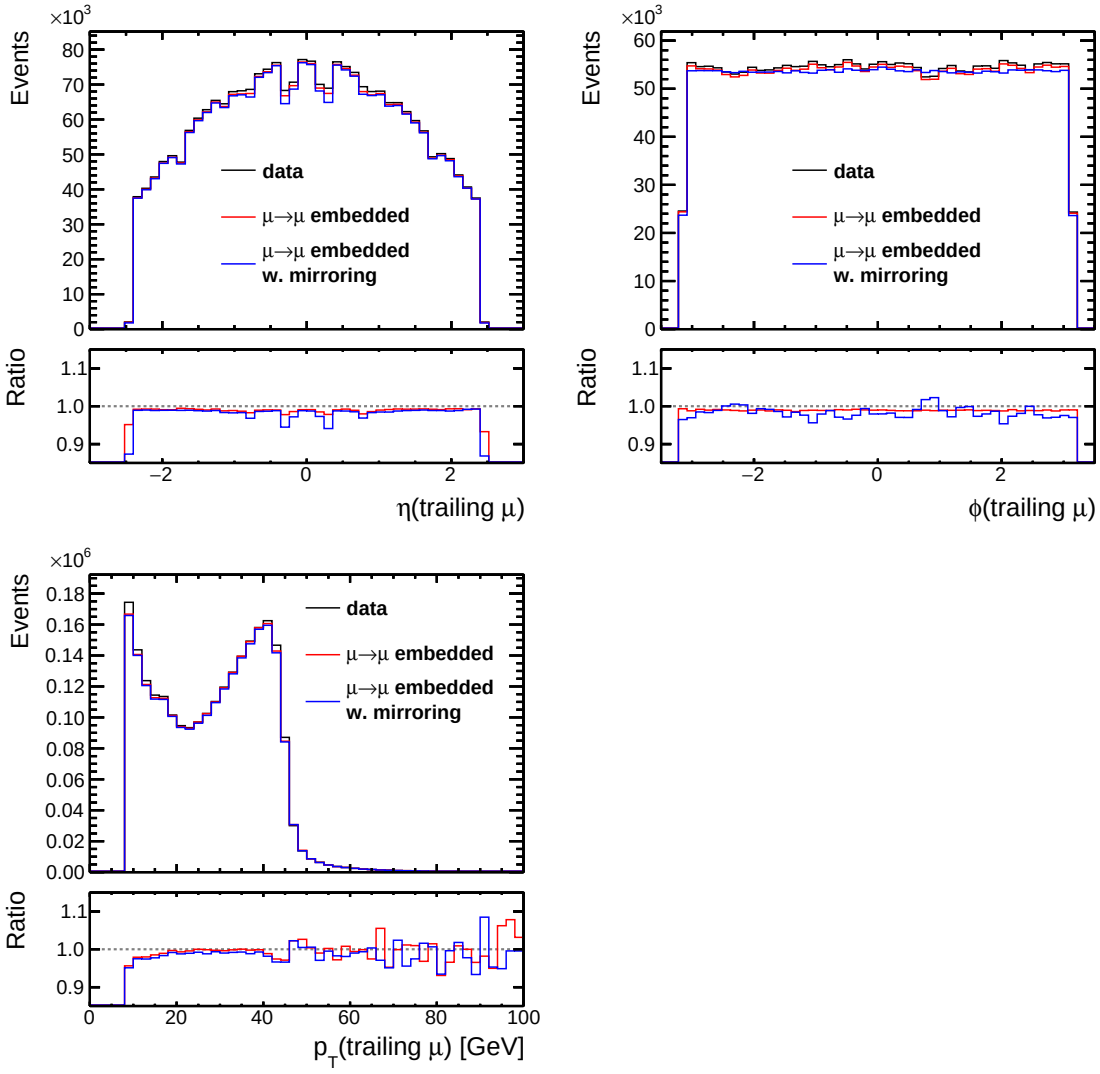

Figure B.3.: Kinematics of the trailing muon.

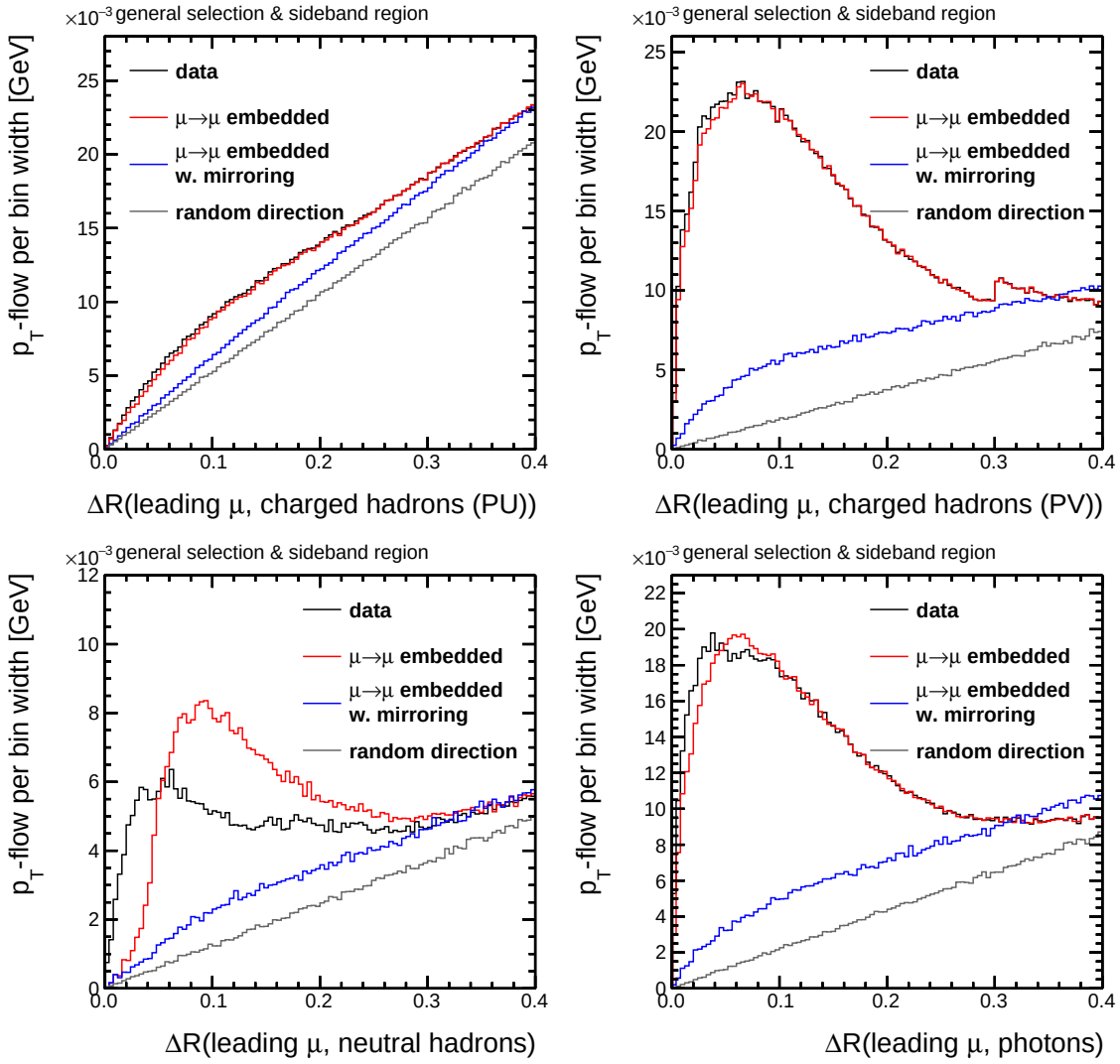

Figure B.4.: The p_T-flow as a function of ΔR for different particle flow candidates computed for the leading muon selected from the sideband region of the invariant mass $m(\mu\mu)$. The comparison is done for four datasets: (i) reference data in black, $\mu \rightarrow \mu$-embedded datasets (ii) without mirroring in red and (iii) with mirroring in blue and (iv) a dataset with a random direction chosen instead of the direction of the leading muon in grey.

Ich versichere hiermit, die vorliegende Arbeit selbstständig und nur unter Verwendung angegebener Hilfsmittel verfasst zu haben. Sämtliche Quellen, die in dieser Arbeit verwendet werden, habe ich an entsprechenden Stellen kenntlich gemacht.

Artur Akhmetshin

Karlsruhe, den 09.11.2016