Measurement of the CP Properties in the Higgs Boson Production at the CMS Experiment

Jordy Jozef Hubert Simon Degens
RWTH Aachen University,
Physikalisches Institut 3b

Referee: Prof. Dr. Achim Stahl,
Co-referee: Prof. Dr. Wolfgang Lohmann

April 12, 2017
## Contents

1 Introduction 5

2 The LHC and the CMS experiment 7
   2.1 Large Hadron Collider ............................................. 7
   2.2 The CMS experiment .............................................. 8

3 Theory introduction 9
   3.1 Production of the Higgs boson .................................... 9
   3.2 Decay of the Higgs boson and \( \tau \)-lepton .................... 10
   3.3 CP state .......................................................... 11

4 Monte Carlo simulation 13
   4.1 Signal samples .................................................... 13
   4.2 Comparison of the kinematic distributions in the POWHEG and JHU standard model event samples ................................. 14
      4.2.1 Momenta of the Higgs boson and \( \tau \)-lepton ......................... 14
      4.2.2 \( \tau \)-lepton decay mode ..................................... 16
   4.3 CP sensitive observable \( \Delta \phi_{jj} \) ............................... 17
   4.4 Rescaling the number of signal events ............................ 19
   4.5 Background Monte Carlo samples ................................ 20
   4.6 Event selection .................................................... 20

5 Statistical analysis 23
   5.1 Distinguishing between the CP hypotheses ....................... 23
   5.2 Luminosity projection ............................................. 26

6 Conclusion and outlook 29

References 30
Chapter 1

Introduction

In 1964 Peter Higgs [1], Francois Englert and Robert Brout [2] predicted the Higgs boson when they introduced the Brout-Englert-Higgs mechanism to give the gauge bosons masses. The ATLAS and CMS experiments have discovered such a boson in 2012 [3, 4]. The discovery of the Higgs boson is one of the biggest achievements of the LHC so far.

The Brout-Englert-Higgs mechanism (BEH mechanism) was introduced in the Standard Model of electroweak interactions because the bosons mediating the weak interaction have observable masses, while according to the gauge symmetry they should be massless. This problem can not be solved by adding bare masses since this would break the gauge invariance. The BEH mechanism introduces an additional field: the Higgs field. The potential of this field is symmetric around zero but does not have its minimum at zero. The coupling of the Higgs field is proportional to the mass of a particle. Therefore a massless particle like the photon does not interact with the Higgs field, and a heavy particle like the $\tau$-lepton couples stronger to it than a lighter particle as an electron. The Higgs boson can be seen as an excitation of the Higgs field and can be used to measure the properties of the Higgs field.

In some theories beyond the Standard Model more than one Higgs boson is predicted. These bosons differ by several quantum numbers. Therefore it is important to measure all the quantum numbers of the Higgs boson, because any disagreement with the Standard Model indicates new physics. Until now there are no deviations to the Standard Model regarding the Higgs bosons’ properties observed. A quantum number which will provide information whether the discovered Higgs boson is Standard Model like is the CP state of the Higgs boson. The Standard Model predicts a spin 0, CP-even Higgs boson, $J^{CP} = 0^+$. The latest data disfavours a pure CP-odd state, but a mixture between CP-even and CP-odd is not excluded yet [5]. Evidence of a CP-odd coupling of the Higgs boson would be a clear signature of physics beyond the Standard Model.

The Higgs boson decaying into two $\tau$-leptons is of great importance due to the fermionic coupling and its large branching fraction. The CP-odd terms are suppressed for bosonic couplings at leading order. Fermions interact with the Higgs boson via so-called Yukawa couplings. Here the CP-odd term is not suppressed and can be studied. Another decay with a large branching fraction is the one in two $b$-quarks. However, since the $b$-quarks are affected by large QCD backgrounds, Higgs bosons decaying into $b$-quarks are hard to measure. Therefore the $\tau$-decay channel is the most promising channel to analyse the CP state of the Higgs boson.
The Higgs CP properties are often proposed to be measured by analysing the decay products of the produced $\tau$-leptons. This requires precise measurement of the impact parameters of the $\tau$-leptons. The impact parameters are however difficult to measure in a hadron collider environment. In this thesis, the CP properties are studied in the Gluon Fusion production channel of the Higgs boson.
Chapter 2

The LHC and the CMS experiment

2.1 Large Hadron Collider

The Large Hadron Collider (LHC) is a proton collider at CERN in Geneva. It is a 27 km long circular accelerator and storage ring, where protons collide with 13 TeV center of mass energy. The protons are created by stripping orbiting electrons from hydrogen atoms. They then are accelerated in the Linac2 to an energy of 50 MeV, and injected into the Booster. The Booster accelerates the protons to 1.4 GeV. The beam then goes into the Proton Synchrotron. Proton bunches leave the Proton Synchrotron with 25 GeV and are sent into the Super Proton Synchrotron, and are injected into the two LHC pipelines after they reach 450 GeV. In the LHC the protons are accelerated to 6.5 GeV corresponding to a center of mass energy of 13 TeV. Inside the pipelines there is a vacuum. The proton beams are bended using dipole magnets, and focused by quadrupole magnets. Every proton beam has 2808 bunches consisting of $1.1 \cdot 10^{11}$ protons [6]. When a collision occurs a fraction of their kinetic energy is turned into mass and new particles are produced. To investigate the new made particles detectors are placed at the collision points. Along the LHC ring there are four detectors: ALICE, ATLAS, CMS and LHCb.

![Figure 2.1: Schematic view of the accelerators and experiments at CERN [6].](image-url)
2.2 The CMS experiment

The Compact Muon Solenoid (CMS) is one of the multipurpose detectors at the LHC. The main goal for the CMS detector is to find and study the Higgs boson and search for physics beyond the Standard Model. CMS is able to measure the momentum and track for charged particles and for neutral particles the energy. To do this it is built in an onion-like structure of subdetectors as can be seen in Figure 2.2. The first subdetector is the tracking system, which consists of silicon pixel and silicon strip detectors. They allow to reconstruct the trajectories of charged particles. Because of the magnetic field inside the detector the trajectories of the charged particles are curved and the curvature radius is used to measure their momentum. The next layer are the calorimeters. Here electrons, photons and jets are fully stopped and their deposited energy is measured. First the particle will hit the electromagnetic calorimeter (ECAL) where electrons and photons are fully stopped and their energy is measured. The ECAL uses PbWO$_4$ crystals for measuring the energy. After that there is another calorimeter, the hadron calorimeter (HCAL), which stops hadrons and measures their energy. The HCAL is a sandwich calorimeter consisting of brass and plastic scintillators. The following object is the superconducting solenoid magnet which provides a magnetic field of 3.8 T inside the detector. The iron return yokes outside the magnet guide the magnetic field.

Muons and neutrinos are not stopped by the calorimeters. Since muons are charged particles they produce tracks in the tracker and in dedicated muon chambers inside the iron yoke. Their curvature in the magnetic field of about 2 T inside the iron yoke gives a second momentum measurement. Neutrinos are not charged and do only weakly interact with the detector material so they are not visible in the detector. By calculating the missing transverse momentum and assigning it to the neutrinos the sum of all neutrinos can be measured [7].

Hadrons are composed of quarks. Quarks are colored objects and can not exist in free form according to the QCD laws. Accordingly they create hadrons, which are always color neutral. Due to color conservation new quarks and antiquarks are produced to make the final state hadrons color neutral. This collimated bunch of particles originating from a quark is a jet. If particles are detected in a small cone in the detector they are considered to be one jet.

![Figure 2.2: Cutaway view of the CMS detector](image-url)
Chapter 3

Theory introduction

3.1 Production of the Higgs boson

In the Standard Model the Higgs boson is produced mainly in three different mechanisms. The dominant production mechanism is Gluon Fusion (ggH). Here, two gluons create a loop of top-quarks which can produce the Higgs boson. The second production mechanism is called Vector Boson Fusion (VBF), where two quarks radiate a W or Z boson that annihilate into a Higgs boson. In the third process, a W or a Z is produced and they radiate off a Higgs boson. This process is called Higgsstrahlung. The Feynman diagrams for the three mechanisms are shown in Figure 3.1.

(a) ggH

(b) VBF

(c) Higgsstrahlung

Figure 3.1: Higgs boson production processes.
3.2 Decay of the Higgs boson and $\tau$-lepton

The Higgs boson decays via various channels, and the branching fraction is dependent on the mass of the Higgs boson. In the Standard Model the branching fractions for the decaying Higgs boson, with a mass of 125 GeV, are listed in Table 3.1.

Table 3.1: Branching fractions for the Higgs boson [9].

<table>
<thead>
<tr>
<th>decay mode</th>
<th>branching fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b\bar{b}$</td>
<td>57.7%</td>
</tr>
<tr>
<td>$\tau\tau$</td>
<td>6.32%</td>
</tr>
<tr>
<td>$WW$</td>
<td>21.5%</td>
</tr>
<tr>
<td>$gg$</td>
<td>8.57%</td>
</tr>
<tr>
<td>$c\bar{c}$</td>
<td>2.91%</td>
</tr>
<tr>
<td>$ZZ$</td>
<td>2.64%</td>
</tr>
<tr>
<td>$\gamma\gamma$</td>
<td>0.228%</td>
</tr>
</tbody>
</table>

This analysis focuses on the Higgs boson decaying into two $\tau$-leptons. Since the CP-odd term in the Lagrangian is suppressed at leading order for the bosonic coupling to the Higgs field, the measurement of the CP-odd state is not possible in the bosonic decay channels. Therefore the $\tau$-decay channel is of importance since the CP-odd term of the Higgs field in the Lagrangian couples to fermions in the same order as the CP-even term, and can be analysed.

The produced $\tau$-lepton can decay into hadrons, an electron or a muon and additional neutrinos. In $H \rightarrow \tau\tau$ decays there are different channels depending on the decay mode of the the two $\tau$-leptons. The fully hadronic channel where both $\tau$ decay into hadrons, the semileptonic where one $\tau$ decays into a hadron and the other decays leptonically and the full leptonic decay mode where both $\tau$ decay into leptons. The branching fractions for the decaying $\tau$-leptons are given in Table 3.2. The Feynman diagram is shown in Figure 3.2.

Table 3.2: Branching fractions for the $\tau$-lepton [10].

<table>
<thead>
<tr>
<th>decay mode</th>
<th>branching fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu + \nu_\mu$</td>
<td>17.4%</td>
</tr>
<tr>
<td>$e + \nu_e$</td>
<td>17.8%</td>
</tr>
<tr>
<td>Hadronic</td>
<td>64.8%</td>
</tr>
</tbody>
</table>

Figure 3.2: Feynman diagram for the $\tau$-lepton decay.
3.3 CP state

The CP state of a particle describes its behaviour under Charge conjugation and Parity transformation. Neutral particles might be eigenstates of the CP operator. An eigenstate with eigenvalue +1 is a CP-even state $|\Psi^+\rangle$ and with eigenvalue -1 to a CP-odd state $|\Psi^-\rangle$.

$$\text{CP } |\Psi^+\rangle = + |\Psi^+\rangle.$$  
$$\text{CP } |\Psi^-\rangle = - |\Psi^-\rangle.$$  

If the particle is not an eigenstate but a mixture of CP-even and CP-odd, the CP symmetry is not conserved.

$$\text{CP } |\Psi\rangle = \text{CP} (a |\Psi^+\rangle + b |\Psi^-\rangle) = a |\Psi^+\rangle - b |\Psi^-\rangle \neq \pm |\Psi\rangle,$$  

with $a, b \in \mathbb{R}$ and $a^2 + b^2 = 1$.

In field theories like the Standard Model the interactions between fields are described by the Lagrangian. For a Higgs boson produced by Gluon Fusion the effective Lagrangian is [11]:

$$L = \cos(\alpha_H) \frac{\alpha_s}{12\pi v} h G_{\mu\nu}^a G^{a,\mu\nu} + \sin(\alpha_H) \frac{\alpha_s}{8\pi v} h G_{\mu\nu}^a \epsilon_{\mu\nu\rho\sigma} G^{a,\rho\sigma}.$$  

Here the CP-even part is described by the first term and the CP-odd part by the second term. The mixing angle is denoted by $\alpha_H$, where $\alpha_H = 0$ corresponds to a pure CP-even state which is predicted in the Standard Model, while $\alpha_H = \frac{\pi}{2}$ corresponds to a CP-odd state. Since the Higgs boson can also be a mixture of the pure CP states, $\alpha_H$ could be any angle between 0 and $\pi$. $h$ denotes the Higgs field and $G_{\mu\nu}^a$ describes the gluon field. $\epsilon_{\mu\nu\rho\sigma}$ is the epsilon tensor, $\alpha_s$ the strong coupling constant and $v$ the vacuum expectation value of the Higgs field. Since the CP-odd part has a factor $\frac{1}{8}$ and the CP-even a factor $\frac{1}{12}$ the yield for the CP-odd Higgs boson is expected to be higher. In this study we consider Higgs boson CP states corresponding to $\alpha_H = 0$, $\alpha_H = \frac{\pi}{4}$ and $\alpha_H = \frac{\pi}{2}$.

In the Gluon Fusion production of the Higgs boson, gluons can be radiated both from gluons or quarks forming higher order QCD terms. These gluons produce jets which can be measured in the detector. The angle between the two jets in the transversal plane, $\Delta\phi_{jj} = \phi_1 - \phi_2$, is sensitive to the CP state of the Higgs Boson [12].
Consider two jets radiated from the initial state in opposite hemispheres and assume that the two initial state gluons have opposite momenta in the beam direction and only small transverse momenta, \( \vec{p}_\pm = \pm E_{CM} \vec{e}_z \) where \( E_{CM} \) is the center of mass energy of the dijet system. The squared matrix element for CP-even(+) and CP-odd(-) is proportional to [11]:

\[
|M|_{ggH\pm}^2 \sim \exp(4y_j) [A \pm B \cos(2\Delta \phi_{jj})],
\]

(3.5)

where \( y_j \) is the rapidity of the jet, \( y_j = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \), \( A = \xi^4 + \xi^{-4} + \frac{1}{2}(\xi^5 + \xi^{-3}) \) and \( B = 2 + \xi^2 \) with

\[
\xi \equiv \frac{E_{CM}}{E_{CM} - m_H}.
\]

(3.6)

In the limit where \( m_h \ll E_{CM} \) we obtain

\[
|M|_{GF\pm}^2 \sim \exp(4y_j) [3 \pm 3 \cos(2\Delta \phi_{jj})].
\]

(3.7)

The cross section is proportional to the squared matrix element and a dependency on \( \cos(2\Delta \phi_{jj}) \) is expected.
Chapter 4

Monte Carlo simulation

Monte Carlo simulations are used in the particle physics analyses to simulate physics processes even with complex events which may occur in collisions and to estimate the detector response. They are based on theoretical calculations for these processes. The theoretical cross sections are often too difficult to be calculated by hand so they are estimated with Monte Carlo simulation. A Monte Carlo generator delivers the four momenta of all produced particles (generator level). These particles are then transported through the detector and using the signals from all subdetectors the events are reconstructed using the same programs applied to recorded data. In the analysis a comparison between theoretical prediction and data can be done.

4.1 Signal samples

As signal samples three Gluon Fusion Monte Carlo samples produced with the JHU generator [13–16] are used, each with a different CP state of the Higgs boson. The three samples have the following CP state for the Higgs boson

- The Standard Model CP-even Higgs boson, $\alpha_H = 0$
- A Higgs Boson with a CP state corresponding to $\alpha_H = \frac{\pi}{4}$
- The CP-odd pseudoscalar Higgs boson, $\alpha_H = \frac{\pi}{2}$

For Standard Model Higgs analysis a different Monte Carlo sample is available, created with the generator POWHEG [17]. The main difference between the POWHEG and the JHU samples is that the JHU simulates $gg \rightarrow ggH$ as leading order process, where gluons are radiated off from the initial state particles. Two extra jets produced in the initial state with a dijet mass $m_{jj} > 200$ GeV are required. POWHEG [17] simulates $gg \rightarrow H$. First a comparison between the POWHEG and JHU Gluon Fusion Monte Carlo samples will be discussed at generator level and second a statistical analysis using the three Monte Carlo samples produced by the JHU generator is done at reconstruction level to estimate the potential to get evidence for a Higgs boson with a CP-odd admixture.

To have a fair comparison we used the same selection criteria for both samples. All jet-related variables are in this analysis always at reconstruction level, because the parton information on generator level is not available in the used Monte Carlo samples. This however is not critical since the jets are well reconstructed. We require that there are at least two jets in the final state, and that the dijet mass, $m_{jj}$, is larger than 200 GeV. The statistical analysis later is performed at reconstruction level and therefore all selection criterion are on reconstruction level.
4.2 Comparison of the kinematic distributions in the POWHEG and JHU standard model event samples

4.2.1 Momenta of the Higgs boson and $\tau$-lepton

Assuming that both Standard Model samples are based on the same physics, energy and angular distributions of the final state particles must be compatible within statistical uncertainties. A comparison between the momenta, the rapidities and the azimuthal angles of the produced Higgs bosons and final state $\tau$-leptons for the POWHEG and JHU standard model samples is done. Since the JHU generator produces at least 2 jets, only events with at least two jets are selected, $N_{jets} \geq 2$. The second selection criterion is applied on the dijet mass $m_{jj} > 200\text{GeV}$ since this is also a generator cut for the JHU samples. A comparison between the Higgs boson’s $p_T$, $\eta^1$ and $\phi$ distributions obtained from the POWHEG and JHU Standard Model Monte Carlo samples is shown in Figure 4.1.

Figure 4.1: Distributions of the transverse momentum $p_T$ (a), the rapidity $\eta$ (b), and the azimuthal angle $\phi$ (c) for produced Higgs bosons obtained from the POWHEG and JHU generators.

---

1pseudorapidity: related to the angle of a particle relative to the beam axis. For relativistic particles the pseudorapidity has the same numeric value as the rapidity.
4.2. Comparison of the kinematic distributions in the POWHEG and JHU standard model event samples

As can be seen in Figure 4.1 (a) the Higgs boson’s $p_T$ spectrum is harder in the JHU sample. Also the Higgs boson is more central in $\eta$ in this sample as shown in Figure 4.1 (b). The distribution for $\phi$ in Figure 4.1 (c) is for both generators compatible within statistical uncertainties. The same differences have been observed between the POWHEG event samples and the JHU samples with different CP state. A further investigation of the distributions of the same variables of the produced \( \tau \)-leptons reveals that the $p_T$ distributions is also harder as expected for the JHU sample, as seen in Figure 4.2 (a). The $\eta$ and $\phi$ distributions are compatible within statistical uncertainties, as Figure 4.2 (b/c) shows. Apparently, the additional jets radiated off from the initial state particles change the kinematics of the events. The larger amount of Higgs bosons at high $p_T$ in the JHU samples leads also to a larger fraction of $\tau$-leptons at higher $p_T$.

Figure 4.2: Distributions of the transverse momentum $p_T$ (a), the rapidity $\eta$ (b), and the azimuthal angle $\phi$ (c) for produced $\tau$-leptons obtained from the POWHEG and JHU generators.
4.2.2 \( \tau \)-lepton decay mode

The decay of the \( \tau \)-leptons is decoupled from the production process. The branching fractions of the \( \tau \)-leptons are given in Table 3.2. In Figure 4.3 the mass of the visible decay products of the \( \tau \)-lepton on generator level is plotted for both Monte Carlo generators. It can be seen that POWHEG correctly models the \( \tau \)-lepton decay into light leptons (\( e, \mu \)) and hadrons (\( \pi, \rho, a_1, \ldots \)). The relative fractions agree with the ones listed in Table 3.2. However for the JHU samples this is not the case. Only a peak at the single pion mass is observed. Apparently, the \( \tau \)-decay is not modelled in the right way. Another unexpected feature is the tail above the pion mass.

Since the POWHEG sample only simulates events with a CP-even Higgs boson, this sample cannot be used for a study of the impact of CP mixing on the \( \Delta \phi_{jj} \) distributions defined in equation 3.7. Therefore special samples have been generated with the JHU Monte Carlo generator for different values of \( \alpha_H \). However, in the JHU samples the decay of the \( \tau \)-leptons is not modelled properly. The \( \tau \)-leptons decay only into one pion and a neutrino. As a temporary solution only the \( \tau \)-leptons decaying into hadrons are used and their number in JHU sample is rescaled to the number of \( \tau \)-leptons decays into hadrons in the Gluon Fusion POWHEG sample for a given luminosity.
4.3 CP sensitive observable $\Delta \phi_{jj}$

The distributions, at reconstruction level, of $\Delta \phi_{jj}$ for CP-even, CP-odd and a CP mixed Higgs boson produced with the JHU Monte Carlo generator are shown in Figure 4.4. We require the jets to be in opposite hemispheres, $(\eta_{j1} \cdot \eta_{j2} < 0)$, have at least a difference in $\eta$ of 3.5, $|\Delta \eta_{jj}| > 3.5$, and the invariant mass of the two jets must be large $m_{jj} > 500$ GeV. These selection cuts are fulfilled by events as described in section 3.3, with two high energy jets in opposite direction of flight.

![Figure 4.4: The $\Delta \phi_{jj}$ distribution for different CP states of the Higgs boson.](image)

The distributions in the quantity $\Delta \phi_{jj}$ for all three CP states have the form as predicted in previous calculations [11]. These distributions are normalized to unity, and one should keep in mind that the yield is different for different CP states. A clear difference between the three CP states is observed in the shape of the $\Delta \phi_{jj}$ distribution, hence $\Delta \phi_{jj}$ is a variable sensitive to the CP state of the Higgs boson.

In Figure 4.5 the distribution of $\Delta \phi_{jj}$ is shown for the three CP states using no requirements on the jets. The sensitivity to the CP state of the Higgs boson almost disappears. The requirement on the jets to be in opposite hemispheres $(\eta_{j1} \cdot \eta_{j2} < 0)$ enhances the sensitivity to the CP state of the Higgs boson, as Figure 4.5 (b) shows. The requirement is applied throughout the remaining part of this work.
Figure 4.5: The $\Delta \phi_{jj}$ distribution for the three CP states with no jet selection criterion (a) and after applying $\eta_{j1} \cdot \eta_{j2} < 0$ (b).

The impact of the cut on $|\Delta \eta_{jj}|$ can be seen in Figure 4.6. The requirement for the jets to have large rapidity difference decreases the number of events and increases the statistical uncertainties on the one hand but gives a more pronounced difference in the shape for CP-even and CP-odd states. For this study a cut for $|\Delta \eta_{jj}| > 3.5$ is made as proposed in reference [11]. This is a compromise between statistical precision and sensitivity to the shape. In Figure 4.7 the same distributions are shown for different masses of the dijet system. Again the difference in shape rises with larger $m_{jj}$. A cut on $m_{jj} > 500\text{GeV}$ is made again as proposed in reference [11]. For large differences in $\eta$ or $m_{jj}$ the radiated jets represent events as described in section 3.3 and have the form as predicted in equation 3.7.

Figure 4.6: The $\Delta \phi_{jj}$ distributions for different $|\Delta \eta_{jj}|$ cuts for the CP-even (a) and CP-odd (b) Higgs boson.
4.4. Rescaling the number of signal events

In section 4.2.2 is shown that the $\tau$-leptons in the JHU samples only decay into pions and a neutrino. Since this does not correspond to the physical truth, the number of events in this channel for the Standard Model JHU sample is scaled down to the number of events in the POWHEG sample after requiring only $\tau$-lepton decays in hadrons. The same scale factor $k = \frac{N_{\text{POWHEG}}}{N_{\text{JHU,SM}}} = 0.1$ is then applied to the other JHU samples with different Higgs boson CP state. The number of events obtained in the different samples shown in Table 4.1. The shape of the $\Delta \phi_{jj}$ distribution for the 3 JHU samples stays the same, since they are modelled properly. The cross sections are different for different CP states, and this affects the number of events for each CP state as listed in Table 4.1.

Figure 4.7: The $\Delta \phi_{jj}$ distribution for different $m_{jj}$ cuts for the CP-even (a) and CP-odd (b) Higgs boson.

From these distributions I concluded that the kinematics of the initial state jets is simulated properly. The CP sensitive variable $\Delta \phi_{jj}$ is modelled as expected in all JHU samples and is not dependent on the decay of the $\tau$-lepton. Therefore rescaling the number of events in the hadronic decay channel for the JHU sample to the number of events in the Gluon Fusion POWHEG sample for a given luminosity is reasonable. This is described in section 4.4 and a statistical analysis to separate the three CP states is based on the rescaled samples.
Table 4.1: The numbers of events after application of the scale factor $k$.

<table>
<thead>
<tr>
<th>Monte Carlo sample</th>
<th>theoretical cross section</th>
<th>events after cut</th>
<th>events after rescaling</th>
</tr>
</thead>
<tbody>
<tr>
<td>PowHeg</td>
<td>0.99</td>
<td>6.9</td>
<td>6.9</td>
</tr>
<tr>
<td>JHU Standard Model</td>
<td>0.99</td>
<td>69.8</td>
<td>6.9</td>
</tr>
<tr>
<td>JHU maximal mixing</td>
<td>1.40</td>
<td>96.0</td>
<td>9.6</td>
</tr>
<tr>
<td>JHU pseudoscalar</td>
<td>2.09</td>
<td>153.0</td>
<td>15.3</td>
</tr>
</tbody>
</table>

4.5 Background Monte Carlo samples

Higgs bosons produced via Gluon Fusion with two initial state jets are considered as signal. The Higgs boson then decays into two $\tau$-leptons where both are required to decay hadronically. A signal event contains at least four reconstructed jets in the final state, where two jets must be identified as $\tau$-leptons. Events that fulfill these criteria are not necessarily Higgs boson events. They can also be produced by several other processes which are considered as backgrounds. These are other processes which also produce $\tau$-leptons in the final state, for example the Drell-Yan process ($Z/\gamma \rightarrow \tau\tau$). It can also happen that a particle is misidentified as a $\tau$-lepton, for example in the $W$+jets events with one $\tau$-lepton from the $W$ decay and the other from a misidentified jet.

To estimate the background, Monte Carlo samples from the spring 2016 simulation campaign with a center of mass energy $E_{cm} = \sqrt{s} = 13$ TeV are used. The following backgrounds are considered in this analyses:

- Drell-Yan background samples ($Z/\gamma \rightarrow \tau\tau/ll$). Using MadGraph [18] for the hard interaction and PYTHIA8 [19] for the hadronization as generators.
- Di-boson background Monte Carlo samples produced with the generators aMC@NLO [20] for the hard interaction and PYTHIA8 for the hadronization:
  1. ZZ
  2. WW
  3. WZ
- top quark (t$\bar{t}$+jets) background samples using POWHEG [17] for the hard interaction and PYTHIA8 for the hadronization.
- $W$+jets Monte Carlo samples, with generators MadGraph for the hard interaction and PYTHIA8 for the hadronization.
- QCD background estimated from data using the ABCD method [21].
- Higgs boson produced by Vector Boson Fusion (qqH) using POWHEG for the hard interaction and PYTHIA8 as generator for the hadronization.

4.6 Event selection

The sensitivity of the statistical analysis to different CP states depends on the difference between the $\Delta \phi_{jj}$ distributions and how well the backgrounds can be suppressed. Therefore cuts are applied in order to reduce the background contributions [22]. First, the hadronically decaying $\tau$-lepton candidates are identified using the HPS-algorithm [23].
4.6. Event selection

Then events are selected that fulfill the following criteria:

- \( n_{\text{jets}} \geq 2 \)
- \((\eta_1 \cdot \eta_2 < 0)\)
- \(|\Delta \eta_{jj}| > 3.5\)
- \(m_{jj} > 500\text{GeV}\)
- \(p_{T_1} > 50\)

The first selection criterion is on events where at least two additional jets are detected, \( n_{\text{jets}} \geq 2 \). The jets must be in opposite hemispheres as has been described in section 4.3, \((\eta_1 \cdot \eta_2 < 0)\) and \(|\Delta \eta_{jj}| > 3.5\). The distributions of the variables the selection is based on are shown in Figure 4.8. Only the dominant backgrounds and a CP-even signal are considered in these plots. The requirement on the invariant mass of the dijet pair, \(m_{jj} > 500\text{GeV}\), increases the sensitivity to the CP state and reduces the Drell-Yan, QCD and \(t\bar{t}\) background contribution. The cut on the momentum of the leading jet, \(p_{T_1} > 50\), further decreases the Drell-Yan and QCD background.

![Figure 4.8](image_url)

Figure 4.8: The distribution of \(|\Delta \eta_{jj}|\) (a), \(m_{jj}\) (b) and leading jet \(p_{T_1}\) (c) for signal and background events.
The QCD, Drell-Yan and $t\bar{t}$ backgrounds are reduced considerably by applying the selection cuts. The VBF background, however, is increased relative to the signal. Since this is not a dominant background the total background is decreased. Using the integrated luminosity of $36.5 \, \text{fb}^{-1}$ recorded in 2016 by the CMS detector. The distributions for $\Delta \phi_{jj}$ are plotted for background and a CP-even Higgs boson signal in Figure 4.9. The signal over background ratio, $\frac{S}{B} = \frac{\text{signal yield}}{\text{background yield}}$, is before cuts 0.017 and 0.04 after applying them, increasing the signal over background ratio by more than a factor 2.

Figure 4.9: The $\Delta \phi_{jj}$ distribution for signal and background events, (a) without the cuts on $|\Delta \eta_{jj}|$, $m_{jj}$ and $p_T$ and (b) with these cuts applied. The number of events corresponds to an integrated luminosity of $36.5 \, \text{fb}^{-1}$. 

(Notations and diagrams related to the figure are not transcribed here as they are visual components of the text.)
Chapter 5

Statistical analysis

5.1 Distinguishing between the CP hypotheses

In order to investigate the separation between the different CP states, the probability that a certain CP hypothesis is realised in nature is estimated based on a likelihood formalism. Since it is a counting experiment the yield follows a Poisson distribution. The likelihood function is then given as:

\[
L(\vec{D}| \mu, \alpha_H, \vec{\theta}) = \prod_{i=1}^{n} \left[ \frac{\mu S_i(\alpha_H) + B_i}{D_i(\alpha_{toy})} \right]^{D_i(\alpha_{toy})} \exp\left[ -\mu S_i(\alpha_H) - B_i \right] \cdot \prod_{j=1}^{m} p(\theta_j|\tilde{\theta}_j). \quad (5.1)
\]

The signal model, \( S_i(\alpha_H) \), determines the number of signal events in each bin \( i \) for a given CP hypothesis \( \alpha_H \). \( B_i \) is the total background contribution for each bin \( i \) and \( D_i \) the number of data events in the same bins. \( B_i \) is independent of the CP state of the Higgs boson. The signal strength parameter \( \mu \) scales the number of signal events. The systematics uncertainties are considered using nuisance parameters, and are included via the term \( p(\theta_j|\tilde{\theta}_j) \). This term is the probability density function (pdf) for a given \( \theta_j \) with a predicted value of \( \tilde{\theta}_j \). The systematic uncertainties considered in the \( H \to \tau\tau \) analysis are used \[22\]. They follow a log normal distribution and are listed in Table 5.1.

Table 5.1: Systematic uncertainties considered in this analysis.

<table>
<thead>
<tr>
<th>Uncertainty</th>
<th>Affected Processes</th>
<th>Change in Acceptance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tau ID (&amp; trigger)</td>
<td>signal &amp; sim. backgrounds</td>
<td>10-16%</td>
</tr>
<tr>
<td>e misidentified as ( \tau_h )</td>
<td>Drell-Yan</td>
<td>10 -15%</td>
</tr>
<tr>
<td>Jet misidentified as ( \tau_h )</td>
<td>Drell-Yan</td>
<td>20%</td>
</tr>
<tr>
<td>Jet energy scale</td>
<td>signal &amp; sim. backgrounds</td>
<td>shape unc.</td>
</tr>
<tr>
<td>( \tau ) energy scale</td>
<td>signal &amp; sim. backgrounds</td>
<td>shape unc.</td>
</tr>
<tr>
<td>MET scale</td>
<td>signal &amp; sim. backgrounds</td>
<td>2 - 3%</td>
</tr>
<tr>
<td>Norm. ( Z \to \tau\tau )</td>
<td>Drell-Yan</td>
<td>4%</td>
</tr>
<tr>
<td>Norm. diboson</td>
<td>di-boson</td>
<td>5%</td>
</tr>
<tr>
<td>Norm. W+jets</td>
<td>W+jets</td>
<td>4%</td>
</tr>
<tr>
<td>QCD uncertainty</td>
<td>QCD</td>
<td>28%</td>
</tr>
<tr>
<td>Theory uncertainties qqH</td>
<td>qqH</td>
<td>0.7 - 3.2%</td>
</tr>
<tr>
<td>Luminosity</td>
<td>signal &amp; sim. backgrounds</td>
<td>6.2%</td>
</tr>
</tbody>
</table>
The likelihood function is maximised for these variables in order to find the values for these nuisance parameters, that are realised in data. After calculating the likelihood a comparison between the different CP hypotheses is done. This is based on the test statistic $q$, where $\alpha_A$ is the null hypothesis and $\alpha_B$ is the alternative hypothesis.

$$q = \ln(L(\vec{D} | \mu, \alpha_A, \vec{\theta})) - \ln(L(\vec{D} | \mu, \alpha_B, \vec{\theta}))$$  \hspace{1cm} (5.2)

In order to obtain a distribution of the test statistics, one needs to generate toy datasets. New datasets are artificially created by generating new signal and background templates based on Poisson statistics. New datasets are created for each CP hypothesis and a new test statistic $q$ is obtained for each of these datasets.

$$D_{toy}(\alpha_H) = S_{toy}(\alpha_H) + B_{toy}$$ \hspace{1cm} (5.3)

In order to distinguish between the two hypotheses the p-value is calculated. The p-value is the probability to identify the wrong alternative hypothesis, when the null hypothesis is correct. This is done by calculating the integral for the alternative hypothesis probability density function of $q$ up until the observed $q$ value

$$p = \int_{-\infty}^{q_{obs}} p_B(q) dq.$$ \hspace{1cm} (5.4)

This is visualised in Figure 5.1.

---

**Figure 5.1:** A visualization of the test statistic $q$ for the null and alternative hypothesis and the determination of the p-value.
The statistical analysis is done in the full hadronic channel based on $\Delta\phi_{jj}$ as observable. As observation an asimov dataset, $D_{\text{asimov}}(\alpha_H) = S(\alpha_H) + B$, is used with a CP-even Standard Model Higgs Boson as signal to evaluate the sensitivity to $\Delta\phi_{jj}$. As null hypothesis the Standard Model Higgs boson is used and for the alternative hypothesis either a CP-odd pseudoscalar Higgs boson or a state with maximal mixing, $\alpha = \frac{\pi}{4}$, is used. The Gluon Fusion distributions generated with the JHU generator are used as a signal and as background we consider the ones mentioned in section 4.5, both after all selections. The likelihood is calculated once for a model dependent Higgs CP state. Here the separation between the different hypotheses is based on the differences in yield and shape. For a model independent CP state of the Higgs boson, the separation is only based on the shape of the $\Delta\phi_{jj}$ distributions. For each CP hypothesis 20000 toy datasets are generated and the resulting test statistic distributions are plotted in Figure 5.2 together with the single test statistic values $q$ obtained for the pseudo-data.

Figure 5.2: Test statistic distribution for $\alpha_H = 0$ as null hypothesis and $\alpha_H = \frac{\pi}{2}$ as alternative hypothesis in (a) and (b) and for $\alpha_H = \frac{\pi}{4}$ in (c) and (d) obtained with a luminosity of $36.5 \text{ fb}^{-1}$. In the left plots the difference in yield and shape is used to distinguish between the two CP hypotheses. In the right plots the signal yield is fixed to the expectation for $\alpha = 0$ which lets the fit only distinguish between the shapes.
For the separation CP states $\alpha_H = 0$ and $\alpha_H = \frac{\pi}{2}$, the obtained p-value is 0.29 taking into account both yield and shape. This corresponds to a significance of 0.54 $\sigma$. Taking into account only the shape a p-value of 0.39 is obtained. Since the separation is better for the model dependent CP state, the main separation therefore comes from different yields.

The separation between the CP states $\alpha_H = 0$ and $\alpha_H = \frac{\pi}{4}$ is also too small to distinguish between the two CP states. A p-value of 0.44 which corresponds to a significance of 0.14 $\sigma$ is obtained if we distinguish between yield and shape. If only the shape is used to separate the CP hypotheses the p-value is 0.45 corresponding to 0.12 $\sigma$.

With the current dataset there is no discrimination possible between the different CP states, but the LHC is still taking data and in the future a larger dataset will be available. The distribution for signal and background in Figure 4.9 (b) shows that the background distribution fluctuates. Especially the bins of the QCD template are not well populated. A smoother background distribution will improve the discrimination between the CP states. Also only $\tau$-lepton decay in hadrons are analyzed. Analyzing also the other decay modes will improve the discrimination between the $\tau$-leptons.

### 5.2 Luminosity projection

The LHC started taking data at $\sqrt{s} = 13$ TeV in 2015 and will run in the coming years taking more and more data. In Figure 5.3 the plan for the LHC for the next years is shown. In the coming years an integrated luminosity of 300 fb$^{-1}$ is expected. It is interesting to investigate the sensitivity of the analysis as a function of the increasing luminosity. An increased luminosity corresponds to a larger dataset to be analysed and this leads to smaller statistical uncertainties. A very conservative assumption is made, that the relative systematic uncertainties remain the same, although an improved measurement is likely. This will increase the separation power between two hypotheses.

![Figure 5.3: Plan for the LHC](image)

The number of events is proportional to the integrated luminosity. For a given luminosity ($L_1$) the expected number of events ($N_1$) for a certain process with cross section $\sigma$ is:

$$N_1 = \sigma \cdot L_1 \Rightarrow \sigma = \frac{N_1}{L_1}. \quad (5.5)$$
Since the cross section $\sigma$ is constant, the number of events for a higher integrated luminosity ($L_2$) is:

$$N_2 = \sigma L_2 = \frac{N_1}{L_1} L_2.$$  

(5.6)

Again only the $\tau$-leptons decaying into hadrons are used for discriminating between the CP hypotheses. The number of events with a luminosity of 36.5 fb$^{-1}$ are used to calculate the expected yields for higher luminosities. The signal to background ratio remains untouched.

![Graphs showing test statistic distributions for a Luminosity of 300 fb$^{-1}$ with null hypothesis $\alpha_H = 0$ and as alternative hypothesis $\alpha = \frac{\pi}{2}$ (a) and $\alpha = \frac{\pi}{4}$ (b).](image)

Figure 5.4: test statistic distributions for a Luminosity of 300 fb$^{-1}$ with null hypothesis $\alpha_H = 0$ and as alternative hypothesis $\alpha = \frac{\pi}{2}$ (a) and $\alpha = \frac{\pi}{4}$ (b).

In Figure 5.4 the test statistic distribution $q$ for an integrated luminosity of 300 fb$^{-1}$ is shown. Here both the yield and the shape are used to discriminate between the two hypotheses.

The separation between CP-even and CP-odd gives a p-value of 0.074 which corresponds to a significance of 1.45 $\sigma$. This means that the probability to falsely identify the alternative CP-odd Higgs boson is 7.4% if a CP-even Higgs boson is realised.

The p-value for CP-even vs. a CP-mixed state, $\alpha_H = \frac{\pi}{4}$, is 0.31. The probability to falsely accept the mixed CP state is 31%. This corresponds to a significance of 0.49 $\sigma$.

In Figure 5.5 the obtained significance as a function of the integrated luminosity is shown. From the statistical uncertainties a dependency on the square root of the luminosity is expected. For the systematic uncertainties a constant dependency of the luminosity for the separation is expected. The square sum of the two terms gives the total dependency of the separation as a function of the luminosity.
Figure 5.5: The significance for the separation of the CP hypotheses as a function of the integrated luminosity.

Increasing the size of the dataset increases the discrimination between the hypotheses. The discrimination, for an integrated luminosity of 300 fb$^{-1}$, between the CP states is obtained only for the $\tau$-leptons decaying into hadrons. Taking into account more $\tau$-decay channels will enhance the statistical precision, and will increases the discrimination. Since the QCD background is obtained from data, a larger dataset will decrease the QCD uncertainties and a smoother distributions is expected.
Chapter 6

Conclusion and outlook

The CP state of the Higgs boson is still an open question. The Higgs boson decaying into \( \tau \)-leptons is the most promising decay channel to measure the Higgs CP state. In this study the angle between two radiated jets in the gluon fusion production of the Higgs boson. As Figure 4.4 shows this variable is indeed sensitive to the CP state of the Higgs boson is considered. A statistical analysis has been performed with only the \( \tau \)-leptons which decay into hadrons. The discriminating power has been analysed for model dependent and independent Higgs boson CP hypotheses. For a model dependent Higgs boson the difference in Gluon Fusion cross sections for different CP states can be used to enhance the sensitivity between the CP states. Although the statistical uncertainties are too large for a discrimination between the CP states at the moment. A projection to higher luminosities shows that a discrimination between CP-even and CP-odd of 1.45 \( \sigma \) has been estimated for an integrated luminosity of 300 fb\(^{-1}\) using only the hadronically decaying \( \tau \)-leptons.

This first analysis to explore the potential to measure the CP properties of the Higgs boson from the radiated jets in the initial states reveals that it will be a challenging topic. Several improvements are, however possible. Taking into account the other decay channels for the produced \( \tau \)-leptons will increase the discrimination between the CP states. An increased statistical precision for the background will decrease its fluctuations and might improve the discrimination between the CP states. This will be the next steps for analysing the CP properties of the Higgs boson in the initial state.
References


[22] CMS Collaboration, “Analysis of the SM 125 GeV Higgs Boson decaying to a pair of $\tau$-leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV”. To be published.
