Data-Driven Validation of the Photon Identification for Top Photon Vertex Investigations with the CMS Experiment

von

Thomas Steeger

Bachelorarbeit in Physik

vorgelegt der
Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen

im
August 2012

angefertigt im
III. Physikalischen Institut B

bei
Priv.-Doz. Dr. Oliver Pooth
Ich versichere, dass ich die Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht habe.


____________________
## Contents

1 Introduction 1

2 The Standard Model 3
2.1 Gauge Bosons .......................................................... 3
2.2 Fermions .................................................................. 3
  2.2.1 Leptons ................................................................. 4
  2.2.2 Quarks ................................................................. 4

3 The Top Quark at the LHC 5

4 The CMS Experiment at the LHC 7
4.1 The Large Hadron Collider .............................................. 7
4.2 The CMS Detector ........................................................ 8
  4.2.1 The Tracking System ............................................. 9
  4.2.2 The Electromagnetic Calorimeter .............................. 10
  4.2.3 The Hadron Calorimeter ........................................ 11
  4.2.4 The Solenoid ........................................................ 11
  4.2.5 The Muon System .................................................. 11

5 Data, Simulation and Selection 13

6 Data-Driven Estimation of the Photon Identification 15
6.1 Photon Radiation of Top Quarks ................................. 15
6.2 Photon Identification .................................................. 16
6.3 Fake Photon Rate from Simulated Events ....................... 23
6.4 Fake Photon Rate with the Data-Driven Ansatz ............... 24
  6.4.1 The Discriminating Variable $\sigma_{\text{spin}}$ .................. 24
  6.4.2 Independence of the Composition of Different Processes 25
  6.4.3 The Template Fit Method ...................................... 25
  6.4.4 Estimation of Systematic Uncertainties: Ensemble Study 28
  6.4.5 Sources of Fake Photons ....................................... 30

7 Conclusion 31

Bibliography 33
Chapter 1

Introduction

Within human living mankind strives for knowledge trying to gain deeper insight into the secrets of nature. This process of understanding the system, everything has to adhere to, is greatly accelerating since the invention of computer technology, which provides great resources of processing power, and thereby facilitates progress in science. Although many discoveries could only just be achieved in the last century, some basic concepts have already arisen in ancient Greece, which is well-known for its great philosophers. The ancient Greek philosopher and atomist Democritus spread the idea that matter consists of indivisible basic units but centuries of scientific progress passed before human mankind has been able to prove Nature’s composition of elementary constituents. Thus, the actuality of conceptual ideas from Democritus shows the motivating fascination of a theoretical idea to experimentalists.

The knowledge in physics has increased since Democritus. It has been proven that the composition of chemical elements determine the characteristics of matter. Furthermore, these elements consist of different types of atoms comprising electrons, protons and neutrons. Later on, it was noticed that also protons and neutrons have a substructure. These particles were called quarks. All this led to the development of the standard model of particle physics describing Nature’s fundamental components and their interaction.

Big effort has to be taken in order to prove or extend the standard model. For this purpose particle accelerators are built, which accelerate particles to let them collide under reproducible conditions. Due to inelastic scattering processes divergent elementary particles are created and can be investigated.

In this bachelor thesis the top quark is studied. Therefore top-antitop pairs with a radiated photon are used. The top-photon radiation was measured recently for the first time \[1\]. It is an important cross check of the standard model because theoretical extensions lead to modifications of this process. Additionally, it is a direct proof of the top quark charge to be \(\frac{2}{3}e\). A key ingredient of the top-photon cross section measurement is the identification efficiency of photons. The aim of this analysis is to estimate this efficiency from data. This makes the cross section analysis more independent of simulation and more robust. These investigations are based on the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC). The LHC is located near Geneva at the border of France and Switzerland and it is the world’s highest-energy particle accelerator. The CMS ex-
Experiment is a general purpose detector for particle physics and explores the collision of proton-proton bunches with the aim to confirm and expand the standard model of particle physics.

Annotation

In this bachelor thesis the system of natural units for particle physics is used frequently. That means that the speed of light as well as the Planck constant \( \hbar \) are set to 1.

\[
c = 1 \quad \hbar = 1 \quad (1.1)
\]

This leads to various conversion factors for other units. Table 1.1 gives a short overview.

<table>
<thead>
<tr>
<th>unit</th>
<th>SI</th>
<th>natural units</th>
</tr>
</thead>
<tbody>
<tr>
<td>meter</td>
<td>( 10^{-15} ) m</td>
<td>5.07 GeV(^{-1} )</td>
</tr>
<tr>
<td>kilogram</td>
<td>( 1.78 \cdot 10^{-27} ) kg</td>
<td>1 GeV</td>
</tr>
<tr>
<td>second</td>
<td>1 s</td>
<td>( 1.52 \cdot 10^{24} ) GeV(^{-1} )</td>
</tr>
<tr>
<td>energy</td>
<td>( 1.60 \cdot 10^{-19} ) J</td>
<td>1 eV</td>
</tr>
</tbody>
</table>

Furthermore, the official coordinate conventions for the CMS detector will be used. This means that the origin of the coordinate system is the nominal collision point. The \( z \)-axis points along the beam direction, the \( y \)-axis points vertically upwards and the \( x \)-axis points to the center of the LHC. The azimuthal angle \( \phi \) and the polar angle \( \theta \) are the conventional angles according to spherical coordinates. Additionally, the pseudorapidity \( \eta \) \( (1.2) \) is used instead of \( \theta \), which is useful due to a more constant particle flux in \( \eta \) \( [3] \).

\[
\eta = - \ln \left( \tan \left( \frac{\theta}{2} \right) \right) \quad (1.2)
\]
Chapter 2

The Standard Model

The standard model of particle physics describes elementary particles and their interaction. Elementary particles can be classified into fermions and bosons. Fermions are particles, which are characterized by the Fermi-Dirac statistics. According to the spin-statistics theorem each fermion is of half-integer spin and thereby obeys the Pauli exclusion principle, which is the reason, why matter is stable and occupies a volume. Bosons follow the Bose-Einstein statistics and are of integer spin. The standard model knows 24 fermions and 13 gauge bosons. These numbers also contain antiparticles. Antiparticles are duplicates of elementary particles with additive quantum numbers of opposite sign, that build up antimatter. Every form of known matter or antimatter consists of fermions, while gauge bosons describe the different interactions.

2.1 Gauge Bosons

In quantum field theory gauge bosons are introduced by the quantization of fields. Thus, they are quanta of their associated fields and responsible for the different interactions. These are the strong interaction, electromagnetism, the weak interaction and the Higgs mechanism. Gravity cannot be explained within the standard model. Each gauge boson couples to every particle, which carries the corresponding charge. The gauge bosons of the strong interaction are gluons, that couple to every particle with a color-charge. There exist 8 different kinds of gluons each carrying a color and an anticolor, which allows them to interact with each other, too. Electromagnetism is conveyed by the photon. The photon interacts with each particle that holds an electric charge. In the weak interaction the three bosons \( Z^0, W^+, W^- \) only couple to left-handed particles and right-handed antiparticles having a weak isospin charge. The Higgs boson is responsible for the Higgs mechanism. Each elementary particle acquires mass by interacting with the Higgs field. This mechanism is necessary to explain the mass of the weak gauge bosons.

2.2 Fermions

The 24 fermions can be subdivided into 12 leptons, namely six leptons and antileptons and 12 quarks, namely six quarks and antiquarks. Furthermore, quarks and leptons can be divided into three different generations. Although the first generation would suffice for explaining common matter, a second and a third generation have been observed. Each
CHAPTER 2. THE STANDARD MODEL

generation comprises two leptons and two quarks arranged as isospin partners. Particles in the second and third generation are exact copies of the corresponding particles of the first generation, except for their masses.

2.2.1 Leptons

The leptons within the standard model are the electron, the electron neutrino, the muon, the muon neutrino, the tauon, the tauon neutrino and their antiparticles. Table 2.1 gives a short overview about the generations of leptons, their masses and charges. Thus, neutrinos only take part in the weak interaction, while electrons, muons and taus participate in electromagnetism and in the weak interaction if they are left-handed.

Table 2.1: Leptons in the standard model [4].

<table>
<thead>
<tr>
<th>generation</th>
<th>particle</th>
<th>electric charge</th>
<th>weak charge</th>
<th>color charge</th>
<th>mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>neutrino $\nu_e$</td>
<td>-</td>
<td>$+\frac{1}{2}$</td>
<td>-</td>
<td>$&lt;2$ eV</td>
</tr>
<tr>
<td>I.</td>
<td>electron $e$</td>
<td>$-e$</td>
<td>$-\frac{1}{2}$</td>
<td>-</td>
<td>$\approx 511$ keV</td>
</tr>
<tr>
<td>II.</td>
<td>neutrino $\nu_\mu$</td>
<td>-</td>
<td>$+\frac{1}{2}$</td>
<td>-</td>
<td>$&lt;190$ keV</td>
</tr>
<tr>
<td>II.</td>
<td>muon $\mu$</td>
<td>$-e$</td>
<td>$-\frac{1}{2}$</td>
<td>-</td>
<td>$\approx 106$ MeV</td>
</tr>
<tr>
<td>III.</td>
<td>neutrino $\nu_\tau$</td>
<td>-</td>
<td>$+\frac{1}{2}$</td>
<td>-</td>
<td>$&lt;182$ MeV</td>
</tr>
<tr>
<td>III.</td>
<td>tauon $\tau$</td>
<td>$-e$</td>
<td>$-\frac{1}{2}$</td>
<td>-</td>
<td>$\approx 1777$ MeV</td>
</tr>
</tbody>
</table>

2.2.2 Quarks

The 12 quarks within the standard model are the up quark, the down quark, the charm quark, the strange quark, the top quark, the bottom quark and their antiparticles. Table 2.2 gives a short overview about the generations of quarks, their masses and charges. Quarks never exist unbound. Due to confinement they are only realized in color-neutral hadrons. Hence, when quarks are separated, at a certain point the gluon string, which holds them together, collapses and a new quark-antiquark pair comes into being, because it is more energetically favorable. So, at first order, hadrons comprise two possibilities of color-neutral quark combination. Mesons consist of a quark-antiquark pair and are bosons, while baryons are fermions and composed of three quarks or three antiquarks. Left-handed quarks participate in all interactions.

Table 2.2: Quarks in the standard model [4].

<table>
<thead>
<tr>
<th>gen.</th>
<th>particle</th>
<th>el. charge</th>
<th>weak charge</th>
<th>color charge</th>
<th>mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>up quark $u$</td>
<td>$\frac{2}{3}e$</td>
<td>$+\frac{1}{2}$</td>
<td>rgb</td>
<td>$1.7 - 3.3$ MeV</td>
</tr>
<tr>
<td>I.</td>
<td>down quark $d$</td>
<td>$-\frac{1}{3}e$</td>
<td>$-\frac{1}{2}$</td>
<td>rgb</td>
<td>$4.1 - 5.8$ MeV</td>
</tr>
<tr>
<td>II.</td>
<td>charm quark $c$</td>
<td>$\frac{2}{3}e$</td>
<td>$+\frac{1}{2}$</td>
<td>rgb</td>
<td>$\approx 1.27$ GeV</td>
</tr>
<tr>
<td>II.</td>
<td>strange quark $s$</td>
<td>$-\frac{1}{3}e$</td>
<td>$-\frac{1}{2}$</td>
<td>rgb</td>
<td>$\approx 101$ MeV</td>
</tr>
<tr>
<td>III.</td>
<td>top quark $t$</td>
<td>$\frac{2}{3}e$</td>
<td>$+\frac{1}{2}$</td>
<td>rgb</td>
<td>$\approx 172$ GeV</td>
</tr>
<tr>
<td>III.</td>
<td>bottom quark $b$</td>
<td>$-\frac{1}{3}e$</td>
<td>$-\frac{1}{2}$</td>
<td>rgb</td>
<td>$\approx 4.19$ GeV</td>
</tr>
</tbody>
</table>
Chapter 3

The Top Quark at the LHC

The top quark is the heaviest quark ($m_{\text{top}} = 172.0 \pm 0.9 \pm 1.3 \text{ GeV} \,[4]$) in the standard model and was discovered in 1995 at Fermilab [5]. It cannot be found in mesons or baryons, because its lifetime ($\tau_{\text{top}} \approx 3.16 \cdot 10^{-25} \text{ s}$) is too short for hadronization processes ($t_{\text{hadronization}} \approx 10^{-23} \text{ s}$) [4]. The electric charge is $q_{\text{top}} = \frac{2}{3}e$ [4]. The main production process of top quarks at the LHC are top-antitop ($t\bar{t}$) events with a cross section of approximately 165 pb at NLO [6]. The dominant production mechanisms are gluon-gluon fusion ($\approx 80\%$) and quark-antiquark annihilation ($\approx 20\%$) [7]. These values refer to the maximum center-of-mass energy of $\sqrt{s} = 7 \text{ TeV}$. The feynman diagrams of these production processes can be seen in Figure 3.1.

![Feynman diagrams for top-antitop production](image)

Figure 3.1: Feynman diagrams for top-antitop production [8].

Top-antitop processes can be divided into the full hadronic, the dileptonic and the semileptonic channel. The full hadronic decay channel offers the advantage of a high branching
Chapter 3. The Top Quark at the LHC

Figure 3.2: Semi-muonic $t\bar{t}$ decay from $q\bar{q}$ annihilation [9].

ratio (BR 45% [4]) and fully kinematically reconstructable final states. But the problem is that on the one hand a high combinatorical ambiguity occurs, because of six QCD jets (6! possibilities) and on the other hand the measurement of the jet-energy scale leads to high uncertainties. The dileptonic decay channel is characterized by a nice signature because of two high energetic leptons. However, two neutrinos lead to a four-fold ambiguity in reconstruction and moreover the branching ratio is very small (BR 10% [4]). In this analysis a semileptonic, namely the semi-muonic $t\bar{t}$ decay (BR 45% [4]) is used, in order to provide top-antitop events. This decay channel is depicted in Figure 3.2 and offers several advantages. On the one hand it allows a good kinematic reconstruction of the final state objects, because only a single neutrino is responsible for the missing transversal energy. Hence, a complete kinematic reconstruction is possible, even though a two-fold ambiguity remains. On the other hand this process has a unique signature, because a high energetic isolated muon as well as 4 high $p_T$-jets, of which 2 jets look bottom-like, are expected. The muonic channel has been chosen because CMS provides a reliable muon reconstruction and identification.

However, there are processes, which have a very similar signature, so that they can hardly be distinguished. Therefore, they have to be considered in the analysis, too. The most dominant background processes for this analysis are $W$+jets, $Z$+jets, QCD multijet and single top events. The following Figure 3.3 depicts them. A misidentification of these background processes has several reasons. For example hadronic jets can fake muons in the case of the QCD multijet event. Especially the signature of $W$+jets events are very similar to $t\bar{t}$ events. Another possibility is that in single top events a bad jet algorithm reconstructs the two bottom-jets as several QCD jets.

Figure 3.3: Background processes with similar signature [9].
Chapter 4

The CMS Experiment at the LHC

In Chapter 1 the LHC and the CMS experiment have already been introduced. In this chapter a more detailed overview will be given on intention, composition and operation.

4.1 The Large Hadron Collider

The LHC is a high energy particle accelerator located in Geneva, which has been operating since September 2008. It was built by harnessing the tunnel of the previous Large Electron Positron Collider (LEP) making the LHC a circular particle accelerator with a circumference of 26.66 km in total. In contrast to the previous LEP accelerator, heavy ions or protons are collided. This allows to increase the beam energy, while still using the old LEP tunnel due to reduced synchrotron radiation. In this bachelor thesis only the collision of protons will be of interest. The protons have to undergo several acceleration steps before they are finally inserted into the LHC. According to the design values the proton beam is split into 2808 bunches per beam. Each bunch holds approximately $1.15 \times 10^{11}$ protons, which have a velocity of $\beta = 0.9999991$ and thereby an energy of 7 TeV, which results in a center-of-mass energy at the collision point of $\sqrt{s_{\text{max}}} = 14$ TeV, in the case that the LHC is operating at its full design luminosity of $L_{\text{design}} = 10^{34}$ cm$^{-2}$ s$^{-1}$. The LHC has not yet been operating at full design values. This year 2012 the LHC operates at a center-of-mass energy of $\sqrt{s} = 8$ TeV with a peak luminosity of about $L = 6.76 \times 10^{33}$ cm$^{-2}$ s$^{-1}$.

The following Figure 4.1 gives a visual impression of the development of the acquired integrated luminosity in 2011 and 2012.

Currently, the proton beam is split into 1380 bunches per beam. Superconducting magnets ensure that the beams stay on their circular trajectory. In order to make the beam follow its appropriate trajectory, 14232 dipolmagnets with a length of 14.3 m each, at present operating at a magnetic field of $B \approx 4.7$ T, have been installed [11–13].

By studying the decay products of particle collisions several goals are focussed to enhance our understanding of physics. These are amongst others:

- Search for the Higgs Boson$^2$

$^1$Most of the information in this chapter is based on [3], which provides further information.
$^2$LHC experiments recently announced the discovery of a new particle resonance at 125 GeV, that might be associated with the Higgs boson.
• Search for supersymmetric particles

• Search for new massive vector bosons like e.g. $Z'$

• Search for extra dimensions

Hence, several experiments are set up at the tunnel. These are ATLAS, CMS, ALICE, LHCb, LHCf and TOTEM. ATLAS and CMS are general purpose detectors and research on a broad spectrum of physics. ALICE is dedicated to study the conditions of quark-gluon plasma in heavy ion collisions and LHCb looks at the asymmetry of matter and antimatter in our universe by searching for CP violation in the B-sector. TOTEM measures particle scattering at very small angles and proton-proton inelastic scattering. This allows on the one hand to determine the beam luminosity and on the other hand the size of the proton can be investigated very precisely. LHCf explores hadronization and deep inelastic scattering physics in the extreme forward region in order to compare these data with cosmic ray airshowers [14]. The data for this analysis are obtained by the CMS detector.

### 4.2 The CMS Detector

The name of the CMS detector (Compact Muon Solenoid) summarizes its characteristic features. At first, the CMS detector is with a length of 21.6 m, a diameter of 14.6 m and a weight of $12.5 \cdot 10^6$ kg a very compact particle detector. The most significant visual feature are the muon chambers, with which one is able to measure the momenta of muons very precisely. Therefore, a solenoid must produce an extremely strong magnetic field. Figure 4.2 gives a look at the different components inside the CMS detector. The different detector subsystems from the innermost to the outermost around the interaction point (IP) are the pixel tracker, the strip tracker, the electromagnetic calorimeter (ECAL), the hadron calorimeter (HCAL) and the muon system.
4.2. The CMS Detector

4.2.1 The Tracking System

In order to provide a good reconstruction of the momenta of particles, the tracking system builds the innerst detector. Due to the Lorentz force, charged particles cross the detector on a bent path in the B-field provided by the solenoid. Thus, momenta can be calculated. The tracking system is based on semiconducting detectors. Generally, the tracking system can be divided into the pixel tracker and the strip tracker.

The Pixel Tracker

The pixel tracker is built as close as possible to the IP so that it guarantees the best possible resolution. For this purpose three layers of pixel detectors with a length of 53 cm, located at mean radii of 4.4 cm, 7.3 cm and 10.2 cm cover the barrel region ($|\eta| < 1.4$). For the endcap area ($1.4 < |\eta| < 3.0$) two end disks are placed at both ends at a $z$-distance of 34.5 cm and 46.5 cm to the IP. Each disk covers a radius length of 10 cm. They are subdivided and each part is tilted by $20^\circ$ to achieve a better resolution capacity and coverage. The size of one pixel is only about $100 \times 150 \mu m^2$. Thus, a spatial resolution of $10 \mu m$ in the $r - \phi$ plane and $20 \mu m$ in $z$-direction can be obtained. Figure 4.3 gives a visual impression.

Figure 4.2: The composition of the CMS detector [3].
The Strip Tracker

The strip tracker surrounds the pixel tracker. It consists of silicon strips and can be divided into four parts. These are the Tracker Inner Barrel (TIB), the Tracker Outer Barrel (TOB), the Tracker End Cap (TEC) and the Tracker Inner Disks (TID). Four layers of these silicon strips build the TIB and cover a z-range of approximately 130 cm. For the inner two layers stereo modules are used, which comprise of two mono layers twisted by an angle of 5.73°. This principle ensures a good resolution capacity. The other three parts are very similar in their structure. In total the strip tracker consists of 15400 modules. It has a slightly smaller resolution capacity because the particle flux also decreases with distance to the IP.

4.2.2 The Electromagnetic Calorimeter

The main purpose of the electromagnetic calorimeter (ECAL) is to measure the energy deposition of photons and electrons. It is made of 75848 lead tungstate crystals, which use the scintillation effect for particle detection. Thereby, in total (encap and barrel) an η-range of |η| < 3.0 is covered. The lead tungstate material guarantees on the one hand a short radiation length and Moliere radius, and on the other hand it is very resistant to radiation. The energy resolution increases for high particle energies and can be approximated to $\frac{\sigma_E}{E} \propto \frac{1}{\sqrt{E}}$ so that above a particle energy of 25 GeV the relative error is smaller than 1%.

The scintillation effect occurs when high energetic particles cross the crystals so that photons are produced. These photons are detected with the help of avalanche photo diodes or vacuum photo triodes. In both cases the principle of a photo multiplier is used in order to intensify the signal. Arriving photons collide with a photocathode so that electrons are emitted and accelerated towards dynodes emitting even more electrons. This avalanche effect ensures a detectable signal as these electrons hit the anode and thereby produce a current. Electrons and photons crossing the ECAL do not instantly deposit their energy.
They produce electromagnetic showers as they go through it. These showers are an interaction of bremsstrahlung \((e \rightarrow e + \gamma)\) and pairproduction \((\gamma \rightarrow e + e)\). Nevertheless, the showerlength is much smaller than those of hadrons allowing an easier reconstruction. This is the reason, why the ECAL can be very compact in contrast to the HCAL.

### 4.2.3 The Hadron Calorimeter

The HCAL is the last detector, whose main part is located in front of the solenoid\(^3\). In general the task of the hadron calorimeter is to measure the energy of hadronic cascades. Thereby, it is also possible to determine the missing transverse energy \(E_{T}^{\text{miss}}\) in combination with the data of the ECAL. Basically, it operates like the ECAL with the difference that tile/fibre technology is used. This means that the active medium consists of plastic scitillators in combination with wavelength-shifting fibres, which lead photons to the photodetection readout. In order to minimize the dimension of the HCAL the absorber material is composed of brass, which has a short interaction length. Thus, it is a sampling calorimeter. Besides it is non-magnetic, which is of great advantage due to the short distance to the solenoid. In general the HCAL can be divided into four parts. The hadron barrel (HB) covers an \(\eta\)-range of \(|\eta| < 1.4\) and consists of 32 towers. The hadron outer (HO) detector covers an \(\eta\)-range of \(|\eta| < 1.26\), is the only part of the HCAL outside the solenoid and comprises 5 rings. Its main purpose is to catch and detect surviving hadrons for protecting the muon system. Besides it helps to increase the \(E_{T}^{\text{miss}}\) resolution. The hadron endcap is made up of 14 towers covering an \(\eta\)-range of \(1.3 < |\eta| < 3.0\). Finally, for the \(\eta\)-range of \(3 < |\eta| < 5\) the hadron forward (HF) calorimeter, which uses steel/quartz fibre technology, is the last calorimeter included by the HCAL. 900 towers constitute the HF, that uses the Cherenkov effect for creating a signal from particle flux. The error in energy resolution is very similar for all parts of the HCAL\(^{15}\) and is given in equation \((4.1)\).

\[
\left(\frac{\sigma_E}{E}\right)^2 \approx \frac{1}{E(\text{GeV})^2} + 0.045^2
\]

### 4.2.4 The Solenoid

The purpose of the solenoid in CMS is forcing charged particles on a bent path as they cross the detector. Momenta of particles can be calculated according to the radius of their trajectory. In the CMS experiment a superconducting solenoid is used, which produces a magnetic field of \(B = 3.8\) T. This strong magnetic field makes sure that even highly relativistic particles describe a sufficiently circular trajectory. The magnet has a length of 12.9 m and a distance of 5.9 m to the beam pipe. A current of 19.5 kA circulates through 2168 coils, producing a stored energy of 2.7 GJ. It consists of high-purity aluminium-stabilized conductors.

### 4.2.5 The Muon System

Muons have a mass of approximately \(m_{\text{muon}} = 106\) MeV, which means that they hardly deposit any energy as they cross the calorimeters. The muon system is able to gauge the

\(^3\)looking radially outwards
momenta of muons. In connection with the tracking system momenta can be measured with a very small relative error $\frac{\Delta p}{p} < 0.01$. Steel plates are used in the muon system as a return yoke for the solenoid. This provides a magnetic field of approximately 2 T in the muon system and by this muons are bent again, which leads to a better resolution of the transversal momentum. The muon system is based on three types of gaseous detectors. In the barrel region drift tube (DT) chambers are used, because of a small neutron background, a low muon rate and a low magnetic field. In the two endcaps these circumstances do not exist, so that cathode strip chambers (CSC) are installed. In both regions (barrel and endcap) the muon detection is supported by resistive plate chambers (RPC), that guarantee a good time resolution and a fast response. Figure 4.4 depicts the muon system.

250 drift tube chambers are deployed in 4 layers and 5 wheels in the barrel region. Each wheel comprises 12 sectors. Chambers of different sectors are staggered in order to improve the resolution capacity for muons crossing near section borders. The muon chambers of the 3 inner layers (fourth layer) consist of 12 (8) planes of aluminium drift tubes. These drift tubes are gas detectors operating in the proportional region. Thus a single point resolution of approximately 200 $\mu$m can be achieved. The endcap region is made up of 468 cathode strip chambers. CSCs are gas detectors, which determine the particle trajectory by finding out the center-of-gravity of the charge distribution. This is possible because the strips of anode and cathode are perpendicular to each other. Most of the CSCs are built overlapping to completely cover the surface. Both DTs and CSCs are supported by resistive plate chambers, which are mainly used as triggers, because they lack in resolution capacity. Therefore, RPCs are sandwiched between DTs and CSCs.
Chapter 5

Data, Simulation and Selection

The data used in this analysis are LHC-data of 2011 with a center-of-mass energy of \( \sqrt{s} = 7 \text{ TeV} \) and a luminosity of \( \mathcal{L}_{\text{int}} = 4.6 \text{ fb}^{-1} \). In order to analyze these data, simulated events are created. The simulation takes account of the quantum mechanic nature of these processes as well as the response of the CMS detector. By comparing the simulation with data it can be validated and adjusted. Additionally, it provides information of the event composition after selection. The simulated events are official samples provided from CMS. The creation of simulated events consists of several parts. At first MadGraph [16] and the leading order parton distribution function set CTEQ6L1 [17] simulate the interaction process according to Feynman rules, giving a list of incoming and outgoing particles and their four momenta. In order to adequately simulate hadronization and fragmentation processes of the strong interaction, Pythia is used. Furthermore, the MLM matching procedure [18] is applied so that double counting in showers from MadGraph and Pythia are avoided. The detector electronics is simulated with Geant4 [19]. Before the simulated events are used for analysis they are pile-up reweighted to increase the similarity to the provided data.

For the reconstruction of physics objects the particle flow algorithm [20] is employed and for the event selection a CMS top reference selection [21] is used. The main characteristics of this selection are the following.

- **Trigger**: the IsoMu24 trigger requests an isolated unprescaled single muon with a minimum transversal momentum \( p_t > 24 \text{ GeV} \).
- **Vertex**: a well identified primary vertex is required.
- **Muon**: one single muon with \( p_t > 26 \text{ GeV} \).
- **Muon Veto**: veto on low energetic “loose” muons with \( p_t > 10 \text{ GeV} \).
- **Electron Veto**: veto on electrons with \( E_T > 15 \text{ GeV} \).
- **Jets**: not less than four well identified jets with \( p_T > 30 \text{ GeV} \).

In this analysis no b-tag is applied, because the stability of the fitting procedure of the data-driven ansatz is reliant on sufficient statistics.
Chapter 6

Data-Driven Estimation of the Photon Identification

6.1 Photon Radiation of Top Quarks

If the top quark is the weak isospin partner of the bottom quark, its electric charge is predicted by the standard model with \( q_{\text{top}}(SM) = \frac{2}{3}e \). The aim is to provide a data-driven photon identification, which can be used in searches for \( t\bar{t}\gamma \)-events, that allow to prove the electric charge of the top quark.

The R-ratio of the two cross sections (6.1) can be used to estimate the charge.

\[
R(q) = \frac{\sigma_{t\bar{t}\gamma}}{\sigma_{t\bar{t}}} \tag{6.1}
\]

The cross section \( \sigma_{t\bar{t}\gamma} \) denotes top-antitop events, where a photon is radiated off either a top quark or an antitop quark, while the cross section \( \sigma_{t\bar{t}} \) refers to inclusive top-antitop events. Measuring this R-ratio, a connection to the electric charge can be made. The following Figure 6.1 illustrates the difference between both events.

Having a look at Equation (6.2), one can see that the cross section \( \sigma_{t\bar{t}\gamma} \) will be obtained by the number of these events \( N(t\bar{t}\gamma) \). The integrated luminosity \( L_{\text{int}} \) refers to the number

![Figure 6.1: Top-antitop production with and without photon radiation.](image)
of obtained data and $\varepsilon$ is a factor, which includes detector acceptance and efficiency. The problem is that the number of top-antitop events with photon radiation \( N(t\bar{t}\gamma) \) is distorted by fake photons \( N(t\bar{t}\not\gamma) \) \[6.3\]. These are particles, which are identified by the detector as photons, but in reality correspond to other particles with very similar signature. \( N_{t\bar{t}\gamma} \) refers to particles, which have been identified as photons by the detector.

\[
\sigma_{t\bar{t}\gamma} = \frac{N(t\bar{t}\gamma)}{L_{\text{int}} \cdot \varepsilon} \tag{6.2}
\]

\[
N(t\bar{t}\gamma) = N_{t\bar{t}\gamma} - N(t\bar{t}\not\gamma) \tag{6.3}
\]

It is experimentally impossible to determine if the detected photons are radiated off the top quark or another particle. The estimation how many of these fake photons are radiated off top quarks cannot be given within this analysis and needs further investigations. Therefore, \( N(\not\gamma) \), the complete number of fake photons, irrespective of their origin, will be of interest as a contribution to fake photons in top-antitop events. Accordingly, \( N(\gamma) \) is the complete number of real photons. Thus, the main aim of this bachelor thesis is to provide a fake photon rate \( \frac{N(\gamma)}{N(\not\gamma) + N(\gamma)} \), which is a first approach for the correction of charge calculations on the top quark, which are based on photon radiation in top-antitop events in the $\mu$+jet channel. The photon identification is applied on top of a reference $\mu$+jet selection, as described in Chapter [5].

### 6.2 Photon Identification

In this analysis real photons will be considered as “signal” and fake photons as “background”. For a good significance of the cross section analysis the signal needs to be well enhanced with respect to the background while a sufficient number of events needs to remain after the selection. Therefore, selection requirements are applied to simulation and data. The implemented photon requirements are motivated in [22]. The following Table 6.1 gives an overview on the implemented photon identification. It will be introduced and studied, although a detailed and deductive explanation, which motivates the optimum of the implemented requirements cannot be given within the scope of this analysis. The red colored requirements, Hollow Cone Track Isolation and $\eta$-width, will not be applied for the whole analysis, as explained later.

In the following N-1 plots are shown very often to motivate an applied requirement. N-1 plots show simulated events, on which all requirements are applied, except for one. The distribution of this investigated observable is plotted, so that one is able to study the benefits of this single cut, that has not yet been applied. In this way the application of correlated requirements is avoided. By the use of generator information a separation between real and fake photons is possible. In all N-1 plots the $\eta$-width cut has not been applied, because it is correlated to the discriminating variable for the data-driven analysis.

#### Requirement on the Pseudorapidity

The pseudorapidity is restricted to $|\eta| < 1.4442$ for the barrel region and $1.566 < |\eta| < 2.5$ for the endcap region. This requirement excludes the transition between barrel and
Table 6.1: Photon Requirements \cite{22}.

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Condition for Barrel</th>
<th>Condition for Endcap</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>\eta</td>
<td>&lt; 1.4442$</td>
</tr>
<tr>
<td>Track Veto</td>
<td>no pixelseeds</td>
<td>no pixelseeds</td>
</tr>
<tr>
<td>Hadronic over EM</td>
<td>$&lt; 0.05$</td>
<td>$&lt; 0.05$</td>
</tr>
<tr>
<td>Transversal Energy</td>
<td>$&gt; 15$ GeV</td>
<td>$&gt; 15$ GeV</td>
</tr>
<tr>
<td>Jurassic ECAL Isolation</td>
<td>$&lt; 4.2 + 0.006E_{T}^{\gamma}$</td>
<td>$&lt; 4.2 + 0.006E_{T}^{\gamma}$</td>
</tr>
<tr>
<td>Tower-Based HCAL Isolation</td>
<td>$&lt; 2.2 + 0.0025E_{T}^{\gamma}$</td>
<td>$&lt; 2.2 + 0.0025E_{T}^{\gamma}$</td>
</tr>
<tr>
<td>Hollow Cone Track Isolation</td>
<td>$&lt; 2 + 0.001E_{T}^{\gamma}$</td>
<td>$&lt; 2 + 0.001E_{T}^{\gamma}$</td>
</tr>
<tr>
<td>$\eta$-width</td>
<td>$&lt; 0.011$</td>
<td>$&lt; 0.03$</td>
</tr>
</tbody>
</table>

$E_{T}^{\gamma}$ refers to the transversal photon energy in GeV

endcap because of many detector gaps, due to the detector electronics, resulting in a bad detector acceptance. The following Figure 6.2 depicts the $\eta$-distribution of real and fake photons. It shows a steep decline in the excluded areas independent of the type of particle.

![eta_distribution.png](image_url)

\textit{Figure 6.2:} $\eta$-distribution of photons within CMS.

**Track Veto Requirement**

The track veto requirement refers to the pixel tracker. It requests that particles do not leave a trace in the tracking system in order to be reconstructed as a real photon. The reason is that photons are uncharged particles, which are not detected by the tracking system. Therefore, energy depositions in the ECAL are matched to tracks in the pixel detector. Figure 6.3 is a boolean N-1 plot of the photon distribution according to the pixelseeds. The amount of real photons is much higher with an applied pixel seed veto. Due to this cut many electrons and positrons faking photons can be identified. It can be seen that the fraction of real photons in the barrel region are larger than in the endcap region.
Hadronic over EM Requirement

Photons are particles, which mainly deposit their energy in the ECAL. In order to ensure that the main energy deposition takes place in the ECAL, this requirement demands that the ratio between the energy deposited in the HCAL in direction of a photon candidate in the ECAL is very small (Hadronic/EM < 0.05). The N-1 plot (Figure 6.4) gives a visual impression. In both plots the y-axis is logarithmic. In this case the requirement cannot be validated completely, because most real photons show an energy deposition ratio, which is much smaller than the requirement. Though, the trend can be affirmed. A possible reason is that the requirements are optimized for a $W\gamma$ and $Z\gamma$ analysis and not for a top pair analysis.

Figure 6.3: N-1 plots of the Pixel Seed Requirement.

0 corresponds to electromagnetic-calorimetric energy depositions, that do not match to pixel tracks, while 1 corresponds to energy depositions, that do.

Figure 6.4: N-1 plots of the ratio of energy deposition in the HCAL and ECAL.
Transversal Energy Requirement

The transversal energy $E_T$ is the energy of a particle projected onto the $\phi$-direction of the detector. In order to increase the amount of real photons, as well as the proportion of photons stemming from the top quark, which have a higher transversal momentum, only photons with a transversal energy of $E_T > 15$ GeV are accepted. Additionally, this requirement takes account of the detector acceptance, which is very bad for low energetic particles. Figure 6.5 shows N-1 plots of the transversal energy in the barrel and in the endcap region. In both regions one observes a validation of the requirement, due to the fact that the proportion of real photons from the hard process increases with higher transversal energies.

Jurassic ECAL Isolation Requirement

The jurassic ECAL isolation requirement (JECAL-Iso < 4.2 + 0.006$E_T^\gamma$) sums up the transversal energy, which is deposited in the ECAL in an annulus of $0.06 < R < 0.40$, excluding a rectangular strip of $\Delta \eta \times \Delta \phi = 0.04 \times 0.40$. Real photons will deposit their main energy in the rectangular strip with a high extent in the $\phi$-direction, because they undergo pairproduction as well as bremsstrahlung processes as they cross the ECAL. Thereby, electrons and positrons are bent by the magnetic field along the $\phi$-direction. The following N-1 plots (Figure 6.6) depict the distribution of this requirement variable. In both regions the benefits of this requirement are confirmed, albeit in the endcap region the requirement is not optimal for this analysis.

Tower-Based HCAL Isolation

This requirement requests, that real photon candidates only deposit a small amount of energy in the HCAL tower located in direction of the corresponding signal in the ECAL (TBHCAL-Iso < 2.2 + 0.0025$E_T^\gamma$). This HCAL tower excludes the energy used for the hadronic over EM requirement. In the N-1 plot (Figure 6.7) one can see, that in the barrel region nearly all events are located in the first bin so that the $y$-axis is logarithmic. The
Hollow Cone Track Isolation Requirement

The hollow cone track isolation requirement only accepts real photon candidates in the ECAL, if the momenta detected around matched tracks in the tracking system are very low. Therefore, a hollow cone is built around the track of a real photon candidate. Real photon candidates with a HCT-Isolation $< 2 + 0.001 E_{T}^{\gamma}$ are accepted. The N-1 plot (Figure 6.8) of the barrel region confirms this requirement, because the rate of real photons increases with this requirement. In the endcap region, one cannot clearly identify this trend, but regarding absolute numbers of real and fake photons this cut aims at the barrel region.
6.2. Photon Identification

Eta-Width Requirement

The eta-width requirement refers to energy depositions in the ECAL and gives a limit on the $\eta$-width for barrel ($\eta$-width($EB$) < 0.011) and endcap ($\eta$-width($EE$) < 0.030). The N-1 plots (Figure 6.9) show this distribution for barrel and endcap. This requirement is very effective in increasing the real photon rate. Nevertheless, it is a very restrictive requirement and reduces the passing rate drastically. Furthermore, a correlation to the discriminating variable $\sigma_{\eta_{\eta}}$ (cf. Section 6.4.1 for further information), which is very similar to the $\eta$-width variable, is implied by comparing the $\eta$-width plots (Figure 6.10) with the plot of the discriminating variable (Figure 6.11). On these grounds the $\eta$-width requirement is not applied in the scope of this analysis.

Figure 6.8: N-1 plots of the Hollow Cone Track Isolation.

Figure 6.9: N-1 plots of the $\eta$-width.
The successive implementation of the previously discussed requirements is referred to as cutflow in the following. The cutflow (Figure 6.12) illustrates the development of the composition of real and fake photons in simulation. One can observe that every requirement increases the photon purity, except for the transversal energy requirement in the barrel region. A possible explanation is, that fake photons with a high transversal energy are sorted out by later applied requirements. The implemented photon identification requirements can successfully enrich the initially fake photon dominated sample with real photons. Fake photons are reduced by a factor of 1000 in the barrel and by a factor of 100 in the endcap region. The hollow cone track isolation requirement will not be included in this analysis. Therefore, it is the last applied requirement in the cutflow, so that they are still of use, even without this requirement.

It is also of interest to compare simulation and data. In Figure 6.13 the simulation has been scaled to the data luminosity. Without photon requirements one observes a simulation overshoot in both regions, but especially in the barrel. The problem with simulation is, that they are based on perturbation theory, which is especially for QCD events very
6.3. Fake Photon Rate from Simulated Events

At first, the photon efficiency is calculated on simulated events, so that it can be compared with the result estimated from data in the following Section 6.4. Accordingly, both results are obtained after the same selection requirements discussed in Section 6.2. With the help of simulated events a fake photon rate \( \hat{R}(\hat{\gamma}) \) can be determined using the number of real photons \( N(\gamma) \) and fake photons \( N(\hat{\gamma}) \) after the application of requirements. Due to complicated and might not model all distributions exact enough. Furthermore, no b-tag is used and errors on data luminosity can have a significant impact. All this might be an explanation for the simulation overshoot. This simulation overshoot turns into an undershoot, when more and more requirements are applied. This indicates that the simulated events do not precisely represent reality, which is even more a motivation to estimate the fake photon rate data-driven.

![Diagram](image1)

\( Figure \ 6.12: \) Cutflows (logarithmic) of real and fake photons in simulation.

![Diagram](image2)

\( Figure \ 6.13: \) Cutflows (logarithmic) of simulation and data.
luminosity scaling the numbers are arbitrary. The errors on the number of real and fake photons originate from the cutflow and have been scaled to data luminosity. The following Table 6.2 summarizes the results. The error on the fake photon rate is calculated with Gaussian error propagation.

Table 6.2: Number of photons and fake photon rate for simulated events.

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>real photons $N(\gamma)$</td>
<td></td>
<td>643.93 ± 20.43</td>
<td></td>
</tr>
<tr>
<td>fake photons $N(\not{\gamma})$</td>
<td></td>
<td>351.54 ± 16.59</td>
<td></td>
</tr>
<tr>
<td>fake photon rate $R(\not{\gamma}) = \frac{N(\not{\gamma})}{N(\gamma) + N(\not{\gamma})}$</td>
<td></td>
<td>0.35 ± 0.02</td>
<td></td>
</tr>
</tbody>
</table>

6.4 Fake Photon Rate with the Data-Driven Ansatz

In contrast to a fake photon rate, which is solely based on simulated events, a data-driven ansatz provides the advantage, that it involves collision data for obtaining the result, so that it better regards reality. The idea behind the data-driven ansatz is to find a discriminating variable for real photons and fake photons. Using simulated events, two different shapes are obtained in this variable, namely one shape for real photons and one shape for fake photons. The same is done for the collision data, with the difference that a distinction between real and fake photons is impossible, so that only one shape is obtained. The aim is to reconstruct the shape of the collision data as good as possible by scaling both shapes obtained from the simulation with free parameters and adding them together. Thus, a data-driven fake photon rate can be calculated.

6.4.1 The Discriminating Variable $\sigma_{\eta \eta}$

The discriminating variable in this analysis is $\sigma_{\eta \eta}$, which is defined in Equation (6.4). It is a shower shape variable, which measures the width of the photon supercluster in the $\eta$-direction [22].

$$
\sigma_{\eta \eta} = \sqrt{\frac{\sum (\eta_i - \bar{\eta})^2 w_i}{\sum w_i}}, \quad \bar{\eta} = \frac{\sum \eta_i w_i}{\sum w_i}, \quad w_i = \max \left( 0, 4.7 + \log \left( \frac{E_i}{E_{5 \times 5}} \right) \right)
$$

(6.4)

At first the center $\bar{\eta}$ of the energy deposition is determined. Then a weighted distance in $\eta$ is calculated for every crystal in the $5 \times 5$ crystal matrix. The weight for every crystal is based on the relative energy deposition in it. Thus, this variable is able to discriminate real photons and fake photons, because fake photons tend to deposit their energy on a larger area. This is illustrated by the following Figure 6.14, which juxtaposes real and fake photons from simulated events in the discriminating variable $\sigma_{\eta \eta}$. For these shapes no requirements are applied in order to provide sufficient statistics to visualize the fundamental difference in the shower shape variable between real and fake photons. Obviously, $\sigma_{\eta \eta}$ is very effective in discrimination.
6.4. FAKE PHOTON RATE WITH THE DATA-DRIVEN ANSATZ

\[ f(\sigma_{\eta_i\eta}) = N_S g(\sigma_{\eta_i\eta}) + N_B g(\sigma_{\eta_i\eta}) \]  

\[ (6.5) \]

Figure 6.14: Photon shapes in $\sigma_{\eta_i\eta}$ without any requirement.

Figure 6.15: $\sigma_{\eta_i\eta}$ for various processes.

6.4.2 Independence of the Composition of Different Processes

The composition of top-antitop, QCD multijet, W+jets, Z+jets and single top events can have a significant influence on the distribution in $\sigma_{\eta_i\eta}$. In order to check if the shape depends on this composition, each process, except for the QCD multijet process, has been plotted normalized in the discriminating variable, using simulated events. The QCD multijet process has been omitted, because the events of this process are nearly completely rejected by the requirements of the photon identification. The individual results have been laid on top of each other (Figure 6.15).

A nearly identical distribution in $\sigma_{\eta_i\eta}$ can be observed. Thus, the fit is considered to be independent of the composition of different processes.

6.4.3 The Template Fit Method

The idea is to use both shapes in $\sigma_{\eta_i\eta}$ as templates and fit these to the data. Equation (6.5) shows the function, which is fitted to the data.

\[ f(\sigma_{\eta_i\eta}) = N_S g(\sigma_{\eta_i\eta}) + N_B g(\sigma_{\eta_i\eta}) \]  

\[ (6.5) \]
\( S(\sigma_{\eta\eta}) \) denotes the signal shape, namely real photons, and \( B(\sigma_{\eta\eta}) \) is the background shape, namely fake photons. \( N_{Sg} \) and \( N_{Bg} \) are factors, with which every bin of the corresponding shape is scaled. They are obtained by the fit as free parameters. The shapes are fitted to the data via an extended Poisson binned likelihood fit. The fitting procedure is carried out by ROOT. Let \( p_i(x_i, f(\sigma_{\eta\eta})) \) be the probability that an event appears in the bin \( i \), under a model, which is dependent on the parameters \( N_{Sg} \) and \( N_{Bg} \). The Likelihood fit maximizes the Likelihood-function, which is given by Equation (6.6). \( m \) is the number of bins, \( k_i \) the number of events in bin \( i \) and \( \nu_i = \nu p_i \) the expectation value for the \( i \)-th bin content. The number of events \( \nu \) is a free parameter.

\[
L = \prod_{i=1}^{m} e^{-\nu_i} \frac{\nu_i^{k_i}}{k_i!}
\]

(6.6)

In the first attempt all requirements except for the \( \eta \)-width are applied on the shapes of real photons and fake photons from simulation and data. In order to improve the fit all photon shapes have been rebinned before fitting in order to avoid bins with very little or without any content. This is helpful for minimizing the statistic error of the fit and makes the algorithm more stable and reliable. Having a look, at the photon shapes from simulation (Figure 6.16) and at the fit (Figure 6.17) one can notice that a large systematic error occurs due to the photon statistics in the shapes. Especially the number of fake photons is too small to guarantee a stable fit. This is illustrated in the penultimate bin, where the data to simulation ratio is extremely bad, because for this bin nearly no fake photons are available for fitting.

![Rebinned photon shapes](a) real photons (b) fake photons)

**Figure 6.16:** Rebinned photon shapes in \( \sigma_{\eta\eta} \) with all requirements except for \( \eta \)-width.

A possible solution to improve the fit and investigate the performance of the method is to leave out requirements. The arising disadvantage is, that the proportion of signal de-
6.4. **Fake Photon Rate with the Data-Driven Ansatz**

Increases and this might lead to an increase in the error on possible charge calculations for the top quark. On the other hand the advantage of omitting requirements is, that the number of real photons and mainly fake photons increases. Thus, the bin contents of the photon shapes increase. This results in less systematic uncertainties due to the fitting procedure. Since this uncertainty is dominating the result a decrease of the signal to background ratio is acceptable and the hollow cone track isolation requirement has been omitted. Figure 6.18 and Figure 6.19 depict the shapes from simulation and the fit to data without the \( \eta \)-width and hollow cone track isolation requirement.

**Figure 6.17:** Likelihood Fit of photon shapes to data with all requirements except for \( \eta \)-width.

The quality of the fit is better than before. The data to simulation ratio improves a lot. In about half of all bins an agreement between data and prediction is established. This is still
below the expected 68%, but the exclusion of more requirements would make the result useless. A possible reason for the unsatisfying fit might be the incorrect modelling of the shapes, like it is intended in the cutflow (Figure 6.13), and still the lack of statistics. This leads to relatively large bin errors.

The results from this fit are used for further analysis. The following Table 6.3 summarizes them. \( N(\gamma) \) and \( N(f_i) \) are the numbers of real and fake photons in the shapes of \( \sigma_{\eta \eta} \) of simulated events. The errors are Poissonian. Hence, a data-driven fake photon rate 
\[
R_{dd}(f_i) = \frac{N_{Bg} \cdot N(f_i)}{N_{Bg} \cdot N(\gamma) + N_{Sg} \cdot N(\gamma)}
\]
can be calculated. The error on \( R_{dd}(f_i) \) is also obtained by Gaussian error propagation.

Table 6.3: Fit Results.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( N(\gamma) )</td>
<td>643.93 ± 20.43</td>
</tr>
<tr>
<td>( N(f_i) )</td>
<td>351.54 ± 16.59</td>
</tr>
<tr>
<td>( N_{Sg} )</td>
<td>1.420 ± 0.067</td>
</tr>
<tr>
<td>( N_{Bg} )</td>
<td>1.064 ± 0.103</td>
</tr>
<tr>
<td>( R_{dd} )</td>
<td>0.29 ± 0.03</td>
</tr>
</tbody>
</table>

Details on the fit procedure can be found in the bachelor thesis of Christian Wichary [23], as well as an additional in-depth discussion on separating the fit procedure into the barrel and the endcap region. As expected, photons in the endcap and the barrel show a very different distribution in \( \sigma_{\eta \eta} \), probably caused by different angles to the detector system. Also the fake photon rates for barrel and endcap slightly differ, which might contribute to the deterioration of the quality of the data-driven ansatz. However, this separation further reduces the available statistics for a single fit, which is why reliance upon a separated fit seems inappropriate.

6.4.4 Estimation of Systematic Uncertainties: Ensemble Study

In order to investigate systematic uncertainties, which are not regarded in the fit procedure, an ensemble study is performed. Hence, pseudodata are created by adding the
simulated real photon shape, scaled with the correspondent fit parameter, to the simulated fake photon shape, which is multiplied by a factor $A$ before. The factor is modelled between $A = 0.5 - 1.5$ in steps of 0.1, in order to include the fit parameter of the fake photon shape. Every bin content of every set of pseudodata is then changed randomly with a Gaussian distribution and afterwards with a Poisson distribution. The Gaussian includes systematic uncertainties and the Poisson distribution simulates statistical errors. Then the real and fake photon shapes from simulation are fitted to this set of pseudodata. This is repeated 6000 times for every single shape of pseudodata. Thus, 6000 different fake photon rates are created. A Gaussian distribution is fitted to them and the mean together with the error is used as a confidence interval for a Neyman plot. One Gaussian fit is shown as an example in Figure 6.20.

In the Neyman plot, the $y$-axis represents the fitted fake rate, the $x$-axis shows the true fake rate. The final fit to real data is projected with its errors onto the $x$-axis, which then indicates the real value with the correct systematic error. Figure 6.21 gives a schematic visualization of the procedure and in Figure 6.22 the result is presented. A straight line with a slope of 1 indicates that the real and the measured value show no systematic shift.

**Figure 6.20:** Example Gaussian distribution.

**Figure 6.21:** Scheme of the ensemble study.
The final result is a data-driven fake photon rate with a systematic and statistic error.

\[ R_{dd} = 0.29 \pm 0.06 \]

Comparing this result with the fake photon rate from simulated events, it might be an indication that the simulation slightly underestimates the content of real photons, even though both results are consistent within their errors.

### 6.4.5 Sources of Fake Photons

In order to have a look at particles, which fake photons, fake photons are matched to corresponding particles at generator level. This means that they are compared with the information of generated events for identifying the real particle. A detailed description of the matching algorithm can be found in [23]. The result is depicted in Figure 6.23. Mostly charged pions (\(\pi^\pm\)) and a few electrons and muons fake photons. Additionally, an irreducible background of \(\pi^0\)-induced photons has been found in the real photon content. This can be explained by the extremely high branching ratio of neutral pions for the decay into two photons (BR(\(\pi^0 \rightarrow 2\gamma\)) \(\approx 0.99\) [4]).

![Figure 6.23: Sources of fake photons.](image-url)
In this bachelor thesis a photon identification is implemented and its performance is investigated. An attempt at providing a data-driven photon identification is made. For this purpose, a semileptonic reference \( \mu + \text{jet} \) selection of top-antitop events and similar background events of the CMS collision data of 2011 at a center-of-mass energy of \( \sqrt{s} = 7 \text{ TeV} \) and an integrated luminosity of \( \mathcal{L}_{\text{int}} = 4.6 \text{ fb}^{-1} \) is used. In order to increase the percentage of real photons, several requirements are applied. The benefits of these requirements are affirmed with the help of N-1 plots and in most cases the explicit magnitude can be verified. For investigating the development of simulation and data under the successive application of requirements, the cutflow has been studied. Furthermore, a comparison between simulated events and data is depicted in the cutflow, which shows an overshoot of simulation. This overshoot turns into an undershoot when requirements are applied. Thus, the simulated events do not model reality adequately. A fit of the two photon shapes to data can be improved if more statistics is provided for simulation and data. An opportunity would be to use LHC data of 2012, which offer a higher luminosity. A data-driven fake photon rate has been determined \((R_{\text{dd}} = 0.29^{+0.06}_{-0.06})\) and is consistent with the fake photon rate based on simulated events. The statistic and especially the systematic error have been regarded with the help of an ensemble study. The dependency of the fit result on the composition of different processes has been excluded. The identification of fake photons by matching to other particles makes clear that the dominant source of fake photons are charged pions, while there is still an irreducible background of \( \pi^0 \)-induced photons in the real photon content. For future charge calculations on the top quark it is important to reduce this background. All in all the photon identification is very good at discrimination and together with the data-driven fake photon rate it is expected that this result can be used in future CMS top photon vertex analyses.
Bibliography


[22] C. M. Kuo, Y. Maravin. Study of $W\gamma$ and $Z\gamma$ production at CMS with $\sqrt{s} = 7$ TeV. (2012).


Acknowledgments

At first, I would like to thank Prof. Dr. Achim Stahl and Priv.-Doz. Dr. Oliver Pooth for the opportunity to write this bachelor thesis in the Institut III B.

I would like to thank the top group, Heiko Geenen, Yvonne Küssel and Heiner Tholen, for everyday support and advice.

Furthermore, I would like to thank Christian Wichary for the great cooperation.

I would like to thank Florian Lenz and Till Arndt for their technical knowledge and the pleasant working atmosphere.

Lastly, I would like to thank my family and my roommates.