

**Search for Supersymmetry in opposite-sign  
same-flavour dilepton events with the CMS  
detector in proton-proton collisions at**

$$\sqrt{s} = 8 \text{ TeV}$$

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von  
**Jan-Frederik Schulte, M.Sc.**  
aus Münster

Berichter:  
**Prof. Dr. Lutz Feld**  
**Prof. Dr. Michael Krämer**

Termin der mündlichen Prüfung: xx.xx.2015

Diese Dissertation ist auf den Internetseiten der Hochschulbibliothek online verfügbar.



## **Zusammenfassung**

### **Abstract**



---

# Contents

Zusammenfassung . . . . .	i
Abstract . . . . .	i
<b>1. Introduction</b>	<b>1</b>
<b>2. The Standard Model and its extension to Supersymmetry</b>	<b>3</b>
2.1. The Standard Model of particle physics . . . . .	3
2.2. Motivation for extending the Standard Model and Supersymmetry . . . . .	3
2.3. Production of lepton pairs in supersymmetric models . . . . .	3
2.4. Kinematic edges in the dilepton invariant mass spectrum . . . . .	3
<b>3. Experimental setup</b>	<b>5</b>
3.1. The CERN Large Hadron Collider . . . . .	5
3.2. The CMS detector . . . . .	6
3.2.1. The tracking system . . . . .	7
3.2.2. The electromagnetic calorimeter . . . . .	8
3.2.3. The hadron calorimeter . . . . .	10
3.2.4. The muon system . . . . .	12
3.3. Trigger and data acquisition . . . . .	14
<b>4. Data analysis and event selection</b>	<b>17</b>
4.1. Object reconstruction . . . . .	17
4.1.1. Muon reconstruction and selection . . . . .	17
4.1.2. Electron reconstruction and selection . . . . .	18
4.1.3. Observables reconstructed with Particle Flow . . . . .	20
4.2. Event processing and datasets . . . . .	25
4.2.1. Primary datasets . . . . .	25
4.2.2. Simulated datasets . . . . .	26
4.3. Event selection . . . . .	29
4.3.1. Event cleaning . . . . .	29
4.3.2. Inclusive dilepton selection . . . . .	31
4.3.3. Selections in $E_T^{\text{miss}}$ and jet multiplicity . . . . .	33
<b>5. Estimation of Standard Model backgrounds</b>	<b>35</b>
5.1. Flavour-symmetric backgrounds . . . . .	35
5.1.1. Direct measurement of $R_{\text{SF/OF}}$ . . . . .	36
5.1.2. Determination of $R_{\text{SF/OF}}$ with the factorization method . . . . .	36
5.1.3. Combined correction factors and resulting background estimates . . . . .	44
5.2. Backgrounds containing a Z boson . . . . .	46
5.2.1. JZB method . . . . .	48
5.2.2. $E_T^{\text{miss}}$ templates method . . . . .	48
5.2.3. Extrapolation for off-Z regions . . . . .	48

5.2.4. Resulting background prediction . . . . .	49
5.3. Investigation of possible further backgrounds . . . . .	49
5.4. Search for a kinematic edge with a fit . . . . .	49
<b>6. Results</b>	<b>53</b>
6.1. Result of the counting experiment . . . . .	53
6.2. Result of the search for a kinematic edge . . . . .	54
<b>7. Interpretation in simplified models</b>	<b>57</b>
7.1. Simplified Models for Supersymmetric Signatures . . . . .	57
7.2. The T6bbledge and T6bbslepton model . . . . .	57
7.3. Interpretation of the counting experiment in simplified models . . . . .	57
<b>8. Outlook to LHC Run II</b>	<b>59</b>
<b>9. Conclusion</b>	<b>61</b>
<b>A. Data and Monte Carlo Samples</b>	<b>63</b>
<b>B. Dependencies of <math>r_{\mu e}</math></b>	<b>65</b>
<b>C. Dependencies of <math>R_T</math></b>	<b>69</b>

---

# **1. Introduction**



---

## **2. The Standard Model and its extension to Supersymmetry**

**2.1. The Standard Model of particle physics**

**2.2. Motivation for extending the Standard Model and Supersymmetry**

**2.3. Production of lepton pairs in supersymmetric models**

**2.4. Kinematic edges in the dilepton invariant mass spectrum**



---

## 3. Experimental setup

### 3.1. The CERN Large Hadron Collider

The Large Hadron Collider (LHC) [1], located at CERN near Geneva and stretching far into the french countryside, is capable of colliding protons and lead ions at higher energies than any of its predecessors. Also the instantaneous luminosity delivered to the experiments exceeds that of any previous machine at the energy frontier. It was constructed in the tunnel formerly inhabited by the LEP electron-positron collider in 100 m depths below the surface with a circumference 27 km. It was designed to collide protons at a centre-of-mass energy of  $\sqrt{s} = 14 \text{ TeV}$  with instantaneous luminosities of  $10^{34} \text{ s}^{-1} \text{ m}^{-2}$ .

The LHC consists of eight arcs, as shown in Figure 3.1, where superconducting dipole magnets are used to provide a magnetic field of up to 8.3 T at the highest planned energies to bend the charged particles along the curvature of the tunnel, while quadrupole and other specialised magnets are used to focus the beams. In straight segments between these arcs, LHC infrastructure and the experiments are located. The infrastructure components include the cooling facilities necessary to reach a temperature of 1.9 K around the ring, the superconducting cavities in which the protons are accelerated by standing electromagnetic waves, collimators for beam cleaning and the beam dump, where the beam is ejected from the LHC at the end of fills. In the other four straight segments the beam is brought into collisions, which are studied by the four large experiments at the LHC. Of these, CMS [2] and ATLAS [3] are multi-purpose detectors with a diverse physics program, while ALICE [3] and LHCb [4] are more narrowly focused on heavy ion collisions and flavour physics, respectively.

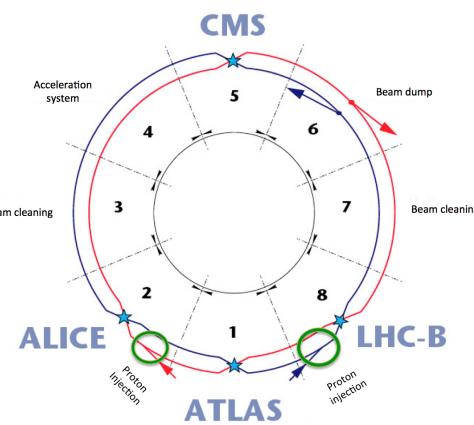


Figure 3.1.: Schematic view of the LHC with its eight arcs. The four interaction points, where the experiments are located, are marked with blue stars. Other important parts of the LHC infrastructure are also indicated [5].

The protons circulating the LHC are injected at an energy of 450 GeV after running through a chain of pre-accelerators, the Linac2, the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS). The proton beams are separated into bunches of about  $10^{11}$  particles. Though being 25 ns by design, corresponding to up to 2808 bunches in the LHC, the smallest spacing between bunches in time was 50 ns throughout the running in 2012 and most of 2011. In these running conditions, after three years of running in the years 2010 to 2012, constituting the so called Run I of the LHC, a centre-of-mass energy of  $\sqrt{s} = 8$  TeV has been reached. The instantaneous luminosities delivered to the experiments reached a maximum of  $7.7 \cdot 10^{33} \text{ s}^{-1} \text{ m}^{-2}$  in late 2012, as can be seen on the left side of Figure 3.2. The integrated luminosity delivered to the CMS experiments in 2012 was  $23.3 \text{ fb}^{-1}$ , exceeding that of 2011 by almost a factor of four [6], as shown on the right side of Figure 3.2.

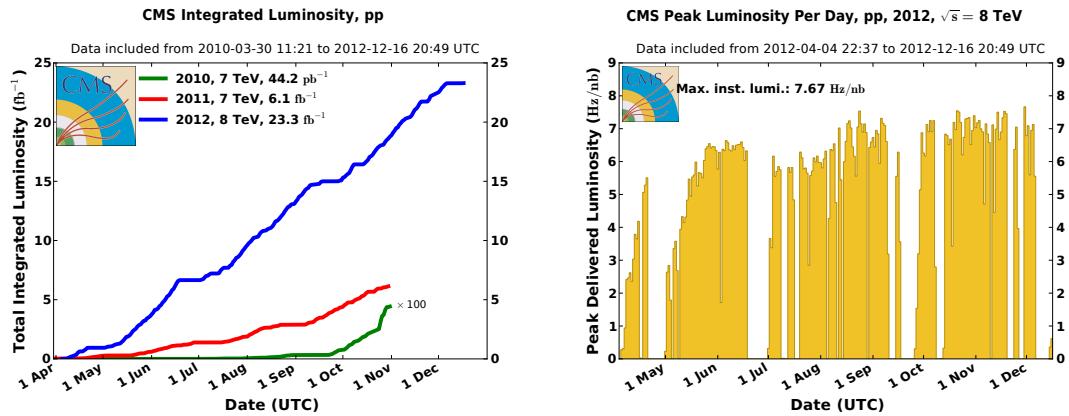


Figure 3.2.: Development of instantaneous (left) and integrated (right) luminosity delivered to the CMS experiment. The left plots shows the results for all three years of data taking, while the right one only shows the 2012 data taking.

## 3.2. The CMS detector

Located in one of the four intersections of the LHC beams, the CMS detector is designed to measure the resulting collisions to high precision. Key ingredients are a high precision measurements of the properties of single particles as well as a good coverage of the  $4\pi$  solid angle. The central element of the CMS detector is a superconducting solenoid. Cooled to 4.45 K, it is able to produce a homogeneous magnetic field of 3.8 T, which allows to measure the momentum of charged particle by bending their trajectories. As shown in Figure 3.3, the different components of the detector are layered in cylindrical shapes around the interaction point. The magnet encompasses most of the main subdetectors, namely the tracking system which measures the trajectories of charged particles and the electromagnetic and hadron calorimeters, designed to measure the energy of particles. Located outside of the volume of the solenoid are the iron return yoke and muon detectors. This cylindrical structure is complemented on both sides by endcaps, which close the solid angle in the direction of the beams and are partly located outside the volume of the solenoid. The different components are described in more detail in the following.

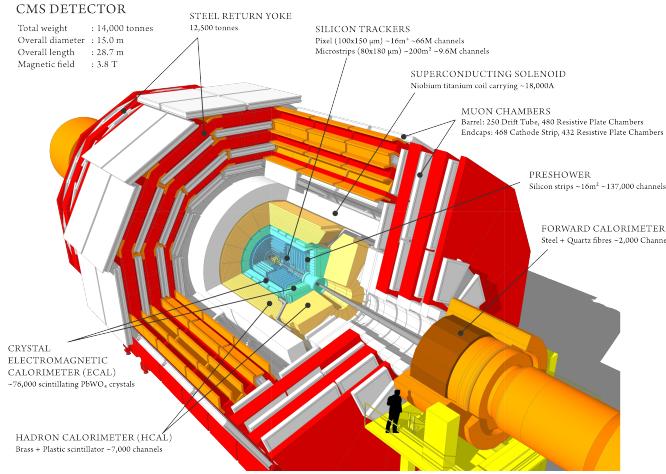


Figure 3.3.: Schematic view of the CMS detector [7]. From the inside out, the tracking system is shown in blue, the electromagnetic calorimeter in green, the hadron calorimeter in light yellow, the superconducting solenoid in white, the return yoke in red and the muon system again in white.

### 3.2.1. The tracking system

The trajectory of charged particles can be determined by measuring the signal of the ionization they cause when traversing matter. The tracking system of the CMS detector consists of many layers of silicon pixels and strips. Combining the points at which a charged particle traverses the different layers, the trajectory of this particle can be measured. The bending of this trajectory under the influence of the magnetic field allows to determine the momentum of the particle. The tracking system has a diameter of 2.5 m and a length of 5.8 m, corresponding to a geometric coverage of  $|\eta| < 2.5$ . The tracking detector consists, as shown in Figure 3.4, of the pixel detector (PIXEL) surrounded by various components of the silicon strip tracker.

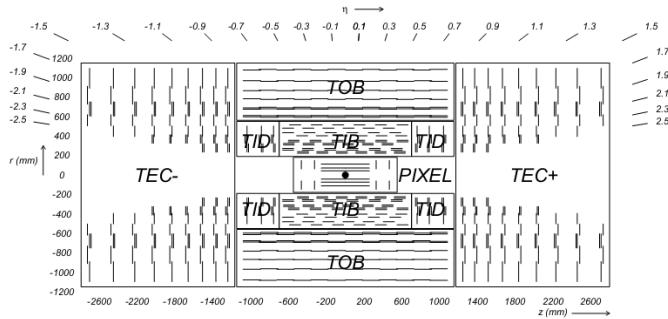


Figure 3.4.: Schematic view of the CMS tracking detector. The innermost part shows the pixel detector (PIXEL), surrounded by the tracker inner barrel (TIB) and tracker inner discs (TID). The outermost parts of the tracking detector are the tracker outer barrel (TOB) and the two tracker endcaps (TEC+ and TEC-).

### The silicon pixel detector

The innermost part of the tracking system in the pixel detector, which consists of three layers in the barrel region at radii between 4.4 cm and 10.2 cm, complemented by two discs perpendicular to the beam axis, located at  $|z| = 34.5$  cm and  $|z| = 46.5$  cm. As the particle density is highest close to the interaction point, a high granularity is needed to maintain a low occupancy of the pixel detector. Therefore the pixel detector consists of roughly 66 million pixels with a combined active area of about  $1\text{ m}^2$ . Each pixel has a size of  $150 \times 100 \mu\text{m}^2$ . The analogue readout of the pixels allows to combine the measurements of neighbouring pixels, bringing the spatial resolution down to 15 to 20  $\mu\text{m}$ . This is especially important for the reconstruction of the interaction vertices and the tagging of the secondary vertices from the decay of b-hadrons.

### The silicon strip detector

Further away from the interaction point, between 20 cm and 116 cm, the granularity of the tracking system is reduced. Silicon strip detectors are used, structured in four layers of the tracker inner barrel (TIB), complemented on each side with three discs of the tracker inner discs (TID). All this is surrounded by the six layers of the tracker outer barrel (TOB). The tracker endcaps (TECs) consist of nine discs each. The individual strips have a length of about 10 cm and a pitch between 80  $\mu\text{m}$  in the two inner layers of the TIB and 183  $\mu\text{m}$  in the four inner layers of the TOB. The single point resolution in TIB and TOB depends on the layout of the specific layer and varies between 23  $\mu\text{m}$  and 53  $\mu\text{m}$ .

Stereo modules, constructed by placing two modules back to back, rotated by 100 mrad, are placed in the first two layers of both TIB and TOB, the first two discs of TID and the first two and the fifth discs of the TECs. These allow for 2-D measurements, with a precision of the  $z$  position measurement of 230  $\mu\text{m}$  in TIB and 530  $\mu\text{m}$  in TOB.

For high momentum tracks of about 100 GeV in the region of  $|\eta| < 1.6$  a  $p_{\text{T}}$  resolution of 1-2% is achieved, while the impact parameter of these tracks can be measured with a resolution of about 10  $\mu\text{m}$ .

Compared to other tracking technologies, an all silicon tracking system as used in CMS consists of significantly more material. The material budget lies between 0.4 and 1.8 radiation length  $X_0$ , as shown in Figure 3.5. For light charged particles such as electrons this leads to a significant probability to emit bremsstrahlung in while traversing the tracking detector, which has to be taken into account in the reconstruction of particles.

#### 3.2.2. The electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) measures the energy of electrons and photons. It uses lead tungstate ( $\text{PbWO}_4$ ) crystals both as absorber and active material. The electromagnetic shower induced by the electron or photon leads to the emission of scintillation light in the crystal, which is measured at the end of the crystals by avalanche photo diodes (APDs) in

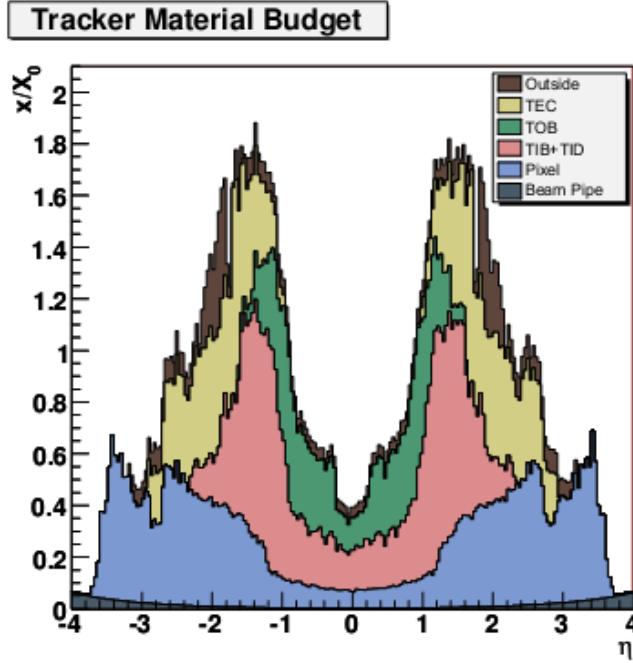


Figure 3.5.: Material budget of the CMS tracking detector in units of radiation length  $X_0$  as a function of  $\eta$ .

the barrel segment of the ECAL and more radiation hard vacuum photo triodes (VTPs) in the endcap region. The choice of lead tungstate was driven by the need for a material that is at the same time dense ( $8.28 \text{ g/cm}^3$ ), has a small Molière radius (2.2 cm) and is fast. About 80% of the scintillation light is emitted after 25 ns, which is the time between two LHC bunch crossing under design conditions. The structure of the ECAL is shown in Figure 3.6. The ECAL barrel (EB) covers the region of  $|\eta| < 1.479$  and consists of 61200 crystals. They have a size of  $2.2 \times 2.2 \text{ cm}^2$  at the front and  $2.6 \times 2.6 \text{ cm}^2$  at the back, with a length of 23 cm, corresponding to  $25.8 X_0$ . In the ECAL endcaps (EE), consisting of 7324 crystal each, they are slightly larger ( $2.862 \times 2.862 \text{ cm}^2$  to  $3.0 \times 3.0 \text{ cm}^2$ ) and shorter (22 cm, corresponding to  $24.7 X_0$ ). The EEs extend the geometric coverage of the ECAL to  $|\eta| = 3.0$ .

In the region of  $1.653 < |\eta| < 2.6$  a preshower detector, consisting of two layers of silicon strips and two layers of lead absorber, is installed to distinguish between prompt photons and those from the decay  $\pi^0 \rightarrow \gamma\gamma$ . The strips, oriented perpendicular to each other, have a pitch of 2 mm, allowing to resolve the two showers of the photons from the  $\pi^0$ .

The production of scintillation photons per energy deposit is temperature depend. Therefore the ECAL is kept at a temperature of  $18 \pm 0.05^\circ\text{C}$ , where it is about 4.5 photons per MeV.

The typical resolution of the ECAL is parametrized as

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{0.12}{E}\right)^2 + (0.30\%)^2, \quad (3.1)$$

with three terms describing different sources of uncertainty. The first term includes statistical

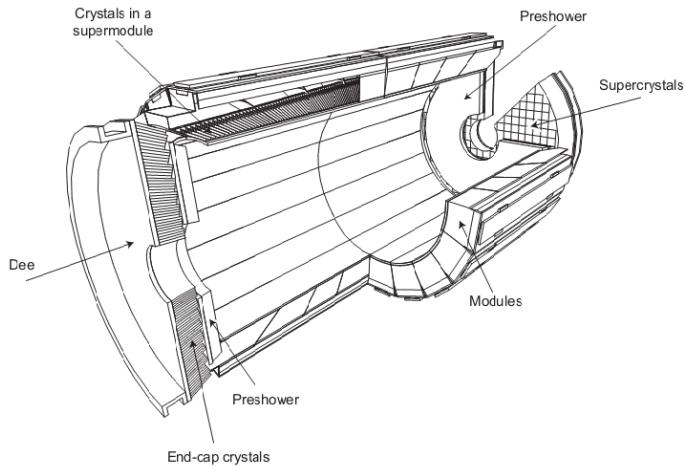


Figure 3.6.: Schematic view of the CMS ECAL.

fluctuation in the production of scintillation light as well as the energy distribution over several crystals. The second term covers such sources of noise as electronic noise or pileup. The constant term accounts for other sources of uncertainties such as calibration errors. The size of the different contributions has been confirmed in test beam measurements [8].

### 3.2.3. The hadron calorimeter

The hadron calorimeter (HCAL) measures the energy of charged and neutral hadrons. In the barrel region of the detector it is situated between the back face of the ECAL and the coil of the solenoid, at radii between 1.7 m and 2.95 m, limiting the amount of material that can be used in its construction. Therefore additional detectors are placed outside of the volume of the magnet, forming the hadron outer calorimeter (HO). The HCAL barrel (HB) is complemented on each side by a HCAL endcap (HE) and the geometric coverage is extended to high values of  $|\eta|$  by dedicated forward calorimeters (HF). The placement of these subdetectors relative to the other components of CMS are shown in Figure 3.7.

#### The HCAL barrel and outer detectors (HB and HO)

The HB covers the geometric region  $|\eta| \leq 1.3$ . It is constructed as a sandwich calorimeter, consisting of plastic scintillator as the active material. For the absorber material, the fourteen inner layers of the HB are made from brass, while steel is used for the front and back plates of the HB to increase the stability of the construction. The scintillator is divided in 144 segments in  $\phi$  and 32 segments in  $\eta$ , resulting in a spatial granularity of 0.087 in both  $\eta$  and  $\phi$ . The scintillation light produced in the active material is transported to hybrid photo diodes using scintillating fibres. As all layers of one tower in  $\eta$  and  $\phi$  are read out by the same photo diode, there is no segmentation in the readout in  $r$ , except for the two towers closest to the HE on each side. The material of the ECAL in front of the HB corresponds to about 1.1

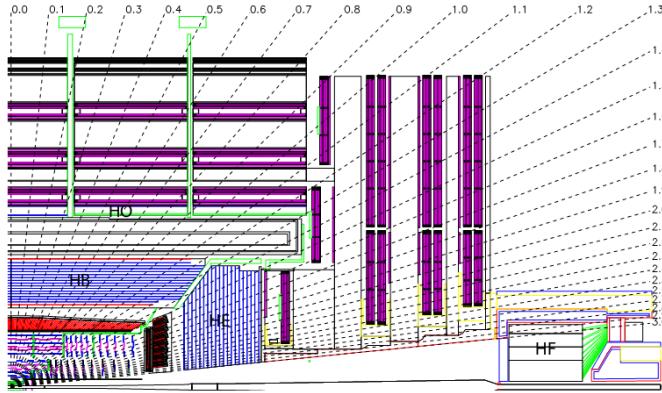


Figure 3.7.: Schematic view of the CMS HCAL.

interaction length  $\lambda_i$ . The absorber material of the HB itself amounts to only  $5.82 \lambda_i$  at  $\eta = 0$ , which increases to  $10.6 \lambda_i$  at  $|\eta| = 1.3$ . To measure the energy of jets not contained in the HB, the HO is placed outside the vacuum containment of the solenoid. It consists of one additional layer of scintillator, with the magnet acting as absorber, except for most central part of the detector, where one additional layer of steel absorber and scintillator are installed. Hereby the material budget of the HCAL is increased to at least  $10 \lambda_i$  over the whole barrel region.

### The HCAL endcaps (HE)

The HE extends the geometric coverage of the HCAL up to  $|\eta| = 3.0$ , coinciding with that of the EEs. It is constructed from the combination of brass absorber and plastic scintillator as the HB and for the region  $1.3 \leq |\eta| \leq 1.6$  also retains the  $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$  granularity in  $\eta$  and  $\phi$ . For  $|\eta| \geq 1.6$  the segmentation is coarser, resulting a granularity of  $\Delta\eta \times \Delta\phi \approx 0.17 \times 0.17$ . This structure again corresponds to about  $10 \lambda_i$ . The longitudinal segmentation of the readout of the towers differs based on the location of the tower. The two towers closest to the beam line are read out in three segments, while most others are read out in two segments. The two towers overlapping with HB are read out without longitudinal segmentation. Multipixel hybrid photo diodes have been chosen for the readout due to their low sensitivity to magnetic fields.

### The hadron forward calorimeter (HF)

Of all subdetectors of CMS, the HF covers the highest values of  $|\eta|$ , reaching up to values of  $|\eta| = 5.2$ . This close to the beam pipe radiation hardness is the key feature of the design, as nearly 90% of the energy deposited in the detector as the result of a proton-proton interaction is allotted to the HF. It is constructed as two 3.5 m long cylinders with a radius of 1.3 m, located at  $|z| = 11.2$  m. The first 1.65 m consist of plates of steel with a thickness of 5 mm, again

corresponding to about  $10 \lambda_i$ . The active material are quartz fibres, which are inserted into grooves in the steel plates. The particles created in showers in the absorber emit Cherenkov radiation in the fibres, which is detected by photomultiplier tubes at the end of the fibres. As the Cherenkov threshold is lowest for electrons, at 190 keV, the HF is more sensitive towards electromagnetic than hadronic showers. To separate the two, half of the fibres start only at a depth of 22 cm inside the absorber. As the electromagnetic shower develop faster, they deposit most of their energy before this point, which distinguishes them from hadronic showers.

### 3.2.4. The muon system

Muons are in general not stopped by any of the subdetectors of the CMS detector inside the solenoid. Therefore they can be measured with high precision in a clean environment outside of it. The muon detectors are therefore placed inside the return yoke of the magnet, both for the muon barrel (MB), covering up to  $|\eta| = 1.2$  and muon endcap (ME) detectors, placed between  $|\eta| = 0.9$  and  $|\eta| = 2.4$ . Being placed so far away from the interaction point, the muon detector have to cover a large area, which requires them to be rather inexpensive compared to other technologies used in the construction of CMS. Three different types of gaseous detectors are used to provide at the same time identification,  $p_T$  measurement and triggering for muons. In the barrel region, drift tubes (DT) are used as the main muon detectors, while in the endcaps cathode strip chambers (CSC) are used because they are faster and better equipped to deal with the larger and inhomogeneous magnetic field in this region of the detector. To provide a very fast muon tagging for the trigger, resistive plate chamber (RPC) complement the other two technologies in both the barrel and the endcaps.

#### Drift tubes (DT)

In the barrel, the TD chambers, each made of 2 or 3 super layers, which in turns are made up of four layers of TDs, are grouped together into four muon stations. Of these the first three contain 2 groups of 4 layers of chambers measuring in the  $r - \phi$  plane and one group measuring the  $z$  coordinate. In the last muon station only the groups measuring in  $r - \phi$  are present. The chambers in each chamber are offset by half of the width of a cell with respect to the next layer to not leave dead spots in the geometric coverage. The DT system consists of about 172000 sensitive wires. The drift tubes are filled with a mixture of 85% Ar and 15%  $CO_2$ , and their structure is shown in Figure 3.8. The  $r\phi$  resolution of a single layer of DT is about 250  $\mu m$ , so that one chamber reaches a precision of 100  $\mu m$ .

#### Cathode strip chambers (CSC)

The CSC are multiwire proportional chambers, consisting of 6 planes of anode wires interleaved with 7 panels of cathode strips. The chambers have trapezoidal shape and are arranged in four discs around the beam axis, each further segmented into two or three rings. The cathode strips measure the  $\phi$  coordinate while the anode wires measure the radial coordinate. Figure 3.9

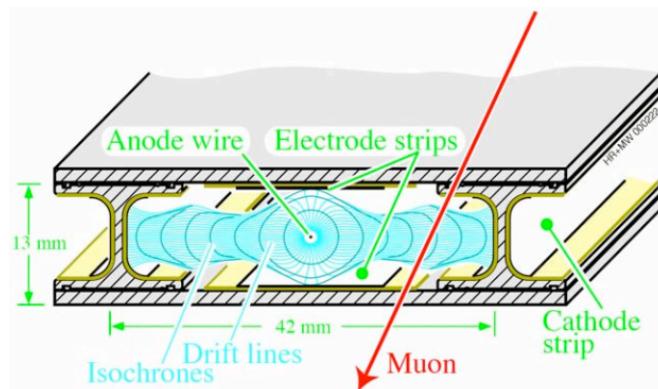


Figure 3.8.: Schematic view of one DT.

shows the structure of one chamber on the left side and the creation of a signal due to a amplification of the initial ionization in an avalanche close to the anode wire.

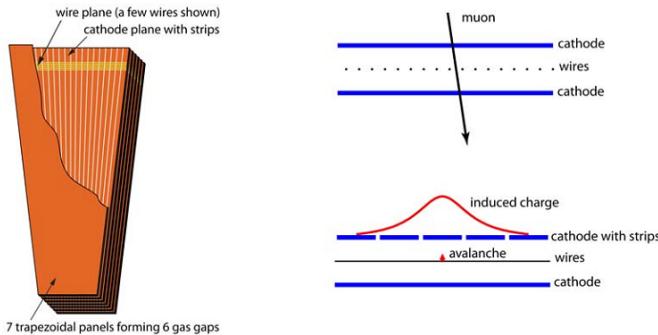


Figure 3.9.: Schematic view of one CSC (left) and the creation of a signal (right).

### Resistive plate chambers (RPC)

The RPCs consist of three layers of Bakelite, which form two small gas filled gaps and between which high voltage is applied. The amplification of the initial signal is very fast in this configuration, with drift times of about 5 ns. Therefore this technology is well suited to associate muon candidates to the LHC bunch crossings. In the barrel region six layers of RPCs are installed, while three layers are used in the endcaps for  $|\eta| \leq 1.6$ .

### Momentum resolution

The  $p_T$  resolution of the muon system alone was expected to be about 10% for muons with  $p_T$  up to 200 GeV. Combined with the information from the inner tracking system, a resolution of about 1% was expected to be achieved in the central region of  $|\eta| \leq 0.8$  for a  $p_T$  of 10 GeV, increasing to about 2% for a  $p_T$  of 200 GeV.

The  $p_T$  resolution for muons has been measured using data collected in 2010 [9]. Using the muon system alone, resolutions better than 10% have been found for the barrel region for muons with  $p_T > 15 \text{ GeV}$ . The muon resolution improves when combining the information from the muon system with those from the inner tracking system. The precision of the tracking system dominates for a wide  $p_T$  range and averaging over  $\eta$  and  $\phi$  resolutions of  $(1.8 \pm 0.3(\text{stat.}))\%$  at  $p_T = 30 \text{ GeV}$  to  $(2.3 \pm 0.3(\text{stat.}))\%$  at  $p_T = 50 \text{ GeV}$  have been achieved.

### 3.3. Trigger and data acquisition

If the LHC is operated at the design bunch spacing of 25 ns, bunch crossings occur with a rate of about 40 MHz in the interactions points. Even if this has been reduced by at least factor of two during Run I of the LHC because of the increased bunch spacing, event rates of this magnitude can not be handled by the available means of data processing. The total event rate is therefore reduced by a factor of about  $10^6$  by two subsequent trigger systems. The Level-1 (L1) trigger consists of programmable electronics, allowing for a coarse reconstruction of physics objects in the calorimeters and the muon system. This system reduces the event rate to a maximum of 100 kHz. Following an L1 accept (L1A), the CMS data acquisition system (DAQ) collects the event information from the readout of the different subdetectors and passes it one to the High-Level trigger (HLT). The HLT is a software trigger and has access to the full detector readout [10]. It can perform a full reconstruction of the events to apply approximate versions of the algorithms used in offline data analysis. It accepts events at a rate of a few  $10^2 \text{ Hz}$ .

#### Level-1 trigger (L1)

The output of the different subdetectors are stored in pipelined buffers inside the readout electronics. This limits the time between the bunch crossing and the distribution of the L1 accept to the subsystems to  $3.2 \mu\text{s}$ . The L1 is therefore constructed from mostly custom-built programmable electronics either directly inside the detector or located close by in the underground facilities. As the readout of the tracker and track reconstruction are feasible on this time scale only calorimeter and muon system information are used. The L1 system is divided into local, regional and global components, as shown in Figure 3.10.

The calorimeter trigger the local component are the Trigger Primitive Generators (TPGs). For  $|\eta| \leq 1.74$  they have an  $(\eta, \phi)$ -coverage of  $0.087 \times 0.087$ , corresponding to one HCAL tower and a  $5 \times 5$  matrix of ECAL crystals in front of it. The TPGs communicate the energy deposits in the trigger tower and the number of the bunch crossing to the Regional Calorimeter Trigger. One calorimeter region consists of  $4 \times 4$  trigger towers. Candidates for electrons or photons ( $e/\gamma$ ) are formed by selecting the towers with the highest  $E_T$  in the ECAL. Based on information about the energy distribution inside the ECAL tower, the ratio of energy in ECAL and HCAL in the trigger tower and the overall distribution of energy in the neighbouring trigger towers the candidates are classified as isolated or non-isolated. Per region four isolated and four non-isolated  $e/\gamma$  candidates, the transverse energy sums of the trigger towers and information to

identify  $\tau$  leptons and muons via their minimum ionizing particle (MIP) signature and isolation are passed to the Global Calorimeter Trigger (GCT). The GCT performs a simple jet clustering algorithm and is able to calculate per event observables such as the number of jets, the total and missing transverse energy and sum of the transverse energy of all jets above a certain threshold ( $H_T$ ). These information are delivered to the global trigger

In the muon trigger all three detector components (DT,CSC, and RPC) contribute. In the local trigger, the DT chambers deliver track segments in the  $\phi$ -projection and hit patterns in the  $\eta$ -projection, while the CSCs produce three-dimensional track segments. Both use timing information to associate this information with the bunch crossing. In the regional trigger, DT and CSC information are processed in separate track finders, which produce muon candidates. These are ordered as a function of  $p_T$  and track quality and up to four candidates are delivered to the global muon trigger from each track finder. The RPCs also deliver muon candidates. With their excellent timing resolution of about 1 ns they deliver an unambiguous association of the muon candidates to the correct bunch crossing. The global muon trigger receives up to four muon candidates each from the DTs, the CSCs and the barrel and endcap RPCs. The information consists of  $p_T$ ,  $\eta$ ,  $\phi$  and information on the quality of the muon reconstruction. Candidates from the RPCs are matched with the ones from DT and CSC and, if matches are found, merged into single candidates. Unmatched candidates with low quality are suppressed. The track of the candidates is extrapolated back into the calorimeters to add the MIP and isolation information from the regional calorimeter trigger. The four best muon candidates are forwarded to the global trigger.

The global trigger collects the information from the global muon and global calorimeter trigger. Up to 128 trigger algorithms can be performed on the trigger objects at the same time, the most basic being simple  $p_T$  thresholds. If the criteria of one of the algorithms is met by the event, it is accepted by the L1 trigger and a signal is sent to the DAQ to read out the event.

## Data acquisition system (DAQ)

Following an L1 accept the DAQ receives the information from the different subdetectors split in about 650 data sources, each delivering about 2 kB of data. These event fragments are assembled into whole events by the event builder. The event is then sent to one Filter Unit in the Event Filter, where the HLT software is running. The DAQ has to deal with input rates of up to 100 kHz and consists of 8 nearly independent slices, each able to take input at a rate of 12.5 kHz. The DAQ includes a back-pressure system, which automatically throttles the L1 trigger in the case where the input rate exceeds what can be processed by the DAQ. This introduces dead times in the detector readout but prevents data corruption and buffer overflows. To fully utilize the capacities of the trigger system and the DAQ, trigger thresholds can be adapted during data taking. The shortest time scale on which the thresholds are kept constant is called lumi section and is defined as  $2^{20}$  LHC orbits, corresponding to about 93 s. The structure of the DAQ is shown in Figure 3.10.

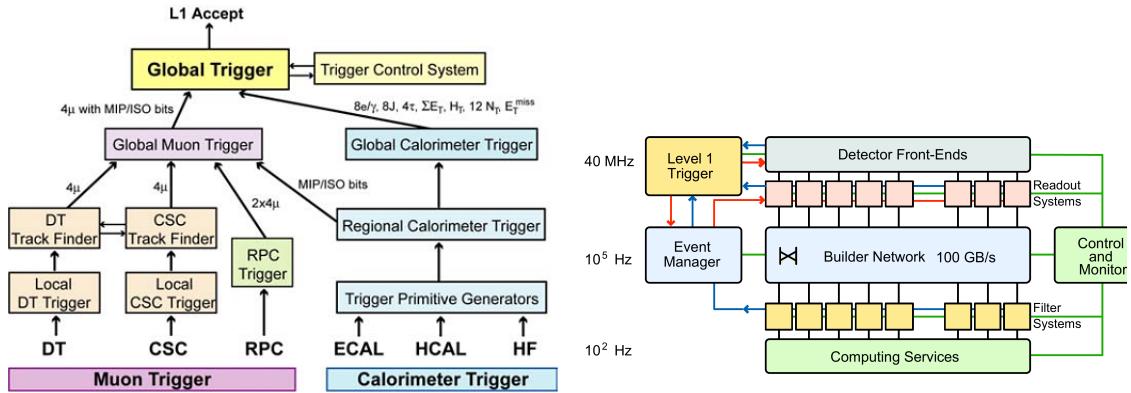


Figure 3.10.: Structure of the CMS Level-1 trigger (left) and data acquisition system (DAQ) (right)

### High level trigger (HLT)

The HLT software is run on a dedicated computing element, the Event Filter Farm, located in the CMS service caverns. During the data taking in 2012 it consisted of about 13200 processor cores [11], allowing for a processing time of about 150 ms at a input rate of 100 kHz. The HLT system reduces input data rates of up to 100 GB/s to several hundred MB/s, which are sent to the CERN computer centre for storage. As a full event reconstruction can be performed at HLT level, even if it is often restricted to small regions of the detector for timing reasons, much more complex quantities can be used to separate potentially interesting signatures from the large backgrounds. However, approximate methods have to be used sometimes to maintain an acceptable processing time per event. Also, some calibration and alignment methods can only be performed after the data taking, making the HLT less precise compared to the offline reconstruction. While a large variety of triggers is used by CMS to select different kinds of events, this description will focus on the ones most relevant to this analysis.

The most important signal and control samples are collected with dilepton triggers. In general, they select events that contain two leptons (electrons or muons), of which one is required to have a reconstructed transverse momentum  $p_T$  of at least 17 GeV, while for the second this is relaxed to 8 GeV. In general, the lepton with the higher (lower)  $p_T$  is referred to as the leading (trailing) lepton. While the algorithms employed to reconstruct muons and electrons are the same among all possible combinations of leptons, the triggers used for preselection at L1 level differ in their thresholds [12]. The dielectron HLT is based on the so called L1 seed L1\_DoubleEG\_13\_7 and the dimuon on a logical OR of L1\_DoubleMu\_10\_Open and L1\_DoubleMu\_10\_3p5. To select events with an electron and a muon, two HLTs are employed, one with the electron and one with the muon as the leading lepton. The L1 seeds are a logical OR of L1\_Mu3p5\_EG12 and L1\_MuOpen\_EG12, and L1\_Mu12\_EG7, respectively. The numbers refer to the  $p_T$  or  $E_T$  thresholds of the muon and e/ $\gamma$  candidates in GeV, while Open indicates a threshold of 0 GeV. As the algorithms to reconstruct physics objects at HLT level are so similar to those used offline, no dedicated description is given here.

---

## 4. Data analysis and event selection

### 4.1. Object reconstruction

The physics objects relevant to this analysis are electrons, muons, jets and the missing transverse energy  $E_T^{\text{miss}}$ . Here the reconstruction of these objects from the information provided by the CMS detector is described. While the electron and muon candidates used here are reconstructed independent of each other with dedicated algorithms, jets and  $E_T^{\text{miss}}$  are provided by the particle flow (PF) algorithm. It combines information from all subdetectors to achieve a consistent description of the full event.

#### 4.1.1. Muon reconstruction and selection

The track of a muon is reconstructed separately in the inner tracker and the muon system, resulting in a *tracker track* and a *standalone muon*.

Tracks in the inner tracker are reconstructed using a method called Combinatorial Track Finder (CTF) [13], which performs pattern recognition and track fitting employing a Kalman filter technique [14]. The track is described by a five-dimensional state vector, whose initial parameters are taken from track seeds, determined from three hits or two hits and a vertex constraint in the pixel detector or the innermost layers of the strip detector. The state vector is extrapolated to the next tracker layer taking into account uncertainties and energy losses due to interactions with the tracker. If tracker hits are found in the modules where they are expected from the extrapolation, they are added to the track candidate. If no hits are found, a ghost hit is added to the track to account for inefficiencies in the hit reconstruction. A track fit is then performed to all hits associated with the track candidate, using again Kalman filtering and smoothing. This procedure is performed iteratively, each time removing the hits already associated to a track candidate and relaxing the requirements on the track seeds to allow for reconstruction of track with low  $p_T$  or not originating from the primary interaction. In the reconstruction of the data taken in 2012, seven iterations were performed [15].

For the reconstruction of *standalone muons* in the muon system, the hits inside the individual muon chambers are fitted to generate track segments, providing first estimates of the track parameters under the hypothesis that the muon was created in the interaction region and was travelling through the muon system from the inside out. These segments are used as starting points for a track reconstruction using all hits from the DTs, CSCs and RPCs, again using the Kalman filtering technique [16].

Tracker tracks are promoted to *tracker muons* when they can be matched to a track segment in the muon detector. *Standalone muons* are matched to tracks from the inner tracker. If a compatible track is found a combined fit to all hits of the track and the *standalone muon*

Criterion	Selection
Acceptance	
$p_T$	$> 10 \text{ GeV}$
$ \eta $	$< 2.4$
Muon ID	
Required to be a	<i>tracker muon</i> <i>global muon</i> <i>particle flow muon</i>
Track quality	
$\chi^2/N_{dof}$	$< 10$
valid muon hits	$> 0$
matched stations	$> 1$
valid pixel hits	$> 0$
tracker layers with hits	$> 6$
Impact parameter	
$d_0 = \sqrt{dx^2 + dy^2}$	$< 0.02 \text{ cm}$
$dz$	$< 0.1 \text{ cm}$

Table 4.1.: Summary of requirements of the muon selection.

is performed, resulting in a *global muon*. The PF algorithm applies further selection requirements to the reconstructed *global* and *track muons*, introducing a fourth category, the *particle flow muon* [17].

Muons selected in this analysis are required to be reconstructed as *tracker*, *global* and *particle flow* muons. The  $\chi^2$  per degrees of freedom of the track fit must not exceed 10. Several requirements on the information available for the different track fits are made: At least one muon chamber hit must be included in the track fit of the *global muon*. For the fit of the *tracker muon* at least one hit in the pixel detector and six layers with hits in the strip detector have to be available. Also the track from the inner tracker has to be matched to at least two track segments in the muon chambers. To ensure that the muon originates from the primary interaction and to suppress backgrounds from cosmic muons the impact parameter of the track with respect to the primary vertex must not exceed 0.02 cm in the  $x$ - $y$  plane and 0.1 cm in  $z$  direction. Selected are muons with a  $p_T$  larger than 10 GeV and  $|\eta|$  less than 2.4. The muon selection is summarized in Table 4.1.

### 4.1.2. Electron reconstruction and selection

The signature of an electron in the CMS detector is a track reconstructed by the tracking detectors that leads to a matching cluster of energy reconstructed in the ECAL. In practice this is complicated by the large material budget of the tracking detectors, resulting in a high probability of an electron to loose energy in form of bremsstrahlung. About 35% of all electrons loose more than 70% of their energy and for 10% the energy loss exceeds 95% [18]. The reconstruction is further complicated by the large solenoidal magnetic field, which bends the

electron's trajectory away from the radiated photons, leading to a spread of the energy in  $\phi$  direction. This has to be taken into account both in the tracking algorithms and the clustering of the energy deposits in the ECAL.

In the ECAL two different algorithms are used to group the energy deposits into clusters and clusters of clusters, called super clusters (SCs), in the barrel and endcap regions of the detector. Both are designed to group together the energy deposits of the electron itself and those of the bremsstrahlung photons. In the range of  $1.6 < |\eta| < 2.6$  the preshower is located in front of the ECAL and electrons will deposit a fraction of their energy there. The energy deposited in the strips of the preshower between an SC in the ECAL and the primary vertex is summed and added to the energy of this SC [19]. ADD SC position calculation and electron charge!  
CHANGE ALL REFS TO EGM-13-001 once submitted to journal!

In the track reconstruction with Kalman filters as discussed above energy losses due to interactions of the particles with the tracker material are considered to be Gaussian. For electrons, however, this is not sufficient because the dominant energy loss due to bremsstrahlung is a non-gaussian contribution. Electron candidate tracks are therefore fitted with a Gaussian Sum Filtern (GSF) algorithm [20], which models the non-Gaussian components as a sum of Gaussian distributions. GSF tracking is initiated in two ways. *ECAL driven seeding* requires the presence of a track seed that matches the position of an SC when extrapolating backwards from the ECAL to track [18]. Alternatively, *tracker driven seeding* is started by tracks that either match the position of ECAL clusters when extrapolated to the ECAL surface, covering the case of no bremsstrahlung, or are of poor quality with only few associated hits [13]. The GSF track and the energy measurement in the ECAL are combined into the final electron candidate.

Electrons are selected requiring  $p_T$  larger than 10 GeV and  $|\eta| < 2.5$ . The gap region between ECAL barrel and endcaps of  $1.442 < |\eta| < 1.566$  is excluded. To suppress background from muons that radiate photons electrons with a distance of  $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$  to the nearest *global or tracker muon* less than 0.1 are rejected. Backgrounds from photon conversion, decays of heavy flavour quarks or charged hadrons are suppressed by a set of selection criteria. The matching of track and supercluster is quantified by the differences between the supercluster position and the parameters of the track extrapolated from the vertex to the ECAL surface in  $\Delta\phi$  and  $\Delta\eta$ . As the energy of the electron is contained in the ECAL, the ratio of hadronic energy deposited in the HCAL behind the electron candidate must be small. The energy spread in the ECAL due to bremsstrahlung occurs in  $\phi$  direction. Therefore no significant spread of the energy in  $\eta$ , parametrized as

$$\sigma_{i\eta i\eta} = \frac{\sum_i^{5\times 5} w_i \cdot (\eta_i - \bar{\eta}_{5\times 5})^2}{\sum_i^{5\times 5} w_i}, \quad (4.1)$$

$$w_i = \max(0, 4.7 + \ln(\frac{E_i}{E_{5\times 5}})), \quad (4.2)$$

is expected, where for  $5 \times 5$  crystals around the seed crystal the distance in  $\eta$  from the mean  $\eta$  of the cluster is summed, weighted by the energy deposit in each crystal. For a well measured electron there is good agreement between the energy deposited in the ECAL

Criterion	Selection at HLT		Selection at Analysis Level	
	EB	EE	EB	EE
Acceptance				
$p_T$	trigger dependent		$> 10 \text{ GeV}$	
$ \eta $	$< 2.5$		$< 2.5, \text{ excluding } 1.442 <  \eta  < 1.566$	
ID variables				
$ \Delta\eta $	0.01	0.01	0.007	0.009
$ \Delta\phi $	0.15	0.10	0.15	0.10
$\sigma_{inj\eta}$	0.011	0.031	0.01	0.03
$H/E$	0.10	0.075	0.12	0.10
$ \frac{1}{E} - \frac{1}{p} $	-		0.05	0.05
Conversion rejection				
missing pixel hits	-		$\leq 1$	$\leq 1$
vertex fit probability	-		$< 10^{-6}$	$< 10^{-6}$
Impact parameter				
$d0 = \sqrt{dx^2 + dy^2}$	-		$< 0.02 \text{ cm}$	$< 0.02 \text{ cm}$
$dz$	-		$< 0.1 \text{ cm}$	$< 0.1 \text{ cm}$

Table 4.2.: Summary of requirements of the electron selection.

and the track momentum measured at the vertex. Therefore the value of  $|\frac{1}{E} - \frac{1}{p}|$  must be small. Requirements on the impact parameter of the track with respect to the vertex are made. Two requirements are applied to reject electrons from converted photons. Only one pixel layer with missing hit is allowed, rejecting most conversions occurring after the first layer of the pixel detectors. To reject also conversion in this first layer and in the beam pipe, vertex fits for the electron track with neighbouring tracks are performed in order to reconstruct the point of conversion. For a prompt electron, the probability of these fits is low. Some of these requirements are already applied on HLT level. In order to select electrons for which the trigger is fully efficient, selections at least as strict are applied at analysis level. The specific requirements are listed in Table 4.2, separately for barrel and endcap when appropriate.

#### 4.1.3. Observables reconstructed with Particle Flow

##### The algorithm

The particle flow (PF) algorithm is designed to combine information from all subdetectors to reconstruct a consistent description of the event, resulting in a list of reconstructed particles. The basic building blocks are PF elements, which are reconstructed in each subdetector separately: Tracks of charged particles in the tracker, energy clusters in the calorimeters and muon tracks in the muon system. A linking algorithm then combines elements into blocks based on their geometrical distance, for example by extrapolating a track into the ECAL and HCAL and searching for compatible clusters. Similarly, calorimeter clusters are linked between the preshower, ECAL, and HCAL and tracks from the tracker are associated with those from the

muon system. The reconstruction of *particle candidates* is performed on the elements inside a block. Muons are reconstructed first, followed by electrons, for which, similar to the standard algorithm described above, a refit of the track with the GSF algorithm is performed and bremsstrahlung photons are collected in the ECAL. Lastly, calorimeter clusters compatible with a track are identified as charged hadron, while clusters without a matching track are either categorized as neutral hadrons, or, if no or only small energy deposits in the HCAL exist, as photons [21].

## Jets

The jets of particles produced in the hadronization of quarks and gluons are grouped together by clustering algorithms. An anti- $k_T$  algorithm [22], performed using a fast implementation [23, 24], is used in this analysis. Input to the clustering are the *particle candidates* reconstructed by the particle flow algorithm.

The anti- $k_T$  algorithm is a sequential clustering algorithm. Two distance measures are introduced, the first between two particles or pseudo-jets  $i$  and  $j$  and the second between particle or pseudo-jet  $i$  and the beam axis:

$$d_{ij} = \min(k_{Ti}^{-2}, k_{Tj}^{-2}) \frac{\Delta_{ij}^2}{R^2}, \quad (4.3)$$

$$d_{iB} = k_{ti}^{-2}, \quad (4.4)$$

with  $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$  and  $k_{Ti}$ ,  $y_i$ , and  $\phi_i$  being the transverse momentum, rapidity and azimuth of a particle. For an entity (particle, pseudo-jet)  $i$  all distances are calculated. If the smallest is a  $d_{ij}$ ,  $i$  and  $j$  are combined in a new pseudo-jet. If the smallest distance if the distance to the beam  $d_{iB}$ , the pseudo-jet is considered a final jet and removed from the list of particles available for clustering. The parameter  $R$  governs the size of the resulting jet and is set to 0.5 in this analysis.

The measured jet energy has to be corrected for energy offsets and the non-uniform and non-linear response of the detector. Each component of the jet's four-momentum vector is corrected by a multiplicative factor [25]

$$p_\mu^{cor} = C \cdot p_\mu^{raw}. \quad (4.5)$$

The correction is applied as a sequence of different factors:

$$C = C_{\text{offset}}^{L1}(p_T^{raw}) \cdot C_{MC}^{L2L3}(p'_T, \eta) \cdot C_{\text{rel}}^{L2\text{Residual}}(\eta) \cdot C_{\text{abs}}^{L3\text{Residual}}(p''_T). \quad (4.6)$$

The L1 correction, applied to the raw jet, corrects for offsets due to the underlying event and pileup using a jet area approach. The jet area  $A_j$  is determined for each jet and the particles in the event are clustered with a  $k_T$  jet clustering algorithm with a distance parameter  $R = 0.6$ , which clusters a large number of soft jets in each event. The median  $p_T$  density  $\rho$  is then defined as the median of the distribution of  $p_{Tj}/A_j$  for all of these jets. Therfore  $\rho$  is not influenced by the presence of hard jets from the primary interaction in the event and is a measure for the pileup activity, the underlying event and electronic noise. Jets are then corrected by the

factor  $C_{\text{offset}}^{L1}(p_T^{\text{raw}}) = 1 - \frac{(\rho - \langle \rho_{UE} \rangle) \cdot A_j}{p_T^{\text{raw}}}$ , where  $\langle \rho_{UE} \rangle$  is the mean  $p_T$  density due to the underlying event, measured in events with no pileup interactions. The L2L3 corrections, derived from Simulation, corrects for the non-linearities and non-uniformities of the detector response to jets of different  $p_T$  and  $\eta$  and is applied to the offset-corrected jet. To correct for the differences between Simulation and real data, jets in data events are further corrected for the residual differences between the two.

In this analysis, the  $p_T$  of a jet is required to exceed 40 GeV and jets are required to lie inside the fiducial volume of the ECAL of  $|\eta| < 3.0$ . A set of loose quality selections is applied to suppress jets reconstructed because of detector noise, ensuring that the jet is reconstructed in more than one subdetector and has more than one constituent. As the jet clustering is performed using all reconstructed *particle candidates*, jets within  $\Delta R = 0.4$  to leptons identified with the criteria described above are rejected.

Because of their long lifetime, b-jets decay at a measurable distance from their production vertex, allowing for the reconstruction of a secondary vertex. In this analysis the *combined secondary vertex* (CSV) algorithm is used. Likelihood ratios based on a variety of variables characterizing the secondary vertices and the tracks inside the jet are used to construct a single discriminator. If its value exceeds a given threshold, the jet is tagged as originating from a b-quark [26]. The performance of the b-tagging algorithms have been measured on the  $\sqrt{s} = 8$  TeV dataset [27]. The average identification efficiency as a function of the discriminator value is shown on the left side of Figure 4.1. For the medium working point of 0.679, chosen in this analysis, the efficiency is about 70% while the probability to misidentify a jet originating from a light quark as a b-jet is between 1% and 3%, depending on the  $p_T$  of the jet, as shown on the right side of Figure 4.1. In this analysis b-jets are considered with a  $p_T$  larger than 30 GeV and  $|\eta| < 2.4$ .

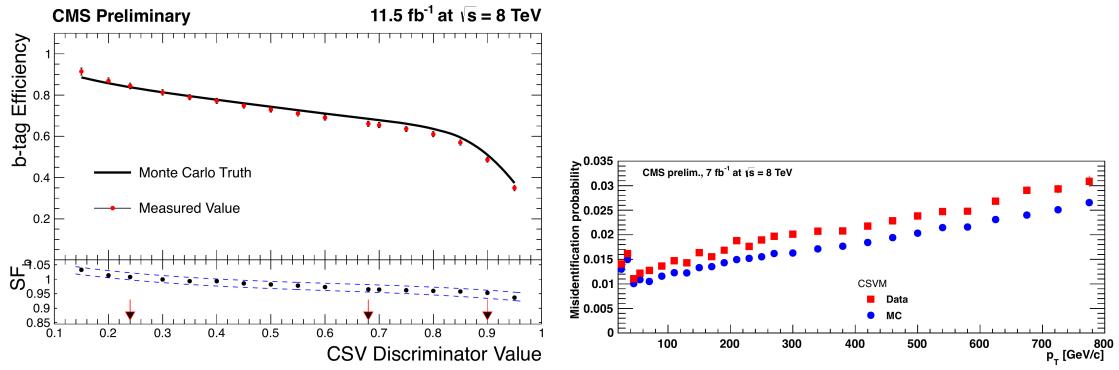


Figure 4.1.: Performance of the CSV b-tagging algorithm. Shown is the identification efficiency as a function of the discriminator value (left) and the probability of misidentifying a jet originating from a light quark as a b-jet (right) [27].

### Missing transverse energy

As the transverse momenta of the initial partons are negligible compared to their large momenta in beam direction, the sum of the transverse momenta of all particles produced in the interaction is essentially zero because of conservation of momentum. For the reconstructed event this is not necessarily the case, leading to a missing transverse energy ( $E_T^{\text{miss}}$ ) different from zero. The measurements in the detector have a finite resolution and are subject to detector noise and are subject to gaps in the detector acceptance. Also, particles that are only weakly interacting, such as neutrinos, are not detected by CMS and cause an imbalance of the transverse momentum sum. As this imbalance is the only experimental signature of this class of particles, a good  $E_T^{\text{miss}}$  resolution is a key factor for the discovery of processes that include the production of new weakly interacting particles.

Several algorithms have been developed in CMS to reconstruct  $E_T^{\text{miss}}$  [28]. Calorimetric (Calo)  $\vec{E}_T^{\text{miss}}$  is calculated as the negative vector sum of the energy deposits in each calorimeter tower. The small energy deposits from muons are replaced by the  $p_T$  measurements of muons. A further correction is introduced in the track-corrected (TC)  $\vec{E}_T^{\text{miss}}$ . For well reconstructed tracks, the track measurement is more precise than the measurement of a hadron's energy in the HCAL. Therefore, for tracks not associated with an electron or muon, the track measurement is used in the calculation of  $\vec{E}_T^{\text{miss}}$ . The energy deposit in the calorimeter is excluded, based on a model of the calorimeter response, treating all hadrons as pions. In contrast to these subdetector-based approaches, the event description of the particle flow algorithm can be used to calculate  $\vec{E}_T^{\text{miss}}$ . It is defined as the negative vector sum over the  $p_T$  vectors of all particle flow candidates:

$$\vec{E}_T^{\text{miss}} = - \sum_{\text{PF candidates}} \vec{p}_T^i. \quad (4.7)$$

Comparing the resolution for the  $E_T^{\text{miss}}$  components in  $x$  and  $y$  direction, as shown in Figure 4.2 for the data collected in 2011, PF  $E_T^{\text{miss}}$  performs best of the three algorithms and is therefore used in this analysis.

Several corrections can be applied to the calculation of  $E_T^{\text{miss}}$ . The *type-I* corrections propagate the corrections to the jet energy to the  $E_T^{\text{miss}}$  calculation for all jets with  $p_T$  larger than 10 GeV and with less than 90% of their energy deposited in the ECAL. The effects of pileup on the  $E_T^{\text{miss}}$  reconstruction can be mitigated by applying *type-0* corrections, which are calculated on minimum bias events to parametrize the effects of such interactions on  $E_T^{\text{miss}}$ . Further corrections can be applied to correct for modulations of the  $E_T^{\text{miss}}$  in  $\phi$  [29]. As this analysis searches for events with a large genuine  $E_T^{\text{miss}}$  and is therefore not very sensitive to  $E_T^{\text{miss}}$  introduced by resolution effect, none of these corrections are applied. However, *type-I* corrected PF  $E_T^{\text{miss}}$  is considered as a cross check.

### Lepton isolation

While the lepton selection criteria described above are sufficient to reject backgrounds from particles misidentified as leptons, they do not suppress real leptons not originating from the primary interaction. As these are often produced in decay of heavy flavour quarks inside a

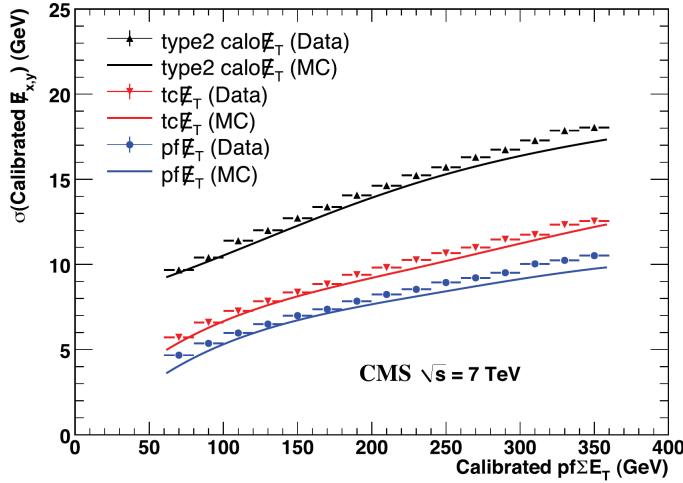


Figure 4.2.: Calibrated  $E_T^{\text{miss}}$  resolution as a function of the sum of the  $E_T$  of all particle flow candidates in an event. Shown are Calo  $E_T^{\text{miss}}$  in black, TC  $E_T^{\text{miss}}$  in red and PF  $E_T^{\text{miss}}$  in blue [28].

jet, are more suitable criterion is to consider the amount of activity in the detector close to the lepton candidate, called lepton isolation. In this analysis, particle based isolation is used. For this the energy deposited by charged hadron, neutral hadron, and photon particle flow candidates in a cone of  $\Delta R = 0.3$  around the lepton is summed:

$$\text{Iso} = \sum_{\text{charged hadrons}} E_T + \sum_{\text{neutral hadrons}} E_T + \sum_{\text{photons}} E_T. \quad (4.8)$$

The calculation of the isolation is distorted by pileup if PF candidates originating from pileup interactions lie within the cone and are counted in the isolation sum. This is easily remedied for charged hadrons, as those originating from a pileup vertex can be excluded from the calculation. For neutral hadrons and photons there is no track that can be associated to a vertex and a direct identification as pileup particles is not possible. Different approaches are pursued to correct for this contribution for electrons and muons. In both cases an estimate for the contribution of neutral pileup is subtracted from the isolation sum, which changes to:

$$\text{Iso} = \sum_{\text{charged hadrons}} E_T + \max(0, \sum_{\text{neutral hadrons}} E_T + \sum_{\text{photons}} E_T - \sum_{\text{neutral PU}} E_T). \quad (4.9)$$

For electrons the correction is similar to the L1 offset correction for jets described above. As a measure for the pileup contribution in the isolation cone the median  $p_T$  density in the event  $\rho$  is multiplied by the effective area of the electron in the detector, which is calculated in bins of  $\eta$ . The pileup correction is therefore defined as  $\sum_{\text{neutral PU}} E_T = \rho \cdot A_{\text{electron}}^{\text{eff}}$ . For muons  $\Delta\beta$  corrections are applied. Here it is utilized that on average the contribution of neutral particles from pileup is half that of charged particles, leading to a correction defined as  $\sum_{\text{neutral PU}} E_T = 0.5 \cdot \sum_{\text{charged PU}} E_T$ . Because of the stochastic nature of these approaches, overcorrection is possible. Therefore, no negative contribution from neutral particles is allowed in equation 4.9. ADD PLOTS SHOWING ISO VS NVTX

For both electrons and muons the isolation sum must not exceed  $0.15 \cdot p_T$  of the lepton candidate. **TO DO:** make plots illustrating the choice

## 4.2. Event processing and datasets

Events accepted by the HLT are reconstructed using the algorithms described above, implemented in the **CMS Software** (CMSSW) framework [30, 31]. While a first reconstruction is performed immediately after it is recorded, making it available to analysis within a few days, the full dataset recorded by CMS in 2012 has been reprocessed in a second reconstruction with updated calibrations and detector alignment in the first months of 2013. The software version used for this purpose was CMSSW\_5\_3\_7\_patch6. The events are stored in the **Analysis Object Data** (AOD) format, which contains mostly high level objects, such as electrons and muons, and does not provide access to detailed detector information such as energy deposits, which are not of interest in many analyses. This allows to reduce the event size to  $\approx 0.1$  MB, compared to about 2 MB for the raw detector output.

The data processing in this analysis is split into two parts. As a first step, the events in AOD format are processed utilizing the resources of the **Worldwide LHC Computing Grid** (WLCG) [32, 33], a system of cross-linked computing centres providing storage and computing capacities to the LHC experiments. Datasets stored on grid sites can be accessed through the **CMS remote analysis builder** (CRAB) [34]. At this stage, dilepton events are selected based on the identification criteria described above and the properties of the lepton pairs, together with other event characteristics, are stored. This is done with version **TO DO: Tag current version and add version number** of the SuSyAachen framework, which utilizes tools provided within the CMSSW framework, notably the **Physics Analysis Toolkit** (PAT) [35]. All datasets used in this analysis have been processed using CMSSW\_5\_3\_8\_patch3. Detector calibrations and alignment constant to be used in the processing of events in CMSSW are specified in so called **Global Tags**. The tags used in this analysis are FT53\_V21A\_AN6 for data and START53\_V27 for simulation.

The second part consists of all further analysis performed on the events selected in the previous processing. As the event size is much reduced, it can be performed using conventional desktop PCs. Throughout the event processing chain, the ROOT framework [36] for data analysis in particle physics is frequently used. In the final analysis steps, ROOT version 5.34.21 is used.

### 4.2.1. Primary datasets

Events are sorted into different primary datasets based on the HLT decisions, grouping together events accepted by related triggers. As this allows for events to appear in several of these datasets, precautions against double counting have to be taken when combining different data streams in one analysis. The primary datasets most relevant to this analysis are DoubleElectron, DoubleMu, and MuEG, containing, amongst others, events triggered by the different dilepton triggers. As auxiliary datasets, events from primary datasets triggered by hadronic activity (HT,JetHT), single leptons (SingleElectron,SingleMu), and the  $\alpha_T$  vari-

Primary dataset	purpose	dataset
DoubleElectron	Signal	/DoubleElectron/Run2012A-22Jan2013-v1/AOD /DoubleElectron/Run2012B-22Jan2013-v1/AOD /DoubleElectron/Run2012C-22Jan2013-v1/AOD /DoubleElectron/Run2012D-22Jan2013-v1/AOD
DoubleMu	Signal	/DoubleMu/Run2012A-22Jan2013-v1/AOD /DoubleMuParked/Run2012B-22Jan2013-v1/AOD /DoubleMuParked/Run2012C-22Jan2013-v1/AOD /DoubleMuParked/Run2012D-22Jan2013-v1/AOD
MuEG	Background prediction	/MuEG/Run2012A-22Jan2013-v1/AOD /MuEG/Run2012B-22Jan2013-v1/AOD /MuEG/Run2012C-22Jan2013-v1/AOD /MuEG/Run2012D-22Jan2013-v1/AOD
HT, JetHT	trigger efficiencies	/HT/Run2012A-22Jan2013-v1/AOD /JetHT/Run2012B-22Jan2013-v1/AOD /JetHT/Run2012C-22Jan2013-v1/AOD /JetHT/Run2012D-22Jan2013-v1/AOD
HTMHT	additional trigger studies	/HTMHTParked/Run2012B-22Jan2013-v1/AOD /HTMHTParked/Run2012C-22Jan2013-v1/AOD /HTMHTParked/Run2012D-22Jan2013-v1/AOD
SingleElectron	additional trigger studies	/SingleElectron/Run2012A-22Jan2013-v1/AOD /SingleElectron/Run2012B-22Jan2013-v1/AOD /SingleElectron/Run2012C-22Jan2013-v1/AOD /SingleElectron/Run2012D-22Jan2013-v1/AOD
SingleMu	additional trigger studies	/SingleMu/Run2012A-22Jan2013-v1/AOD /SingleMu/Run2012B-22Jan2013-v1/AOD /SingleMu/Run2012C-22Jan2013-v1/AOD /SingleMu/Run2012D-22Jan2013-v1/AOD

Table 4.3.: List of primary datasets used in the analysis. Additionally, the main purpose of the dataset and datasetpaths in DBS are given.

able TO DO: Explain  $\alpha_T$ , restructure sentence (HT,HTMHT) are used. Each primary dataset is split into four subsets, labelled Run2012A to Run2012D, each run defined by the run period of the LHC between two technical stops. The primary datasets are summarized in Table 4.3, where also the datasetpaths by which the samples can be accessed in the CMS bookkeeping system (DBS) [37] is given.

#### 4.2.2. Simulated datasets

Simulated datasets of Standard Model processes and SUSY models are used throughout the analysis in the design and validation of methods and the interpretation of the results in terms of potential signals. Dedicated methods are used for the different steps needed to achieve a complete model of the proton-proton interactions and the detector response.

### Simulation of the physical processes

Monte Carlo methods are used to generate events according to the properties of physics processes [38]. At the beginning of the description of a process stands the calculation of the cross section for the given hard scattering of fundamental particles, using perturbation theory (see for example [39]). For many Standard Model processes and also some BSM models, calculations in next-to-leading (NLO) or next-to-next-to-leading (NNLO) order have been performed. The automated calculations performed in the event generators used for the simulation in this analysis are however mostly restricted to leading-order (LO) accuracy.

At a hadron collider, the total cross section for a process is given by the cross section for the hard scattering  $\hat{\sigma}$ , convolved with the parton density functions (PDFs)  $f_i^p(x, Q^2)$ , which give the probability that a parton  $i$  with a fraction  $x$  of the proton's momentum at the momentum scale  $Q^2$  of the interaction will take part in the interaction. Considering all possible combinations of partons (three valence quarks, sea quarks and gluons), the total cross section is given by

$$\sigma(pp \rightarrow C) = \sum_{i,j} \int dx_1 dx_2 f_i^p(x_1, Q^2) f_j^p(x_2, Q^2) \hat{\sigma}(ij \rightarrow C). \quad (4.10)$$

The PDFs have to be inferred from data and have been studied in numerous fixed-target experiments and, most importantly, in deep-inelastic electron-proton scattering at the HERA collider [40]. Different approaches are used by several groups to parametrize the PDFs based on the available data. In the generation of simulated datasets for CMS analyses, the CTEQ6L1 [41] PDF set has been used. To study systematic effects introduced by the choice of PDF set, the NNPDF2.3 [42], MSTW2008 [43], and CT10 [44] PDF sets are used. The dependence of the PDFs on the momentum scale is described by the DGLAP (Dokshitzer, Gribov, Lipatov, Altarelli, Parisi) evolution equations [45, 46, 47], which are used to extrapolate them to the regime of the LHC.

For most processes, Madgraph [48] is used to calculate the hard scattering process, together with additional emissions of partons as part of initial and final state radiation (ISR and FSR). The inclusion of these emissions at matrix element level allows for the modelling of the radiation of hard partons that are well separated from other final state particles. However, this treatment breaks down for soft or collinear emissions, which can in turn be described by dedicated parton shower models. For this, Pythia [49] has been used for all samples relevant to this analysis. To achieve a consistent description of the parton shower, events are rejected in which the parton shower in Pythia produces jets in the phase space already covered by the emissions in Madgraph, using the MLM matching scheme [50].

The production of single top quarks is simulated using Powheg [51, 52, 53] at NLO in perturbative QCD. For these samples, a similar matching of the parton showers in Powheg and Pythia is applied.

The hadronization of colour-charged particles produced in the hard scattering or the parton shower is a non-perturbative processes which can only be described by phenomenological models. The *string fragmentation* model, as used in Pythia, is based on the idea of colour strings connecting the colour-charged particles. The energy stored in the strings increases linearly with the distance between the particles, until the string breaks and a  $q\bar{q}$  pair is created,

allowing for the formation of colour-singlets. These singlets may in turn break, until there no longer enough energy available to continue with this process [49]. The hadronization model, as well as the description of the underlying event and multi-parton interactions, has to be tuned to best describe existing data. For all samples used in this analysis, the tune  $Z2*$  [54] is used.

The decays of  $\tau$  leptons are simulated with the dedicated software Tauola [55], which includes polarization and spin correlations effects.

To simulate the effects of pileup, several simulated proton-proton interactions from a sample containing mostly soft QCD processes are added to the simulated events, including pileup interactions with a time distance to the event of  $\pm 50$  ns to emulate the effects of out-of-time pileup. The distribution of the number of additional interactions had to be estimated before the data taking took place and therefore differs from that contained in the recorded dataset.

### Simulation of the detector response

A model of the CMS detector has been created using the GEANT4 toolkit [56]. It allows for a detailed description of the detector geometry and material budget and simulates the interactions of particles with the detector material. It also models the propagation of the particles inside different materials, taking into account for example the magnetic field inside the CMS solenoid. The energy deposits created by the interactions of the particles with the detector are converted into detector hits on which the full event reconstruction is performed. The simulation also includes a modelling of detector noise and dead readout channels.

As this detailed simulation is quite time consuming, a fast simulation of the CMS detector has been developed [57]. Trading some accuracy for large gains in processing time, the fast simulation is used in cases where large numbers of events have to be generated, for example in scans of the parameter space of a physics model. Simplifications include for example an approximation of the tracker geometry, where the modelling of millions of individual modules has been replaced by thin cylinders of active and non-active material placed around the interaction point. A charged particle traversing these layers deposits some energy at the point at which it crosses an active layer with a predefined probability. Also the reconstruction algorithms for tracks have been simplified. Similar approaches have been applied to all subdetectors. Also the simulation of detector noise is reduced. An decrease of processing time per event by a factor of  $\approx 100$  has been observed.

Simulated events are stored in the AODSIM data format, which is identical to the AOD format but also includes Monte Carlo truth information about the simulated particles and their production and decay history. This allows for a processing of simulated datasets with the same software as used in the analysis of data events.

### Background Samples

Possible background contributions in the analysis arise from all Standard Model processes producing lepton pairs or one lepton with the possibility for other parts of their signature

to be misidentified as a second lepton. The properties of these processes can be studied in simulation. Therefore, a extensive list of simulated processes is considered in the analysis. In many cases, processes are divided into different samples based on the different possible final states. This allows to produce larger sample sizes for decays with small branching fraction without having to generate enormous amounts of more abundant final states. The full list of considered processes is shown in Table 4.4. The samples have to be scaled according to the appropriate integrated luminosity, taking into account the number of generated events and the cross section of the process. The weight is then given by  $w = \frac{L \cdot \sigma}{N_{\text{events}}^{\text{gen}}}$ . The top-pair production is normalized to the cross section measured by CMS in the dileptonic decay channel [58]. Cross sections for the production and dileptonic decays of W and Z bosons have been calculated using FEWZ 3.1 [59], including corrections in N(N)LO in electronweak theory (QCD). MCFM 6.6 [60] is used for the calculation of cross sections for diboson production. Cross sections for single top production have been calculated at approx. NNLO [61]. Cross sections at NLO in QCD for triboson production have been calculated using aMC@NLO [62], while for  $t\bar{t}$  production in association with one additional vector boson, MCFM 6.6 has been used [63], while for  $t\bar{t}WW$  the cross section calculated by Madgraph is used. The cross section for top-pair production in association with a photon has been measured by CMS [64].

### Signal Samples

## 4.3. Event selection

A series of selection criteria are applied to the events to select signal-like topologies and reduce the contributions from Standard Model backgrounds to the final sample. Also requirements are defined to select control regions enriched in certain SM processes for the purpose of background prediction and the validation of methods. Additionally, events are rejected that exhibit signs of detector noise or are otherwise not suited for analysis.

### 4.3.1. Event cleaning

As a first step in the selection of reconstructed events, a series of quality requirements is applied. The quality of the data recorded by the CMS detector is assessed in several automated or manual steps, summarised as *Data quality monitoring (DQM)* [65]. For each lumi section this results in a binary decision, flagging it as either *good* or *bad*, accepting only those lumi sections for which all subdetectors were fully operational during data taking and no known problems occurred in the reconstruction of the events.

To reject non-collision events, vertex information is used. To reconstruct vertices, tracks fulfilling certain quality requirements are clustered into vertices with a deterministic annealing algorithm [66, 13]. The vertex position is fitted using an adaptive vertex fitter [67], where a weight  $w_i$  between 0 and 1 is assigned to every track, based on the likelihood of that track being correctly associated with the vertex. The presence of at least one primary vertex is required whose distance to the interaction point is less than 24 cm in  $z$  direction and 2 cm in

category	process	generator	cross section [pb]	processed events	weight
$t\bar{t}$	$t\bar{t} \rightarrow b\bar{b}l\nu l\nu$	Madgraph	22.40	11952631	0.04
	$t\bar{t} \rightarrow b\bar{b}q\bar{q}l\nu$	Madgraph	88.60	24913744	0.07
	$t\bar{t} \rightarrow b\bar{b}q\bar{q}q\bar{q}$	Madgraph	88.80	31172356	0.06
Drell-Yan	$Z/\gamma^* \rightarrow l^+l^-$ $10 \text{ GeV} < m_{ll} < 50 \text{ GeV}$	Madgraph	876.80	7132223	2.43
	$Z/\gamma^* \rightarrow l^+l^-$ $m_{ll} > 50 \text{ GeV}$	Madgraph	3532.80	30000624	2.33
W	$W \rightarrow l\nu$	Madgraph	37509.00	55996720	13.26
WW,WZ,ZZ	$ZZ \rightarrow l^+l^-q\bar{q}$	Madgraph	2.45	1936727	0.03
	$ZZ \rightarrow l^+l^-\nu\nu$	Madgraph	0.36	954911	0.01
	$ZZ \rightarrow l^+l^-l^+l^-$	Madgraph	0.18	4789250	0.00
	$WZ \rightarrow l\nu l^+l^-$	Madgraph	1.06	2017979	0.01
	$WZ \rightarrow q\bar{q}'l^+l^-$	Madgraph	2.32	3205557	0.01
	$WW \rightarrow l\nu l\nu$	Madgraph	5.81	1933235	0.06
single top	$t$ s-Channel	Powheg	3.79	259961	0.29
	$t$ t-Channel	Powheg	56.40	3746457	0.30
	$t$ tW-Channel	Powheg	11.10	497658	0.44
	$\bar{t}$ s-Channel	Powheg	1.76	139974	0.25
	$\bar{t}$ t-Channel	Powheg	30.70	1935072	0.31
	$\bar{t}$ tW-Channel	Powheg	11.10	493460	0.45
Other SM	$WWW$	Madgraph	0.08	220549	0.01
	$WW\gamma$	Madgraph	0.53	215121	0.05
	$WWZ$	Madgraph	0.06	222234	0.01
	$WZZ$	Madgraph	0.02	219835	0.00
	$t\bar{t}\gamma$	Madgraph	2.17	71598	0.60
	$t\bar{t}W$	Madgraph	0.23	196046	0.02
	$t\bar{t}Z$	Madgraph	0.21	210160	0.02
	$t\bar{t}WW$	Madgraph	0.00	217820	0.00
$t\bar{t}$ Systematics	$t\bar{t}$	Madgraph	227.00	6923750	0.65
	$t\bar{t}, m_{top} = 166.5 \text{ GeV}$	Madgraph	227.00	4469095	1.01
	$t\bar{t}, m_{top} = 169.5 \text{ GeV}$	Madgraph	227.00	5202817	0.86
	$t\bar{t}, m_{top} = 175.5 \text{ GeV}$	Madgraph	227.00	5186494	0.87
	$t\bar{t}, m_{top} = 178.5 \text{ GeV}$	Madgraph	227.00	4723379	0.95
	$t\bar{t}$ , Matching scale up	Madgraph	227.00	5393645	0.83
	$t\bar{t}$ , Matching scale down	Madgraph	227.00	5467170	0.82
	$t\bar{t}$ , Factorization scale up	Madgraph	227.00	5009488	0.90
	$t\bar{t}$ , Factorization scale down	Madgraph	227.00	5377388	0.84

Table 4.4.: Simulated datasets used in the analysis. The samples are grouped by physics processes and information about the generator, the cross section of the processes, the number of processed events, and the resulting weight used to scale the simulation to the recorded luminosity are given.

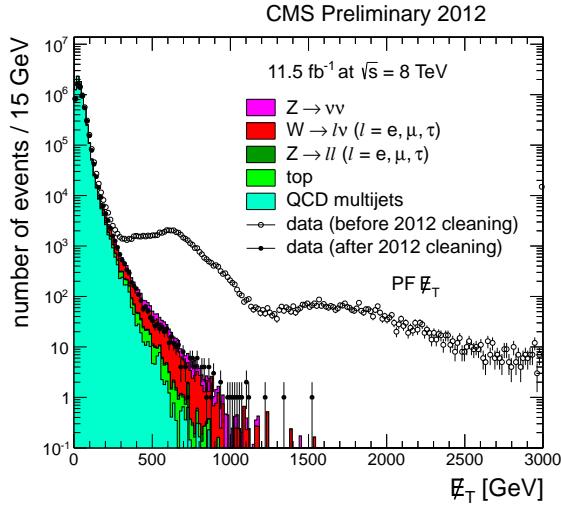


Figure 4.3.: Distribution of  $E_T^{\text{miss}}$  in dijet events in 2012 data. The open data points show all events, while the black points show the data after application of  $E_T^{\text{miss}}$  filters. Simulated Standard Model processes are shown as filled histograms. [29].

the  $x$ - $y$  plane. Also the number of degrees of freedom, defined as [13]:

$$n_{\text{dof}} = -3 + 2 \sum_{i=1}^{N_{\text{tracks}}} w_i, \quad (4.11)$$

is required to be greater than four.

As it relies on the balance of all reconstructed objects,  $E_T^{\text{miss}}$  is especially sensitive to distortions of the event reconstruction by noise or particles not originating from the proton proton collisions. Several sources of these distortions have been identified during the data taking and filters have been developed in CMS to reject events matching their signatures [29]. This includes filters for signal produced by interactions of the beam with gas molecules in the beam pipe or of protons in the beam halo with the LHC infrastructure, anomalous noise in the HCAL or ECAL, dead ECAL cells, calibration lasers mistakenly firing during collision events, or failures of the tracking algorithms. The effect of these filters on the tails of the  $E_T^{\text{miss}}$  distribution in dijet events is shown in Figure 4.3, where it can be seen that it is dominated by events that are rejected by the filters for  $E_T^{\text{miss}}$  larger than 300 GeV.

#### 4.3.2. Inclusive dilepton selection

Events are selected containing two leptons with opposite electric charge,  $p_T$  larger than 20 GeV, and  $|\eta|$  smaller than 2.4. The  $p_T$  requirement is driven by the thresholds of the dilepton triggers, as discussed in Section ??, while the  $|\eta|$  restriction is imposed by the coverage of the muon system. The acceptance for electrons could in principle be extended to  $|\eta| = 2.5$ , but is chosen

to be the same for both lepton flavours. Lepton pairs are required to be selected by the corresponding trigger, e.g. a pair of electrons has to have fired the dielectron trigger. If there is more than one pair of leptons fulfilling this basic requirements in one event, the pair with the largest sum of lepton  $p_T$  is chosen.

As the symmetry between lepton flavours is a key ingredient of the methods to estimate the backgrounds from Standard Model processes, events for which these symmetries are potentially violated are rejected.

As the efficiency to reconstruct electrons is reduced in the overlap region between the barrel and endcap detectors of the ECAL, the relative event yield for events with electrons with  $|\eta|$  between 1.4 and 1.6 is reduced compared to those with muons in this range. This distribution of the  $|\eta|$  of the leading lepton in  $e^\pm e^\mp$  and  $\mu^\pm \mu^\mp$  events is shown in Figure 4.4 (left), illustrating the greatly increased difference between the event yields for electrons and muons in the overlap region. Events containing a lepton with a pseudorapidity of  $1.4 < |\eta| < 1.6$  are therefore rejected. Also an increasing difference between electrons and muons can be seen for events where the leading leptons is in the endcap region of the detector. This is one of the reasons for splitting the event sample in two categories: *central*, where both leptons are reconstructed with  $|\eta| < 1.4$  and *forward*, where at least one lepton has to be reconstructed with  $|\eta| > 1.6$ .

Leptons with small spatial separation can interfere with each other's reconstruction and isolation. These effects are different for electrons and muons, which can be seen in Figure 4.4 (right). The ratio of electrons to muons first rises for lower values of  $\Delta R(l\bar{l})$  before dropping for values below 0.1. The leptons are therefore required to be separated in  $\Delta R(l\bar{l})$  by more than 0.3. Some differences between electrons and muons can also be observed for very high values of  $\Delta R(l\bar{l})$ , but they are less pronounced and this region is less populated.

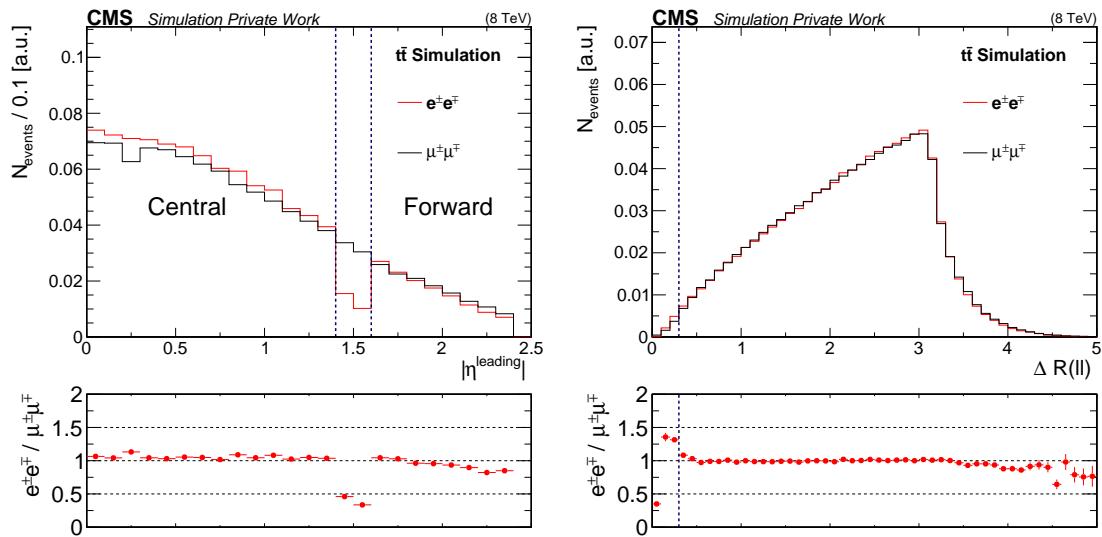


Figure 4.4.: istribution of  $|\eta|$  (left) and  $\Delta R(l\bar{l})$  (right) for the leading lepton for  $\mu^\pm \mu^\mp$  (black dots) and  $e^\pm e^\mp$  (red dots) events in a simulation of  $t\bar{t}$  events. Both distributions are normalized to the same area.

As an additional requirement, to avoid possible reconstruction problems events with low lepton momenta and to avoid contamination from dilepton production in the decay of the bottomonium resonances, the dilepton invariant mass  $m_{\ell\ell}$  is required to be greater than 20 GeV.

### 4.3.3. Selections in $E_T^{\text{miss}}$ and jet multiplicity

Three subsets of the event sample obtained with the inclusive dilepton selection are defined, resulting in samples enriched in different processes. The variables used in the definitions of these regions are  $E_T^{\text{miss}}$  and the number of selected jets  $N_{\text{jets}}$ . The selections are illustrated in the plots of Figure 4.5, which also show the distribution of  $t\bar{t}$  (left) and Drell-Yan (right) events in the  $E_T^{\text{miss}}\text{-}N_{\text{jets}}$  plane.

The signal region, in which the search will be performed, is defined by requiring either  $N_{\text{jets}} \geq 3$  and  $E_T^{\text{miss}} > 100$  GeV or  $N_{\text{jets}} \geq 3$  and  $E_T^{\text{miss}} > 100$  GeV. This definition allows to select signal events for points in the parameter space where more energy is distributed to the jets and less to the invisible component of the signature and vice versa. At the same time the rejection of background events with both lower  $N_{\text{jets}}$  and  $E_T^{\text{miss}}$  is maintained. A control region dominated by flavour-symmetric processes is defined by selecting events with  $N_{\text{jets}} = 2$  and  $100 \text{ GeV} < E_T^{\text{miss}} < 150 \text{ GeV}$ .

To study lepton pairs produced via the Drell-Yan process and to obtain a high statistics sample of leptons for efficiency measurements, events with  $N_{\text{jets}} \geq 2$  and  $E_T^{\text{miss}} < 50$  GeV are selected. The  $N_{\text{jet}}$  requirement greatly reduces the statistics and the purity of this event sample. However, because of the large cross section of the Drell-Yan process, the event yield is still sufficient for the purposes of this analysis and Drell-Yan events dominate over those from  $t\bar{t}$  production by two orders of magnitude. This allows to select events with kinematics close to those of the signal selection in terms of jet multiplicity.

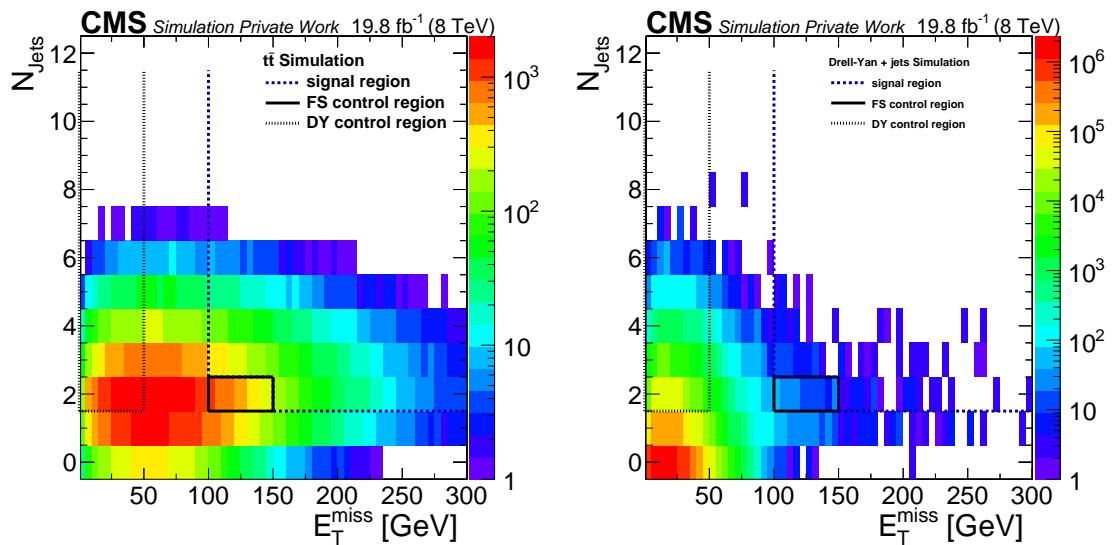


Figure 4.5.: Distribution of backgrounds events in the  $E_T^{\text{miss}}\text{-}N_{\text{jets}}$  plane for  $t\bar{t}$  (left) and Drell-Yan (right) events from Simulation. The events are weighted according to the cross section of the process and the size of the generated event sample, assuming an integrated luminosity of  $19.5\text{fb}^{-1}$ . The three regions defined in the plane are indicated by lines.

---

## 5. Estimation of Standard Model backgrounds

As indicated in Figure 4.5, different Standard Model processes contribute to the event sample in the signal region. To distinguish a potential signal from these backgrounds, a precise estimation of the background contributions is mandatory. While the simulation of these processes and the response of the CMS detector gives a good description of the data for the majority of the phase space, a large number of uncertainty sources are introduced in the modelling of the physical process and the detector. Therefore a higher precision can be achieved by deriving the background estimates directly from the recorded data. The background processes are categorized as either being flavour-symmetric or as containing the production of a Z boson. A dedicated method is applied for each of the two categories. The different background processes are:

### 5.1. Flavour-symmetric backgrounds

Processes that are symmetric in the production of same-flavour and opposite-flavour lepton pairs allow for the estimation of their contribution to the SF event sample from the OF one. The most dominant of these processes is the dileptonic decay of top-pair production, where the leptons are produced uncorrelatedly in the decay of the W bosons. Other examples are the decays of two  $\tau$  leptons, which are in turn produced in the decay of a Z boson or the dileptonic decay of W pairs. Another contribution to this class of backgrounds are misidentified leptons, as will be demonstrated later.

No significant deviation from flavour-symmetry has been observed in the decays of the W boson, with a measured ratio of the branching fractions into  $e + \nu$  and  $\mu + \nu$  of  $1.007 \pm 0.021$ . In the decays of the  $\tau$  lepton the different masses of electron and muon have a noticeable effect, resulting in a slightly favoured decay into electrons. Here the ratio of branching fractions is  $1.0241 \pm 0.0032$  [38]. As backgrounds with  $\tau$  leptons are a sub-dominant contribution to the flavour-symmetric backgrounds, these can be considered to be fully flavour-symmetric on particle level. However, distortions of the flavour-symmetry are introduced by the different efficiencies for triggering, reconstructing, and identifying electrons and muons in CMS. The background estimation from OF events therefore has to include a correction for this deviation, which is applied as a multiplicative factor:

$$N_{SF}^{pred} = R_{SF/OF} \cdot N_{OF}. \quad (5.1)$$

Two independent methods are utilized to measure  $R_{SF/OF}$  on data. In the first approach is is directly measured as the ratio of SF to OF events in the control region for flavour-symmetric backgrounds. The second approach studies the lepton efficiencies and derives  $R_{SF/OF}$  factorized into the effects of trigger efficiencies and reconstruction and identification efficiencies.

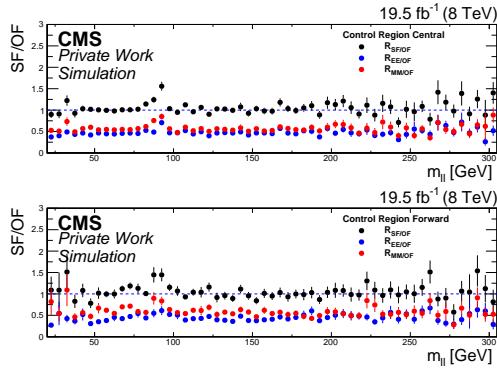


Figure 5.1.: Ratio of SF to OF events as a function of  $m_{\ell\ell}$  in the  $t\bar{t}$  control region for the dileptonic  $t\bar{t}$  sample. Shown are the results for the central (top) and forward (bottom) lepton selection.

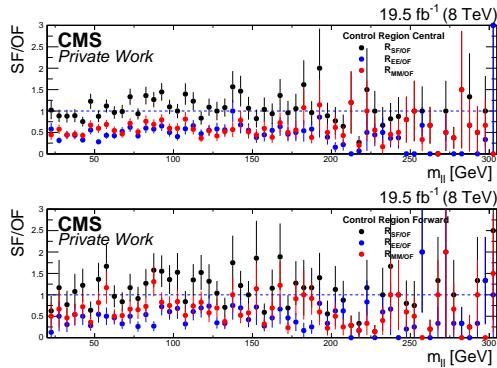


Figure 5.2.: Ratio of SF to OF events as a function of  $m_{\ell\ell}$  in the  $t\bar{t}$  control region on data. Shown are the results for the central (top) and forward (bottom) lepton selection.

### 5.1.1. Direct measurement of $R_{SF/OF}$

The ratio of SF to OF events as a function of  $m_{\ell\ell}$  in simulation is shown in Figure 5.1, separately for the central and forward lepton selection. The ratio is close to one, independent of  $m_{\ell\ell}$ , allowing for the same value of  $R_{SF/OF}$  to be applied over the full mass range. In data however there are contributions from non-flavour-symmetric processes containing Z bosons at the Z peak. Therefore the on-Z mass bin is excluded from the calculation and only events in the mass regions  $20 \text{ GeV} < m_{\ell\ell} < 70 \text{ GeV}$  and  $m_{\ell\ell} > 120 \text{ GeV}$  are considered. The observed ratio on data as a functions of  $m_{\ell\ell}$  is shown in Figure 5.2. No dependence on  $m_{\ell\ell}$  is observed. The numerical results are summarized in Table ??

### 5.1.2. Determination of $R_{SF/OF}$ with the factorization method

Asymmetries between the lepton flavours introduced by differing reconstruction and selection efficiencies can be corrected for if the ratio of efficiencies for muons and electrons  $r_{\mu e} = \frac{\epsilon_\mu}{\epsilon_e}$  is

Table 5.1.

	$N_{SF}$	$N_{OF}$	$R_{SF/OF} \pm \sigma_{stat}$	Transfer factor $\pm \sigma_{stat}$
Central				
Data	1458	1448	$1.007 \pm 0.037$	–
MC	1448.2	1437.2	$1.008 \pm 0.015$	$1.016 \pm 0.021$
Forward				
Data	545	537	$1.015 \pm 0.062$	–
MC	552.6	548.4	$1.008 \pm 0.025$	$0.996 \pm 0.035$
	$N_{SF}$	$N_{OF}$	$R_{ee/OF} \pm \sigma_{stat}$	Transfer factor $\pm \sigma_{stat}$
Central				
Data	663	1448	$0.458 \pm 0.021$	–
MC	651.9	1437.2	$0.454 \pm 0.008$	$1.017 \pm 0.026$
Forward				
Data	239	537	$0.445 \pm 0.035$	–
MC	231.5	548.4	$0.422 \pm 0.013$	$1.023 \pm 0.045$
	$N_{SF}$	$N_{OF}$	$R_{\mu\mu/OF} \pm \sigma_{stat}$	Transfer factor $\pm \sigma_{stat}$
Central				
Data	795	1448	$0.549 \pm 0.024$	–
MC	796.3	1437.2	$0.554 \pm 0.010$	$1.015 \pm 0.026$
Forward				
Data	306	537	$0.570 \pm 0.041$	–
MC	321.1	548.4	$0.586 \pm 0.017$	$0.977 \pm 0.041$

known. Under the assumption that the efficiencies for the two leptons in the event factorize, i.e.  $\epsilon_{ll} = \epsilon_l \cdot \epsilon_l$ , the number of dielectron and dimuon events can be estimated from the opposite-flavour events using the relations

$$n_{ee}^* = \frac{1}{2} \cdot \frac{n_{OF}^*}{r_{\mu e}^*} \quad (5.2)$$

and

$$n_{\mu\mu}^* = \frac{1}{2} \cdot r_{\mu e}^* \cdot n_{OF}^*, \quad (5.3)$$

the \* indicating that these are the values unaffected by trigger efficiencies. The prediction of the combined same-flavour yield is therefore given by

$$n_{SF}^* = \frac{1}{2} \cdot \left( r_{\mu e}^* + \frac{1}{r_{\mu e}^*} \right) n_{OF}^*. \quad (5.4)$$

In practice, all measured quantities are affected by the efficiencies of the different dilepton triggers. The measured number of SF events will therefore be

$$n_{SF} = \epsilon_{ll}^T \cdot n_{ee}^* + \epsilon_{\mu\mu}^T \cdot n_{\mu\mu}^*, \quad (5.5)$$

where  $\epsilon_{ll}^T$  denotes the trigger efficiency for the given dilepton combination.

Also the predictions for  $n_{ee}^*$  and  $n_{\mu\mu}^*$  have to include the trigger efficiencies. Here  $r_{\mu e}^*$  is expressed in terms of the measured value  $r_{\mu e}$ , which is derived from the  $e^\pm e^\mp$  and  $\mu^\pm \mu^\mp$  event yields in the Drell-Yan control region (see Section 5.1.2) as

$$r_{\mu e} = \sqrt{\frac{N_{\mu\mu}}{N_{ee}}} \approx \sqrt{\frac{\epsilon_\mu^2 \epsilon_{\mu\mu}^T}{\epsilon_e^2 \epsilon_{ee}^T}} = r_{\mu e}^* \cdot \sqrt{\frac{\epsilon_{\mu\mu}^T}{\epsilon_{ee}^T}}. \quad (5.6)$$

Taking into account also that the measured event yield in the OF channel is  $n_{OF} = \epsilon_{e\mu}^T \cdot n_{OF}^*$ , the estimate for the yields in the same-flavour channel becomes

$$n_{ee} = \frac{1}{2r_{\mu e}} \cdot \frac{\sqrt{\epsilon_{ee}^T \epsilon_{\mu\mu}^T}}{\epsilon_{e\mu}^T} n_{OF} = R_{ee/OF} n_{OF} \quad (5.7)$$

and

$$n_{\mu\mu} = \frac{1}{2} r_{\mu e} \cdot \frac{\sqrt{\epsilon_{ee}^T \epsilon_{\mu\mu}^T}}{\epsilon_{e\mu}^T} n_{OF} = R_{\mu\mu/OF} n_{OF}. \quad (5.8)$$

Finally, the combined prediction of the SF yield is

$$n_{SF} = \frac{1}{2} \left( r_{\mu e} + \frac{1}{r_{\mu e}} \right) \cdot \frac{\sqrt{\epsilon_{ee}^T \epsilon_{\mu\mu}^T}}{\epsilon_{e\mu}^T} n_{OF} = R_{SF/OF} n_{OF}. \quad (5.9)$$

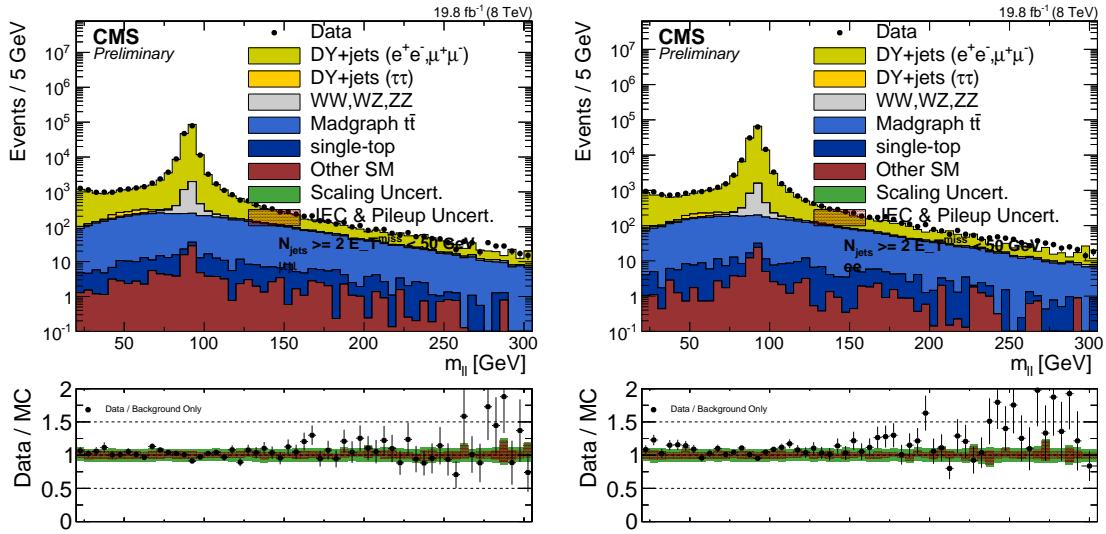


Figure 5.3.: Distribution of  $m_{ll}$  in the Drell-Yan control region for  $\mu^{\pm}\mu^{\mp}$  events (left) and  $e^{\pm}e^{\mp}$  events (right). The data is shown as the black dots, while the contributions from Standard Model processes, estimated from simulation, are shown as the stacked histograms.

### Measurement of $r_{\mu e}$

The measurement of  $r_{\mu e}$  is performed in the Drell-Yan control region as the ratio of  $\mu^{\pm}\mu^{\mp}$  to  $e^{\pm}e^{\mp}$  events on the Z peak, requiring  $60 \text{ GeV} < m_{ll} < 120 \text{ GeV}$ . A comparison of the recorded data to the different contributions from Standard Model processes, estimated from simulation, is shown in Figure 5.3. The Drell-Yan process is the dominating source of events in this selection. Good agreement between data and simulation is observed, indicating a good understanding of this kinematic region by CMS. The results of the calculation of  $r_{\mu e}$  are shown in Table ???. Given are the observed yields for  $\mu^{\pm}\mu^{\mp}$  and  $e^{\pm}e^{\mp}$  events and the resulting value of  $r_{\mu e}$  with statistical and systematic uncertainties. In the central lepton selection, the  $\mu^{\pm}\mu^{\mp}$  yield is about 18% higher than the  $e^{\pm}e^{\mp}$  yield. Similar results are observed on Drell-Yan simulation. For events with leptons in the forward region, a larger asymmetry between muons and electrons is observed, here the  $\mu^{\pm}\mu^{\mp}$  yield is about 40% higher than the  $e^{\pm}e^{\mp}$  yield.

Table 5.2.

	$N_{\mu\mu}$	$N_{ee}$	$r_{\mu e} \pm \sigma_{\text{stat.}} \pm \sigma_{\text{syst.}}$
Central			
Data	98284	83035	$1.09 \pm 0.01 \pm 0.11$
MC	102173	78811	$1.14 \pm 0.01 \pm 0.11$
Forward			
Data	62212	44437	$1.18 \pm 0.01 \pm 0.24$
MC	68090	43665	$1.25 \pm 0.01 \pm 0.25$

The systematic uncertainties of the measurement 10% for the central and 20% for the forward lepton selection. These values are obtained from studies of the dependency of  $r_{\mu e}$  on relevant observables. These are on the one hand properties of the lepton pairs, while on the other hand the jet multiplicity and  $E_T^{\text{miss}}$  are studied to asses potential biases introduced when applying  $R_{\text{SF/OF}}$  in the signal region. The dependencies of  $r_{\mu e}$  on  $m_{\ell\ell}$ ,  $E_T^{\text{miss}}$ , and  $N_{\text{jets}}$  are shown in Figure 5.4. Some dependency is observed in the case of  $m_{\ell\ell}$ , where the values are higher for low  $m_{\ell\ell}$  below the Z peak. This can be traced back to a dependency on the  $p_T$  of the leptons, where the efficiency for muons has a sharper turnon compared to electrons, for example for the dilepton triggers. In data, there is a strong effect visible around the Z peak for forward leptons. This is caused by a systematic shift of the position of the Z peak between electron and muons and is not a general property of  $r_{\mu e}$ , as is evident from the comparison with  $t\bar{t}$  simulation shown in the upper right of Figure 5.4. No strong dependencies can be observed for  $E_T^{\text{miss}}$  and  $N_{\text{jets}}$  within the statistical uncertainties. All observed deviations from the central values are covered by the systematic uncertainties assigned to the measurement. Further information on the dependency studies can be found in Appendix B

### Measurement of $R_T$

The trigger efficiencies are measured utilizing an event sample collected with PF  $H_T$  triggers. Lepton pairs corresponding to the flavour combination of the trigger in question are selected. To ensure that only correctly reconstructed events are considered in the calculation of the trigger efficiencies, the events are required to have  $H_T > 200 \text{ GeV}$ , which corresponds to the lowest threshold applied in one of the PF  $H_T$  trigger. To ensure that the factorization method is performed on event samples complete orthogonal to the signal region and the  $t\bar{t}$  enriched control region, events with  $N_{\text{jets}} \geq 2$  and  $E_T^{\text{miss}} > 100 \text{ GeV}$  are rejected. As the offline lepton selection has more strict requirements compared to the one applied at HLT level, the dilepton trigger should have accepted all these events. The trigger efficiency is therefore defined as the ratio of these event that were accepted by the trigger by the total number of events:

$$\epsilon_{ll}^T = \frac{N_{\text{events}}(\text{PF } H_T \text{ trigger} \cap \text{dilepton selection} \cap \text{dilepton trigger})}{N_{\text{events}}(\text{PF } H_T \text{ trigger} \cap \text{dilepton selection})}. \quad (5.10)$$

In the case of  $\mu^\pm\mu^\mp$  and OF events, two triggers are used in the event selection. Therefore their combined efficiency is measured, requesting the logical OR of both. The resulting efficiencies are summarized in Table ??, separately for the central and forward lepton selection. The trigger efficiencies for  $e^\pm e^\mp$  and  $\mu^\pm\mu^\mp$  are about 97% in both the central and forward selection. The efficiency for OF events is lower, about 95% in the central and 90% in the forward selection. This shows how the flavour symmetry is broken at trigger level by the HLT\_Mu17\_TkMu8 trigger, which recovers efficiency for the trailing muon leg of the trigger and is not present in the OF triggers.

To asses the systematic uncertainties of the trigger efficiency measurement, potential biases due to the choice of the orthogonal trigger are studied on simulation, using dileptonic  $t\bar{t}$  events. On simulation the efficiency can be measured without the requirement of the PF  $H_T$  triggers, as no trigger is needed to select the events. The ratio of this true trigger efficiencies to those measured with the PF  $H_T$  triggers as a function of  $m_{\ell\ell}$  is shown in Figure 5.5, separately for

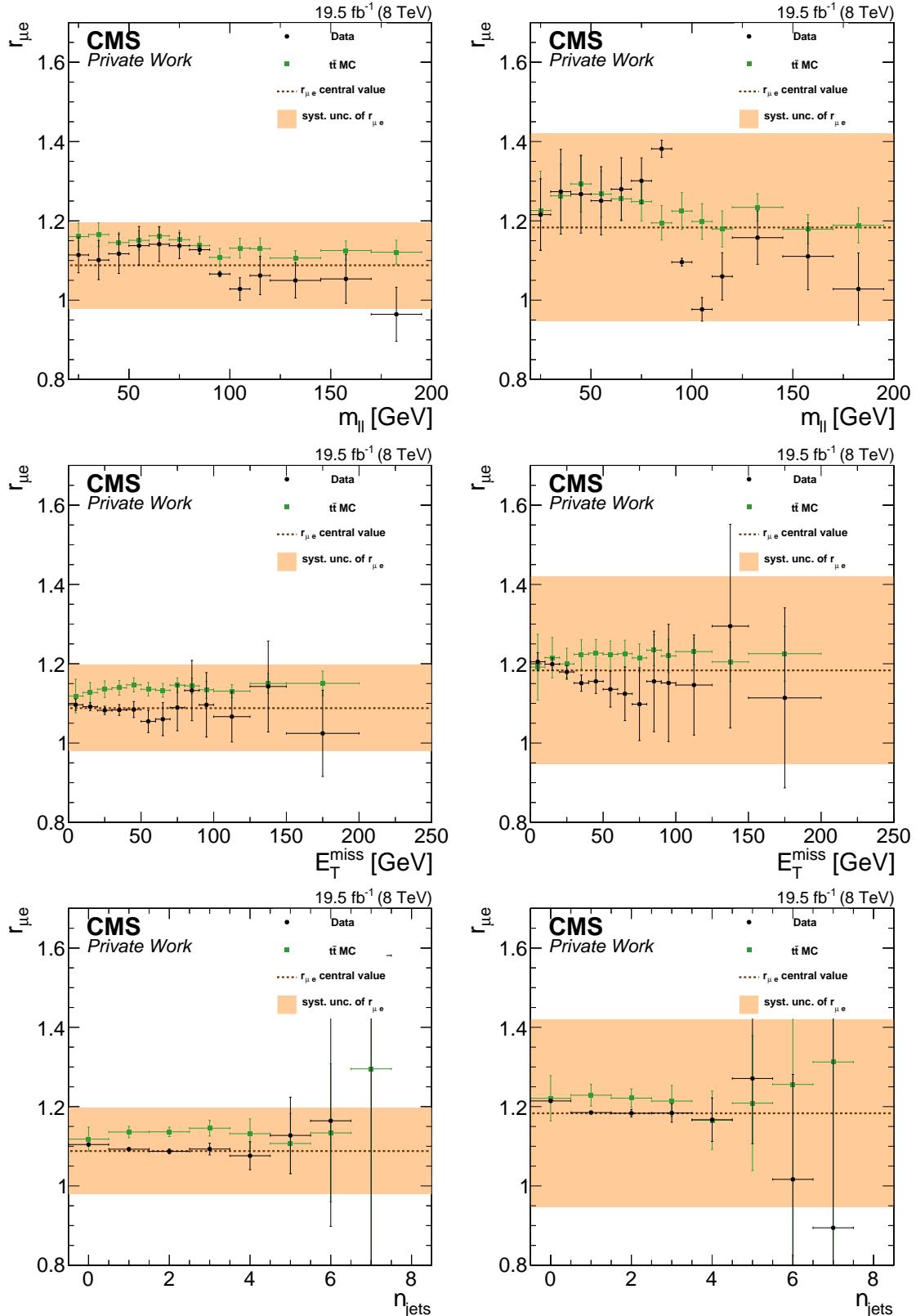


Figure 5.4.: Dependencies of  $r_{\mu e}$  on  $m_{ll}$  (top),  $E_T^{\text{miss}}$  (middle), and  $N_{\text{jets}}$  (bottom) for the central (left) and forward (right) lepton selection. The results on data are shown in black while  $t\bar{t}$  simulation is shown in green. The central value is shown as a brown dashed line while the systematic uncertainty is shown as an orange band.

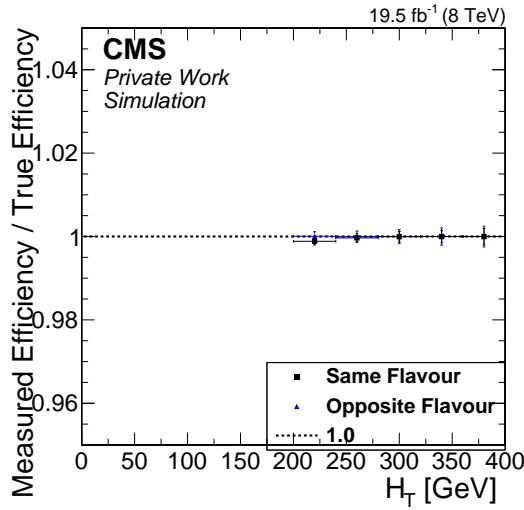


Figure 5.5.: Ratio of true trigger efficiencies to those calculated using the PF  $H_T$  triggers separately for SF and OF events.

SF and OF events. In both cases all deviations from one are in the order of 0.1% and well compatible with it inside the statistical uncertainties. No bias of the efficiency measurement due to the PF  $H_T$  triggers is observed and therefore no systematic uncertainty due to this choice is assigned to the measurement. The turnon of the triggers at low lepton  $p_T$  is of particular interest because asymmetries between the lepton flavours are more likely to occur in this difficult environment. The dependence of the trigger efficiency on the  $p_T$  of the trailing lepton can be studied using datasets trigger by single lepton triggers. Triggers with  $p_T$  thresholds of 24(27) GeV for muons (electrons) are available. These thresholds are relatively low for single lepton triggers, resulting in very strict selection criteria having to be applied on HLT level. Dilepton events are selected in which the leading lepton can be matched geometrically to the trigger object that fired the single lepton trigger. The trailing lepton is matched to trigger objects that have fired the trailing leg of the dilepton trigger. The efficiency is defined as the number of events in which such a match can be found to the total number of dilepton events. The resulting efficiency curves are shown in Figure 5.6 for five different trailing lepton trigger legs and combining the central and forward selection. It can be seen that the efficiency for trailing electrons are virtually the same for the dielectron and OF triggers (black and light blue markers). The same is true for trailing muons in the dimuon and OF triggers (brown and dark blue markers). However, the efficiency for trailing muons in the `HLT_Mu17_TkMu8` trigger is significantly higher than those in the trigger paths not using the tracker information, as expected. Trailing muons have a much sharper turnon than trailing electrons, being fully efficient already at a  $p_T$  of 10 GeV. For electrons the plateau is reached only above  $\approx 30$  GeV. The turnon is steeper for trailing electrons in the dielectron trigger compared to those in the OF trigger, resulting in an increased deviation from flavour symmetry below 20 GeV. As in the design of the analysis the precision and stability of the background prediction was given priority over lepton acceptance, events with trailing leptons in this  $p_T$  range are rejected. To asses the systematic uncertainties of the trigger efficiency measurement, the dependency of  $R_T$  as used in the calculation of  $R_{SF/OF}$  on different observables is tested. Here a dataset

Table 5.3.: Trigger efficiency-values for data and MC with OS,  $p_T > 20(20)$  GeV and  $H_T > 200$  GeV for central and forward region separated.

	nominator	denominator	$\epsilon_{trigger} \pm \sigma_{stat}$	nominator	denominator	$\epsilon_{trigger} \pm \sigma_{stat}$
Data						
	Central			Forward		
ee	3592	3692	$0.973 \pm 0.003$	954	980	$0.973 \pm 0.006$
$\mu\mu$	1375	1420	$0.968 \pm 0.005$	547	566	$0.966 \pm 0.009$
e $\mu$	493	521	$0.946 \pm 0.012$	102	114	$0.895 \pm 0.037$
MC						
	Central			Forward		
ee	2912.8	2994.6	$0.973 \pm 0.001$	943.0	969.9	$0.972 \pm 0.001$
$\mu\mu$	3241.6	3287.9	$0.986 \pm 0.000$	1141.6	1183.4	$0.965 \pm 0.001$
e $\mu$	6279.1	6523.1	$0.963 \pm 0.000$	2138.8	2249.0	$0.951 \pm 0.001$

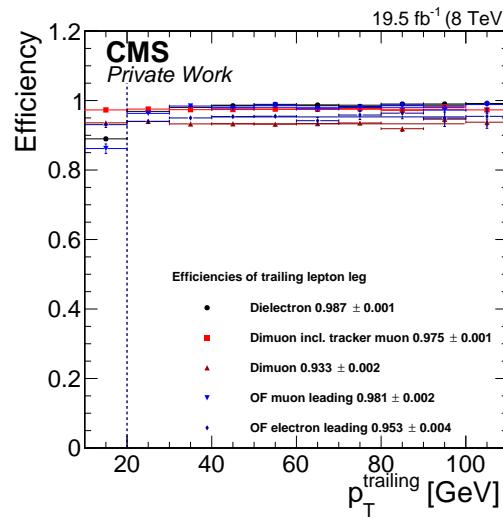


Figure 5.6.: Ratio of true trigger efficiencies to those calculated using the PF  $H_T$  triggers separately for SF and OF events.

not triggered by  $\alpha_T$  triggers is used, as the PF  $H_T$  triggers do not provide enough events for these studies. A 5% systematic uncertainty is assigned to each trigger efficiency, resulting in an uncertainty of 6.4% on  $R_T$ , covering all observed deviations from the measured value of  $R_T$  within the statistical uncertainties, as shown in Figure 5.7. Studies of further variables can be found in Appendix C

### Results of the factorization method

The results of the factorization methods are summarized in Table 5.4. The factor  $R_T$  is calculated from the trigger efficiencies in Table 5.3 as  $R_T = \sqrt{\frac{\epsilon_{ee}^T \cdot \epsilon_{\mu\mu}^T}{\epsilon_{e\mu}^T}}$ . The resulting correction factors  $R_{SF/OF}$ ,  $R_{ee/OF}$ , and  $R_{\mu\mu/OF}$  are calculated as described in equations 5.7-5.9.

Table 5.4.

	Central		Forward	
	Data	MC	Data	MC
$r_{\mu e}$	$1.088 \pm 0.109$	$1.139 \pm 0.114$	$1.183 \pm 0.237$	$1.249 \pm 0.250$
$R_T$	$1.026 \pm 0.066$	$1.017 \pm 0.064$	$1.084 \pm 0.079$	$1.018 \pm 0.065$
$R_{SF/OF}$	$1.029 \pm 0.067$	$1.026 \pm 0.066$	$1.099 \pm 0.088$	$1.044 \pm 0.081$
$R_{ee/OF}$	$0.471 \pm 0.116$	$0.447 \pm 0.119$	$0.458 \pm 0.259$	$0.408 \pm 0.256$
$R_{\mu\mu/OF}$	$0.558 \pm 0.117$	$0.579 \pm 0.122$	$0.641 \pm 0.261$	$0.636 \pm 0.258$

As for the direct measurements in the control region, the observed deviations of  $R_{SF/OF}$  from one are small and inside the uncertainties of the method. Because the uncertainty on  $r_{\mu e}$  cancels out to a large degree in the calculation  $R_{SF/OF}$ , the total uncertainty is dominated by the uncertainty on  $R_T$ , while for  $R_{ee/OF}$  and  $R_{\mu\mu/OF}$  the uncertainty of  $r_{\mu e}$  is the dominant one. This results in much larger uncertainties on the latter two factors. The dependency of  $R_{SF/OF}$  on  $r_{\mu e}$  and  $R_T$  is illustrated in Figure 5.8. The measured values of  $r_{\mu e}$  and  $R_{SF/OF}$  are shown as straight lines, surrounded by bands indicating their uncertainties. The red line shows the value of  $R_{SF/OF}$  corresponding to a given  $r_{\mu e}$ , under the assumption that the trigger efficiencies stay constant. The dashed black lines show the impact that the uncertainties on the trigger efficiencies have on the resulting  $R_{SF/OF}$ . It can be seen that variations of  $r_{\mu e}$  inside its systematic uncertainties have only little effect on  $R_{SF/OF}$ , while changes in the trigger efficiencies have a much larger impact.

### 5.1.3. Combined correction factors and resulting background estimates

An overview of the resulting corrections factors of the measurement in the control region and the factorization method are shown in Table 5.5. In all cases the results of the two methods agree very well within their uncertainties. Given the fact that they are performed on exclusive datasets and under the assumption that their uncertainties follow a Gaussian distribution,

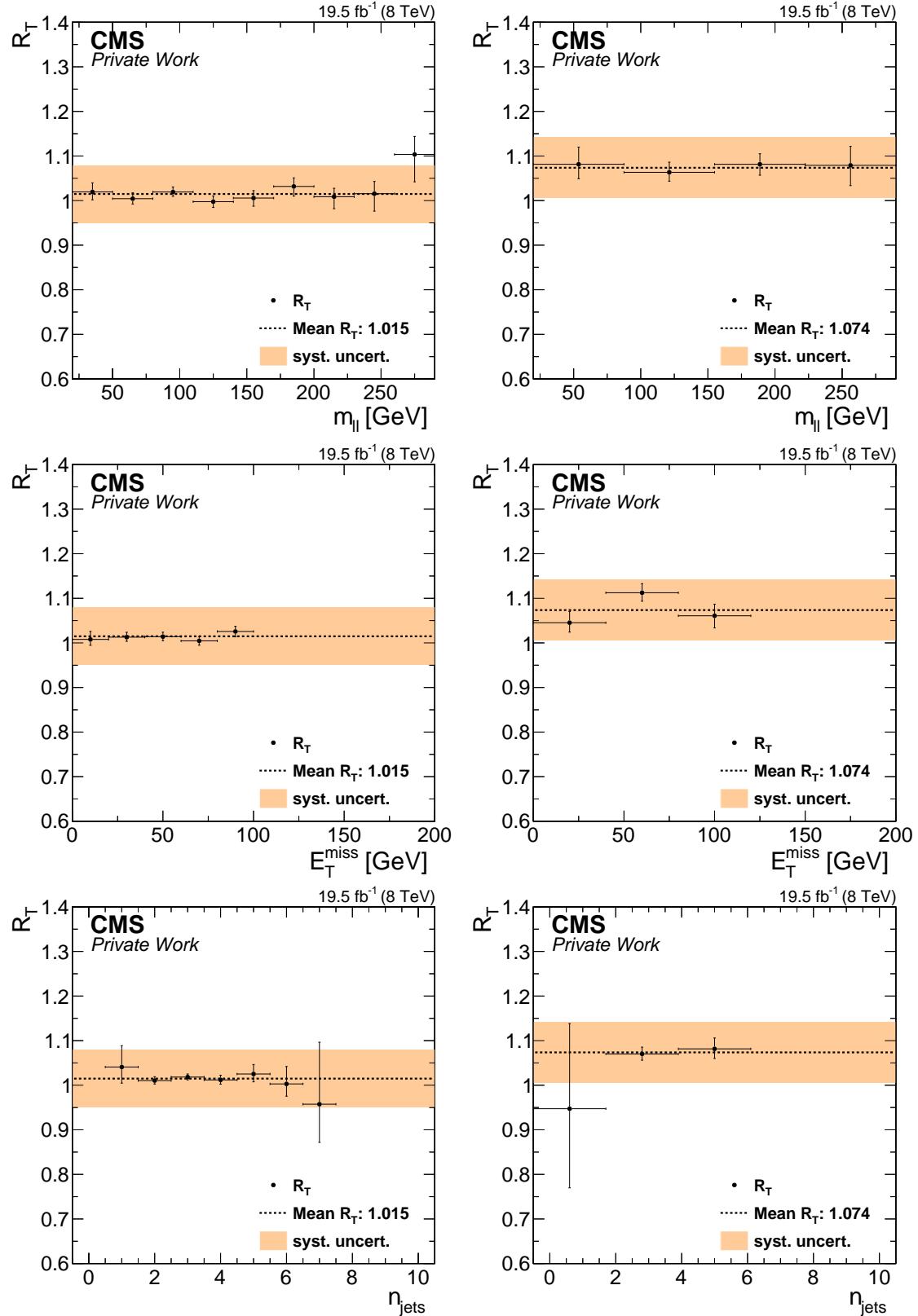


Figure 5.7.: Dependencies of  $R_T$  on  $m_{ll}$  (top),  $E_T^{\text{miss}}$  (middle), and  $N_{\text{jets}}$  (bottom) for the central (left) and forward (right) lepton selection. The results on data are shown in black. The central value is shown as a black dashed line while the systematic uncertainty is shown as an orange band.

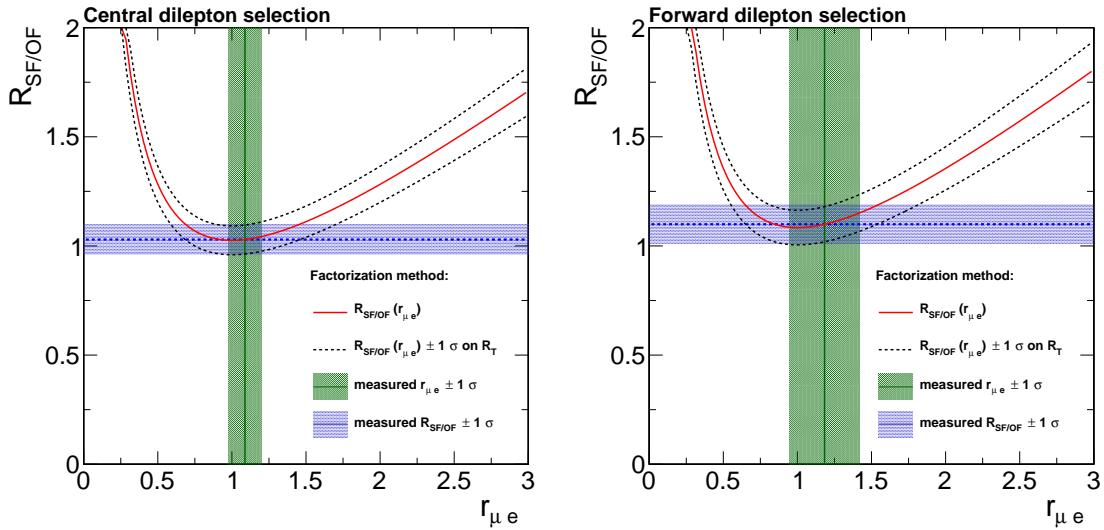


Figure 5.8.: Dependency of  $R_{SF/OF}$  on  $r_{\mu e}$  in the factorization method. The measured central values of  $R_{SF/OF}$  and  $r_{\mu e}$  and their uncertainty are shown as the blue and green lines and bands. The red line illustrates the value of  $R_{SF/OF}$  for a given  $r_{\mu e}$  if  $R_T$  is kept constant. The impact of a variation of  $R_T$  within its uncertainties on  $R_{SF/OF}$  is shown by the dashed black lines.

which is justified by the fact that they are either statistical in nature or assigned to cover statistical fluctuations in the dependency studies of  $r_{\mu e}$  and  $R_T$ , they can be combined using a weighted average.

A precision of about 4% and 6% on the translation from opposite flavour to same flavour is reached in the central and forward dilepton selection, respectively. Separating the lepton flavours, the uncertainty in  $R_{ee}/OF$  and  $R_{\mu\mu}/OF$  is about 6% for central and 9-10% for forward leptons. This increase is caused by the missing cancellation of the uncertainty on  $r_{\mu e}$  and the increased statistical uncertainty in the direct measurement in the control region. Given that the weighted average is calculated separately for each flavour combination,  $R_{ee}/OF$  and  $R_{\mu\mu}/OF$  do not add up exactly to  $R_{SF/OF}$ , which in turn means that the sum of the background estimates for flavour-symmetric backgrounds in the  $e^\pm e^\mp$  and  $\mu^\pm \mu^\mp$  channels does not equal that in the SF channel. The OF yields in the  $m_{\ell\ell}$  bins of the counting experiment together with the resulting background estimates in the different dilepton channels are shown in Table 5.6.

## 5.2. Backgrounds containing a Z boson

To estimate the contribution of backgrounds containing a Z boson to the event sample in the signal region, both the Jet-Z-balance (JZB) and  $E_T^{\text{miss}}$  templates methods [68] are used. The first studies the balance of the Z boson against the jets in the event. In the  $E_T^{\text{miss}}$  template method, the  $E_T^{\text{miss}}$  distribution of  $\gamma + \text{jets}$  events are used to estimate that of Z+jets events. As the development and application of these methods have not been part of the work covered in this thesis, only a short description will be given.

Table 5.5.

	$R_{\text{SF}/\text{OF}}$			
	Central		Forward	
	Data	MC	Data	MC
from factorization method	1.029±0.067	1.026±0.066	1.099±0.088	1.044±0.081
from direct measurement	1.007±0.043	1.008±0.015	1.015±0.106	1.008±0.025
weighted avarage	1.013±0.036	1.009±0.015	1.048±0.055	1.011±0.024

	$R_{ee/\text{OF}}$			
	Central		Forward	
	Data	MC	Data	MC
from factorization method	0.471±0.116	0.447±0.119	0.458±0.259	0.408±0.256
from direct measurement	0.458±0.034	0.454±0.008	0.445±0.092	0.422±0.013
weighted avarage	0.459±0.027	0.454±0.008	0.445±0.045	0.422±0.013

	$R_{\mu\mu/\text{OF}}$			
	Central		Forward	
	Data	MC	Data	MC
from factorization method	0.558±0.117	0.579±0.122	0.641±0.261	0.636±0.258
from direct measurement	0.549±0.035	0.554±0.010	0.570±0.095	0.586±0.017
weighted avarage	0.550±0.030	0.554±0.010	0.572±0.050	0.586±0.017

Table 5.6.: Resulting estimates for flavour-symmetric backgrounds. Given is the observed event yield in  $e^\pm\mu^\mp$  events and the resulting estimate after applying the correction, separately for the SF,  $e^\pm e^\mp$ , and  $\mu^\pm\mu^\mp$  channels. Statistical and systematic uncertainties are given separately. Low-mass refers to  $20 < m_{\ell\ell} < 70$  GeV, on-Z to  $81 < m_{\ell\ell} < 101$  GeV and high-mass to  $m_{\ell\ell} > 120$  GeV.

	Low-mass		On-Z		High-mass	
	Central	Forward	Central	Forward	Central	Forward
Observed OF events	737	138	364	131	779	393
Estimate in SF channel	746 ± 27 ± 26	144 ± 12 ± 7	368 ± 19 ± 13	137 ± 11 ± 7	789 ± 28 ± 28	411 ± 20 ± 21
Estimate in $e^\pm e^\mp$ channel	337 ± 12 ± 19	61 ± 5 ± 6	166 ± 8 ± 9	58 ± 5 ± 5	357 ± 12 ± 21	175 ± 8 ± 17
Estimate in $\mu^\pm\mu^\mp$ channel	405 ± 14 ± 21	79 ± 6 ± 6	200 ± 10 ± 10	74 ± 6 ± 6	428 ± 15 ± 23	224 ± 11 ± 19

### 5.2.1. JZB method

The JZB variable is defined as the balance of the  $p_T$  of the jets in the event with the  $p_T$  of the Z candidate. To avoid biases due to jet selection,  $\vec{E}_T^{\text{miss}}$  is used as a measure of the hadronic recoil of the Z:

$$\text{JZB} = \left| \sum_{\text{jets}} \vec{p}_T \right| - |\vec{p}_T^Z| \approx |\vec{E}_T^{\text{miss}} - \vec{p}_T^Z| - |\vec{p}_T^Z|. \quad (5.11)$$

For SM processes such as  $Z + \text{jets}$ , where  $E_T^{\text{miss}}$  is caused by mismeasurements of the jets, the JZB distribution is symmetric around 0. For BSM processes, where the Z boson is produced correlated with invisible particles, the JZB distribution is expected to be asymmetric, favouring positive sign especially for large values of JZB and by extension high  $E_T^{\text{miss}}$ . Therefore it is possible to predict the contribution of SM processes containing a Z boson at high  $E_T^{\text{miss}}$  from the events in that region with negative values of JZB, after subtraction of flavour-symmetric processes from OF events. The assumption that the JZB distribution is symmetric around 0 for SM processes with only instrumental  $E_T^{\text{miss}}$  is studied in MC and 20% systematic uncertainty are assigned. **FIXME:** Add reference to 2012 paper.

### 5.2.2. $E_T^{\text{miss}}$ templates method

The  $E_T^{\text{miss}}$  templates method utilizes the similarity of  $Z + \text{jets}$  and  $\gamma + \text{jets}$  events, especially mismeasurement as only source of  $E_T^{\text{miss}}$ . Therefore, after corrections for residual kinematic differences, the contribution of  $Z + \text{jets}$  events at high  $E_T^{\text{miss}}$  can be estimated from  $\gamma + \text{jets}$  events passing the same selection. The dominant systematic uncertainties are assigned based on tests of the method on simulation and are in the order of 15-100% for  $E_T^{\text{miss}}$  values falling into the signal selection of this analysis. The large uncertainties for higher  $E_T^{\text{miss}}$  values are driven by the available MC statistics to validate the method.

In contrast to the JZB method, the  $E_T^{\text{miss}}$  templates method does account only for  $Z + \text{jets}$  events. Other backgrounds containing Z bosons, such as WZ, ZZ or more rare SM processes like  $t\bar{t}Z$  or Triboson production are estimated from simulation and assigned 50% uncertainty.

### 5.2.3. Extrapolation for off-Z regions

Both methods described above result in background estimates for the on-Z region of 81 GeV  $< m_{\ell\ell} < 101$  GeV. Contributions to the low-Mass and high-Mass selections from off-Shell Z bosons or the Drell–Yan continuum are estimated by applying an extrapolation factor  $R_{\text{out/in}}$  to the on-Z prediction.  $R_{\text{out/in}}$  is measured in the Drell–Yan control region, separately for the low-Mass and high-Mass region as well as  $e^\pm e^\mp$ ,  $\mu^\pm \mu^\mp$  and SF leptons. It is defined as the ratio of the event yield in the mass region in question (*out*) by that in the on-Z mass region (*in*), after subtraction of contribution from flavour-symmetric processes from the OF sample:

$$R_{\text{out/in}} = \frac{N_{\text{out}}^{\text{SF}} - N_{\text{out}}^{\text{OF}} R_{\text{SF/OF}}}{N_{\text{in}}^{\text{SF}} - N_{\text{in}}^{\text{OF}} R_{\text{SF/OF}}}, \quad (5.12)$$

where the SF is substituted for  $e^\pm e^\mp$  or  $\mu^\pm \mu^\mp$  according to the desired lepton flavour. The  $m_{\ell\ell}$  distribution in the Drell–Yan control region is shown in Figure 5.9. The resulting values range from about 6–8% for the low-Mass region to 2–3% for the high-Mass region, as summarized in Table 5.8.

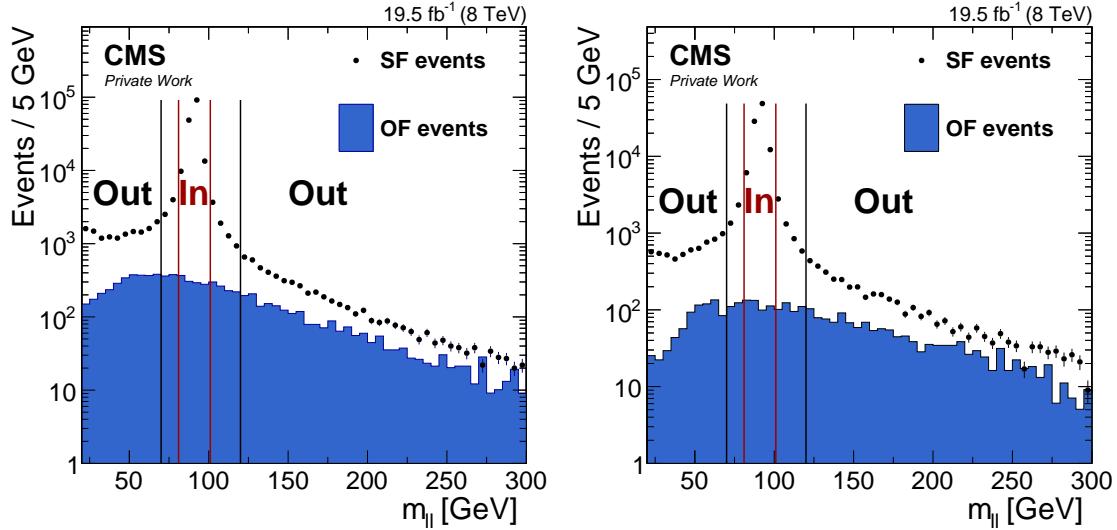


Figure 5.9.: The  $m_{\ell\ell}$  distribution in the Drell–Yan control region in the central (left) and forward (right) dielepton selection. SF data is shown as the black points while OF data is shown as the blue histogram. The black lines indicate the boundaries of the two *out* regions and the dark red lines those of the *in* region.

The validity of applying the  $R_{\text{out/in}}$  as measured in the Drell–Yan control region in the signal region is checked by studying the behaviour of the quantity as a function of  $E_T^{\text{miss}}$  and  $N_{\text{Jets}}$ . The results for the low-Mass selection and the case of SF leptons is shown in Figure 5.10 for both the central and forward lepton selections. In neither selection a clear dependency on  $E_T^{\text{miss}}$  is observed. However, above 70 GeV there is not enough statistics left after subtraction of the flavour-symmetric backgrounds, so it is not possible to judge the behaviour all the way up to the signal region. The value of  $R_{\text{out/in}}$  clearly increases with the number of jets. As the requirement on  $N_{\text{Jets}}$  is the same between the Drell–Yan control region and the signal region, the events do not differ much in terms of jet multiplicity and this dependency does not significantly restrict the applicability of the  $R_{\text{out/in}}$  factors in the signal region. In total, an systematic uncertainty of 25% is assigned to cover the observed effects. Consistent results are observed also for the high-Mass region and for the split into  $e^\pm e^\mp$  and  $\mu^\pm \mu^\mp$ , as can be seen in Appendix **FIXME: Make routin appendix and refer to it.**

#### 5.2.4. Resulting background prediction

The predictions for the on- $Z$  region from the JZB and  $E_T^{\text{miss}}$  templates methods agree within their uncertainties. As for the  $R_{\text{SF/OF}}$  factor, a weighted average is used to combine the two estimates. The results of both methods as well as the combination are shown in Table 5.7.

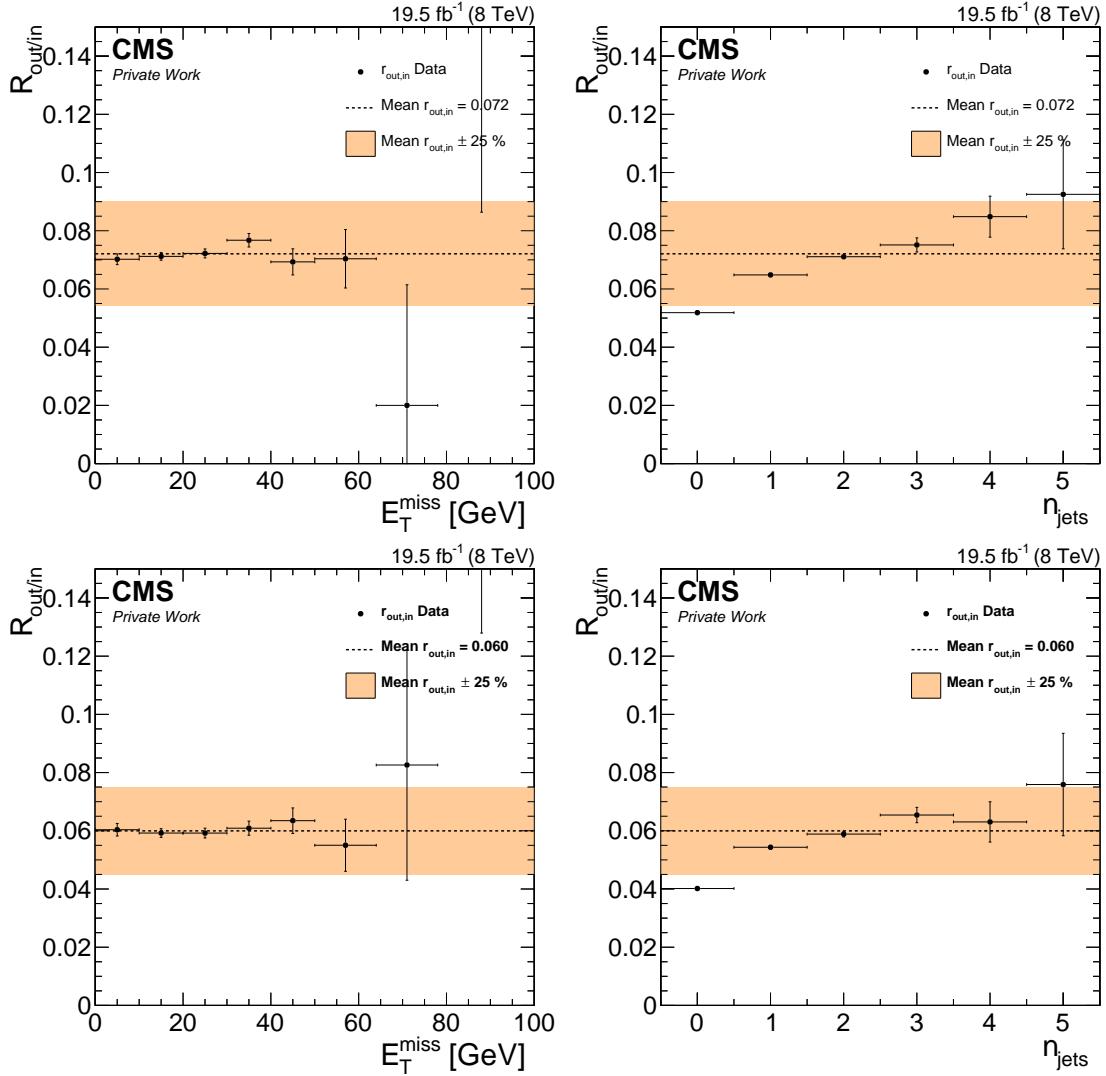


Figure 5.10.: Dependencies of  $R_{\text{out/in}}$  on  $E_T^{\text{miss}}$  (left) and  $N_{\text{jets}}$  (right) for the central (top) and forward (bottom) lepton selection. The results on data are shown in black. The central value is shown as a black dashed line while the systematic uncertainty is shown as an orange band.

Table 5.7.

	$N_{\text{out}}$	$N_{\text{in}}$	$R_{\text{out/in}}(SF) \pm \sigma_{\text{stat}}$
Central			
Low mass			
Data	$11608.6 \pm 131.7$	$160645.8 \pm 403.9$	$0.072 \pm 0.001 \pm 0.018$
MC	$10725.8 \pm 128.7$	$167291.6 \pm 412.0$	$0.064 \pm 0.001 \pm 0.016$
high Mass			
Data	$3571.3 \pm 91.2$	$160645.8 \pm 403.9$	$0.022 \pm 0.001 \pm 0.006$
MC	$3243.0 \pm 88.9$	$167291.6 \pm 412.0$	$0.019 \pm 0.001 \pm 0.005$
Forward			
Low mass			
Data	$5657.6 \pm 84.2$	$94407.6 \pm 308.7$	$0.060 \pm 0.001 \pm 0.015$
MC	$5695.8 \pm 85.5$	$103407.4 \pm 323.0$	$0.055 \pm 0.001 \pm 0.014$
high Mass			
Data	$2672.5 \pm 75.8$	$94407.6 \pm 308.7$	$0.028 \pm 0.001 \pm 0.007$
MC	$2459.1 \pm 72.3$	$103407.4 \pm 323.0$	$0.024 \pm 0.001 \pm 0.006$

Shown there are also the  $R_{\text{out/in}}$  values and the resulting estimates for the low- and high-Mass regions. For this, the on-Z predictions, which are derived on the PromptReco, are scaled by  $1.03 \pm 0.03$  to take into account the slight increase in jet multiplicity in the ReReco dataset.

### 5.3. Investigation of possible further backgrounds

### 5.4. Search for a kinematic edge with a fit

Table 5.8.: Estimate of the Z background yields in the Z peak region and extrapolation to the signal mass region for the full dataset.

	central		
	$e^\pm e^\mp$	$\mu^\pm \mu^\mp$	SF
Z bkgd estimate (JZB)	$57.9 \pm 13.8 \pm 10.1$	$46.1 \pm 13.8 \pm 8.0$	$104 \pm 21 \pm 18$
Z bkgd estimate ( $E_T^{\text{miss}}$ templates)	$63.2 \pm 4.3 \pm 15.3$	$69.5 \pm 4.0 \pm 16.9$	$133 \pm 7 \pm 32$
Z bkgd estimate (Combined)	$60.7 \pm 11.6$	$56.8 \pm 11.7$	$116 \pm 21$
$R_{\text{out/in}}$ low-Mass	$0.069 \pm 0.001 \pm 0.017$	$0.075 \pm 0.001 \pm 0.019$	$0.072 \pm 0.001 \pm 0.018$
low-Mass estimate	$4.3 \pm 1.3$	$4.4 \pm 1.4$	$8.6 \pm 2.7$
$R_{\text{out/in}}$ high-Mass	$0.025 \pm 0.001 \pm 0.006$	$0.020 \pm 0.001 \pm 0.005$	$0.022 \pm 0.001 \pm 0.006$
high-Mass estimate	$1.5 \pm 0.5$	$1.2 \pm 0.4$	$2.7 \pm 0.8$
	forward		
	$e^\pm e^\mp$	$\mu^\pm \mu^\mp$	SF
Z bkgd estimate (JZB)	$15.6 \pm 8.3 \pm 2.9$	$13.8 \pm 8.3 \pm 2.8$	$29 \pm 11 \pm 6$
Z bkgd estimate (MET templates)	$24.4 \pm 1.8 \pm 6.0$	$32.3 \pm 2.2 \pm 7.9$	$56.9 \pm 3.6 \pm 14.0$
Z bkgd estimate (Combined)	$21 \pm 5$	$25 \pm 6$	$42 \pm 9$
$R_{\text{out/in}}$ low-Mass	$0.055 \pm 0.001 \pm 0.014$	$0.064 \pm 0.001 \pm 0.016$	$0.060 \pm 0.001 \pm 0.015$
low-Mass estimate	$1.2 \pm 0.4$	$1.6 \pm 0.6$	$2.6 \pm 0.8$
$R_{\text{out/in}}$ high-Mass	$0.031 \pm 0.001 \pm 0.008$	$0.026 \pm 0.001 \pm 0.007$	$0.028 \pm 0.001 \pm 0.007$
high-Mass estimate	$0.7 \pm 0.2$	$0.7 \pm 0.2$	$1.2 \pm 0.4$

# 6. Results

## 6.1. Result of the counting experiment

The distribution of the dilepton invariant mass in the central and forward signal regions are shown in Figure 6.1. The resulting event yields are compared to the expectation from SM backgrounds in Table 6.2. A maximum likelihood fit is performed in each region to find the best estimator for the difference of expected and observed yield. The significances of deviations of this difference from zero are evaluated using the profile likelihood ratio of the signal and signal+background hypothesis [69]. In general, the observed data is in agreement with the background estimation within about one standard deviation, except for the low-mass region in the central dilepton selection. Here the observed yield exceeds the expectation by  $109^{+48}_{-49}$  events. The size of this excess corresponds to a significance of  $2.2\sigma$ .

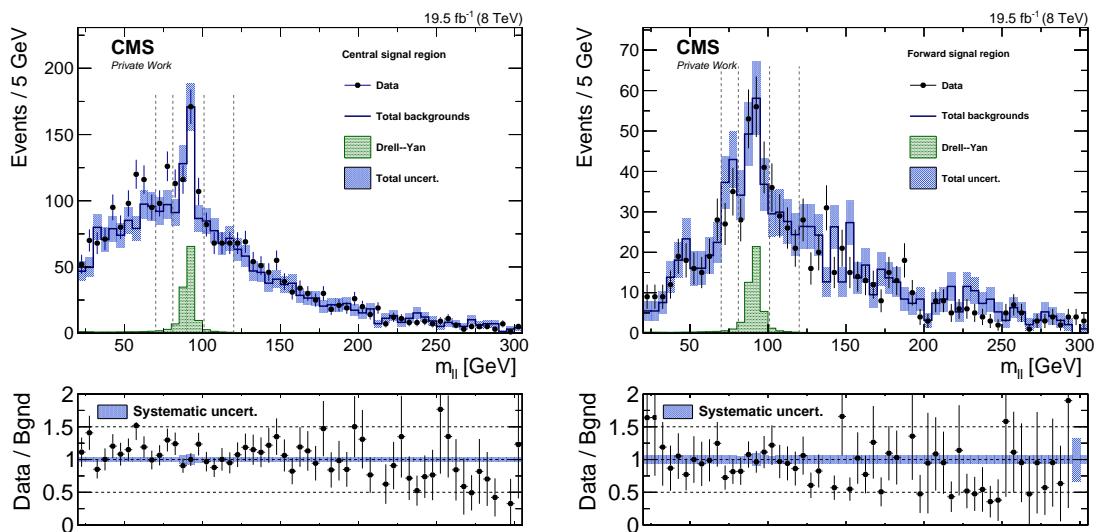


Figure 6.1.: Distribution of  $m_{ee}$  in the signal region for the central (left) and forward (right) dilepton selection. The data is shown as black dots, while the total background prediction from data is shown as a blue histogram. The blue error bars indicate the combined statistical and systematic background uncertainty in each bin. The contribution from backgrounds containing a Z boson is shown as a green histogram. The dashed lines indicates the boundaries of the three mass bins. Beneath the plot the ratio of data to the background prediction is shown. The error bars include the statistical uncertainties of data and background, while the blue band indicates the systematic uncertainties on the background.

In the Tables 6.3 and ??, the results are shown separately for  $e^\pm e^\mp$  and  $\mu^\pm \mu^\mp$  events. As expected from the fact that  $r_{\mu e}$  is larger than one, the yields in the  $\mu^\pm \mu^\mp$  channel are slightly larger than in the  $e^\pm e^\mp$  channel. For the flavour-symmetry backgrounds, and therefore also

Table 6.1.: Results of the edge-search counting experiment for event yields in the signal regions. The statistical and systematic uncertainties are added in quadrature, except for the flavor-symmetric backgrounds. Low-mass refers to  $20 < m_{\ell\ell} < 70$  GeV, on-Z to  $81 < m_{\ell\ell} < 101$  GeV and high-mass to  $m_{\ell\ell} > 120$  GeV.

	Low-mass		On-Z		High-mass	
	Central	Forward	Central	Forward	Central	Forward
Observed	865	154	494	176	849	381
Flavor-symmetric	$746 \pm 27 \pm 26$	$144 \pm 12 \pm 7$	$368 \pm 19 \pm 13$	$137 \pm 11 \pm 7$	$789 \pm 28 \pm 28$	$411 \pm 20 \pm 21$
Drell-Yan	$8.6 \pm 2.7$	$2.6 \pm 0.8$	$119 \pm 21$	$43 \pm 9$	$2.7 \pm 0.8$	$1.2 \pm 0.4$
Total estimated	$755 \pm 38$	$147 \pm 14$	$488 \pm 31$	$180 \pm 16$	$792 \pm 39$	$413 \pm 30$
Observed - estimated	$109^{+48}_{-48}$	$7^{+19}_{-19}$	$6^{+37}_{-38}$	$-5^{+21}_{-21}$	$57^{+49}_{-50}$	$-32^{+35}_{-37}$
Significance	$2.2 \sigma$	$0.4 \sigma$	$0.1 \sigma$	$<0.1 \sigma$	$1.1 \sigma$	$<0.1 \sigma$

for the total background estimates and the difference of observation and estimation, the yields in the  $e^\pm e^\mp$  and  $\mu^\pm \mu^\mp$  channels do not exactly add up to those in the combined SF channel.

**FIXME:** Reference explanation once written

Table 6.2.: Results of the edge-search counting experiment for event yields in the signal regions. The statistical and systematic uncertainties are added in quadrature, except for the flavor-symmetric backgrounds. Low-mass refers to  $20 < m_{\ell\ell} < 70$  GeV, on-Z to  $81 < m_{\ell\ell} < 101$  GeV and high-mass to  $m_{\ell\ell} > 120$  GeV.

	Low-mass		On-Z		High-mass	
	Central	Forward	Central	Forward	Central	Forward
Observed	389	53	232	86	401	195
Flavor-symmetric	$337 \pm 12 \pm 19$	$61 \pm 5 \pm 6$	$166 \pm 8 \pm 9$	$58 \pm 5 \pm 5$	$357 \pm 12 \pm 21$	$175 \pm 8 \pm 17$
Drell-Yan	$4.3 \pm 1.3$	$1.2 \pm 0.4$	$62 \pm 11$	$21 \pm 5$	$1.5 \pm 0.5$	$0.7 \pm 0.2$
Total estimated	$342 \pm 23$	$62 \pm 8$	$229 \pm 17$	$79 \pm 9$	$358 \pm 24$	$175 \pm 19$
Observed - estimated	$47^{+25}_{-25}$	$-10^{+9}_{-9}$	$3^{+21}_{-21}$	$6^{+12}_{-12}$	$42^{+25}_{-26}$	$19^{+18}_{-18}$
Significance	$1.9 \sigma$	$<0.1 \sigma$	$0.1 \sigma$	$0.5 \sigma$	$1.7 \sigma$	$1.1 \sigma$

## 6.2. Result of the search for a kinematic edge

Table 6.3.: Results of the edge-search counting experiment for event yields in the signal regions. The statistical and systematic uncertainties are added in quadrature, except for the flavor-symmetric backgrounds. Low-mass refers to  $20 < m_{\ell\ell} < 70$  GeV, on-Z to  $81 < m_{\ell\ell} < 101$  GeV and high-mass to  $m_{\ell\ell} > 120$  GeV.

	Low-mass		On-Z		High-mass	
	Central	Forward	Central	Forward	Central	Forward
Observed	476	101	262	90	448	186
Flavor-symmetric	$405 \pm 14 \pm 21$	$79 \pm 6 \pm 6$	$200 \pm 10 \pm 10$	$74 \pm 6 \pm 6$	$428 \pm 15 \pm 23$	$224 \pm 11 \pm 19$
Drell-Yan	$4.4 \pm 1.4$	$1.6 \pm 0.6$	$58 \pm 10$	$25 \pm 6$	$1.2 \pm 0.4$	$0.7 \pm 0.2$
Total estimated	$409 \pm 26$	$80 \pm 9$	$258 \pm 18$	$100 \pm 11$	$429 \pm 27$	$225 \pm 22$
Observed - estimated	$67^{+29}_{-29}$	$20^{+13}_{-13}$	$3^{+22}_{-23}$	$-11^{+13}_{-13}$	$19^{+29}_{-29}$	$-40^{+21}_{-21}$
Significance	$2.3 \sigma$	$1.6 \sigma$	$0.2 \sigma$	$<0.1 \sigma$	$0.6 \sigma$	$<0.1 \sigma$



---

## **7. Interpretation in simplified models**

**7.1. Simplified Models for Supersymmetric Signatures**

**7.2. The T6bblledge and T6bbslepton model**

**7.3. Interpretation of the counting experiment in simplified models**



---

## **8. Outlook to LHC Run II**



---

## **9. Conclusion**



---

## A. Data and Monte Carlo Samples

process	sample
$t\bar{t} \rightarrow b\bar{b}\ell\nu h\nu$	<code>/TTJets_FullLeptMGDecays_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V7C-v2/AODSIM</code>
$t\bar{t} \rightarrow b\bar{b}q\bar{q}\ell\nu$	<code>/TTJets_SemileptMGDecays_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V7C-v1/AODSIM</code>
$t\bar{t} \rightarrow b\bar{b}q\bar{q}\bar{q}\bar{q}$	<code>/TTJets_HadronicMGDecays_8TeV-madgraph/Summer12_DR53X-PU_S10_START53_V7A_ext-v1/AODSIM</code>
$Z/\gamma^* \rightarrow l^+l^- 10 \text{ GeV} < m_{ll} < 50 \text{ GeV}$	<code>/DYJetsToLL_M-10To50filter_8TeV-madgraph/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$Z/\gamma^* \rightarrow l^+l^- m_{ll} > 50 \text{ GeV}$	<code>/DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarball/Summer12_DR53X-PU_S10_START53_V7A-v2/AODSIM</code>
$W \rightarrow l\nu$	<code>/WJetsToLNu_TuneZ2Star_8TeV-madgraph-tarball/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$ZZ \rightarrow l^+l^- q\bar{q}$	<code>/ZZJetsTo2L2Q_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$ZZ \rightarrow l^+l^- \nu\bar{\nu}$	<code>/ZZJetsTo2L2Nu_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V7A-v3/AODSIM</code>
$ZZ \rightarrow l^+l^- l^+l^-$	<code>/ZZJetsTo4L_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$WZ \rightarrow ll l^+l^-$	<code>/WZJetsTo3LNu_TuneZ2_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$WZ \rightarrow qq l^+l^-$	<code>/WZJetsTo2L2Q_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$WW \rightarrow ll\nu\nu$	<code>/WWJetsTo2L2Nu_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$t$ s-Channel	<code>/T_s-channel_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$t$ t-Channel	<code>/T_t-channel_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$t$ tW-Channel	<code>/T_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$\bar{t}$ s-Channel	<code>/Tbar_s-channel_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$\bar{t}$ t-Channel	<code>/Tbar_t-channel_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$\bar{t}$ tW-Channel	<code>/Tbar_tW-channel1-DR_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$WWW$	<code>/WWWWJets_8TeV-madgraph/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$WW\gamma$	<code>/WGJets_8TeV-madgraph/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$WWZ$	<code>/WWZNloGstarJets_8TeV-madgraph/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$WZZ$	<code>/WZZNloGstarJets_8TeV-madgraph/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$t\bar{t}\gamma$	<code>/TTGJets_8TeV-madgraph/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$t\bar{t}W$	<code>/TWJets_8TeV-madgraph/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$t\bar{t}Z$	<code>/TTZJets_8TeV-madgraph_v2/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$t\bar{t}WW$	<code>/TTWWJets_8TeV-madgraph/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$t\bar{t}$	<code>/TTJets_MassiveBinDECAY_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$t\bar{t}$ , $m_{top} = 166.5 \text{ GeV}$	<code>/TTJets_mass166_5_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$t\bar{t}$ , $m_{top} = 169.5 \text{ GeV}$	<code>/TTJets_mass169_5_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$t\bar{t}$ , $m_{top} = 175.5 \text{ GeV}$	<code>/TTJets_mass175_5_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$t\bar{t}$ , $m_{top} = 178.5 \text{ GeV}$	<code>/TTJets_mass178_5_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$t\bar{t}$ , Matching scale up	<code>/TTJets_matchingdown_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$t\bar{t}$ , Matching scale down	<code>/TTJets_matchingup_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$t\bar{t}$ , Factorization scale up	<code>/TTJets_scaleup_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>
$t\bar{t}$ , Factorization scale down	<code>/TTJets_scaledown_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</code>

Table A.1.: Bla

---

## B. Dependencies of $r_{\mu e}$

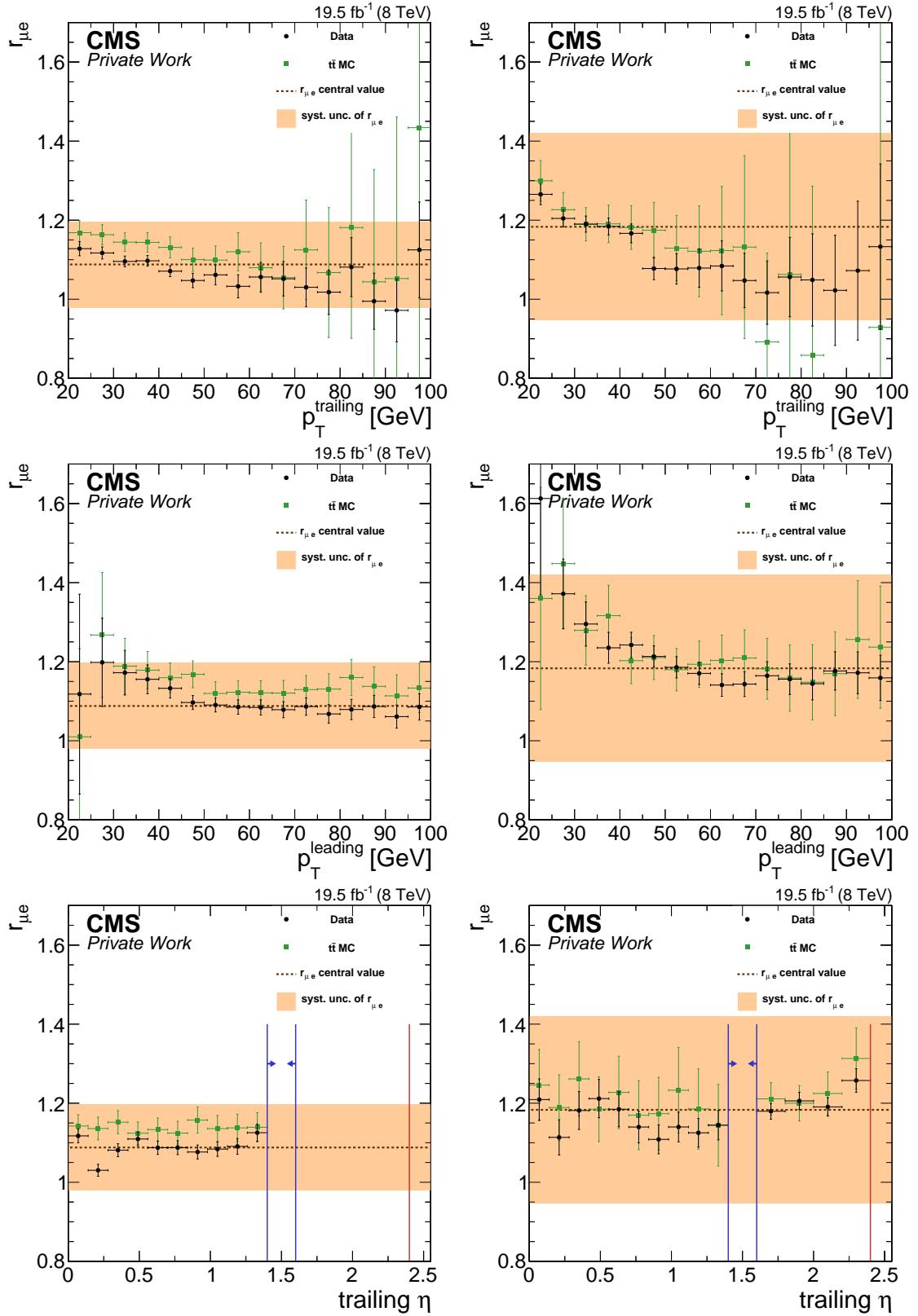


Figure B.1.: Dependencies of  $r_{\mu e}$  on the  $p_T$  of the trailing (top) and the leading (middle) lepton, as well as  $|\eta|$  of the trailing lepton (bottom) for the central (left) and forward (right) lepton selection. The results on data are shown in black while  $t\bar{t}$  simulation is shown in green. The central value is shown as a brown dashed line while the systematic uncertainty is shown as an orange band.

## B. Dependencies of $r_{\mu e}$

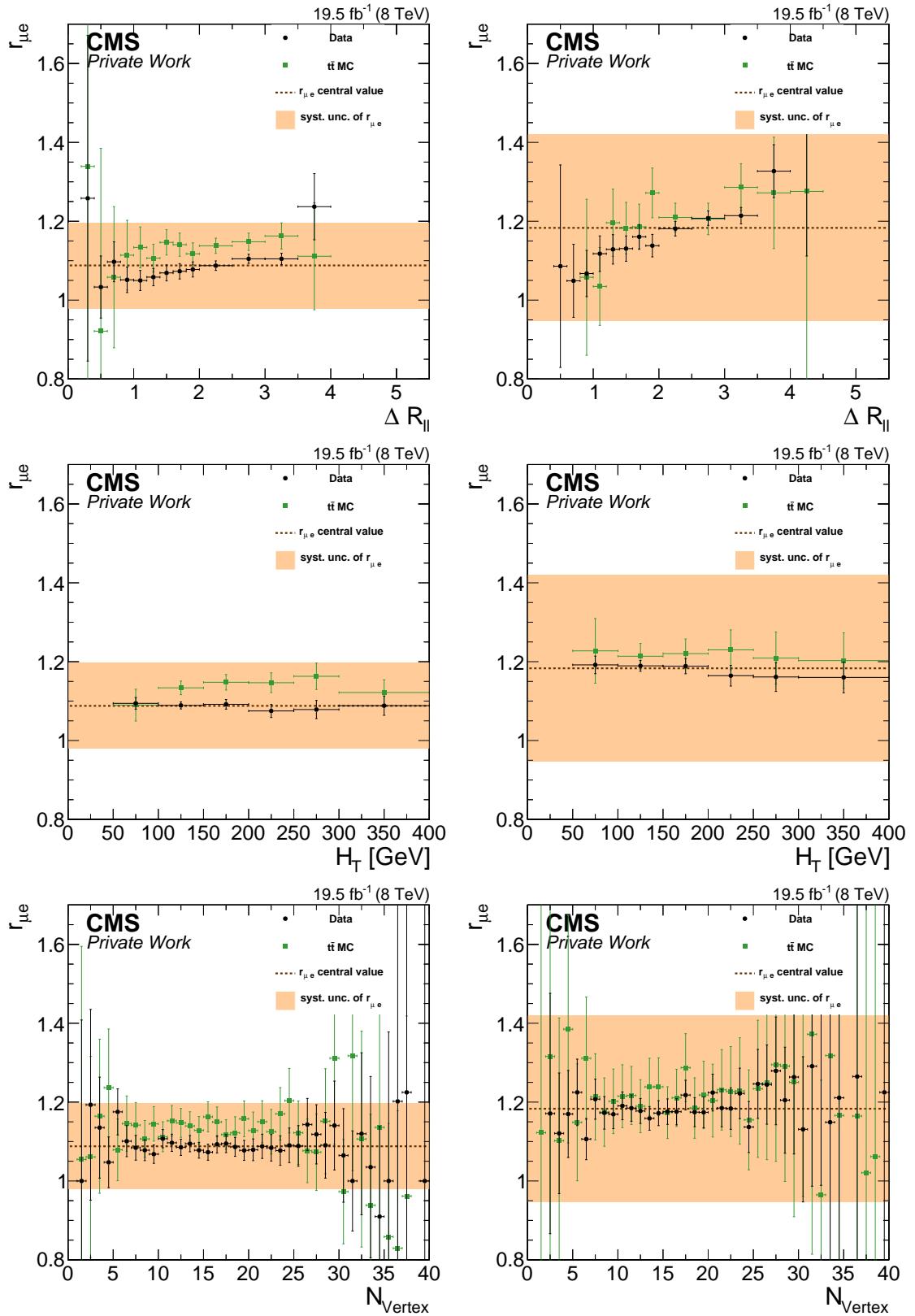


Figure B.2.: Dependencies of  $r_{\mu e}$  on  $\Delta R(l l)$  (top),  $H_T$  (middle), and  $N_{\text{vertex}}$  (bottom) for the central (left) and forward (right) lepton selection. The results on data are shown in black while  $t\bar{t}$  simulation is shown in green. The central value is shown as a brown dashed line while the systematic uncertainty is shown as an orange band.



---

## C. Dependencies of $R_T$

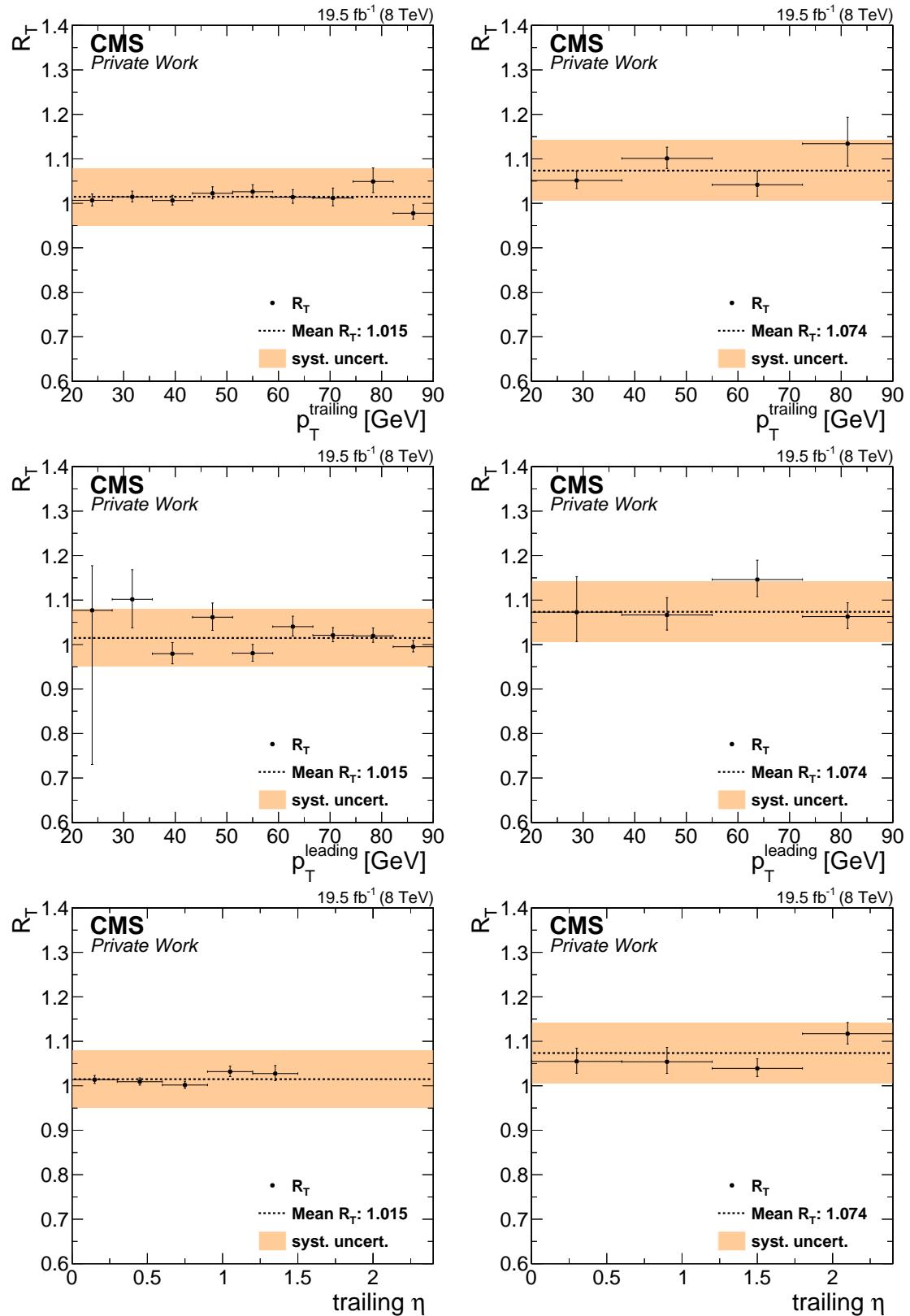


Figure C.1.: Dependencies of  $R_T$  on leading lepton  $p_T$  (top), trailing lepton  $p_T$  (middle), and  $|\eta|$  of the trailing lepton (bottom) for the central (left) and forward (right) lepton selection. The results on data are shown in black. The central value is shown as a black dashed line while the systematic uncertainty is shown as an orange band.

### C. Dependencies of $R_T$

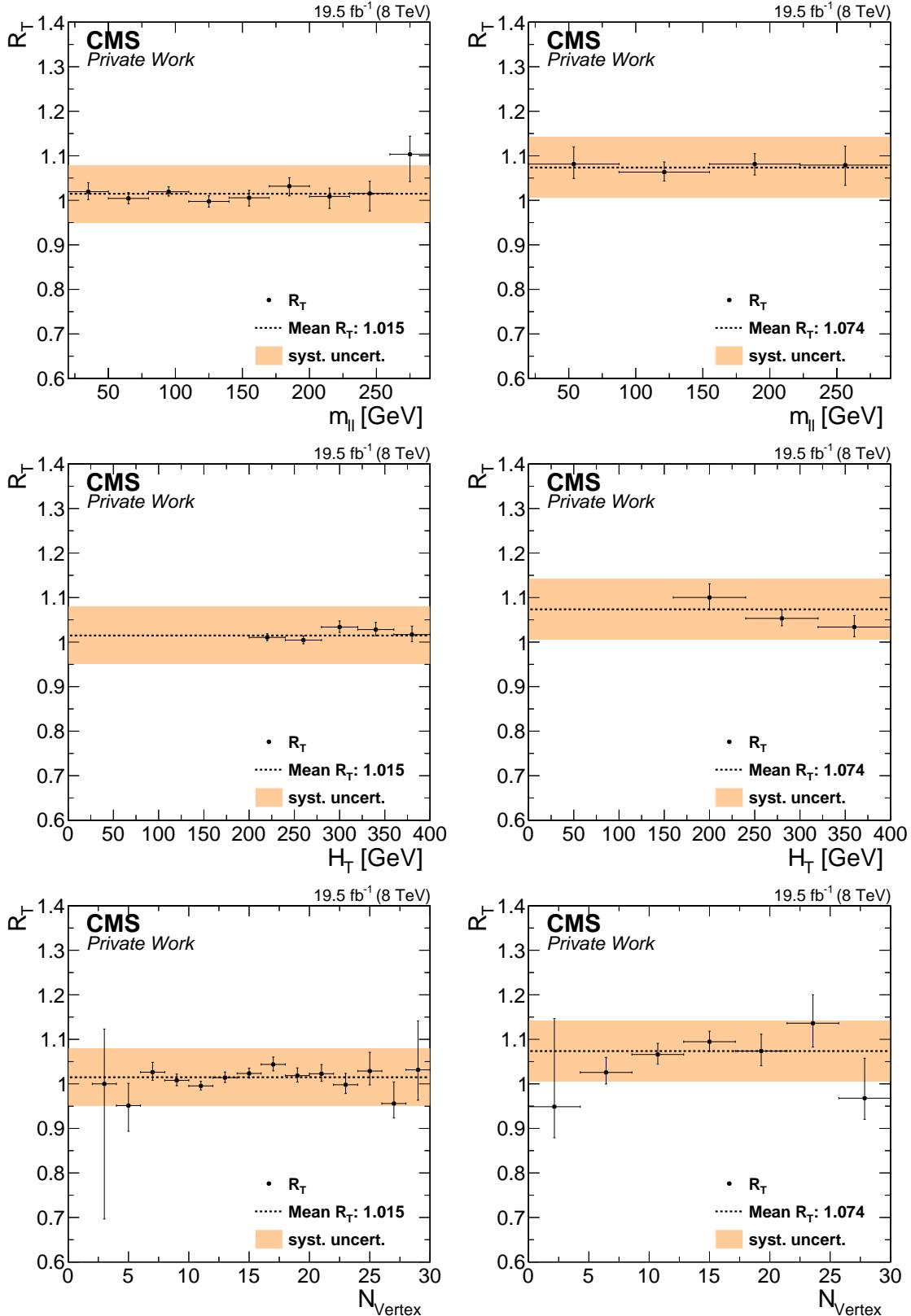


Figure C.2.: Dependencies of  $R_T$  on  $m_{ll}$  (top),  $H_T$  (middle), and  $N_{vertex}$  (bottom) for the central (left) and forward (right) lepton selection. The results on data are shown in black. The central value is shown as a black dashed line while the systematic uncertainty is shown as an orange band.



---

# Bibliography

- [1] Lyndon Evans and Philip Bryant. LHC Machine. *Journal of Instrumentation*, 3(08):S08001, 2008.
- [2] CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):S08004, 2008.
- [3] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3(08):S08003, 2008.
- [4] LHCb Collaboration. The LHCb Detector at the LHC. *Journal of Instrumentation*, 3(08):S08005, 2008.
- [5] Particle kickers. <https://cds.cern.ch/record/1706606/>. Accessed: 2014-08-06.
- [6] CMS Collaboration. Public CMS Luminosity Information. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults?rev=101>. Accessed: 2014-08-06.
- [7] CMS detector design. <http://cms.web.cern.ch/news/cms-detector-design>. Accessed: 2014-08-08.
- [8] CMS Collaboration. Electromagnetic calorimeter calibration with 7 TeV data. 2010. CMS Physics Analysis Summary EGM-10-003.
- [9] CMS Collaboration. Performance of CMS muon reconstruction in pp collision events at  $\sqrt{s} = 7$  TeV CHECK FOR JOURNAL PUBLICATION. *J. Instrum.*, 7(arXiv:1206.4071. CMS-MUO-10-004. CERN-PH-EP-2012-173):P10002. 81 p, Jun 2012. Comments: Submitted to the Journal of Instrumentation.
- [10] W. Adam et al. The CMS high level trigger. *Eur.Phys.J.*, C46:605–667, 2006.
- [11] G. Bauer et al. The CMS High Level Trigger System: Experience and Future Development. *Journal of Physics: Conference Series*, 396(1):012008, 2012.
- [12] CMS HLT config browser. <http://j2eeps.cern.ch/cms-project-confdb-hltdev/browser/>. Accessed: 2014-10-09.
- [13] Serguei Chatrchyan et al. Description and performance of track and primary-vertex reconstruction with the CMS tracker CHECK FOR JOURNAL PUBLICATION. 2014.
- [14] R. Fröhwirth. Application of kalman filtering to track and vertex fitting. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 262(2.3):444 – 450, 1987.
- [15] CMS Software Guide on Iterative Tracking. <https://twiki.cern.ch/twiki/bin/>

- [view/CMSPublic/SWGuideIterativeTracking?rev=9](https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideIterativeTracking?rev=9). Accessed: 2014-10-11.
- [16] CMS Collaboration. Performance of cms muon reconstruction in cosmic-ray events. *Journal of Instrumentation*, 5(03):T03022, 2010.
  - [17] Particle-flow commissioning with muons and electrons from J/Psi and W events at 7 TeV. Technical Report CMS-PAS-PFT-10-003, CERN, 2010. Geneva, 2010.
  - [18] S. Baffioni, C. Charlot, F. Ferri, D. Futyan, P. Meridiani, et al. Electron reconstruction in CMS. *Eur.Phys.J.*, C49:1099–1116, 2007.
  - [19] CMS Collaboration. Review of clustering algorithms and energy corrections in ECAL. *CMS Internal Note IN-10-008*, Mar 2010.
  - [20] W. Adam, R. Fröhwirth, A. Strandlie, and T. Todorov. Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC. *Journal of Physics G: Nuclear and Particle Physics*, 31(9):N9, 2005.
  - [21] Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. Technical Report CMS-PAS-PFT-09-001, CERN, 2009. Geneva, Apr 2009.
  - [22] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Anti- $k_t$  jet clustering algorithm. *JHEP*, 0804:063, 2008.
  - [23] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet User Manual. *Eur.Phys.J.*, C72:1896, 2012.
  - [24] Matteo Cacciari and Gavin P. Salam. Dispelling the  $N^3$  myth for the  $k_t$  jet-finder. *Phys.Lett.*, B641:57–61, 2006.
  - [25] The CMS collaboration. Determination of jet energy calibration and transverse momentum resolution in cms. *Journal of Instrumentation*, 6(11):P11002, 2011.
  - [26] Serguei Chatrchyan et al. Identification of b-quark jets with the CMS experiment. *JINST*, 8:P04013, 2013.
  - [27] Results on b-tagging identification in 8 TeV pp collisions. Feb 2013.
  - [28] The CMS collaboration. Missing transverse energy performance of the cms detector. *Journal of Instrumentation*, 6(09):P09001, 2011.
  - [29] MET performance in 8 TeV data. Technical Report CMS-PAS-JME-12-002, CERN, Geneva, 2013.
  - [30] CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS. CERN, Geneva, 2006. There is an error on cover due to a technical problem for some items.
  - [31] CMS Software Guide on the CMSSW Application Framework. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCMSSWFramework?rev=13>. Accessed: 2014-

10-26.

- [32] Ian Bird. Computing for the large hadron collider. *Annual Review of Nuclear and Particle Science*, 61(1):99–118, 2011.
- [33] Worldwide LHC Computing Grid . <http://wlcg.web.cern.ch/>. Accessed: 2014-10-26.
- [34] D. Spiga, S. Lacaprara, W. Bacchi, M. Cinquilli, G. Codispoti, M. Corvo, A. Dorigo, A. Fanfani, F. Fanzago, F. Farina, M. Merlo, O. Gutsche, L. Servoli, and C. Kavka. The cms remote analysis builder (crab). 4873:580–586, 2007.
- [35] W Adam, V Adler, B Hegner, L Lista, S Lowette, P Maksimovic, G Petrucciani, S Rappoccio, F Ronga, R Tenchini, and R Wolf. PAT: The CMS Physics Analysis Toolkit. *Journal of Physics: Conference Series*, 219(3):032017, 2010.
- [36] R. Brun and F. Rademakers. ROOT - An Object Oriented Data Analysis Framework. *Nucl. Inst. & Meth. in Phys. Res. A*, 389:81–86, 1996. See also <http://root.cern.ch/>.
- [37] A Afaq, A Dolgert, Y Guo, C Jones, S Kosyakov, V Kuznetsov, L Lueking, D Riley, and V Sekhri. The cms dataset bookkeeping service. *Journal of Physics: Conference Series*, 119(7):072001, 2008.
- [38] K.A. Olive et al. (Particle Data Group). *Chin. Phys. C*, 38(090001), 2014.
- [39] F. Halzen and A. Martin. *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. John Wiley & Sons, 1984.
- [40] F.D. Aaron et al. Combined Measurement and QCD Analysis of the Inclusive e+- p Scattering Cross Sections at HERA. *JHEP*, 1001:109, 2010.
- [41] J. Pumplin, D.R. Stump, J. Huston, H.L. Lai, Pavel M. Nadolsky, et al. New generation of parton distributions with uncertainties from global QCD analysis. *JHEP*, 0207:012, 2002.
- [42] Richard D. Ball, Valerio Bertone, Stefano Carrazza, Christopher S. Deans, Luigi Del Debbio, et al. Parton distributions with LHC data. *Nucl.Phys.*, B867:244–289, 2013.
- [43] A.D. Martin, W.J. Stirling, R.S. Thorne, and G. Watt. Parton distributions for the LHC. *Eur.Phys.J.*, C63:189–285, 2009.
- [44] Hung-Liang Lai, Marco Guzzi, Joey Huston, Zhao Li, Pavel M. Nadolsky, et al. New parton distributions for collider physics. *Phys.Rev.*, D82:074024, 2010.
- [45] V. N. Gribov and L. N. Lipatov. Deep inelastic electron scattering in perturbation theory. *Sov. J. Nucl. Phys.*, 15:438–450, 1972.
- [46] G. Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl.Phys.*, B126:298, 1977.
- [47] Y. L. Dokshitzer. Calculation of the Structure Functions for Deep Inelastic Scattering

- and  $e^+e^-$  Annihilation by Perturbation Theory in Quantum Chromodynamics. *Sov. Phys. JETP*, 46:641–653, 1977.
- [48] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5 : Going Beyond. *JHEP*, 1106:128, 2011.
- [49] Torbjorn Sjöstrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006.
- [50] Stefan Hoeche, Frank Krauss, Nils Lavesson, Leif Lonnblad, Michelangelo Mangano, et al. Matching parton showers and matrix elements. 2006.
- [51] Paolo Nason. A new method for combining nlo qcd with shower monte carlo algorithms. *Journal of High Energy Physics*, 2004(11):040, 2004.
- [52] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. NLO single-top production matched with shower in POWHEG: s- and t-channel contributions. *JHEP*, 0909:111, 2009.
- [53] Emanuele Re. Single-top Wt-channel production matched with parton showers using the POWHEG method. *Eur.Phys.J.*, C71:1547, 2011.
- [54] Rick Field. Min-Bias and the Underlying Event at the LHC. *Acta Phys.Polon.*, B42:2631–2656, 2011. Tune  $Z2^*$  is a variation of Tune  $Z2$  described in this reference.
- [55] S. Jadach, Z. WaĘĘs, R. Decker, and J.H. KĀžhn. The  $\tau$  decay library tauola, version 2.4. *Computer Physics Communications*, 76(3):361 – 380, 1993.
- [56] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl.Instrum.Meth.*, A506:250–303, 2003.
- [57] S Abdullin, P Azzi, F Beaudette, P Janot, A Perrotta, and the CMS Collaboration. The fast simulation of the cms detector at lhc. *Journal of Physics: Conference Series*, 331(3):032049, 2011.
- [58] S. et al Chatrchyan. Measurement of the  $t\bar{t}$  production cross section in the dilepton channel in pp collisions at  $\sqrt{s} = 8$  TeV. *Journal of High Energy Physics*, 2014(2), 2014.
- [59] Ye Li and Frank Petriello. Combining QCD and electroweak corrections to dilepton production in FEWZ. *Phys.Rev.*, D86:094034, 2012.
- [60] John M. Campbell, R. Keith Ellis, and Ciaran Williams. Vector boson pair production at the LHC. *JHEP*, 1107:018, 2011.
- [61] Nikolaos Kidonakis. Differential and total cross sections for top pair and single top production. pages 831–834, 2012.
- [62] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower

- simulations. *Journal of High Energy Physics*, 2014(7), 2014.
- [63] M.V. Garzelli, A. Kardos, C.G. Papadopoulos, and Z. Trocsanyi.  $t\bar{t}W^{+-}$  and  $t\bar{t}Z$  Hadroproduction at NLO accuracy in QCD with Parton Shower and Hadronization effects. *JHEP*, 1211:056, 2012.
  - [64] Measurement of the inclusive top-quark pair + photon production cross section in the muon + jets channel in pp collisions at 8 TeV. Technical Report CMS-PAS-TOP-13-011, CERN, Geneva, 2014.
  - [65] L Tuura, A Meyer, I Segoni, and G Della Ricca. Cms data quality monitoring: Systems and experiences. *Journal of Physics: Conference Series*, 219(7):072020, 2010.
  - [66] K Rose. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *Proceedings of the IEEE*, 8, 1998.
  - [67] R Fröhwirth, Wolfgang Waltenberger, and Pascal Vanlaer. Adaptive Vertex Fitting. Technical Report CMS-NOTE-2007-008, CERN, Geneva, Mar 2007.
  - [68] Serguei Chatrchyan et al. Search for physics beyond the standard model in events with a  $Z$  boson, jets, and missing transverse energy in  $pp$  collisions at  $\sqrt{s} = 7$  TeV. *Phys.Lett.*, B716:260–284, 2012.
  - [69] CMS and ATLAS Collaborations. Procedure for the LHC Higgs boson search combination in Summer 2011. *CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-011*.