



1 **A search for flavour changing neutral currents**  
2 **involving a top quark and a Z boson, using the**  
3 **data collected by the CMS collaboration at a**  
4 **centre-of-mass energy of 13 TeV**

5 **Isis Van Parijs**

6 **Proefschrift ingediend met het oog op het behalen van de academische graad**  
7 **Doctor in de Wetenschappen.**

Published in Faculteit Wetenschappen & Bio-ingenieurswetenschappen  
Vrije Universiteit Brussel

At 1. June 2017.

8

Responsible Contact: I. Van Parijs  
Intitute for High Energy Physics  
Promotor: Prof. Jorgen D'Hondt



11	<b>1 Theoretical basis</b>	<b>1</b>
12	1.1 Getting to the nature of things . . . . .	1
13	1.2 Standard Model Lagrangian, connecting fields with particles . . . . .	2
14	1.3 Flavours in the SM . . . . .	5
15	1.4 The top of the SM . . . . .	6
16	1.5 Hunting down the SM top quark . . . . .	9
17	1.6 Why to look beyond the SM . . . . .	13
18	1.7 An effective approach beyond the SM: FCNC involving a top quark . . . . .	14
19	1.8 The top-FCNC constrained . . . . .	16
20	<b>2 Experimental set-up</b>	<b>19</b>
21	2.1 The Large Hadron Collider . . . . .	19
22	2.2 The Compact Muon Solenoid . . . . .	22
23	2.2.1 CMS coordinate system . . . . .	23
24	2.2.2 Towards the heart of CMS . . . . .	24
25	2.2.3 Data acquisition . . . . .	32
26	2.2.4 Phase 1 upgrades . . . . .	32
27	2.2.5 CMS computing model . . . . .	34
28	<b>3 Analysis techniques</b>	<b>37</b>
29	3.1 Hadron collisions at high energies . . . . .	37
30	3.2 Event generation . . . . .	40
31	3.2.1 Fundamentals of simulating a proton collision . . . . .	40
32	3.2.2 Programs for event generation . . . . .	40
33	3.2.3 Generating FCNC top-Z interactions . . . . .	42
34	3.2.4 Generating SM background events . . . . .	44
35	3.3 Multivariate analysis techniques: Boosted Decision Trees . . . . .	46
36	3.4 Statistical methodology . . . . .	48
37	3.4.1 The absence of signal: limits . . . . .	49
38	3.4.2 Adding sources of uncertainty . . . . .	51
39	3.4.3 Asymptotic approximation of the CL method . . . . .	52
40	3.4.4 Extracting the signal model parameters . . . . .	52
41	<b>4 Event reconstruction and identification</b>	<b>53</b>

42	4.1	Object Reconstruction . . . . .	53
43	4.1.1	Charged particle tracks . . . . .	54
44	4.1.2	Following the Muon's Footsteps . . . . .	55
45	4.1.3	The path of the Electron . . . . .	56
46	4.1.4	Primary Vertex Reconstruction . . . . .	57
47	4.1.5	Calorimeter clusters . . . . .	57
48	4.2	Putting the pieces together . . . . .	58
49	4.3	Particle flow identification . . . . .	59
50	4.3.1	Muons . . . . .	59
51	4.3.2	Electrons and isolated photons . . . . .	59
52	4.3.3	Hadrons and non-isolated photons . . . . .	59
53	4.3.4	Post processing . . . . .	60
54	4.4	Pile up mitigation and luminosity measurement . . . . .	60
55	4.5	Physics object reconstruction and identification . . . . .	61
56	4.5.1	Muons . . . . .	61
57	4.5.2	Electrons . . . . .	64
58	4.5.3	Jets . . . . .	66
59	4.5.4	Jets from b fragmentation . . . . .	68
60	4.5.5	Missing transverse energy . . . . .	70
61	4.6	Summary of corrections . . . . .	71
62	<b>5</b>	<b>Event selection and categorisation</b>	<b>73</b>
63	5.1	Baseline event selection and filters . . . . .	73
64	5.2	Event selection . . . . .	73
65	5.3	Effect of the corrections in dilepton events . . . . .	73
66	5.4	Regions and channels . . . . .	73
67	5.5	Data driven background simulation . . . . .	73
68	<b>6</b>	<b>The search for FCNC involving a top quark and a Z boson</b>	<b>75</b>
69	6.1	Construction of template distributions . . . . .	75
70	6.2	Systematic uncertainties . . . . .	75
71	6.3	Limit setting procedure . . . . .	75
72	6.4	Result and discussion . . . . .	75
73	<b>7</b>	<b>Denouement of the top-Z FCNC hunt at 13 TeV</b>	<b>77</b>
74		<b>Bibliography</b>	<b>79</b>

# Event reconstruction and identification

# 4

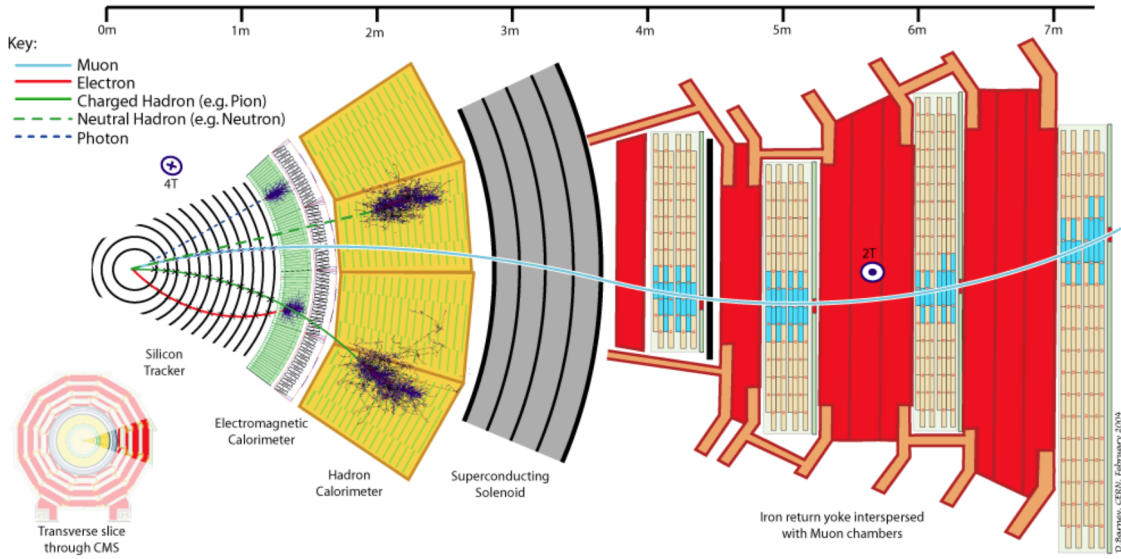
917

918 After the detector simulation described in [Section 3.2](#), the simulated data has the exact same  
919 format as the real collision data recorded at the CMS experiment. Therefore the same software  
920 can be used for the reconstruction of both simulation and real data. In [Section 4.1](#), the object  
921 reconstruction is explained. After reconstructing the objects, they are connected ([Section 4.2](#)) to  
922 physics objects need to be identified ([Section 4.3](#)) and corrected for pile up ([Section 4.4](#)). The  
923 objects used for physics analysis have extra requirements as shown in [Section 4.5](#). A summary  
924 of all the corrections applied to data and simulation is given in [Section 4.6](#).

## 925 4.1 Object Reconstruction

926 In [Figure 4.1](#), the particle interaction in a transverse slice of the CMS detector is shown. The  
927 particles enter first the tracker where charged particle trajectories, so-called tracks, and origins  
928 or vertices are reconstructed from signals (hits) in the sensitive layers. Charged particles get  
929 bent by the magnetic field making it able to measure the electric charges and momenta of  
930 charged particles. In the ECAL, the electron and photons are absorbed and the corresponding  
931 electromagnetic showers are detected as clusters of energy in adjacent cells. From this, the  
932 energy and the direction of the particles can be determined. The charged and neutral hadrons  
933 can initiate a hadronic shower in the ECAL that is fully absorbed in the HCAL. The clusters  
934 from these showers are also used to estimate the energy and direction. Muons and neutrino's  
935 pass through the calorimeters without little to no energy loss. The neutrino's escape the CMS  
936 detector undetected while muons produce hits in the muon detectors.

937 The traditional hadron colliders reconstruction is as follows. The reconstruction of isolated  
938 photons and electrons is primarily done by the ECAL, while the identification of muons is based  
939 on the muon detectors. Hadrons and photons form jets which are measured by the calorimeters  
940 without any contribution from the tracker or muon detectors. Jets can be tagged using the  
941 tracker as coming from hadronic  $\tau$  decays or  $b$  hadronisation based on the properties of the  
942 relevant charged particle tracks. The missing transverse energy  $\vec{p}_T$  is defined as the vectorial  
943 sum of the undetectable particle transverse momenta, and can be reconstructed without any  
944 information from the tracker. The particle flow (PF) [[144](#)] reconstruction correlates the tracks  
945 and clusters from all detector layers with the identification of each final state particle, and



**Figure 4.1:** Cross-section of the CMS detector with all parts of the detector labelled. This sketch shows the specific particle interactions from a beam interaction reign to the muon detector. The muon and charged pion are positively charged, the electron is negatively charged. Figure taken from [144].

combines the corresponding measurements to reconstruct the properties. In the particle flow algorithm, the muon is identified by a track in the inner tracker connected to a track in the muon detector as described in Section 4.1.2. The electrons are identified by a track and an ECAL cluster, not connected to an HCAL cluster as described in Section 4.1.3. The ECAL and HCAL clusters without a track link identify the photons and neutral hadrons, while the addition of the tracker determines the energy and direction of a charged hadron.

Coarse-grained detectors can cause signals of different particles to merge and reduce the ability of identifying and reconstructing the particles. Therefore, particle flow identification requires sufficiently segmented subdetectors such that a global event description is possible. The CMS detector is built to meet to requirements of the particle flow reconstruction. It has an efficient and pure muon identification system, a hermetic HCAL with coarse segmentation, a higher segmented ECAL, a fine-grained tracker and a large magnetic field to separate the calorimeter deposits of charged and neutral particles in jets.

#### 4.1.1 Charged particle tracks

An iterative tracking algorithm is responsible for the reconstruction of the tracks made by charged particles in the inner tracking system. Each iteration consists of four steps [64]: the track-seed generation, the pattern recognition algorithm, removal of track-hit ambiguities and a final track fit.

The seed generation is the first step. It consists of finding reconstructed hits that are usable for seeding the subsequent track-finding algorithm. They are identified from a group of at least three reconstructed hits in the tracker, or from a pair of hits while requiring the origin of the track segment to be compatible with the nominal beam-collision point. Since the pixel

has a higher granularity compared to the strip tracker, its seed generation efficiency is higher. The overall efficiency exceeds 99%. The second step of each iteration, the pattern recognition algorithm, uses the seeds as a starting point for a Kalman filter method [145, 146]. This algorithm extrapolates the seed trajectory towards the next tracker layer taking into account the magnetic field and multiple scattering effects. The track parameters are updated when a compatible hit in the next layer is found. This procedure continues until the outermost layer is reached. Since the Kalman filter method can result in multiple tracks associated to the same seed, or different tracks sharing the same hits, a removal of ambiguities is necessary. This ambiguity resolving is done by removing tracks that are sharing too many hits from the list of track candidates. The tracks with the highest number of hits or with the lowest  $\chi^2$  in the track fit is kept. The updated track parameters are then refitted using the Kalman filter method, where all hits found in the pattern recognition step are taken into account. The fit is done twice - once outwards from the beam line towards the calorimeters, and inwards from the outermost track hit to the beam line -, improving the estimation of the track parameters.

All hits that are unambiguously associated to the final track are removed from the list of available hits. In order to associate the remaining hits, the procedure is repeated with looser track reconstruction criteria. The use of the iterative track reconstruction procedure has a high track finding efficiency, where the fake track reconstruction rate is negligible. For muons, this results in a global track reconstruction efficiency exceeding 98%, and 75-98% for charged hadrons.

#### 4.1.2 Following the Muon's Footsteps

The muon reconstruction [147] has three subdivisions: local reconstruction, regional reconstruction and global reconstruction. The local reconstruction is performed on individual detector elements such as strip and pixel hits in the inner tracking system, and muon hits and/or segments in the muon chambers. Independent tracks are reconstructed in the inner tracker - called tracker tracks - and in the muon system, called standalone muon tracks. Based on these tracks, two reconstructions are considered.

The outside-in approach is referred to as Global Muon reconstruction. For each standalone muon track, a inner tracker track is found by comparing the parameters of the two tracks propagated onto a common surface. Combining the hits from the tracker track and the standalone track, gives a fit via the Kalman filter technique [145, 146] for a global muon track.

The second approach is an inside-out reconstruction, creating tracker muons. All candidate tracker tracks with a  $p_T > 0.5$  GeV and total momentum  $p > 2.5$  GeV are extrapolated to the muon system taking into account the magnetic field, the average expected energy losses, and multiple Coulomb scattering in the detector material. The extrapolated track and the muon segments are considered matched when the difference in the position in the x coordinates is smaller than 3 cm, or when the ratio of this distance to its uncertainty is smaller than four. When at least one muon segment - DT or CSC hits - matches the extrapolated track, the corresponding tracker track is indicated as a tracker muon.

For low transverse momenta ( $p_T \lesssim 5$  GeV), the tracker muon reconstruction is more efficient than the global muon approach. This is due to the fact that tracker muons only require a

single muon segment in muon system, while the global muon approach requires typically segments in at least two muon stations. These tracker muons are used for identifying muons from the hadronisation of b or c quarks. The global muon approach typically improves the tracker reconstruction for  $p_T \gtrsim 200$  GeV. These are labelled isolated when in a cone of  $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} = 0.3$  around the muon, the sum of the transverse momenta of additional tracker tracks and energy deposits in the calorimeter is less than 10% of the muon's transverse momentum.

### 4.1.3 The path of the Electron

The electrons in CMS radiate more than 70% of their energy in the inner track through bremsstrahlung before reaching the ECAL. This has as consequence that the electron tracks are increasingly curved in the magnetic field as a function of its flight distance. Standard tracking algorithms are based on Kalman filtering which assume that the energy loss is Gaussian distributed, and are therefore not suitable to fit the electron tracks. A different filtering algorithm, the Gaussian sum filter (GSF) [148] is used in the electron track reconstruction instead.

In CMS, the electrons are reconstructed in two ways. The older ECAL based tracking is developed to identify high energetic isolated electrons. This tracking algorithm starts from ECAL clusters with a transverse energy above 4 GeV and extrapolates from these cluster the position of the hits in the tracker. In order to account for bremsstrahlung, neighbouring clusters in  $\eta$  and  $\phi$  are grouped together into a supercluster from which then the direction is determined to find the position of the particles in the tracker. This has as consequence that for electrons or positrons in jets, energy deposits of surrounding particles will be entering the supercluster leading to a wrong position of the electron/positron in the tracker. Another disadvantage of the ECAL based tracking is that for low  $p_T$  electrons, the trajectories will be very curved and the supercluster will not contain all of the energy deposit, leading to a higher misconstruction rate.

The faults of the ECAL based tracking are lifted by adding a tracker based algorithm. This algorithm uses all the tracks with a  $p_T$  higher than 2 GeV found with iterative tracking as seeds. Iterative tracking uses the Kalman Filter algorithm several times with an average track reconstruction efficiency but high purity. In contrary with a global combinatorial fit, the iterative tracking accepts tracks with a small transverse momentum that are not leaving any energy in the ECAL, and tracks from particles that only interact with the inner tracker layers. When the electron or positron radiated a small amount of energy, the corresponding track can be reconstructed across the whole tracker and safely propagated to the ECAL surface. When there is a larger amount of energy radiated however, the pattern recognition might fail to accommodate for the change in the electron momentum leading to a track reconstructed with a small number of hits. The solution for this is a preselection based on the  $\chi^2$  and number of hits and the selected tracks are fitted again with Gaussian-Sum-Filter which can accommodate substantial energy losses across the trajectory.

The electron seeds from the ECAL- and tracker-based procedures are merged into a unique collection and are then refitted by using the summed Gaussian distributions as uncertainty per hit in the track fit.



1049 The electron efficiency is measured in 8 TeV proton collision data to be better than 93%  
 1050 for electrons with an ECAL supercluster energy of  $E_T > 20$  GeV [149]. For electrons with an  
 1051  $E_T > 25$  GeV in 13 TeV proton collision data, the efficiency is about 96%[150].

#### 1052 4.1.4 Primary Vertex Reconstruction

1053 The primary vertex (PV) reconstruction should be able to measure the location of all proton  
 1054 interaction vertices in each event: the signal vertex and all vertices from pile up events. It  
 1055 consists of a vertex finding and a vertex fitting algorithm and happens in three steps. Tracks are  
 1056 selected to be consistent with being produced promptly in the primary interaction by imposing  
 1057 requirements on the track parameters [74]. By grouping reconstructed tracks according to the  
 1058  $z$  coordinate of their closest approach to the beam line, vertices for all interaction in the same  
 1059 beam crossing are found, at CMS this is done by a deterministic annealing algorithm [151] based  
 1060 on a statistical mechanics model. On top of this, a vertex fitting algorithm like the Adaptive  
 1061 Vertex fitter [152], is performed. This creates the three-dimensional primary-vertex position.  
 1062 With this fit, the contribution from long-lived hadron decays is reduced by down weighting the  
 1063 tracks with a larger distance to the vertex. The primary vertex corresponding to the highest  
 1064 sum of squared track transverse momenta is noted as the point of the main interaction. The  
 1065 resolution on the primary vertex is about 14  $\mu\text{m}$  in  $r\phi$  and about 19  $\mu\text{m}$  in the  $z$  direction for  
 1066 primary vertices with the sum of the track  $p_T > 100$  GeV for 2016 data taking.

#### 1067 4.1.5 Calorimeter clusters

1068 The cluster algorithm in the calorimeter detects and measures the energy and direction of stable  
 1069 neutral particles such as photons and neutral hadron. Additionally, it is used to separate neutral  
 1070 particles from charged hadron energy deposits, and reconstructs and identifies electrons and  
 1071 their bremsstrahlung photons. Furthermore, the cluster algorithm contributes to the energy  
 1072 measurements of charged hadrons that don't have accurate track parameters, e.g. for low  
 1073 quality and high transverse momentum tracks. The clustering is performed separately in each  
 1074 subdetector: ECAL barrel and endcaps, HCAL barrel and end caps, and the two preshower  
 1075 layers. The HF has no clustering algorithm since the electromagnetic or hadronic components  
 1076 give rise to an HF EM or HF HAD cluster.

1077 The clustering algorithm consists of different steps. First seeds are identified when cells have  
 1078 an energy larger than the seeding threshold and larger than their neighbouring cells. Then  
 1079 topological clusters are made by accumulating cells that share at least a corner with a cell  
 1080 already in the cluster and an energy above a cell threshold set to twice the noise level. The third  
 1081 step is an expectation maximization algorithm that reconstructs the cluster [144]. This algorithm  
 1082 assumes that the energy deposits are Gaussian distributed and is an iterative algorithm with  
 1083 two steps at each iteration. The first step calculates the expected fraction of the energy, while  
 1084 the second step performs a maximum likelihood fit. The positions and energies of the Gaussian  
 1085 functions are then taken as cluster parameters.

1086 The calorimeter clusters are used for reconstructing photons and neutral hadrons. The clusters  
 1087 that are not in the vicinity of the extrapolated charged tracks are easily identified as neutral  
 1088 hadrons or photons. For the energy deposits that overlap with charged hadrons however, the  
 1089 neutral particle energy deposit can only be detected as an excess over the charged particle

deposit. For this reason, a good calibration of the electromagnetic and hadronic calorimeter is vital.

The ECAL calibration is performed before the hadron cluster calibration or particle identification. For run 1, the ECAL response to electrons and photons as well as the cell-to-cell relative calibration is determined with test beam data, radio active sources, and cosmic ray measurements. For run 2, the collision data collected at 7 and 8 TeV was used to refine the calibration. The effect of the thresholds in the clustering algorithm are estimated from simulated single photons with energies varying from 0.25 to 100 GeV. The photons used for the calibration should not have a conversion prior to their entrance to ensure the calibration of single clusters. In all ECAL regions and for all energies, the calibrated photon energies agree with the true photon energies within 1%.

In contrary to the photons, the hadrons deposit in general energy in both ECAL and HCAL. Since the calorimeter response in the HCAL depends on the fraction of shower energy deposited in the ECAL, the ECAL and HCAL cluster energies are recalibrated together to get an estimate of the true hadron energy. Since the calibration is done for hadrons, single neutral hadrons such as  $K_L^0$  are used for determining the calibration constants. The hadrons interaction with the tracker material are rejected for the calibration purposes. This calibration is checked with isolated charged hadron selected from early data recorded at  $\sqrt{s} = 0.9, 2.2$  and 7 TeV.

## 4.2 Putting the pieces together

A link algorithm connects the several PF elements from the various CMS subdetectors. It tests any pair of elements in an event and is restricted to considering nearest neighbours in the  $\eta\phi$ -plane. The quality of the link is determined via the distance between the two elements and PF blocks of elements are formed from elements with a direct link or indirect link through common elements.

The link between a central tracker track and a calorimeter clusters is made by extrapolating the tracker track to the two layers of the preshower, the ECAL, and the HCAL. If this extrapolated position is within the cluster area, the two are linked. When there are several ECAL or HCAL clusters for the same track, the link with the smallest distance is kept. A dedicated cluster algorithm accounts for the energy of the photons emitted through bremsstrahlung at for photons that have converted to an electron-positron pair.

The ECAL to HCAL cluster and ECAL to preshower cluster links are established when the cluster position in the more granular calorimeter, ECAL or preshower, is in accordance with the cluster envelope of the less granular calorimeter, HCAL or ECAL. When there are multiple HCAL clusters linked to the same ECAL cluster, the link with the smallest distance is kept. This is also true for multiple ECAL clusters with the same preshower clusters. The ECAL supercluster is linked with the ECAL cluster when they share at least one ECAL cell.

Nuclear interactions in the tracker can lead to kinks in hadron trajectories as well as the production of secondary particles. This leads to charged particle tracks linked together via a common displaced vertex. The displaced vertices considered should have at least three tracks, with at most one incoming track, and the invariant mass of the outgoing tracks should exceed 0.2 GeV. The link between a track and the muon detectors is done via local, regional, and global reconstruction as explained in [Section 4.1.2](#).

### 1132 **4.3 Particle flow identification**

1133 The identification and reconstruction follows a particular order in each PF block. After each  
 1134 identification and reconstruction the corresponding PF elements (tracks and clusters) are  
 1135 removed from the PF block. The muons are the first to be identified and reconstructed. These  
 1136 are reconstructed if their momenta are compatible with corresponding track only momenta. Then  
 1137 the electron and its corresponding brehmstrahlung photons, are identified and reconstructed by  
 1138 using of the GSF tracking. At the same time, the energetic and isolated photons are identified as  
 1139 well. The remaining elements in the PF block are subjected to a cross identification of charged  
 1140 hadrons, neutral hadrons, and photons that arise from parton fragmentation, hadronisation,  
 1141 and decays in jets. The charged hadron candidate is made from the remaining candidates that  
 1142 have a charged particle track associated with them. Then the charged particle energy fraction  
 1143 is subtracted from the calibrated energy of the linked calorimeter clusters and the remaining  
 1144 energy is assigned to the neutral energy. Depending on the excess of neutral energy in the ECAL  
 1145 and HCAL clusters, a photon or a neutral hadron is assigned respectively. The pseudorapidity  
 1146 range of the inner tracker limits the information on the particles charge to  $|\eta| < 2.4$ . Outside  
 1147 this range a simplified identification is done for hadronic and electromagnetic candidates only.

#### 1148 **4.3.1 Muons**

1149 A set of selection requirements based on the global and tracker muon properties is responsible  
 1150 for muon identification. The muons are considered isolated when the additional inner tracks  
 1151 and calorimeter energy deposits within a distance to the muon direction in the  $\eta\phi$ -plane is  
 1152 smaller than 0.3. The muons coming from charged hadron decays or heavy flavour decays need  
 1153 more stringent criteria. This due to the fact that charged hadrons can be misidentified as muons  
 1154 because of e.g. punch-through, or muons can be seen as charged hadrons, and will absorb the  
 1155 energy deposits of nearby particles.

#### 1156 **4.3.2 Electrons and isolated photons**

1157 The electrons and photons are reconstructed together as discussed before. An electron candidate  
 1158 seeded from a GSF track is considered an electron when the linked ECAL cluster is not linked  
 1159 to three or more additional tracks. The photon seeds are ECAL superclusters with transverse  
 1160 energies above 10 GeV that have no links with a GSF track. After associating photons from  
 1161 brehmstrahlung with the associated electrons, the remaining energy is associated to the photons  
 1162 and the photon direction is taken to be that of the supercluster. The electron direction is  
 1163 chosen to be that of the GSF track and its energy is a combination of the ECAL energy with  
 1164 the momentum of the GSF track. Photons are retained if they are isolated, while electrons  
 1165 should satisfy additional criteria based on a multivariate analysis for isolated and non-isolated  
 1166 electrons.

#### 1167 **4.3.3 Hadrons and non-isolated photons**

1168 After muon, electron and isolated photon identification, the remaining particles are hadrons  
 1169 from jet fragmentation and hadronisation. These can show up as charged hadrons (e.g.  $\pi^\pm$ ,  
 1170  $K^\pm$ , or protons), neutral hadrons (e.g.  $K_L^0$  or neutrons), non isolated photons (e.g. from  $\pi^0$   
 1171 decays), and additional muons from early decays of charged hadrons.

The photons and neutral hadrons are assigned to calorimeter clusters without any link to tracks. When the calorimeter clusters between the ECAL and HCAL are linked, the clusters are assumed to arise from the same hadron shower. If there is not such a link, HCAL clusters are assigned to neutral hadrons, while the ECAL clusters are assigned to photons based on the fact that neutral hadrons leave only 3% of their energy in the ECAL. The HCAL clusters linked with tracks, that are not linked with other HCAL clusters, are assigned to charged hadrons. These tracks are then linked with remaining ECAL clusters.

Hadron interactions can result in the creation of extra particles originating from a secondary vertex. These extra particles are identified by having a common secondary vertex and replaced in the PF list as one single primary charged hadron.

#### 4.3.4 Post processing

After identification and reconstruction of all particles as described above. An artificial large missing transverse momentum  $\vec{p}_T$  can be reconstructed. The cause of the  $\vec{p}_T$  is mostly misidentified or misreconstructed high- $p_T$  muons originating from cosmic rays, misconstruction of the muon's momentum, or punch-through charged hadrons. A post processing step is applied to solve this  $\vec{p}_T$ . Events with genuine large  $\vec{p}_T$  due to the presence of neutrino's are unaffected by this post processing.

### 4.4 Pile up mitigation and luminosity measurement

The particle flow algorithm is designed without taking pile up into account. For the 8 TeV dataset, an average of about 21 pile up interactions happen per bunch crossing. For the dataset taken at 13 TeV, the number of pile up interactions increases to about 27 interactions per bunch crossing. These interactions are spread around the beam axis around the centre of the CMS coordinate system and follow a normal distribution with a standard deviation of about 5 cm [144]. The number of pile up interactions is estimated from the number of interaction vertices reconstructed from charged particle tracks, or from the instantaneous luminosity of the given bunch crossing with dedicated detectors and the inelastic proton-proton crossing.

The luminosity of the CMS interaction point is estimated from measuring certain process rates with luminometers such as the pixel detector, HF calorimeter, and the pixel luminosity telescope [153]. The instantaneous luminosity from recorded process rate  $R$  is then determined as

$$Ldt = \frac{Rdt}{\sigma_{\text{fid}}}, \quad (4.1)$$

where  $\sigma_{\text{fid}} = \sigma \times A$  corresponds to the fiducial cross section recorded in the luminometer acceptance  $A$  which is determined using van der Meer scans [154]. The overall uncertainty on the luminosity measurement is estimated to be 2.5%.

The luminosity is used to infer the number of pile up interactions in data, which can be used to correct the predefined pile up interactions in simulation. The average number of pile up interactions in data is estimated per luminosity section by multiplying the luminosity times

the total inelastic cross section. Then an event weight can be derived from the ratio of the distributions of pile up interactions in data and simulation. For 13 TeV collisions, the inelastic cross section is measured to be  $71.3 \pm 3.5$  mb [154]. However a better agreement in data and simulation for the pile up sensitive variables, such as the number of primary vertices, is found with a lower cross section of 69 mb.

The pile up vertices are separated from the primary vertex by requiring that the primary vertex is the vertex with the highest quadratic sum of the transverse momenta of the corresponding tracks. The charged hadrons coming from these pile up vertices are identified via their tracks and are removed from the list of reconstructed particles to be used for physics analysis. This method is the so-called pile up charged hadron subtraction and denoted as CHS [155]. For the reconstructed particles outside the tracker acceptance as well as photons and neutral hadrons, the CHS method doesn't work. Therefore, the transverse density from pile up interactions is estimated using jet clustering techniques and their effect is subtracted from the particles transverse momenta. Additionally, the pile up contribution can be estimated locally as described for the muons and electrons described in Section 4.5.1 and Section 4.5.2.

## 4.5 Physics object reconstruction and identification

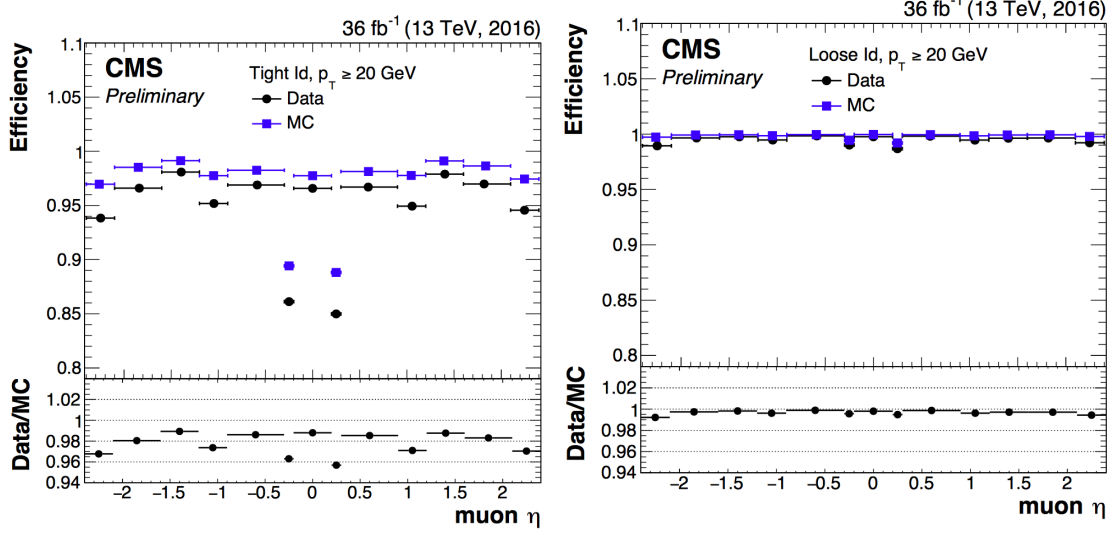
The particle flow objects are used for building physics objects that are used for analysis. These objects are jets, muons, electrons, photons, taus and missing transverse momentum  $\vec{p}_T$ . They are used to compute other quantities such as particle isolation and have extra requirements that are analysis dependent. In the following section, only the physics objects used throughout this thesis are discussed.

### 4.5.1 Muons

The muon candidates used for analysis in this thesis correspond to the tight working point (WP). The tight working point yields the most genuine muons and rejects falsely reconstructed ones. Detailed reports on the performance can be found in [156].

In order to reject objects wrongly reconstructed as muons from hadron showers that reach the muon system (punch-throughs), the global muon fit is required to include at least one valid hit in the muon chambers and for which at least two muon segments in two muon stations is present. Additionally, the muon tracks should have a global fit yielding a goodness-of-fit of  $\chi^2/\text{ndof} < 10$ . The decay of muons in flight is suppressed by requiring at least one pixel hit in the muon track. Furthermore, a minimum of five hits in the tracker is required. Cosmic muons and muons originating from pile up interactions are rejected by constricting the distance of the muon with respect to the primary vertex by putting limits on  $d_{x,y} < 2$  mm and  $d_z < 5$  mm. Also muons according to the loose muon working point will be used in the thesis. These are either global muons or tracker muons reconstructed from the particle flow muon object. In Figure 4.2, the muon efficiencies for data and simulation is presented. These efficiencies are estimated from tag-and-probe methods that select  $Z \rightarrow \mu^- \mu^+$  and tag one muon that passes the identification criteria. The other muon is used as probe and one measures how many times it passes the identification criteria to get the efficiency. Overall, the efficiency is about 95-100%, with two drops due to the crack between the wheels of the DT system. The differences between

1244 data and simulation are corrected by applying  $p_T$  - and  $\eta$ -dependent scale factors ( $\epsilon_{\text{data}}/\epsilon_{\text{MC}}$ )  
 1245 to simulated events. In Table 4.1, the muon requirements for the muons used throughout this  
 thesis are summarised.



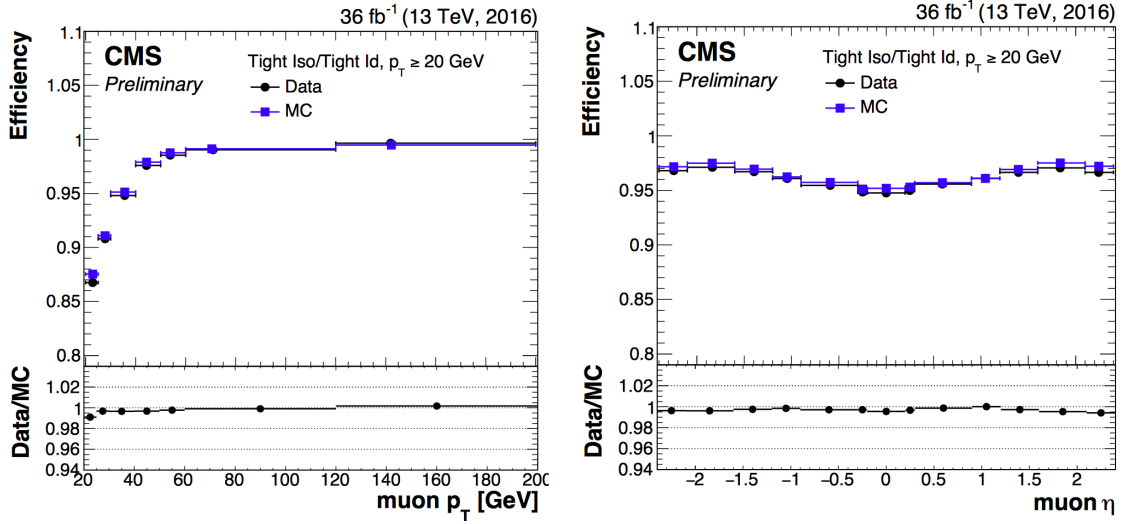
**Figure 4.2:** Comparison of the muon tight ID (left) and loose ID (right) efficiencies in data and simulation as a function of the pseudorapidity of the muon using the full 2016 dataset. Figure taken from [156].

1246

In addition to the identification criteria, the muons are required to be spatially isolated from EM and hadronic activity. The lepton isolation is defined as estimating the total transverse energy of the particles emitted around the direction of the lepton by defining a cone of radius  $\Delta R$  in  $\eta\phi$  plane around the lepton direction. Then a summed energy is calculated from the charged hadrons (CH), neutral hadrons (NH), photons ( $\gamma$ ), excluding the lepton itself. This sum is then corrected to remove the energy coming from pile up interactions. The relative isolation for muons  $\mathcal{I}_\mu$  is defined as [144]:

$$\mathcal{I}_\mu = \frac{\sum p_T(\text{CH}) + \max(0, \sum E_T(\text{NH}), \sum E_T(\gamma) - 0.5 \times \sum E_T(\text{CH}))}{p_T(\mu)}, \quad (4.2)$$

1247 where a cone of  $\Delta R = 0.4$  is adopted and the pile up mitigation is based on the  $\Delta\beta$  correction.  
 1248 The  $\Delta\beta$  correction estimates the pile up energy as half of the contribution coming from charged  
 1249 hadrons. For tight ID muons, this relative isolation should  $\mathcal{I}_\mu < 0.15$ , while for loose muons  
 1250 this should be  $\mathcal{I}_\mu < 0.25$ . In Figure 4.3, the isolation efficiencies as a function of the pseudo  
 1251 rapidities using the tag and probe method are shown for the tight muon ID. The efficiencies are  
 1252 85-100% with a decline for low- $p_T$  muons since they are most likely coming from hadronic  
 1253 or heavy flavour decays. The differences between data and simulation are accounted for by  
 1254 applying  $\eta$ - and  $p_T$ -dependent scale factors on the simulation.



**Figure 4.3:** Comparison of the muon tight isolation requirement with the muon tight ID efficiencies in data and simulation as a function of the transverse momentum (left) or pseudorapidity (right) of the muon using the full 2016 dataset. Figure taken from [156].

**Table 4.1:** Muon requirements for the tight and loose working points, used throughout this thesis.

Property	Loose Muons	Tight Muons
Global muon or Tracker Muon	One or the other	Both
Particle Flow muon	Y	Y
$\chi^2/ndof$ of global muon track fit	N/A	$< 10$
Nb. of hit muon chambers	N/A	$> 0$
Nb. of muon stations contained in the segment	N/A	$> 1$
Size of the transverse impact parameter of the track wrt. to the PV	N/A	$d_{xy} < 2 \text{ mm}$
Longitudinal distance wrt. the PV	N/A	$d_z < 5 \text{ mm}$
Nb. of pixel hits	N/A	$> 0$
Nb. of tracker layers with hits	N/A	$> 5$
Relative Isolation	$< 0.25$	$< 0.15$



### 4.5.2 Electrons

The electrons candidates used correspond to the tight and veto working points. The study of the electron reconstruction and identification performance can be found in [150].

Starting from an electron PF candidate with a GSF track that is outside the barrel-endcap transition region ( $1.4443 < |\eta| < 1.5660$ ), several requirements are set. The electrons from photon conversions are dismissed by requiring the electron track to have not have more than one (two or three) missing hit in the innermost layer for the tight (veto) working point. Additionally, a photon conversion veto is applied by testing if a pair of electron tracks is originating from a common displaced vertex. For the 8 TeV datasets more refined cuts are placed on the electron object using a multivariate analysis. For the 13 TeV dataset this is replaced with more refined cuts on the shower shape variables such as the difference in  $\eta$  or  $\phi$  between the energy weighted supercluster position in the ECAL and the track direction in at the innermost tracker position ( $\Delta\eta_{\text{in}}, \Delta\phi_{\text{in}}$ ), and the ECAL crystal based shower covariance in the  $\eta$  direction ( $\sigma_{\eta\eta}$ ). These cuts also include energy related variables such as the absolute difference between the inverse electron energy measured in the ECAL and the inverse momentum measured in the tracker ( $|1/E - 1/p|$ ), and the ratio of the energy measured in the HCAL and ECAL (H/E). Unlike the muon case, the identification criteria also contain requirements on the isolation of the electrons.

Similar to the muons, the electron relative isolation is determined from the sum of the particles in a cone around the electron itself. The cone radius used for electrons is  $\Delta R = 0.3$  and a  $\rho$  correction for pile up mitigation is applied. For this correction, the expected pile up energy inside the isolation cone is estimated from the median density energy per area of pile up contamination ( $\rho$ ), computed event by event, and the effective area ( $A_{\text{eff.}}$ ) [144]. This effective area is estimated from simulation and denotes the expected amount of neutral energy from pile up interactions per  $\rho$  within the isolation cone as a function of the pseudo rapidity of the associated ECAL superclusters Table 4.2 shows the values used for 13 TeV data. The relative electron isolation  $\mathcal{I}_e$  is calculated as

$$\mathcal{I}_e = \frac{\sum p_T(\text{CH}) + \max(0., \sum E_T(\text{NH}), \sum E_T(\gamma) - \rho \times A_{\text{eff.}})}{p_T(e)}. \quad (4.3)$$

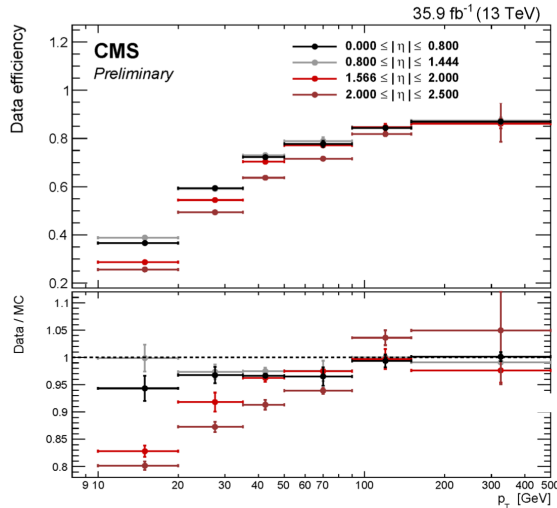
**Table 4.2:** The effective areas  $A_{\text{eff.}}$  used for the electron relative isolation [157].

$\eta$ region	$A_{\text{eff.}}$
$0 <  \eta  < 0.1752$	0.1703
$1.0 <  \eta  < 0.1479$	0.1715
$1.479 <  \eta  < 2.0$	0.1213
$2.0 <  \eta  < 2.2$	0.1230
$2.2 <  \eta  < 2.3$	0.1635
$2.3 <  \eta  < 2.4$	0.1937
$2.4 <  \eta  < 2.5$	0.2393

The efficiency of electron identification is estimated from  $Z \rightarrow e^-e^+$  events via the tag-and-probe method and is shown in Figure 4.4 for the tight working point. The efficiencies reach



1275 95 – 100%. The difference between data and simulation are corrected by dedicated  $p_T$ - and  $\eta$   
 1276 dependent scale factors as well.



**Figure 4.4:** Electron identification efficiency as function of the electron transverse momentum from the full 2016 dataset. Figure taken from [150].

1276

**Table 4.3:** Electron requirements used in this analysis. The requirements are set in the barrel ( $|\eta_{\text{supercluster}}| \leq 1.479$ ) and the end caps ( $|\eta_{\text{supercluster}}| > 1.479$ ).

Properties	$ \eta_{\text{supercluster}}  \leq 1.479$		$ \eta_{\text{supercluster}}  > 1.479$	
	Veto electron	Tight electron	Veto electron	Tight electron
$\sigma_{\eta\eta}$	$< 0.0115$	$< 0.00998$	$< 0.037$	$< 0.0292$
$ \Delta\eta_{\text{in}} $	$< 0.00749$	$< 0.00308$	$< 0.00895$	$< 0.00605$
$ \Delta\phi_{\text{in}} $	$< 0.228$	$< 0.0816$	$< 0.213$	$< 0.0394$
H/E	$< 0.356$	$< 0.0414$	$< 0.211$	$< 0.0641$
relative isolation	$< 0.175$	$< 0.0588$	$< 0.159$	$< 0.0571$
$ 1/E - 1/p $	$< 0.299 \text{ GeV}^{-1}$	$< 0.0129 \text{ GeV}^{-1}$	$< 0.15 \text{ GeV}^{-1}$	$< 0.0129 \text{ GeV}^{-1}$
expected missing inner hits	$\leq 2$	$\leq 1$	$\leq 3$	$\leq 1$
pass conversion veto	Y	Y	Y	Y

### 4.5.3 Jets

Jets are reconstructed from all reconstructed particles without the charged hadrons associated to pile up vertices (PF+CHS jets). The clustering is done via the anti- $k_T$  algorithm [158] with a radius parameter for the cone size of the resulting jet of  $R = 0.5$  for 8 TeV data and  $R = 0.4$  for the more boosted 13 TeV dataset. The initial step of the anti- $k_T$  algorithm considers all candidates as protojets and starts to calculate the distances for protojets  $i$  and  $j$  as

$$d_{ij} = \min\left(\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2}\right),$$

$$d_i = \frac{1}{p_{T,i}^2}.$$
(4.4)

For each iteration the two distances are calculated. When  $d_{ij} < d_i$ , the two protojets are merged and their four momentum is summed. If  $d_i$  is the smallest distance, the protojet is renamed as final jet and ignored in the subsequent steps. More information about the jet algorithm performance can be found in [159].

The jets used for the analysis discussed in this thesis uses the loose identification working point given. The criteria on the constituents of the jets are given in Table 4.4. These requirements find their origin on the assumption that a proper jet originating from the hadronisation of a quark or gluon consists of multiple PF particles and types. The jet should consists of more than one constituent and the neutral hadron fraction and neutral EM energy fractions should be less than 99%. For the jets within the tracker acceptance ( $|\eta| < 2.4$ ), at least one constituent has to be a charged hadron resulting in a charged hadron energy fraction above 0%. Additionally the charged EM energy fraction should be less than 99%. On top of these requirements, objects that are labelled as jets and found in vicinity of any isolated lepton,  $\Delta R < 0.3$ , are removed from the jet collection in that event.

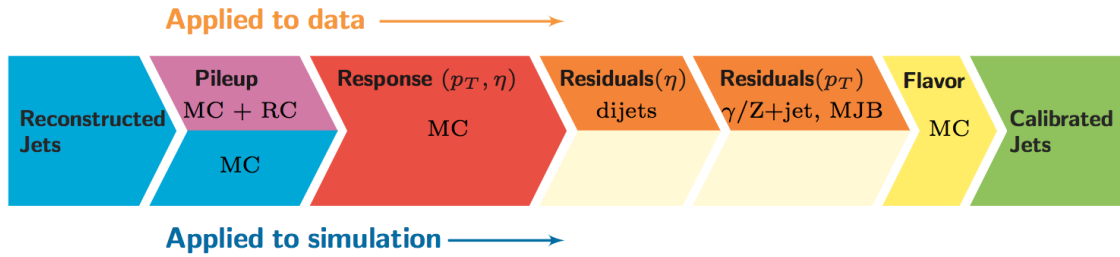
**Table 4.4:** Jet criteria used throughout the thesis. The last three requirements are only for jets within the tracker acceptance.

Properties	Loose Jet ID
Neutral hadron fraction	$< 0.99$
Neutral EM fraction	$< 0.99$
Number of constituents	$> 1$
Charged hadron fraction	$> 0$
Charged multiplicity	$> 0$
Charged EM fraction	$< 0.99$

The energy of the reconstructed jets deviate from the energies of the corresponding jets clustered from the hadronisation products of true partons from simulations due to non linear subdetector response and efficiencies. The jet energy corrections (JEC) calibrate the jets in order to have the correct energy scale and resolution. Jet energy scale corrections (JES) are determined as a function of pseudorapidity and the transverse momentum from data and simulated events by combining several channels and methods. This is extensively described

in [160]. These corrections account for the effects of pile up, the uniformity of the detector response, and residual data-simulation jet energy scale differences. Furthermore, the jet energy resolution (JER) is measured in data and simulation as function of pile up, jet size and jet flavour. A detailed understanding of both the energy scale and the transverse momentum resolution of the jets is crucial for many physics analysis, and these are commonly the main source of systematic uncertainty. The performance of the jet energy corrections for the 13 TeV dataset can be found in [161].

The JEC are factorised and subsequently correct for the offset energy due to pile up, the detector response to hadrons, and residual differences between data and simulation as a function of the jet pseudorapidity and transverse momentum. The consecutive steps of JEC are shown in Figure 4.5. The off set corrections remove the dependence of the jet energy response of



**Figure 4.5:** The sequence of the JEC for data and simulations. The corrections marked with MC are derived from simulation studies, while RC stands for random cone, and MJB for the analysis of multi-jet events. Figure taken from [160].

additional pile up activity. It is based on the jet area method, which uses the effective area of the jets multiplied by the average density in the event to calculate the offset energy to be subtracted from the jets. The correction factors are derived by comparing the jet response with and without pile up events overlaid. The residual differences between data and detector simulation are determined using the random cone method (RC). For this method, many jets are reconstructed in each event by clustering particles through placing random cones. This provides a mapping of the  $\eta\phi$ -space and the average  $p_T$  of those jets gives the average energy offset due to pile up [160]. The next level of corrections have as goal to have an uniform energy response independent of the transverse momentum or pseudorapidity of the jet. These corrections are determined from simulated events by matching the reconstructed to true particle jets and comparing their momenta. The residual corrections between data and simulation are determined by comparing the transverse momentum balance in various types of events (multi-jet, Z + jets, and  $\gamma$  + jets), using a reference jet in the barrel region. The jet flavour corrections are optional and not used for this thesis. More information on the jet flavour corrections can be found in [160]. For jets with a transverse momentum above 30 GeV, the uncertainties from the various corrections are 3-5% for the 13 TeV dataset [161].

After applying JEC, the transverse momentum resolution of the jet is extracted from data and simulated events. There are two methods used to rescale the reconstructed four momentum chosen based on whether or not the simulated jet can be matched to a true jet in simulation.

The factors are defined as

$$c_{\text{matched}} = 1 + \frac{p_T^{\text{reco.}} - p_T^{\text{true}}}{p_T^{\text{reco.}}} (s_{\text{JER}} - 1),$$

$$c_{\text{unmatched}} = 1 + N(0, \sigma_{\text{JER}}) \sqrt{\max(s_{\text{JER}}^2 - 1, 0)},$$
(4.5)

where  $N(0, \sigma_{\text{JER}})$  denotes a sample value from a normal distribution centred at zero with as standard deviation the relative resolution in simulation  $\sigma_{\text{JER}}$ , and  $s_{\text{JER}}$  the  $\eta$ -dependent resolution scale factors. These scale factors are derived from data from di-jet or  $\gamma$  + jets events and analysing the  $p_T$  balance. The resolution scale factors (data/simulation) are found to be 1.1-1.2 except for the transition regions around  $|\eta| = 3$  and  $|\eta| = 1.4$  [161].

#### 4.5.4 Jets from b fragmentation

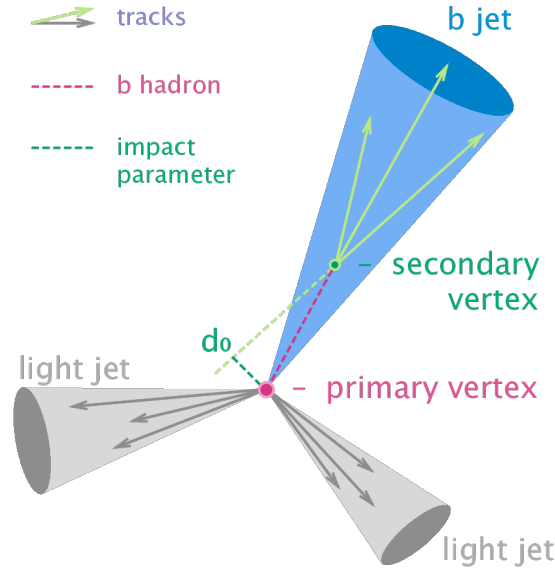
Jets originating from the hadronisation of bottom quarks can be discriminated from jets from gluons and light-flavour quarks as well as charm quark fragmentation through the use of b-tagging. There are a multitude of algorithms developed within CMS to perform b-tagging [162, 163] on jets that fall within the pseudorapidity acceptance of the trackers. These algorithms exploit the properties of the b quark to identify the jets formed by its fragmentation. These hadrons have relative large masses, long lifetimes and daughter particles with hard momentum spectra. Additionally, their semi-leptonic decays can be exploited as well. To use b-jet identification in an analysis, one needs to know its efficiency and misidentification probability. In general these are function of the pseudorapidity and transverse momentum of the considered jet. Their performances are directly measured from data by use of b-jet enriched jet samples (multi-jet or top-quark decays).

This thesis uses b-jets identified by the Combined Secondary Vertex version 2 (CSVv2) algorithm [162]. This algorithm combines secondary vertices together with track based lifetime information by use of a multivariate technique. The secondary vertex is reconstructed from displaced tracks within a jet, as illustrated in Figure 4.6. The final state b quark is encapsulated in a B meson (e.g.  $B^\pm$ ,  $B_0$ ,  $B_s$ ) after the hadronisation. This B meson has relatively long lifetime and can travel a measurable distance from the primary vertex before decaying<sup>1</sup>. After reconstruction, the secondary vertices are required to be in accordance with the B meson hypothesis bases on the amount of shared tracks with the primary vertex, the invariant vertex mass to reject kaon decays, and the direction of the tracks compared to the jet axis.

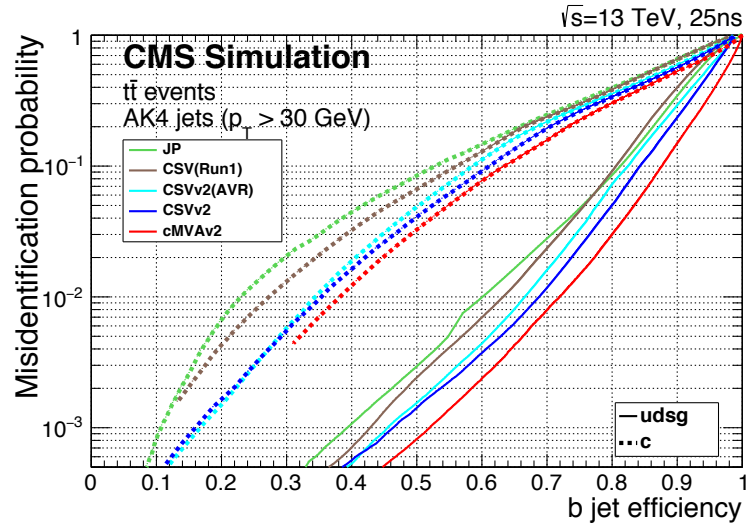
The b-tagging algorithm performances are evaluated taking into account two cases: discrimination of b-tagged jets originating from charm quarks, and discrimination of b-tagged jets against jets coming from gluons or light (u, d, s) quarks. In Figure 4.7, the misidentification probabilities for different b-tagging algorithms within CMS are shown. Based on the misidentification probabilities for a certain threshold on the CSVv2 discriminator, different working points (WP) are defined. These are shown in Table 4.5. The analysis presented in this thesis uses the loose working point which has an average efficiency of 81% and a misidentification rate of 10%.

<sup>1</sup>For example,  $B^\pm$  mesons have a lifetime of about 1.6 ps [4] and travel 4-9 mm before decaying if their momenta is 40-100 GeV.

**NOTE:** Figure 4.6 shows reason why I am not using cMVA, omdat CSVv2 op ttbar events is gemeten en cMVA op multijet?



**Figure 4.6:** Sketch showing the common principle of the identification of b-jets. Figure taken from [164]



**Figure 4.7:** Misidentification probabilities of various b-tagging algorithm in simulation. Figure taken from [163].

**Table 4.5:** Working points used for tagging jets as coming from b quarks for the CSVv2 discriminant.

WP	CSVv2 discr cut	b-tag eff.	misid. prob.
Loose (L)	$> 0.5426$	$\approx 81\%$	$\approx 10\%$
Medium (M)	$> 0.8484$	$\approx 66\%$	$\approx 1\%$
Tight (T)	$> 0.9535$	$\approx 46\%$	$\approx 0.1\%$

The efficiency of tagging a jet as coming from a bottom quark in simulation typically deviates somewhat from data. Efficiency scale factors  $\epsilon_b^{\text{data}}/\epsilon_b^{\text{MC}}$  are derived from data to account for those differences. These scale factors are  $\eta$ -,  $p_T$ -, and flavour dependent, where the flavour of the jet is determined from matched generated hadrons. For cut based analyses these scale factors are applied to the b-tagging efficiencies and mistag rates according to the chosen working point [163]. For shape based analysis however, such as the one presented in this thesis, the scale factor are applied on the distribution of the b-tagging discriminator. This is the so-called IterativeFit method [165]. It uses a tag and probe method to measure the scale factors for both b, c, and light flavoured jets simultaneously. The scale factors to account for the differences in simulation and data for the probe jet are determined iteratively to account for the impact of the b-, c-flavour, and light flavour scale factors on eachother. In a first step, no scale factors are applied. Then the scale factor is measured by applying the scale factors of the previous iteration to simulation until the scale factors become stable. Throughout the procedure, the scale factor for charm jets are set unity with an uncertainty that is twice the one of the b scale factor. The scale factors obtained in  $\eta$ -,  $p_T$  -, and CSVv2 discriminant values are determined with the bin content N of the considered  $(\eta, p_T, \text{discriminant})$  bin as

$$\begin{aligned} \text{SF}_b &= \frac{N_b^{\text{data}} - N_b^{\text{MC}}}{N_b^{\text{MC}}}, \\ \text{SF}_{g,u,d,s} &= \frac{N_{g,u,d,s}^{\text{data}} - N_{g,u,d,s}^{\text{MC}}}{N_{g,u,d,s}^{\text{MC}}}, \\ \text{SF}_c &= 1. \end{aligned} \tag{4.6}$$

1359 The uncertainties related to the IterativeFit method cover possible shape discrepancies between  
 1360 data and simulation. The uncertainty coming from the purity of the sample is subdivided into  
 1361 two uncorrelated uncertainties based on the purity of the light flavoured (lf) and heavy flavoured  
 1362 (hf) jet contributions in the sample. Furthermore, the jet energy scale results in jets migrating  
 1363 from one  $p_T$  bin to an other, having an influence on bin dependent scale factors. The statistical  
 1364 fluctuations of the limited amount of entries in each bin are also accounted for and have an  
 1365 influence on the scale factor uncertainties. These have four uncorrelated sources: two for heavy  
 1366 flavour and two for light flavour jets. Since the uncertainty on the scale factors for the jets  
 1367 originating from a charm quark (cf) is determined from the uncertainty on the b scale factors  
 1368 resulting in two independent uncertainties [165].

#### 1369 4.5.5 Missing transverse energy

The missing transverse momentum  $\vec{p}_T^{\text{miss}}$  and energy  $E_T^{\text{miss}}$  resulting from particles that do not interact with the detector material, are calculated to balance the vectorial sum of the transverse momenta of all particles:

$$\begin{aligned} E_T &= |\vec{p}_T|, \\ \vec{p}_T &= - \sum_{i=1}^{N_{\text{particles}}} \vec{p}_{T,i}. \end{aligned} \tag{4.7}$$

1370 The  $z$ -component can not be calculated from the momentum imbalance since the boost along  
 1371 the  $z$ -axis, determined by the momentum fraction, can not be reconstructed.

The missing transverse energy is influenced by the minimum thresholds in calorimeters, the inefficiencies in the tracker, and the non-linear response of the calorimeter to hadronic particles. The bias is reduced by correcting the transverse momentum of the jets too particle jet  $p_T$  via the JEC and propagating it to the missing transverse momentum taking into account the energy

$$\begin{aligned}\vec{p}_T^{\text{corr}} &= -\sum_{i=1}^{N_{\text{jets}}} \vec{p}_{T,i}^{\text{corr}} - \sum_{i=1}^{N_{\text{unclustered}}} \vec{p}_{T,i}^{\text{raw}}, \\ \vec{p}_T^{\text{corr}} &= \vec{p}_T^{\text{raw}} - \sum_{i=1}^{N_{\text{jets}}} (\vec{p}_{T,i}^{\text{JEC}} - \vec{p}_{T,i}^{\text{PU-only}}).\end{aligned}\tag{4.8}$$

1372 The  $\vec{p}_{T,i}^{\text{PU-only}}$  denotes the transverse momentum of the jet, where only the pile up related  
 1373 corrections are applied. The performance of the missing transverse energy reconstruction can  
 1374 be found in [166].

## 1375 4.6 Summary of corrections

1376 Throughout the chapter several corrections are introduced to improve the agreement between  
 1377 data and simulation. These corrections are sources of systematic uncertainties for the anal-  
 1378 ysis presented in this thesis. Therefore a summary of the corrections and their associated  
 1379 uncertainties is provided.

1380 **Lepton scale factors** The systematic uncertainty on the lepton scale factors consist of three  
 1381 sources: identification, isolation and tracking. The applied scale factors are varied  
 1382 independently within one standard deviation of their measured uncertainties to account  
 1383 for their systematic impact on the measurements.

1384 **Jet energy corrections** The momenta of the reconstructed jets are corrected to match the  
 1385 expected true energy derived from the hadronisation products of partons in simulation.  
 1386 Furthermore, residual corrections and smearing is applied to match the overall energy  
 1387 scale and resolution for simulation and data. These corrections are also propagated to  
 1388 the missing transverse energy. The systematic uncertainties due to these scale factors are  
 1389 estimated by varying them within their uncertainties and repeating the measurements  
 1390 with recalibrated jets and missing transverse energy.

1391 **CSVv2 discriminant shape reweighting** There are three sources of uncertainty contributing  
 1392 to the measurement of the scale factors: statistical uncertainties, jet energy scale and  
 1393 the purity of the sample. The jet energy scale uncertainty is 100% correlated to the jet  
 1394 energy uncertainties and is evaluated simultaneously. The uncertainty coming from the  
 1395 purity of the sample is subdivided into two uncorrelated uncertainties based on the purity  
 1396 of the light flavoured (lf) and heavy flavoured (hf) jet contributions in the sample. A  
 1397 one sigma shift in each of the two purity contributions corresponds to a higher/lower  
 1398 contribution in the purity of the considered flavours. The statistical uncertainties has

1399 four uncorrelated sources, two for heavy flavour and two for light flavour jets. One of  
1400 the uncertainties correspond to the shift consistent with the statistical uncertainties of  
1401 the sample, while the other is propagated in a way that the upper and lower ends of the  
1402 distribution are affected with respect to the centre of the distribution. The uncertainty  
1403 on the charm jet scale factors (cf) is obtained from the uncertainty on the heavy flavour  
1404 scale factors, doubling it in size and constructing two nuisance parameters to control the  
1405 charm flavour scale factors and treating them as independent uncertainties.

1406 **Pile up** Varying the minimum bias cross section, used to calculate the pile up distribution by  
1407  $\pm 4.6\%$ , results in a systematic shift in the pile up distribution. The uncertainty is estimated  
1408 by recalculating the pile up weights to the distributions associated to the minimum bias  
1409 cross sections.

1410 **Luminosity** The luminosity is measured with a global uncertainty of 2.5%, affecting the ex-  
1411 pected number of events.



# Event selection and categorisation

# 5

1412

1413 A basic event selection is made for selecting signal like events. The necessary event requirement  
1414 are discussed in [Section 5.2](#).

1415 The analysis uses signal and background regions to constrain the huge SM background  
1416 compared to the expected signal. [Section 5.4](#) discusses each region that is entering the analysis.  
1417 On top of the use of background estimation from control regions, backgrounds that have prompt  
1418 leptons contaminated by real leptons either from decays of tau leptons or from hadronized  
1419 mesons or baryons (collectively commonly referred as “non-prompt leptons”) as well as by  
1420 hadrons or jets misidentified as leptons<sup>1</sup> are evaluated with a data-driven method discussed in  
1421 [Section 5.5](#).

## 1422 5.1 Baseline event selection and filters

1423 Trigger and filters

## 1424 5.2 Event selection

## 1425 5.3 Effect of the corrections in dilepton events

## 1426 5.4 Regions and channels

## 1427 5.5 Data driven background simulation

---

<sup>1</sup>These two classes of contamination will be referred to as not prompt-lepton (NPL) samples.