SEARCH FOR CHARGINOS, NEUTRALINOS AND SLEPTONS IN OPPOSITE-SIGN DI-LEPTON FINAL STATES AT CENTER-OF-MASS ENERGY OF 13 TEV WITH THE CMS DETECTOR.

By

LEONORA VESTERBACKA OLSSON

A DISSERTATION Doctor of Sciences
DOCTOR OF PHILOSOPHY

ETH Zurich

2019

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation
Doctor of Philosophy

SEARCH FOR CHARGINOS, NEUTRALINOS AND SLEPTONS IN OPPOSITE-SIGN DI-LEPTON FINAL STATES AT CENTER-OF-MASS ENERGY OF 13 TEV WITH THE CMS DETECTOR.

By

Leonora Vesterbacka Olsson

March 2019

Chair: Professor Rainer Wallny
Major: Physics

This document presents two searches for physics beyond the Standard Model (SM), each using 35.9fb^{-1} of proton–proton collision data collected with the CMS detector at a center-of-mass energy of 13 TeV, at the CERN Large Hadron Collider (LHC). The two searches for new phenomena is targeting electroweak production of Supersymmetric (SUSY) particles, so called Charginos, Neutralinos and sleptons, in a production mode that results in two leptons of opposite-sign and same-flavor, large missing transverse momentum, p_T^{miss}.

The document contains a summary of the theoretical framework that make up the SM and SUSY, along with a comprehensive description of the CMS experiment at the LHC accelerator complex. The two searches presented in this thesis both target the production of electroweak SUSY particles, but are divided into two types, according to the production mode. The search for Charginos and Neutralinos result in final states where two or more jets resulting from hadronization are produced, while the search for the direct production of sleptons is chrarcterized by the fact that no hadronization is expected, and thus results in a final state without any jets. The search strategies thus differ slightly, and the two strategies are presented, along with a description of the SM background processes that govern these final states.

Since no excess of collision data is observed with respect to the predicted SM backgrounds in neither of the searches, a statistical interpretation of the results yielding

upper limits in the production cross sections on the SUSY particles, is performed. These limits greatly extend the limits set using 8 TeV collision data during the LHC Run 1.

Concluding remarks commenting on the current absence of evidence for physics beyond the SM (BSM) are given, and an outlook highlighting the unprecedented instantaneous luminosity expected at the LHC, and the window of opportunity for searches for BSM physics that it presents.

CHAPTER 1

EVENT RECONSTRUCTION AND SIMULATION

The basis of any LHC data analysis relies on the concept of an 'event'. An event refers to a succesful collision of two protons that results in the full readout of the CMS detector, and subsequently the combination of subdetector information that forms physics objects. The LHC delivers proton bunches at a rate of 40MHzbut only a small fraction of the protons in the bunches result in collisions of interest to the CMS physics program. Additionally, if all collisions would lead to a full readout of the detector, there would not be enough bandwidth to readout the information, nor space to store it for offline analysis. With these limitations in mind, the CMS has developed a two-tier system to select events of physics interest, the first level (L1) trigger and the high level trigger (HLT). The trigger system is presented in the following section. Furthermore, the way in which the subdetector information is combined to form physics objects, namely the Particle Flow technique, is presented in this section. Apart from only analyzing the recorded events that pass the two-tier trigger system, simulated events of the CMS detector are used to validate the performance of the physics object reconstruction, for the estimation of SM background processes and for visualizing how a potential BSM signal could be found. The last part of this section is presenting how the physics processes are generated and how their interaction with the detector material is simulated.

1.1 Trigger

During 2016 proton-proton run of the LHC, the LHC delivered proton bunches with a time separation of 25ns, and a peak luminosity reaching the unprecedented value of $10^{-34}\text{cm}^{-2}\text{s}^{-1}$. At this peak luminosity, the pp interaction rate exceeds 1GHz, as the mean of the number of interaction per bunch is 25, which is a rate impossible to readout with the technology to date. A two tier triggering system helps differentiate which of the 1GHzcontain interesting events, while discarding the rest. This enables a reduction of the rate down to 400Hz, which is feasible for offline storage.

1.1.1 L1 trigger

The L1 trigger is a hardware system that uses input from the calorimeters and the muon detectors to make a decision to keep the event or not **??**. Tracking information is not included at this step, as it is too time consuming. Instead, information from the
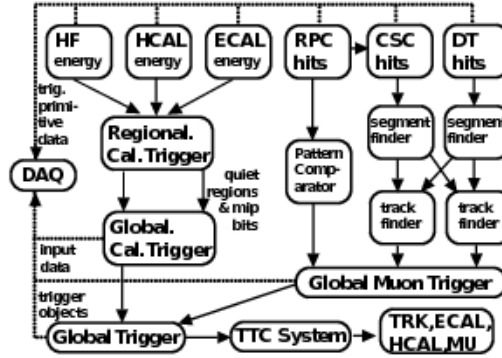
Figure 1-1. L1

calorimeters and the muon detectors is combined into a global trigger, as illustrated in Fig. 1-1. The L1 calorimeter trigger uses inputs like transverse energy and quality flags of the ECAL, HCAL and HF, in the form of Trigger Primitives (TP) from coarsely grouped trigger towers of the calorimeters. This information is provided to the regional calorimeter trigger (RCT), where it is combined to form $e\gamma$ candidates. The next step is the global calorimeter trigger (GCT), where jets are formed using the sum of transverse energy E_T, and the information on the pseudorapidity and τ veto from the RCT is used

to label them as central, forward and τ jets. A crude calculation of the $p_\mathrm{T}^\mathrm{miss}$ can also be performed at L1, by summing the x and y components of the transverse energy in quadrature and rotating the vector by 180°. The objects returned by the GCT, isolated and non-isolated $e\gamma$ candidates, central, forward and τ jets and $p_\mathrm{T}^\mathrm{miss}$, are passed on to the global trigger (GT) and the information is used, together with that of the muon triggers, to decide if an event is kept or not. At this time, there has also been a collection of information gathered from the muon system. Again, no information from the tracker is used at L1, but muons can still be more or less efficiently identified by the muon system. Using the CSC and DT track finders, the tracks of muons can be identified and their p_T. This information, together with muon trigger candidate hits in the RPC, is sent to the global muon trigger (GMT). The information from the GCT and GMT is combined and a decision is made whether to keep the event or not, so called L1-accept. This trigger system has now brought down the rate of 1GHzto 100kHz, and the L1-accept is passed to all subdetectors that are read out and passed on the HLT. A way to further reduce the rates is to scale them down. Some processes with large cross sections, such as QCD, produces single photon events at rates higher than managable, especially at low photon p_T. For this reason, the single photon triggers are *prescaled*, meaning only a fraction of the events are recorded, and the fraction is evolving with the luminosity during data-taking. At analysis level, the events recorded with the prescaled triggers are scaled up according to whatever value they were prescaled with. In this thesis, these kinds of prescaled triggers are used when collecting the single photon sample used for the $p_\mathrm{T}^\mathrm{miss}$ performance studies.

1.1.2   The HLT

While the L1 is completely hardware based and process the information from the detector underground in the experimental cavern, the HLT is both software and hardware based and located in computer farms on the ground level, running on 13,000 CPU cores. At this level, a more thorough object reconstruction is performed with the L1 information. The so called HLT path is a set of algorithms executed in a sequence of steps.

As the tracking is more computing expensive, the first steps is to make a requirement on information from the calorimeters and the muon detectors, before performing the track reconstruction. The basic idea is to enable triggering on high quality objects that eventually can be reconstucted offline, while keeping the rates to a minimum. For this reason, variables such as the p_T and isolation of an object is used in different combination. The rate for triggering on a low p_T muon would be very high if no other requirements are imposed, but if one further requires a well isolated muon, this reduces the rate and possible misidentification of the muon. Conversely, as the production rate of higher p_T muons is lower, one can afford to only impose a p_T requirement and still keep the rate low. As will be seen later in the thesis, this is the reason for the use of the different HLT paths such as `HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL` (involving isolation requirement on both muons) and `HLT_Mu30_TkMu11` (involving higher p_T requirements and no isolation requirements), to ensure triggering on all possible events with interesting physics. Further, HLT paths can involve objects like the Calo $p_\mathrm{T}^\mathrm{miss}$ (computed using only calorimeter deposits), PF $p_\mathrm{T}^\mathrm{miss}$ (computed using only PF jets), H_T (scalar sum of all jet p_T above a threshold) and $\not{H}_\mathrm{T}$ (the missing H_T). The $p_\mathrm{T}^\mathrm{miss}$ triggers are most sensitive to triggering on anomalous events where the large $p_\mathrm{T}^\mathrm{miss}$ is originating from noise, beam halo or other sources, which will be discussed in **??**. In order to keep low rates for these triggers, noise cleaning algorithms are applied at the HLT, and energy deposits associated to beam halo or HB/HE noise is removed from the energy sum and the calorimeter based $p_\mathrm{T}^\mathrm{miss}$ is recomputed. The noise cleaning algorithms are fully efficient in Run 2 and reduce the rate by a factor of 2.5. Additionally, there is even the possibility to get a better estimation of the $p_\mathrm{T}^\mathrm{miss}$ at HLT, by propagating the JECs to the jets and in the computation of the PF $p_\mathrm{T}^\mathrm{miss}$. This results in an improved efficiency and a further rate reduction. This thesis contains a detailed study on the performance of the $p_\mathrm{T}^\mathrm{miss}$ in Run 2, and related to this, Fig. 1-2 is showing the efficiencies of the different $p_\mathrm{T}^\mathrm{miss}$ algorithms at L1 and at HLT compared to the offline $p_\mathrm{T}^\mathrm{miss}$. The HLT reduces the rate from the L1 output of 100kHzto
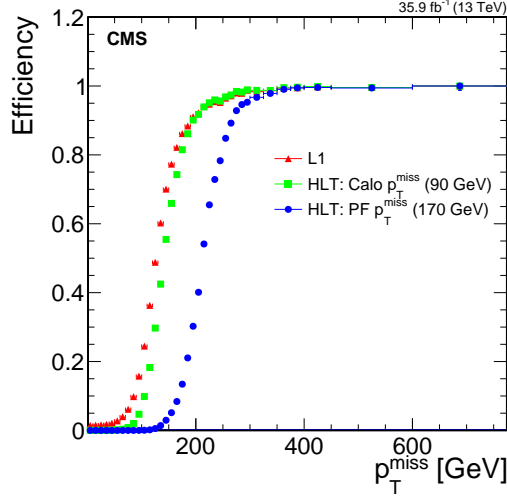
Figure 1-2. L1

1kHz. The data that now have passed the L1 and HLT will be sent of the CERN Tier-0, described in the subsequent section. As the following reconstruction of the data is very computing intense and time consuming, there is the possibility to only use the cores to process some of the data immediately, while some is left for later reconstruction. This concept of data-parking has been used during both Run 1 and 2 of the LHC, and is at the time of writing used to park some of the b-physics data collected during the end of the Run 2.

1.2 Data reconstruction

The computing of the LHC experiments is in large parts made in a four tiered computing system, with decending order of importance. The Tier-0 is the CERN Data center, located onsite in Geneva, and is the first line of tiers that the data collected by the detectors at the LHC experiment is processed at. The Tier-1 consists of 13 computing centers connected to the Grid located all over the world, and share the data reconstruction with the Tier-0, along with providing storage. The next computing tiers are the Tier-2 and Tier-3, which consists of computing resources at universities and institutes where analysis work is for example performed and stored. Once the CMS HLT has decided what events to keep, they are sent to the Tier-0 at CERN, where the reconstruction of the

event is initialized. The CMS software, CMSSW, is a centrally maintained code base that reconstructs objects out of calorimeter deposits, hits in the muon chambers and tracks in the pixel and silicon trackers, and stores them in *Event Data Model* format (EDM).

1.3 Simulated events

Any search or measurement at the LHC are relying on simulated events, so called Monte Carlo (MC) generated events. Physical processes are simulated in a chain, starting from the foundations dictated by the theoretical framework as described in Chapter**??**, followed by the decay, radiation and hadronization of the particles produced, and finally, the simulation of the interaction of the generated process with the detector material. Events are generated from both known SM processes to predict some backgrounds in the searches, and in order to validate the data-driven backgrounds. As the SUSY signals analyzed in this thesis are mulitple, and include numerous assumptions on the masses of the SUSY particles, the full chain of generation for all signal scenarious would be too computationally heavy. For this reason, a faster simulation package is used for the SUSY signals on the level of the detector simulation step, that greatly reduced the computational time by paramtrizing the interactions. The gain in the reduction of the computational time is traded off with a slight decrease in precision, but is taken into account in the statistical analysis. The subsequent chapters contain the description of the three levels of simulation, and the software packages used.

1.3.1 Hard parton scattering

This section relies on the theoretical framework presented in Chapter **??**. At hadron colliders, as opposed to lepton colliders, the energy available in the collisions is distributed over the three valence quarks of the hadrons, the sea quarks and the gluons. In order to properly model the possible interactions, parton distribution functions (PDFs) are used, that dictates the probability of finding a parton within a proton, with a certain energy fraction x of the proton. Further, there are several ways of calculating the matrix element (ME). The square of the MEs are the transition probabilities of the gluons or

(anti-)quarks in to the physical process of interest. The different packages utilizes different order of the strong coupling constant α_s, and are choice of package depends on the requested precision. These packages are known as Monte Carlo (MC) generators, which utilizes the MC random sampling technique.

1.3.1.1 Parton distribution functions

Parton distribution functions explain the probability to find a parton in a proton with some given energy fraction. Due to the non-perturbative nature of partons, i.e. that partons can not be freely observed, the parton distribution functions can not be computed from first principles, however, they can determined by fitting observables to experimental data. One way of determining the PDFs is through the so-called NNPDF3.0, that uses Neural Networks (NN) to model datasets collected by the ATLAS, CMS, HERA-II and LHCb experiments**??**. The NNPDP3.0 tool is developed to be used for searches and measurements during the LHC Run 2, and is used for all generated samples used in this thesis. To demonstrate the behaviour of the parton distributions, the NNPDF3.1 NNLO are displayed in Fig.1-3. N.B that the NNPDF3.1 is a successor of NNPDF3.0, and not used for the MC in this thesis, but lends itself well for a demonstration. As can be seen in Fig.1-3, the red curve, corresponding to the gluon PDF scaled down by a factor of 10, is dominating at the low energies, for both resolution scales. The corresponding interpretation is that the LHC pp collisions are dominated by low energy gluon induced processes, with a sub-dominant contribution of valence quarks at higher energy fractions. This feature is explaining the necessity of the very high collision rate of 40MHz, as most of the interactions are between low-energy carrying gluons, which do not have enough energy to creat "interesting" physics events.

1.3.1.2 Hard scattering processes

The generation of a physics process relies on the calculation of the matrix elements, which states the transition probabilites of the partons involved in the interaction to particles. The MEs can be calculated in many different ways, and each calculation

18
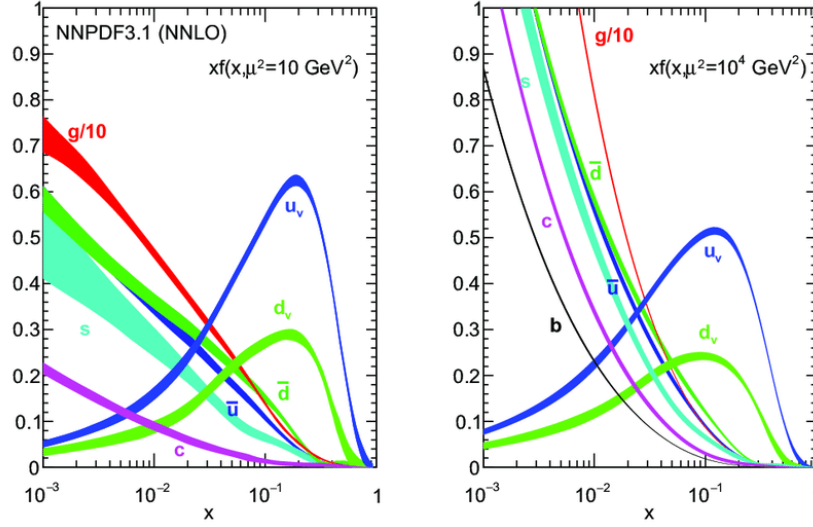
Figure 1-3. The NNPDF3.1 NNLO PDFs, evaluated at two resolution scales; $\mu^2 = 10 \, \text{GeV}^2$ (left) and $\mu^2 = 10^4 \, \text{GeV}^2$ (right)**??**.

corresponds to a MC generator, with the most significant difference being the order of α_s used. Leading order (LO) corresponds to the first order in α_s. In this thesis, a generator that uses LO ME calculation is Madgraph [**?**], which includes contributions with higher jet multiplicities via MLM [**?**] multileg matching and is used to simulate e.g. single γ events. Higher order calculations, such as next-to LO (NLO) and next-to-next-to LO (NNLO) include higher orders in α_s and are therefore more accurately describing the physical process. Madgraph can also be combined with a NLO generator, MC@NLO, into MADGRAPH5_aMC@NLO, and is used when generating e.g. the $WZ \rightarrow 3l1\nu$ process. A full NLO generator is POWHEG [**? ?**], that is used to simulate e.g. the $t\bar{t} \rightarrow 2l2\nu$ process. The complete list of samples used in this thesis will be presented in the analysis section.

1.3.1.3 Multiparton interactions and parton showers

Multiparton interactions (MPI) are a result of the hadron collisions busy nature, where partons other than those involved in the hard scatter event interact. Initial state radiation (ISR) and final state radiation (FSR) are, as the names suggest, an emission of a γ, either before or after the main interaction in the event. Apart from both types

of radiation are adding a photon to the event, ISR and FSR result in slight differences. To illustrate this, we assume a Drell-Yan production, $Z \to e^+e^-$, with ISR and FSR. In the case of the ISR, the photon is radiated by the partons before the production of the Z boson, with a result of carrying off some of the available energy in the collision. Conversely, in FSR, the photon is radiated off of the electron or positron decayed from the Z boson, carrying off some of the energy of that lepton. So if one would reconstruct the invariant mass of the leptons, it would not completely form the Z boson mass of 91 GeV. Instead, one would have to construct the invariant mass of the leptons and the photon in order to recover the Z boson mass. The simulation of both the MPI and ISR anf FSR is done with PYTHIA [?].

1.3.1.4 Hadronization

Colored quarks and gluons produced in the hard scatter process can not exist freely due to color confinement. The mechanism of transforming the colored partons to colorless particles is known as hadronization. The hadronization step is challenging, as is all QCD calculations, and a phenomenological approach is taken in generators such as PYTHIA 8.2 [?]. The Lund string model [?] is the basis of the phenomenological approach, where the QCD field lines can be interpreted as being compressed into a tube like structure, a string. The fragmentation in this model can be viewed as starting in a middle and spreading outwards by repeated breaks of these strings, forming new quark anti-quark pairs.

1.3.1.5 Decay

Finally, PYTHIA is also taking care of the production of resonances and the decay of these unstable particles into stable particles or partons. All resonances are decayed sequentially as part of the hard process, and so the total cross-section as calculated by PYTHIA is dependent upon the available decay channels of the resonance, with the effect that not including a decay channel will decrease the cross-section accordingly. Conversely,

particle decays are performed after hadronisation, and changing the decay channels of a particle will not affect the total cross-section.

1.3.2 Detector simulation

Now that the simulation of the physics processes in hadron collisions have been described, one final step remains in order to be able to compare the data to the simulation, namely how the produced particles interact with the matter in the detector. GEANT4 is a simulation toolkit used to describe the interaction of particles with matter and can simulate everything from tracking of particles bending in a magnetic field to the response of detector components. In GEANT4, a detailed model of CMS is implemented that is taking care of the ionization, multiple scattering, and nuclear interactions and outputs a data format that is similar to that of the collision data. It is this step of the MC simulation that is most computationally heavy, and for this reason, a faster, simpler package is used to simulate the many SUSY scenarios, with the trade off of less precision. Worth noting is that if there would be any hints of SUSY in a particlular scenario, a full simulation would be performed.

1.4 Datasets

Once the collision data and the simulated data has been introduced, the datasets analyzed in this thesis is presented, along with a short introduction of the data formats and weighting of the MC samples.

1.4.1 Dataformats

The collision data reconstruction processed at the Tier-0 is stored in RECO format, a format containing much of the detector information but very large event sizes of 1.2 MB. In order to make the storage and analysis as easy as possible, a set of data tiers are processed that keeps the event sizes to a minimum. The first step in the data tier format processing is the Analysis Object Data (AOD) which has discarded much of the RAW detector information not needed for analysis, and decreased the event size greatly to around 300kB. The last step is the MINIAOD format, that was introduced for the Run

21

II of the LHC. All high level physics objects are included with high level corrections for jets and $p_\mathrm{T}^\mathrm{miss}$, all particles returned by Particle Flow algorithm, all MC truth information for simulation, and all trigger information. By only saving the above information needed for mainstream analyses, the event sizes kept at 30–50 kB. In order to cope with the huge amounts of data collected during Run II, and the various year specific set of MC samples, a further reduction in the event size is needed, and to this end the so called nanoAOD is developed, that manages to keep the event size at 1 kB. The MINIAOD data and MC samples are stored on the Tier-0 and Tier-1, and analysis specific frameworks are used to create subsets of samples containing the particlular datasets needed. As the analyses presented in this thesis are based on different dileptonic datasets, the analysis specific framework used is aimed at picking data collected with dileptonic triggers and the various SM processes containing leptons, and store this in a format called Trees. These trees are used on analysis level for plotting, counting, fitting and statistical analysis.

1.4.2 Event weights

In order to be able to compare the simulted data to the real collision data, each MC sample need to be reweighted according to a so called event weight. The weight takes the form of

$$w = \frac{\sigma \cdot \mathcal{L}}{N} \tag{1--1}$$

where σ is the cross section of that process, $\mathcal{L}$ is the integrated luminosity of the collision data the simulation is compared to, and N is the number of generated events. This weight is then multiplied to each event. From the above equation, one can see that for a process with large cross section, the more simulated events are needed to keep the weight to a desirable small value. The motivation to keep the weight as small as possible, is the statistical error on the MC sample, which goes as the square root of the sum of squared event weights. So the larger the N, the smaller the w and thus the smaller the statistical error.

CHAPTER 2

EVENT SAMPLES AND SELECTIONS

sec:samplesObjects

2.1 Datasets

The data analyzed throughout this thesis consists mainly of the dileptonic streams of 13 TeV pp collision data. Additionally, datasets collected with MET and JetHT triggers are analyzed in the context of the FS background prediction method, and single photon samples are used for the $p_\mathrm{T}^\mathrm{miss}$ study. All datasamples are summarized in Table 2-1.

Several Monte Carlo (MC) event generators are used to simulate the background and signal processes in this analysis, and the detector response is simulated using the GEANT package that provides a detailed description of the CMS detectora and the event reconstruction is perfromed similarly for data as for simulation. The MC samples are weighted to match the data in the number of pileup. The simulation is normalized to luminosity using cross sections from `https://twiki.cern.ch/twiki/bin/viewauth/CMS/SummaryTable1G25ns`. The `PYTHIA8` [?] package is used for parton showering, hadronization and undelying event simulation with the tune `CUETP8M1`, as described in Section 1.1. In Tables 2-2, 2-3, 2-4 and 2-5, the various MC simulations used for the different background prediction methods are listed.

Table 2-1. Datasets used in the slepton/EWK searches and $p_{\mathrm{T}}^{\mathrm{miss}}$ study

Signal events for slepton/EWK search and $p_{\mathrm{T}}^{\mathbf{miss}}$ study
/DoubleEG/Run2016B-03Feb2017_ver2-v2/MINIAOD
/DoubleEG/Run2016(C-G)-03Feb2017-v1/MINIAOD
/DoubleEG/Run2016H-03Feb2017_ver2-v1/MINIAOD
/DoubleEG/Run2016H-03Feb2017_ver3-v1/MINIAOD
/DoubleMuon/Run2016B-03Feb2017_ver2-v2/MINIAOD
/DoubleMuon/Run2016(C-G)-03Feb2017-v1/MINIAOD
/DoubleMuon/Run2016H-03Feb2017_ver2-v1/MINIAOD
/DoubleMuon/Run2016H-03Feb2017_ver3-v1/MINIAOD
Datasets for background prediction
/MuonEG/Run2016B-03Feb2017_ver2-v2/MINIAOD
/MuonEG/Run2016(C-G)-03Feb2017-v1/MINIAOD
/MuonEG/Run2016H-03Feb2017_ver2-v1/MINIAOD
/MuonEG/Run2016H-03Feb2017_ver3-v1/MINIAOD
/JetHT/Run2016B-03Feb2017_ver2-v2/MINIAOD
/JetHT/Run2016(C-G)-03Feb2017-v1/MINIAOD
/JetHT/Run2016H-03Feb2017_ver2-v1/MINIAOD
/JetHT/Run2016H-03Feb2017_ver3-v1/MINIAOD
/MET/Run2016B-03Feb2017_ver2-v2/MINIAOD
/MET/Run2016(C-G)-03Feb2017-v1/MINIAOD
/MET/Run2016H-03Feb2017_ver2-v1/MINIAOD
/MET/Run2016H-03Feb2017_ver3-v1/MINIAOD
Datasets for $p_{\mathrm{T}}^{\mathbf{miss}}$ study
/SinglePhoton/Run2016B-03Feb2017_ver2-v2/MINIAOD
/SinglePhoton/Run2016(C-G)-03Feb2017-v1/MINIAOD
/SinglePhoton/Run2016H-03Feb2017_ver2-v1/MINIAOD
/SinglePhoton/Run2016H-03Feb2017_ver3-v1/MINIAOD

2.1.1 Higher order corrections

For the $WZ \to l^+l^-l\nu$ and $WZ \to l^+l^-2q$ processes, a NLO to NNLO correction factor of 1.109 is applied [**?**]. For the qq→ZZ process, a QCD NLO to NNLO correction factor is applied as a function of generator-level p_{T} of the diboson system, which is described more in depth in Section **??**.

Table 2-2. Simulated SM datasets used for the flavor symmetric (FS) background prediction. All samples are of the `MINIAOD` data format and of the version `RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6/` and `Tune` is short for the pythia8 tune `CUETP8M1`.

Process	Dataset	σ (pb)
ttbar		
$t\bar{t} \rightarrow l^+\nu b + l^-\bar{\nu}\bar{b}$	`/TTTo2L2Nu_TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8`	$831.76\times0.1086^2\times9$
$t\bar{t} \rightarrow l^-\bar{\nu}$+jets	`/TTJets_SingleLeptFromTbar_TuneCUETP8M1_13TeV-madgraphMLM-pythia8`	182.2
$t\bar{t} \rightarrow l^+\nu$+jets	`/TTJets_SingleLeptFromT_TuneCUETP8M1_13TeV-madgraphMLM-pythia8`	182.2
Single Top		
$W^+ \rightarrow t\bar{b}$	`/ST_s-channel_4f_leptonDecays_13TeV-amcatnlo-pythia8`	3.36
$\bar{b} \rightarrow \bar{t}W^+$	`/ST_tW_antitop_5f_NoFullyHadronicDecays_13TeV-powheg`	11.7
$b \rightarrow tW^-$	`/ST_tW_top_5f_NoFullyHadronicDecays_13TeV-powheg`	11.7
$q\bar{b} \rightarrow q'\bar{t}$	`/ST_t-channel_antitop_4f_inclDecays_13TeV-powhegV2-madspin-pythia8`	124.0
$qb \rightarrow q't$	`/ST_t-channel_top_4f_inclDecays_13TeV-powhegV2-madspin-pythia8`	208.0
Diboson (FS)		
$WW \rightarrow l^+\nu l^-\bar{\nu}$	`/WWTo2L2Nu_13TeV-powheg-pythia8`	$(118.7\text{-}3.974)\times0.1086^2\times9$
$gg \rightarrow WW \rightarrow l^+\nu l^-\bar{\nu}$	`/GluGluWWTo2L2Nu_MCFM_13TeV`	$(3.974\times0.1086^2\times9\times1.4$
$gg \rightarrow H \rightarrow WW$	`/GluGluHToWWTo2L2Nu_M125_13TeV_powheg_JHUgen_pythia8`	1.002
WW	`/WW_DoubleScattering_13TeV-pythia8`	1.617
WW	`/WpWpJJ_EWK-QCD_TuneCUETP8M1_13TeV-madgraph-pythia8`	0.037
$q\bar{q} \rightarrow l\nu\gamma$	`/WGToLNuG_TuneCUETP8M1_13TeV-madgraphMLM-pythia8`	405.3
Triboson (FS)		
WWW	`/WWW_Tune_13TeV-amcatnlo-pythia8`	0.209
WWγ	`/WWG_Tune_13TeV-amcatnlo-pythia8`	0.215
Rare (FS)		
$t\bar{t}W$	`/TTWJetsToLNu_Tune_13TeV-amcatnlo-madspin-pythia8`	0.204
$t\bar{t}W$	`/TTWJetsToQQ_Tune_13TeV-amcatnloFXFX-madspin-pythia8`	0.406
$t\bar{t}H$	`/ttHToNonbb_M125_TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8`	0.215
VH	`/VHToNonbb_M125_13TeV_amcatnloFXFX_madspin_pythia8`	0.952
$tttt$	`/TTTT_TuneCUETP8M2T4_13TeV-amcatnlo-pythia8`	0.009
W+jets	`/WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8`	61527

Table 2-3. Simulated SM datasets used for the ZZ to 4 lepton control regions. All samples are of the `MINIAOD` data format and of the version `RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6/` and `Tune` is short for the pythia8 tune `CUETP8M1`. The k-factor referred to is specified in Subsection 2.1.1

Process	Dataset	σ (pb)
ZZ → 4l		
$ZZ \to 4l$	/ZZTo4L_13TeV_powheg_pythia8	1.256× k-factor
$gg \to H \to ZZ$	/GluGluHToZZTo4L_M12513TeV_powheg2_JHUgenV6_pythia8	0.013
$qq \to H \to ZZ$	/VBF_HToZZTo4L_M125_13TeV_powheg2_JHUgenV6_pythia8	0.001
$gg \to ZZ \to 4e$	/GluGluToContinToZZTo4e_13TeV_MCFM701_pythia8	0.001586×2.3
$gg \to ZZ \to 4\mu$	/GluGluToContinToZZTo4mu_13TeV_MCFM701_pythia8	0.001586×2.3
$gg \to ZZ \to 4\tau$	/GluGluToContinToZZTo4tau_13TeV_MCFM701_pythia8	0.001586×2.3
$gg \to ZZ \to 2e2\tau$	/GluGluToContinToZZTo2e2tau_13TeV_MCFM701_pythia8	0.003194×2.3
$gg \to ZZ \to 2e2\mu$	/GluGluToContinToZZTo2e2mu_13TeV_MCFM701_pythia8	0.003194×2.3
$gg \to ZZ \to 2\mu2\tau$	/GluGluToContinToZZTo2mu2tau_13TeV_MCFM701_pythia8	0.003194×2.3
ZZ → 2l2ν		
$ZZ \to 2l2\nu$	/ZZTo2L2Nu_13TeV_powheg_pythia8	0.564× k-factor
$gg \to ZZ \to 2e2\nu$	/GluGluToContinToZZTo2e2nu_13TeV_MCFM701_pythia8	0.001720×2.3
$gg \to ZZ \to 2\mu2\nu$	/GluGluToContinToZZTo2mu2nu_13TeV_MCFM701_pythia8	0.001720×2.3
Others		
$ZZ \to 2l2q$	/ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	3.28
ZZZ	/ZZZ_Tune_13TeV-amcatnlo-pythia8	0.0139
VH	/VHToNonbb_M125_13TeV_amcatnloFXFX_madspin_pythia8	0.952
$q\bar{q} \to l^+l^-\gamma$	/ZGTo2LG_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	123.9

Table 2-4. Simulated SM datasets used for the WZ control regions. All samples are of the `MINIAOD` data format and of the version `RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6/` and `Tune` is short for the pythia8 tune `CUETP8M1`.

Process	Dataset	σ (pb)
$WZ \to l^+l^-l\nu$	/WZTo3LNu_TuneCUETP8M1_13TeV-powheg-pythia8	4.429×1.109
$WZ \to l^+l^-2q$	/WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	5.595×1.109

Table 2-5. Various non flavor symmetric processes. All samples are of the `MINIAOD` data
format and of the version
`RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6/`
and `Tune` is short for the pythia8 tune CUETP8M1.

Process	Dataset	σ (pb)
Drell-Yan		
$Z/\gamma^* \to l^+l^-(50)$	/DYJetsToLL_M-50_Tune_13TeV-madgraphMLM-pythia8	1921.8×3
$Z/\gamma^* \to l^+l^-(10-50)$	/DYJetsToLL_M-10to50_Tune_13TeV-madgraphMLM-pythia8	18610
Various non FS		
$t\bar{t}Z$	/TTZToLL_M-1to10_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.049
$t\bar{t}Z$	/TTZToLLNuNu_M-10_Tune_13TeV-amcatnlo-pythia8	0.253
$t\bar{t}Z$	/TTZToQQ_Tune_13TeV-amcatnlo-pythia8	0.530
tZq	/tZq_ll_4f_13TeV-amcatnlo-pythia8	0.076
tWZ	/ST_tWll_5f_LO_13TeV-MadGraph-pythia8	0.011
WWZ	/WWZ_Tune_13TeV-amcatnlo-pythia8	0.165
WZZ	/WZZ_Tune_13TeV-amcatnlo-pythia8	0.056
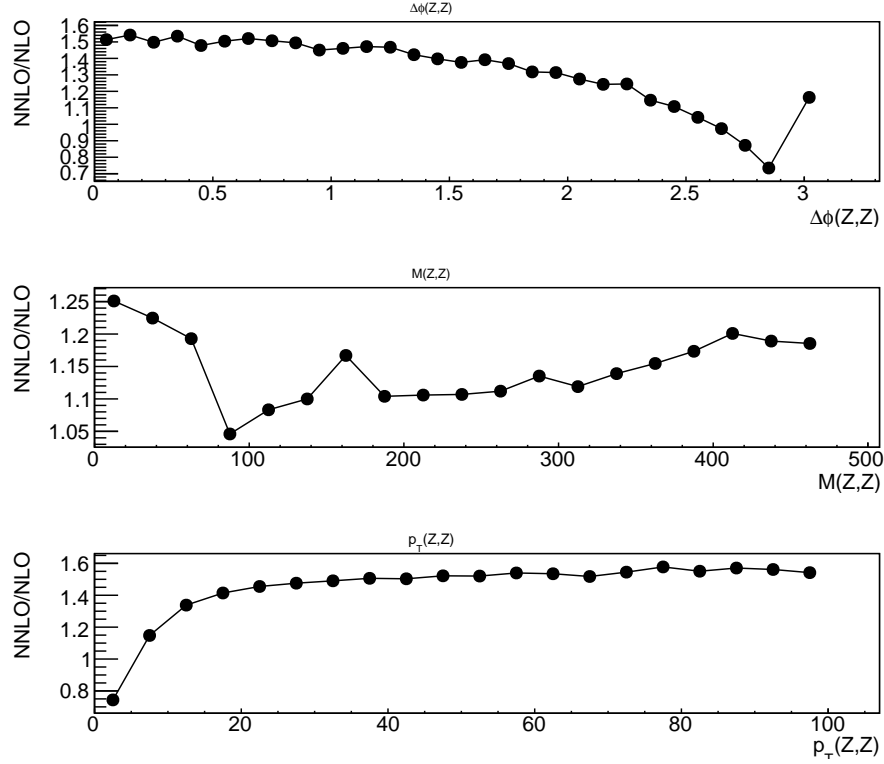
Figure 2-1. QCD NNLO/NLO k factors for the qq→ZZ process in generator level variables
of the diboson system.

2.2 Triggers

The trigger selection for the slepton and EWK analyses is driven by the requirement of at least two leptons at the HLT. More precisely, the signal events in both analyses and the $p_\mathrm{T}^\mathrm{miss}$ study are collected using dielectron and dimuon triggers. Additionally, electron-muon triggers are used to collect a sample dominated by $t\bar{t}$ events for the FS prediction method.

Various requirements on the isolation are imposed on the The triggers are chosen as a combination of These triggers are a combination of non-isolated and isolated requirements on the leptons to ensure The p_T requirements are asymmetric and depend on the flavor composition of the triggers.

Concerning triggers, the selection is dominated by the di-lepton and high $p_\mathrm{T}^\mathrm{miss}$ signatures. Trigger paths requiring isolated leptons are the main paths. Due to the changes in instantaneous luminosity, different di-lepton triggers were active at different times and with varying prescales. This results in the need for a variety of triggers with slighlty different requirements. Non isolated double lepton paths are included to increase the efficiency in events with large dilepton system p_T. Supporting triggers are used for the study of the trigger efficiencies used in the flavor symmetric background prediction method, taken from hadronic events. All triggers which are used one way or the other are documented in Table 2-6.

Table 2-6. Triggers used in the analysis. The first section are the triggers used in most control and signal regions, while the supporting triggers are mostly for the calculation of the trigger efficiencies of the signal triggers.

Signal triggers
Di-muon triggers
HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_v*
HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v*
HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_v*
HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_v*
HLT_Mu27_TkMu8_v*
HLT_Mu30_TkMu11_v*
Di-electron triggers
HLT_Ele17_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*
HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*
HLT_DoubleEle33_CaloIdL_GsfTrkIdVL_v*
HLT_DoubleEle33_CaloIdL_GsfTrkIdVL_MW_v*
Electron-muon triggers
HLT_Mu8_TrkIsoVVL_Ele17_CaloIdL_TrackIdL_IsoVL_v*
HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_v*
HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v*
HLT_Mu17_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v*
HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL_v*
HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL_DZ_v*
HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v*
HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*
HLT_Mu30_Ele30_CaloIdL_GsfTrkIdVL_v*
HLT_Mu33_Ele33_CaloIdL_GsfTrkIdVL_v*
Supporting triggers
HLT_PFHT125_v*
HLT_PFHT200_v*
HLT_PFHT250_v*
HLT_PFHT300_v*
HLT_PFHT350_v*
HLT_PFHT400_v*
HLT_PFHT475_v*
HLT_PFHT600_v*
HLT_PFHT650_v*
HLT_PFHT800_v*
HLT_PFHT900_v*

2.3 Physics Objects

Since the principle behind the main background prediction method, described in a separate AN ([**?**]), relies on the lepton flavor symmetry of the W decay, the identification and isolation requirements of the leptons are chosen so that they are as similar as possible between the flavors. This principle is reflected in the selections of the trigger requirement of the leptons of $p_\mathrm{T} > 23, 17, 12$, and 8 GeV depending on the exact path. Full efficiency for any of these three values is reached at a p_T of 25(20) GeV for the leading (trailing) lepton , which is the cut applied to all leptons. One specific selection of this analysis is that not only the electrons are rejected if they appear in the transition region between the barrel and the endcap, but also the muons. The reason for this is simply that the flavor symmetric background is taken from $e\mu$ events, thus the necessety of having symmetric cuts not only on the efficiency but also on the fiducial regions. For this reason any lepton within the $|\eta|$ region of 1.4 to 1.6 is rejected. In addition, selection criteria on jets and $p_\mathrm{T}^{\mathrm{miss}}$, are presented, which all follow standard CMS SUSY recommendations (including the leptons).

2.3.1 Electrons

Table 2-7 summarizes briefly the most important selection criteria for electrons. Besides kinematical selections, an MVA-trained identification discriminator is used. The discriminator is optimized on electrons from prompt W-boson decays in $t\bar{t}$ versus leptons stemming from so-called "fakes"[1] in $t\bar{t}$. The used working point for the MVA identification discriminator corresponds to the SUSY-PAG-recommended "tight" value, developed at the end of 2016. The actual MVA cut value depends on the lepton p_T and $|\eta|$. In each $|\eta|$ bin the lower value is used for electrons with $p_\mathrm{T} > 25$ GeV while the cut decreases linearly from the upper to the lower value for p_T between 15 and 25 GeV. Additionally, conversion rejection cuts are applied.

---

[1] "Fake" leptons are mostly semi-leptonic b-quark decays where $b \rightarrow cW \rightarrow c\ell\nu$

Table 2-7. Electron selection criteria.

cut	value
Kinematics	
p_T	> 10 GeV
$\|\eta\|$	< 2.4
$\|\eta\|$	$\ni [1.4,\ 1.6]$
Identification	
MVA tight	2016 Working point
conversion rejection	maxLostHits $== 0$
	passConversionVeto()
Isolation	
mini Isolation	< 0.1
Impact parameter	
d_{xy}	0.05
d_z	0.10
SIP3D	< 8

As isolation variable, the mini-isolation is used which features a shrinking cone-size with increasing p_T of the lepton. Thus, the cone size in which the PF particles are summed to calculate the relative isolation is no longer constant, but a function of the p_T of the lepton

$$R = \frac{10.}{min\left[max\left(p_T, 50\right), 200\right]} \quad .$$

For p_T values below 50 GeV, this leads to a constant cone size of 0.2. For p_T values between 50 GeV and 200 GeV, the cone size shrinks from 0.2 to 0.05 at which it remains for higher p_T leptons.

Corrections to the isolation are applied by subtracting the average energy density ρ from the effective geometrical area of the lepton's isolation cone. The variable cone size is taken into account for this correction.

2.3.2 Muons

Selection for the muons follow POG recommendations and are summarized in
Table 2-8. All the variables used are standard variables, and the isolation variable is also
mini-isolation.

Table 2-8. Muon selection criteria.

cut	value
Kinematics	
p_{T}	> 10 GeV
$|\eta|$	< 2.4
$|\eta|$	$\ni [1.4,\ 1.6]$
Identification	
medium Muon ID (POG)	
Isolation	
mini Isolation	< 0.2
Impact parameter	
d_{xy}	0.05
d_z	0.10
SIP3D	< 8

2.3.3 Lepton pair selection

Since there are some events with multiple lepton pairs, it is important to define
an unambiguous way of selecting the "relevant" opposite-sign, same-flavor lepton
pair. The implemented algorithm selects the two highest p_{T} leptons which are fully
identified and that have a distance between them of 0.1 in ΔR. This is to say, there is no
cross-cleaning or prioritization between electrons and muons applied, and non-identified
leptons (including the crack region) do not enter in the consideration of the lepton pair
selection. This is motivated by the expected SUSY signal in which real leptons stemming
from the decay of heavy particles are expected to be "clean" (i.e. well identified) and have
rather large p_{T}.

2.3.4 Isotrack veto

Events with additional leptons are vetoed by using isolated tracks. These tracks are defined using charged PF candidates with different requirements depending on the flavor. PF electrons and PF muons, are required to pass $p_T > 5$ GeV, $|dz| < 0.1$ cm, as well as being associated to the primary vertex with the requirement of fromPV > 1, and track isolation cuts of iso/$p_T < 0.2$ and iso < 8 GeV. The track isolation sum is computed from all charged PF candidates within a cone of $\Delta R < 0.3$, requiring them to pass $|dz| < 0.1$ cm with respect to the primary vertex. Charged PF hadrons are required to pass $p_T > 10$ GeV, $|dz| < 0.1$ cm, be associated to the PV with fromPV > 1, and track isolation cuts of iso/pT < 0.1 and iso < 8 GeV. The track isolation is computed in the same way as for PF leptons above.

2.3.5 Jets and p_T^{miss}

PfCHSjets are considered and standard selections and corrections are followed. The p_T of the jets used for the control regions used for the flavor symmetric background prediction methods is 35 GeV and the $|\eta|$ is required to be < 2.4, in various multiplicity depending on the control region. The jets are required to be seperated from selected leptons by 0.4 in ΔR. Jet energy corrections are applied, namely the `Spring16_23Sep2016V2` set for data and for MC, which is the recommended recipe from the JERC group. In the signal region, a veto on jets of p_T greater than 25 GeV is applied, and the same $|\eta|$ requirement as for the jets used in the background prediction methods. The p_T^{miss} used in this analysis is the Type 1 corrected, which corresponds to the negative of the vectorial sum of all Particle Flow (PF) candidates where the jets are corrected according to the jet energy corrections. p_T^{miss} filters are applied that are designed to reject anomalous p_T^{miss} events due to misreconstruction, detector noise and non-collision backgrounds. These filters are: primary vertex, CSC beam halo, HBHE noise, HBHEiso noise, eebadSC, Ecal

TP, bad muon. The detailed description of these filters can be found in the official page of the MET POG[2] .

2.3.5.1 M_{T2}

The leptonic M_{T2} variable is used to define the signal region. It is a generalization of the transverse mass for pair-produced particles which decay into visible and invisible objects, as described in references [**? ?**]. The visible objects (in this analysis) are the two selected leptons and the invisible objects either the two LSPs from the slepton decay, or neutrinos in the case of di-leptonic $t\bar{t}$ of leptonic WW production. In events where a pair of particles is produced and where the decay products from each are one visible and one invisible particle, the M_{T2} has an endpoint at the mass of the mother particle. Since the main background in this analysis is $t\bar{t}$ and leptons from WW production, this variable is efficiently reducing this background with a cut at the W mass, and is therefor used to define the signal regions in this analysis.

2.3.6   Event selection

To summarize, the following list gives all the details of the minimal event selection used for the analysis.

- two selected leptons of opposite charge with $p_{\mathrm{T}} > 25$ GeV for the leading and $p_{\mathrm{T}} > 20$ GeV for the trailing lepton

- ΔR between the two leptons of > 0.1

- the two highest p_{T} leptons are selected for this pair

- $m_{\ell\ell} > 20$ GeV and a wide Z veto where events with $76 < m_{\ell\ell} < 106$ GeV are rejected.

- $p_{\mathrm{T}}^{\mathrm{miss}} > 100$ GeV

---

[2] https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETOptionalFiltersRun2

CHAPTER 3

PERFORMANCE OF MISSING TRANSVERSE MOMENTUM

At the forefront of SUSY physics program are searches where R-parity is conserved, resulting in one or more lightest SUSY particles that have no SM particles to decay into. The result is LSPs that are unable to interact with the detector material and thus escape detection. The existence of such particles can be inferred by the momentum imbalance in the transverse plane, $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$, with its magnitude denoted $p_{\mathrm{T}}^{\mathrm{miss}}$. When the LSPs are massive, the $p_{\mathrm{T}}^{\mathrm{miss}}$ provides an excellent search tool for SUSY. But other sources can contribute to a large momentum imbalance. Any process with a leptonically decaying W-boson produces a neutrino that escape the detector similarly as the LSP. Additionally, as jets are complex objects to measure, and their energy are corrected through JECs, any over or under measurement in the jets will result in $p_{\mathrm{T}}^{\mathrm{miss}}$. In order to perform a SUSY search where R-parity is conserved, a deep understanding of the $p_{\mathrm{T}}^{\mathrm{miss}}$ object is needed to distinguish the $p_{\mathrm{T}}^{\mathrm{miss}}$ originating from LSPs from SM neutrinos and jet mismeasurements and detector inefficiencies. A challenge for reconstructing physics objects is differentiating tracks from the primary vertex with tracks from overlapping bunch crossings in multiple pp collisions (pileup). A detailed study of the performance of two commonly used $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction algorithms is presented in this chapter, along with a specific study analyzing the performance of the algorithms under extreme pileup conditions, as is expected in the High Luminosity phase of the LHC.

3.1 Missing transverse momentum algorithms in CMS

In collision events, the transverse momentum of the partons is small compared to the energy available in the center of mass, and does not depend on their longitudinal energy. Thus, an assumption that the initial transversal momentum of the system formed by the partons is zero can be made. As an effect of this assumption, if particles escape detection, a transverse energy inequilibrium is created. The final states containing one or more neutrinos therefore result in a significant missing energy corresponding to the vectorial sum of the neutrino momenta. When no neutrinos are created in the event, all missing transverse energy is due to detector inefficiencies and reconstruction issues. Therefore, final states without neutrinos are ideal for the study of the performance of the missing transverse momentum reconstruction algorithms originating from detector effects. The $p_\mathrm{T}^\mathrm{miss}$ is defined as the negative vectorial sum of the particles in the event

$$\vec{p}_\mathrm{T}^{\,\mathrm{miss}} = - \sum \vec{p}_\mathrm{T} \tag{3–1}$$

and its magnitude is denoted $p_\mathrm{T}^\mathrm{miss}$. In CMS, two algorithms for the $p_\mathrm{T}^\mathrm{miss}$ reconstruction are used, PF $p_\mathrm{T}^\mathrm{miss}$ and PUPPI $p_\mathrm{T}^\mathrm{miss}$, and both will be introduced in the following sections.

3.1.1   Particle Flow $p_\mathrm{T}^\mathrm{miss}$ reconstruction

The first reconstruction algorithm is PF $p_\mathrm{T}^\mathrm{miss}$, which is the magnitude of the negative of the vectorial sum of all PF candidates in an event:

$$\vec{p}_\mathrm{T}^{\,\mathrm{miss}} = - \sum_{i \in PF} \vec{p}_\mathrm{T,i} \tag{3–2}$$

As will be shown in the following, the PF $p_\mathrm{T}^\mathrm{miss}$ algorithm is highly performant and is therefore used in the majority of CMS analyses.

3.1.2   PUPPI $p_\mathrm{T}^\mathrm{miss}$ reconstruction

The PUPPI $p_\mathrm{T}^\mathrm{miss}$ algorithm uses the 'pileup per particle identification' method [?]. This method has been developed to reduce the dependence of pileup on physics objects.

3.1.2.1 PUPPI algorithm

In this section, the PUPPI algorithm is summarized, for further clarification please refer to [**?**]. The idea is to estimate how likely the PF candidates are to be originating from pileup, and reweight the particle four-momentum accordingly, with a weight, w_i, close to 1 if the candidate is from the hard scatter and close to 0 for particles from pileup. The procedure to calculate the w_i starts with defining a shape α_i for each particle,

$$\alpha_i = \log \sum_{\substack{j \in \text{event} \\ j \neq i}} \left(\frac{p_{T,j}}{\Delta R_{ij}} \right)^2 \times \Theta(\Delta R_{ij} - R_{min}) \times \Theta(R_0 - \Delta R_{ij}), \qquad (3\text{--}3)$$

where Θ is the Heaviside step function. The α of the i-th particle is thus depending on the p_T of the surrounding particles, and the distance between them in $\eta - \phi$ space, defined as the cone ΔR_{ij}. Only particles within some R_0 around particle i are considered. Surrounding particles j are discarded that are within some minimum radius R_{min} close to the particle i, to reduce the effect from collinear splittings. When the particle i is from hard scattering, the surrounding particles tend to be close in ΔR because of the collinear singularity of the parton shower, resulting in a relatively larger α_i. On the other hand, a wider separation in ΔR is expected for particles originating from pileup, as they should have no correlation with the direction of particle i, resulting in a smaller value for α_i. The p_T of the j-th particles is also used in the calculation of α_i, and the characteristic of this variable is that it is generally softer for particles originating from pileup, yielding the desired smaller value of α_i, and the opposite for when the particle is from the hard scattering. Now that the α_i is defined, the question of what particles should be summed over arise. For this, two regions are used, reflecting the design of the detector; the central region ($|\eta| \leq 2.4$), in which the tracking can distinguish charged tracks from the primary vertex from charged tracks from pileup vertices, and the forward region ($|\eta| > 2.4$) where this discrimination is not possible. Where tracking is available, the PF algorithm can provide the following PF candidates; neutral particles, charged hadrons from the primary vertex and chanrged hadrons from pileup vertices. This results in two different

computations of α_i, namely

$$\alpha_i^C = \log \sum_{\substack{j \in \text{Ch, LV} \\ j \neq i}} \left(\frac{p_{T,j}}{\Delta R_{ij}}\right)^2 \times \Theta(\Delta R_{ij} - R_{min}) \times \Theta(R_0 - \Delta R_{ij}), \qquad (3\text{--}4)$$

$$\alpha_i^F = \log \sum_{\substack{j \in \text{event} \\ j \neq i}} \left(\frac{p_{T,j}}{\Delta R_{ij}}\right)^2 \times \Theta(\Delta R_{ij} - R_{min}) \times \Theta(R_0 - \Delta R_{ij}), \qquad (3\text{--}5)$$

where α_i^C is the sum over all PF candidates, whereas the α_i^F is the sum over all particles in the event. The difference between these two computations is that in the central case, a particle j originating from pileup is discarded from the event, whereas this distinction can not be done in the forward region. However, both methods calculate the w_i from the α_i similarly for each particle, that is used to rescale its four momentum. The actual translation to a weight ranging from 0 to 1 is done by introducing the following quantity

$$\chi_i^2 = \Theta(\alpha_i - \bar{\alpha}_{\text{PU}}) \times \frac{(\alpha_i - \bar{\alpha}_{\text{PU}})^2}{\sigma_{\text{PU}}^2} \qquad (3\text{--}6)$$

where the $\bar{\alpha}_{\text{PU}}$ and $\bar{\sigma}_{\text{PU}}$ are used to characterize the distributions on an event-by-event basis, and defined as

$$\bar{\alpha}_{\text{PU}}^\eta = \text{median}\{\alpha_{i \in \text{Ch,PU}}^\eta\} \qquad (3\text{--}7)$$

and

$$\bar{\sigma}_{\text{PU}}^\eta = \text{RMS}\{\alpha_{i \in \text{Ch,PU}}^\eta\}. \qquad (3\text{--}8)$$

The super script η is C for central or F for forward regions, indicating what region is used for the computation. As can be seen in Eq. 3–6, the χ_i^2 distribution quantifies how much the α_i value is fluctuating from the pileup median $\bar{\alpha}_{\text{PU}}$. Any value of α_i below the $\bar{\alpha}_{\text{PU}}$ is considered pileup like, and due to the definition involving the Heaviside function, these values will result in a χ_i^2 of 0. Conversly, large values of α_i that are far from the $\bar{\alpha}_{\text{PU}}$ will result in a large χ_i^2. Finally, the w_i is defined by

$$w_i = F_{\chi^2, \text{NDF}=1}(\chi_i^2) \qquad (3\text{--}9)$$

with $F_{\chi^2,\mathrm{NDF}=1}$ being the cumulative distribution function of the χ^2 distribution. As a result, whenever the χ_i^2 is 0, the final w_i is 0, whenever the χ_i^2 is large, the final w_i is 1, and all values of χ_i^2 in between results in a fractional weight between 0 and 1.

3.1.2.2 PUPPI $p_\mathrm{T}^\mathrm{miss}$ reconstruction

Now that the particle weights, that are the foundations of the PUPPI algorithms, have been defined the actual rescaling of the particles and how they enter the $p_\mathrm{T}^\mathrm{miss}$ calculation will be covered. For each event, the value of α_i^η are computed for all charged pileup, and the corresponding median and RMS distributions $\bar{\alpha}_\mathrm{PU}^\eta$ and $\bar{\sigma}_\mathrm{PU}^\eta$. When using PF algorithm, the particles available are neutral particles, charged hadrons from the primary vertex, and charged hadrons from pileup. Where tracking is available, the particles originating from pileup can be easily distinguished, and those receive a weight of 0, and are completely discarded in the remainder of the calculation, whereas the charged hadrons from the primary vertex receive a weight of 1. The weights w_i of the remaining particles are calculated, and the four-momentum of these particles is rescaled by this w_i. The charged hadrons from the primary vertex and the rescaled remaining particles used to reinterpret the event, in the context of jet clustering, or in this case, the $p_\mathrm{T}^\mathrm{miss}$ calculation, according to:

$$\mathrm{PUPPI}\ \vec{p}_\mathrm{T}^\mathrm{miss} = -\sum_{i \in PF} w_i \times \vec{p}_\mathrm{T,i} \qquad (3\text{–}10)$$

3.2 Calibration of $p_\mathrm{T}^\mathrm{miss}$

As the $p_\mathrm{T}^\mathrm{miss}$ reconstruction is depending on the accurate measurement of all the reconstructed physics objects, any inefficiency in the reconstruction or minimum energy or p_T thresholds will bias the energy scale of the $p_\mathrm{T}^\mathrm{miss}$. As described in Section **??**, the energy of the jets are corrected with JECs. If these corrections are not taken into account in the computation of the $p_\mathrm{T}^\mathrm{miss}$, there will be a significant bias and imbalance in the event. Therefore, the energy scale of $p_\mathrm{T}^\mathrm{miss}$ is improved by propagating the correction of the p_T of

the jets, $\vec{p}_{T,\text{jet}}^{\text{corr}}$ to p_T^{miss} in the following way:

$$\text{Type 1}\, \vec{p}_T^{\text{miss}} = \vec{p}_T^{\text{miss}} - \sum_{\text{jets}}(\vec{p}_{T,\text{jet}}^{\text{corr}} - \vec{p}_{T,\text{jet}}) \tag{3--11}$$

In the rest of the thesis, the "Raw p_T^{miss} " is the uncorrected p_T^{miss}, and the corrected p_T^{miss}, commonly known as the "Type-1 p_T^{miss}", will be referred to as just p_T^{miss}. As jets are complex objects to measure, the choice of the jets whose corrections should be taken into account is a question of optimization, and subject to future improvement. The motivation behind the 15 GeV p_T threshold of the jets is to reduce the contribution of jets from pileup. As will be shown in the next section, this choice of the p_T threshold gives a response very close to unity. Further, ambiguity can arise if a jet is very close to a recontructed muon, or ressembles an electron or a photon. If a muon reconstructed using the outer tracking system overlaps with a jet, its four momentum is subtracted from the four momentum of the jet, and the JES correction appropriate for the modified jet momentum is used in the p_T^{miss} calculation. Jets are reconstructed from energy deposited in both the HCAL and ECAL, with various fractions of the energy in each calorimeter. In order to not correct jets that are in fact an electron or a photon, i.e. with a large electromagnetic (EM) energy fraction, a choice is made to only correct jets with an EM fraction of less than 90%. The choice of 90% has been providing a well calibrated p_T^{miss} object but can be improved, as will be shown in the next section. The p_T^{miss} relies on the accurate measurement of the reconstructed physics objects, namely muons, electrons, photons, hadronically decaying τ leptons, jets, and unclustered energy (UE). By factorizing the p_T^{miss} into these physics objects, and vary each object within its momentum scale and resolution uncertainties, provides a good estimate of the uncertainty that each of the object contribute to the p_T^{miss}. In the rest of this chapter, the uncertainty in the p_T^{miss} is evaluated by comparing the recalculated p_T^{miss} to the nominal, not varied, p_T^{miss}. As already hinted, the uncertainty related to the energy measurement of the jets is the dominant uncertainty in these measurements. These uncertainties are splitted into those

relating to the JES uncertainties, which are up to 3% (12%) for jets inside (outside) the tracker acceptance, and the JER uncertainties that range between 5–20%. A subdominant uncertainty is that related to the measurement of the muon energy scale, which amuonts to 0.2%, and electron and photon energy scale, which amounts to 0.6% (1.5%) in the barrel (endcap). The UE uncertainty is evaluated based on the momentum resolution of each PF candidate, which depends on the type of the candidate. A detailed description of the PF candidate calibration can be found in Refs. [? ? ?]. The p_T measurement of PF charged hadrons is dominated by the tracker resolution. For PF neutral hadrons, the p_T resolution is dominated by the resolution of the HCAL. The ECAL resolution dominates the PF photon p_T measurement, whereas HF intrinsic resolution dominates that for the PF particles in the HF.

3.3   Event selection

Dilepton and single-photon samples are used to study the $p_\mathrm{T}^\mathrm{miss}$ response and resolution. These samples are chosen as they contain events where no genuine $p_\mathrm{T}^\mathrm{miss}$ from neutrinos is expected, and serves as a good tool to measure the performance of the $p_\mathrm{T}^\mathrm{miss}$ originating from detector inefficiencies or jet mismeasurements.

3.3.1   Dilepton event samples

The dilepton samples are subdivided into two categories based on the flavor of the lepton, namely $Z \rightarrow \mu^+\mu^-$ and $Z \rightarrow \mathrm{e}^+\mathrm{e}^-$ samples. The events for the $Z \rightarrow \mu^+\mu^-$ and $Z \rightarrow \mathrm{e}^+\mathrm{e}^-$ samples are recorded using dimuon and dielectron triggers that select events where the p_T of the two leading leptons are above asymmetric thresholds. Candidate events are required to have both the leading (subleading) lepton p_T greater than 25 (20) GeV and an invariant mass in the range of 80 to 100 GeV, compatible with the mass of the Z boson. In order to have a pure sample of dilepton events originating from Drell–Yan production, a veto is applied on any event containing a third lepton of $p_\mathrm{T} > 20$ GeV. The spectrum of the Z boson transverse momentum, q_T, is shown in

Fig. 3-1 where only the statistical uncertainty in the simulated samples is considered as the dilepton energy resolution is very good.
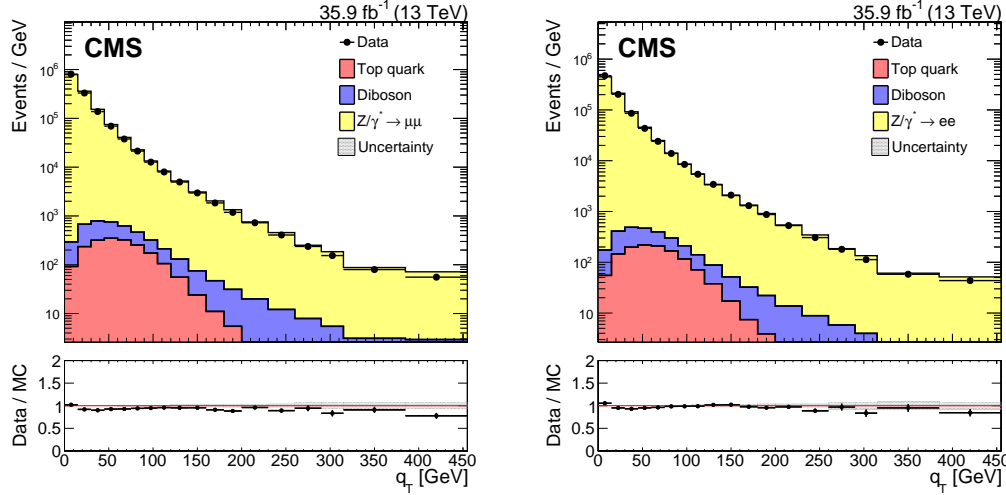

Figure 3-1. Upper panels: Z boson q_T in $Z \to \mu^+\mu^-$ (left) and $Z \to e^+e^-$ (right) samples. The Diboson contribution corresponds to processes with two electroweak bosons produced in the final state. The Top quark contribution corresponds to the top anti-top pair and single top production processes. Lower panel: Data to simulation ratio. The band corresponds to the statistical uncertainty in simulated samples.

3.3.2 Single-photon event sample

The events in the single-photon sample are selected using a set of isolated single-photon triggers with varying thresholds. The p_T thresholds of the triggers are 30, 50, 75, 90, 120, and 165 GeV, and the first five triggers had different L1 accept rate (prescale) during the data taking periods, following the luminosity. Candidate events are weighted based on the prescale values of the triggers. One tight ID photon with $p_T > 50$ GeV is required and events are vetoed that contain leptons of $p_T > 20$ GeV. In order to isolate the events needed for this study, a requirement to have at least one jet with p_T greater than 40 GeV that recoils off of the photon. To match the trigger conditions, the leading photon is further required to have the ratio of the energy deposited in a 3×3 crystal region of the ECAL, centered around the crystal containing an energy deposit greater than all of its immediate neighbors, to the energy of the entire deposit of the photon greater than 0.9.

The photon q_T spectrum is shown in Fig. 3-2. Similarly to Fig. 3-1 only the statistical uncertainty in the simulated samples is considered as the photon energy resolution is very good.
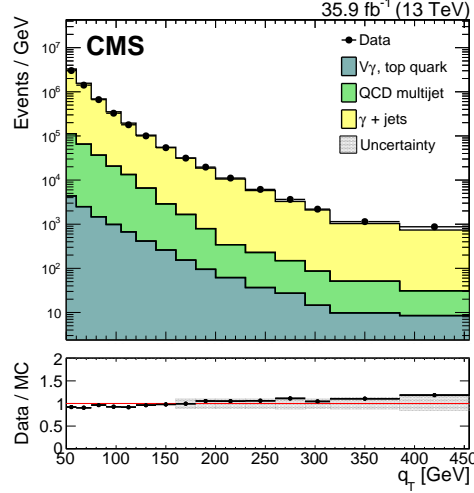

Figure 3-2. Upper panel: Distribution of the photon q_T in the single-photon sample. The Vγ+Top quark contribution corresponds to the Zγ, Wγ, top anti-top pair and single top production processes. Lower panel: Data to simulation ratio. The band corresponds to the statistical uncertainty in the simulated samples.

3.4 p_T^{miss} filters

Non-physical, anomalous high-p_T^{miss} events can arise because of a variety of reconstruction failures or malfunctioning detectors. In the p_T^{miss} study presented below, and in the two analyses covered in this thesis, these anomalous p_T^{miss} events are filtered away on an event basis. In the following, the various filters are presented and the underlying source of the anomalous p_T^{miss}.

3.4.1 HCAL filters

The geometrical patterns of HPD or RBX channels as well as the pulse shape and timing information are utilized by various HCAL barrel and endcap (HBHE) algorithms to identify and eliminate noise. These filter algorithms operate both in "noise filtering" and "event filtering" modes. In the noise filtering mode, the anomalous energy deposits are removed from the event reconstruction; in the event filtering mode, the bunch crossing is

removed from the data set. In addition, there is an isolation-based noise filter that utilizes a topological algorithm, where energy deposits in HCAL and ECAL are combined and compared with measurements from the tracker to identify isolated anomalous activity in HBHE.

3.4.2 ECAL filters

One large source of anomalous $p_\mathrm{T}^\mathrm{miss}$ signals can be created if a hadron hits the front-end electronics of the ECAL super crystals, creating a large false signal. Additionally, anomalously high energetic deposits in the supercrystals, and the lack of information for channels that have nonfunctioning readout electronics, are removed through dedicated noise filters. During the datataking relevant for this thesis, five ECAL endcap supercrystals produced large, anomalous pulses, leading to spurious $p_\mathrm{T}^\mathrm{miss}$. These crystals were removed from the readout, and their energies were not considered. Furthermore, in about 0.7% of ECAL towers, the crystal-by-crystal information is not available. The trigger primitive (TP) information, however, is still available, and is used to estimate the energy. The TP information saturates above $127.5\,\mathrm{GeV}$. Events with a TP close to saturation in one of these crystals are removed.

3.4.3 Beam halo filter

A final large source of anomalous large $p_\mathrm{T}^\mathrm{miss}$ is due to so called machine-induced backgrounds or beam halo, meaning the production of muons when beam protons undergo collisions upstream of the detector. The characterisitcs of beam halo particles is that they travel parallel to the collision axis. If they leave energy deposits in the calorimeters, those will be along a line with constant ϕ. Similarly, any interaction of the beam halo particles in the CSC, will be in line with the calorimeter deposits. The filter exploits information from both the CSC and the calorimeters, and an example event display for a beam halo event is shown in Fig. 3-3, where collinear hits in the CSC are visible. To visualize the need of these event filters, dijet and monojet events are used. Figure 3-4 (left) shows a comparison of the $p_\mathrm{T}^\mathrm{miss}$ before and after the application of the event filters for the dijet

CMS Experiment at LHC, CERN
Data recorded: Thu May 12 03:24:00 2016 CEST
Run/Event: 273158 / 1069617920
Lumi section: 725

Jet 0,
et = 232.25
eta = -1.371
phi = 3.015

MET 0,
pt = 240.94
eta = 0.000
phi = -0.134
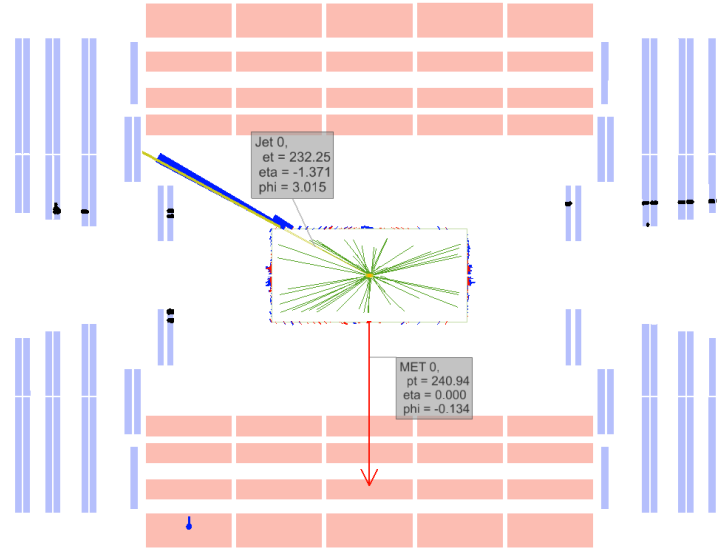
Figure 3-3. Event display for a beam halo event with collinear hits in the CSC (black), $p_{\mathrm{T}}^{\mathrm{miss}}$ of $\sim 241\,\mathrm{GeV}$, and a jet with $p_{\mathrm{T}} =\sim 232\,\mathrm{GeV}$.

sample, where the events with large p_T^{miss} are found to mostly to be due to electronic noise in the calorimeters. Figure 3-4 (right) shows a comparison of the jet ϕ distribution before and after the application of the event filters for the monojet sample, where the excess of events with jet $\phi \approx 0$ or $\phi \approx \pi$ are typical for beam halo events.
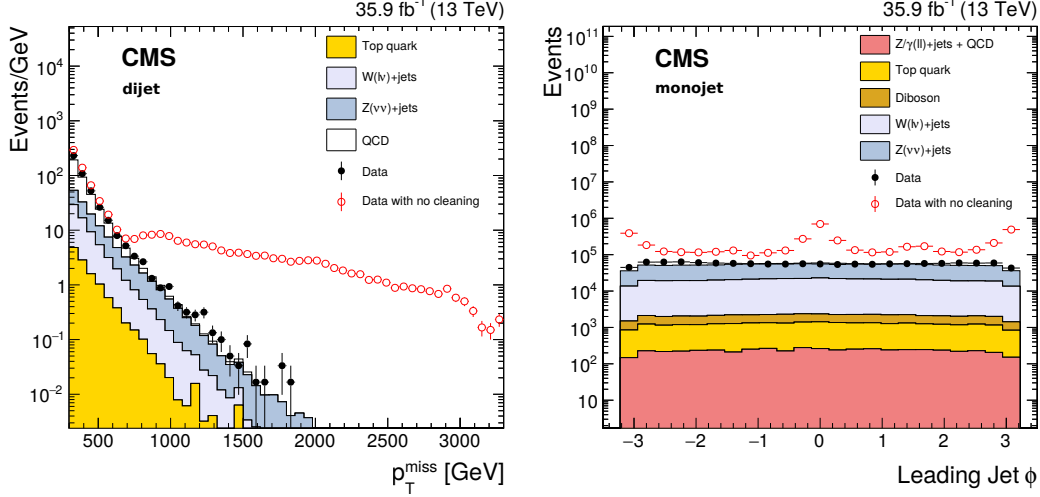

Figure 3-4. The p_T^{miss} (left) and jet ϕ (right) distributions for events passing the dijet (left) and monojet (right) selection with the event filtering algorithms applied, including that based on jet identification requirements (filled markers), without the event filtering algorithms applied (open markers), and from simulation (solid histograms).

3.5 p_T^{miss} performance

After having covered the p_T^{miss} reconstruction algorithms, their calibration and the filtering of anomalous p_T^{miss} events, the main part of the study will be presented. As has already been conveyed, the p_T^{miss} is a very sensitive observable, and is relying on accurate object reconstruction and efficient sub-detectors. Therefore it is a great tool for monitoring the detector and reonstruction performance during data-taking. The performance of the p_T^{miss} can be summarized in three parts. By monitoring the data and simulation of the p_T^{miss} in events with no genuine p_T^{miss} from neutrinos, any anomalies will show up as p_T^{miss} tails, or an overall disagreement between the data and the simulation. The p_T^{miss} response is crucial for validating the JECs or any issues related to the muon, electron or photon energy scale. The p_T^{miss} resolution is valuable for monitoring how the performance degrades

as a function of some variable, such as the number of vertices. Initially, the studies presented in this thesis were performed during the 2016 data-taking, and was crucial in uncovering various reconstruction inefficiencies, and was later refined and summarized in a publication.

3.5.1 p_T^{miss} performance using hadronic recoil

A well-measured Z/γ boson provides a unique event axis and a precise momentum scale. Such events should have little or no genuine p_T^{miss}, and the hadronic recoil is projected onto the axis of this well measured boson, as illustrated in Fig. 3-5. Formally, the hadronic recoil (u) is defined as the vector p_T sum of all PF candidates except for the vector boson (or its decay products in the case of the Zboson decay). The assumption of momentum conservation in the transverse plane imposes $\vec{q}_T + \vec{u}_T + \vec{p}_T^{\text{miss}} = 0$. The
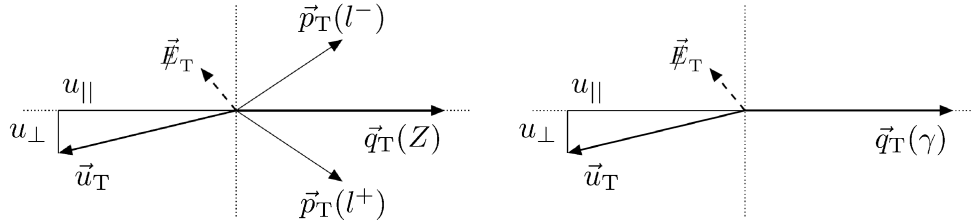

Figure 3-5. Illustration of the Zboson (left) and photon (right) event kinematics in the transverse plane. The vector $\vec{u}_T$ denotes the vectorial sum of all particles reconstructed in the event except for the two leptons from the Zdecay (left) or the photon (right).

hadronic recoil can be split into its parallel and perpendicular components with respect to the boson axis, and these quantities, $u_\parallel$ and $u_\perp$, are used to study the p_T^{miss} response and resolution. The p_T^{miss} response is defined as $-\langle u_\parallel \rangle / \langle q_T \rangle$ where $\langle \rangle$ indicates the mean, and reflects how well balanced the boson is to the hadronic recoil. The p_T^{miss} resolution is estimated through the RMS of the $u_\parallel + q_T$ and $u_\perp$ distributions, and are denoted by $\sigma(u_\parallel)$ and $\sigma(u_\perp)$, respectively.

3.5.2   Performance of PF p_T^{miss} algorithm

The p_T^{miss} in selections with no genuine p_T^{miss} from neutrinos are shown in Fig. 3-6, and show a good agreement between the data and the simulation. The lower pads of Fig. 3-6 show the magnitude of the uncertainties in these events, where the p_T^{miss} resolution is dominated by the resolution of the hadronic activity, ranging up to 10–15% for the jet momentum resolution [?]. Since the momentum resolution for leptons (photons) is σ_{p_T}/p_T of 1–4% (1–3%) [? ?] is sub-dominant to the resolution of the hadronic actitvity, these uncertainties are not taken into account in the ratios, and the final uncertainty shown in the figures include uncertainties in the JES, the JER, and the UE, added in quadrature. The increase in the uncertainty band around 40 GeV is related to the JES and the JER sources in events with at least one jet and no genuine p_T^{miss}. For higher values of p_T^{miss}, where processes with genuine p_T^{miss}, e.g. top quark background, become more dominant, the uncertainty decreases. Distributions of $u_\parallel + q_T$ and $u_\perp$ in $Z \to \mu^+\mu^-$ $Z \to \mathrm{e}^+\mathrm{e}^-$
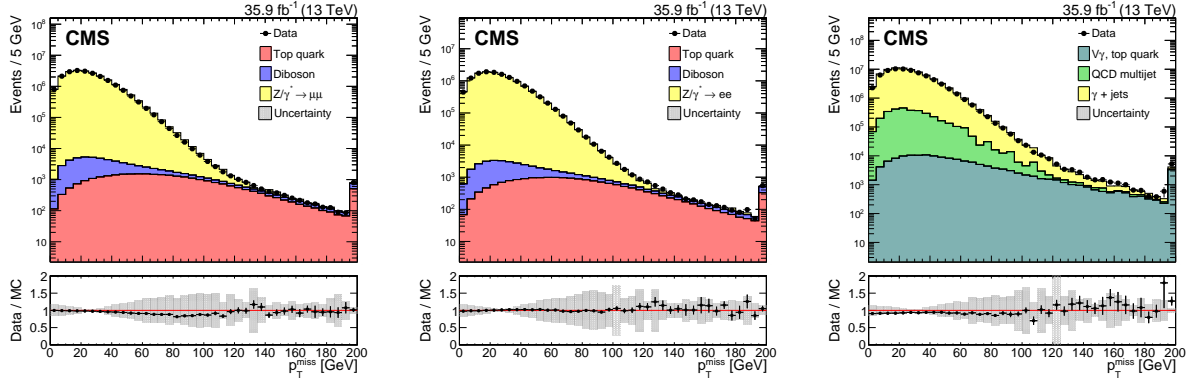

Figure 3-6. The p_T^{miss} for events passing the dimuon (left), dielectron (middle) and single photon (right) selections, in data (black markers) and simulation (solid histograms). The lower bands show the data to simulation ratio with the systematic uncertainties due to the JES, the JER, and variations in the UE are added in quadrature.

and γ+jets events are shown in Fig. 3-7. The kinematic definition of $u_\parallel$ dictates that for processes with no genuine p_T^{miss}, $u_\parallel$ is balanced with the boson q_T. Therefore, the vectorial sum of $u_\parallel$ and q_T results in a symmetric distribution, centered at zero; any deviations from this behavior imply imperfect calibration of p_T^{miss}. In events with genuine

p_T^{miss}, due to the presence of the neutrinos, $u_\parallel$ and q_T are not balanced, leading to an asymmetric distribution as can be seen as the Top quark and Electroweak contribution in the distributions. Due to the assumed isotropic nature of the energy fluctuations of the detector noise and underlying event, the $u_\perp$ distribution is symmetric with a mean value of 0. Figure 3-8 shows the p_T^{miss} response as a function of q_T, in data and simulation,
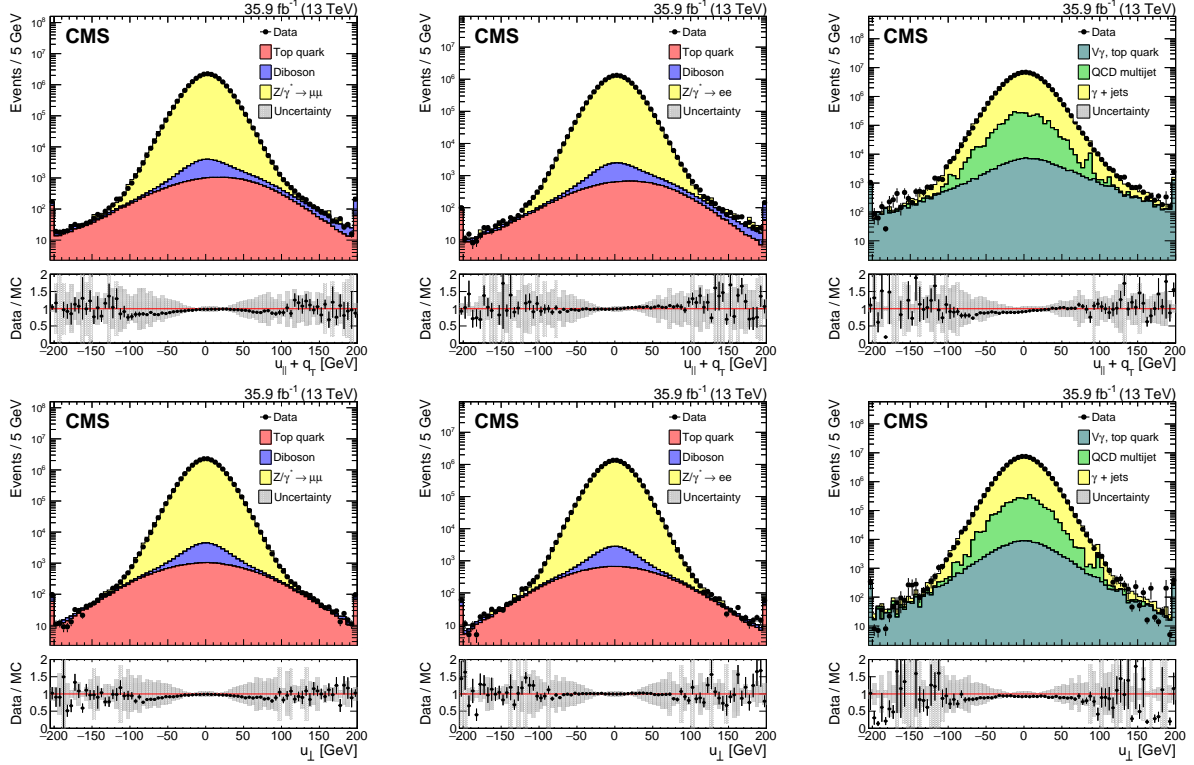

Figure 3-7. Distribution of $u_\parallel + q_T$ (upper) and $u_\perp$ (lower) components of the hadronic recoil, in data (filled markers) and simulation (solid histograms), in the $Z \to \mu^+\mu^-$ (left), $Z \to e^+e^-$ (middle), and γ+jets (right) samples.

in $Z \to \mu^+\mu^-$ $Z \to e^+e^-$ and photon events. Two types of response are displayed. On the left is the response of the uncorrected p_T^{miss} (Raw p_T^{miss}), in which the JECs have not been propagated to the p_T^{miss}. The effect of correcting the p_T^{miss}, the so called "Type-1" correction, is shown on the right, where the response is now much closer to unity, proving that the corrrected jets results in a well balanced event. The underestimation of the hadronic response observed at smaller $q_T \leq 100\,\text{GeV}$ is due to the significant contribution of the uncalibrated component of p_T^{miss}, which mainly consists of jets with $p_T < 15$ GeV

and unclustered particles. The response of p_T^{miss} is found to agree between all three samples within 2%. The residual response difference between the samples stems from the different mechanism used to disambiguate muons, electrons, and photons from jets used in the correction of the p_T^{miss}, as discussed in Section 3.2. Simulation studies has shown that in the case of the electrons and photons, a small fraction ($\lesssim 10\%$) of jets survive the disambiguation criteria yet overlap with prompt electrons and photons. As a result, these jets wrongly contribute to the p_T^{miss} calibration, leading to a 1–2% lower response in the electron and photon channels. The uncorrected p_T^{miss} response show a perfect agreement between the electron and muon events, further proving that the disagreement in the corrected response is due to the propagation of the JECS and the disambiguation criteria of what jets to correct. The resolution of the p_T^{miss} is evaluated as a function of the boson
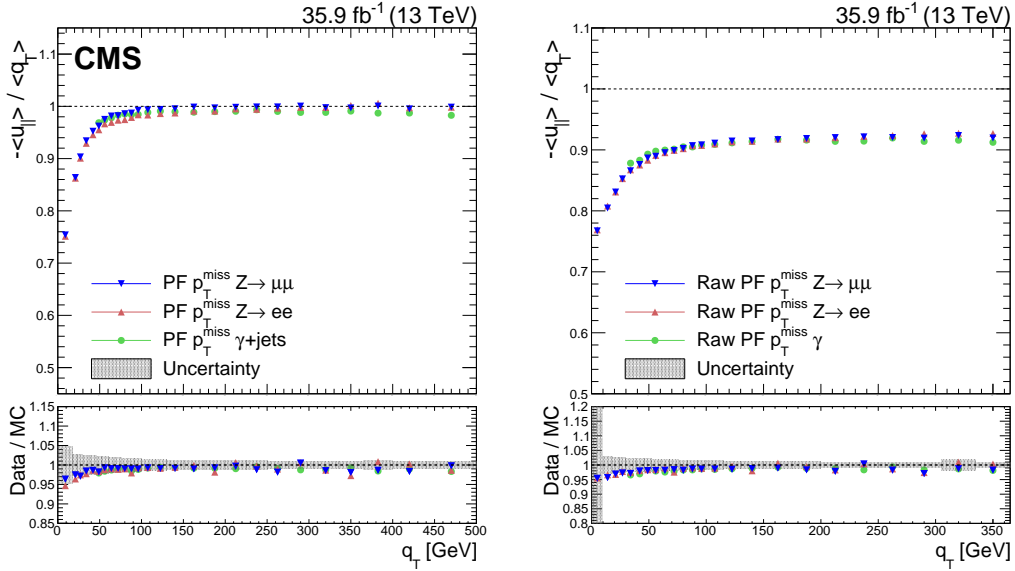

Figure 3-8. Upper panels: Response of the corrected (left) and uncorrected (right) p_T^{miss} in data in $Z \to \mu^+\mu^-$ $Z \to e^+e^-$ and γ+jets events. Lower panels: Ratio of the p_T^{miss} response in data and simulation. The band corresponds to the systematic uncertainties due to the JES, the JER, and variations in the UE added in quadrature, estimated from the $Z \to e^+e^-$ sample.

p_T (q_T), the number of vertices and of the scalar p_T sum of all PF candidates ($\sum E_T$). The upper row of Fig. 3-9 shows the resolution as a function of q_T. In order to compare

the resolution of p_T^{miss} consistently across the samples, the resolution in each sample is corrected for the differences observed in the response, with a resulting negligible impact on the results. The relative resolution both in $u_\parallel$ and $u_\perp$ is found to improve as a function of q_T because of the improved energy resolution in the calorimeters. Furthermore, due to the isotropic nature of energy fluctuations stemming from detector noise and the underlying event, the dependence of the resolution of $u_\perp$ on q_T is smaller than for $u_\parallel$. For $q_T > 200\,\text{GeV}$, the p_T^{miss} resolution is $\approx 13\%$ and $\approx 9\%$, for the $u_\parallel$ and $u_\perp$, respectively. The middle row of Fig. 3-9 displays the resolution as a function of N_{vtx}, showing a significant dependence on N_{vtx}, since only pileup mitigation techniques are employed to the PF jets and not the PF p_T^{miss} algorithm. The resolution is parametrized as:

$$f(N_{vtx}) = \sqrt{\sigma_c^2 + \frac{N_{vtx}}{0.70}\sigma_{PU}^2}, \qquad (3\text{–}12)$$

where σ_c is the resolution term induced by the hard scattering interaction and σ_{PU} is the resolution term induced on average by an additional pileup interaction. The factor 0.70 accounts for the vertex reconstruction efficiency [?]. Results of the parametrization for the $u_\parallel$ and $u_\perp$ components are given in Table 3-1. Each additional pileup vertex is found to degrade the resolution of each component by 3.8–4.0 GeV. Lastly, Fig. 3-9 (lower row) shows the resolution of $u_\parallel$ and $u_\perp$ as a function of the scalar p_T sum of all PF candidates ($\sum E_T$). The resolutions measured in different samples, and in data and simulation, are found to be in good agreement. The relative p_T^{miss} resolution improves with increasing $\sum E_T$, driven by the amount of the activity in the calorimeters. The resolution in different samples is parametrized as:

$$\sigma_{u_\perp, u_\parallel} = \sigma_0 + \sigma_s\sqrt{\sum E_T}, \qquad (3\text{–}13)$$

where σ_0 is the resolution term induced by intrinsic detector noise and σ_s is the stochastic resolution term. Results of the parametrization for the $u_\parallel$ and $u_\perp$ components are given in Table 3-2.
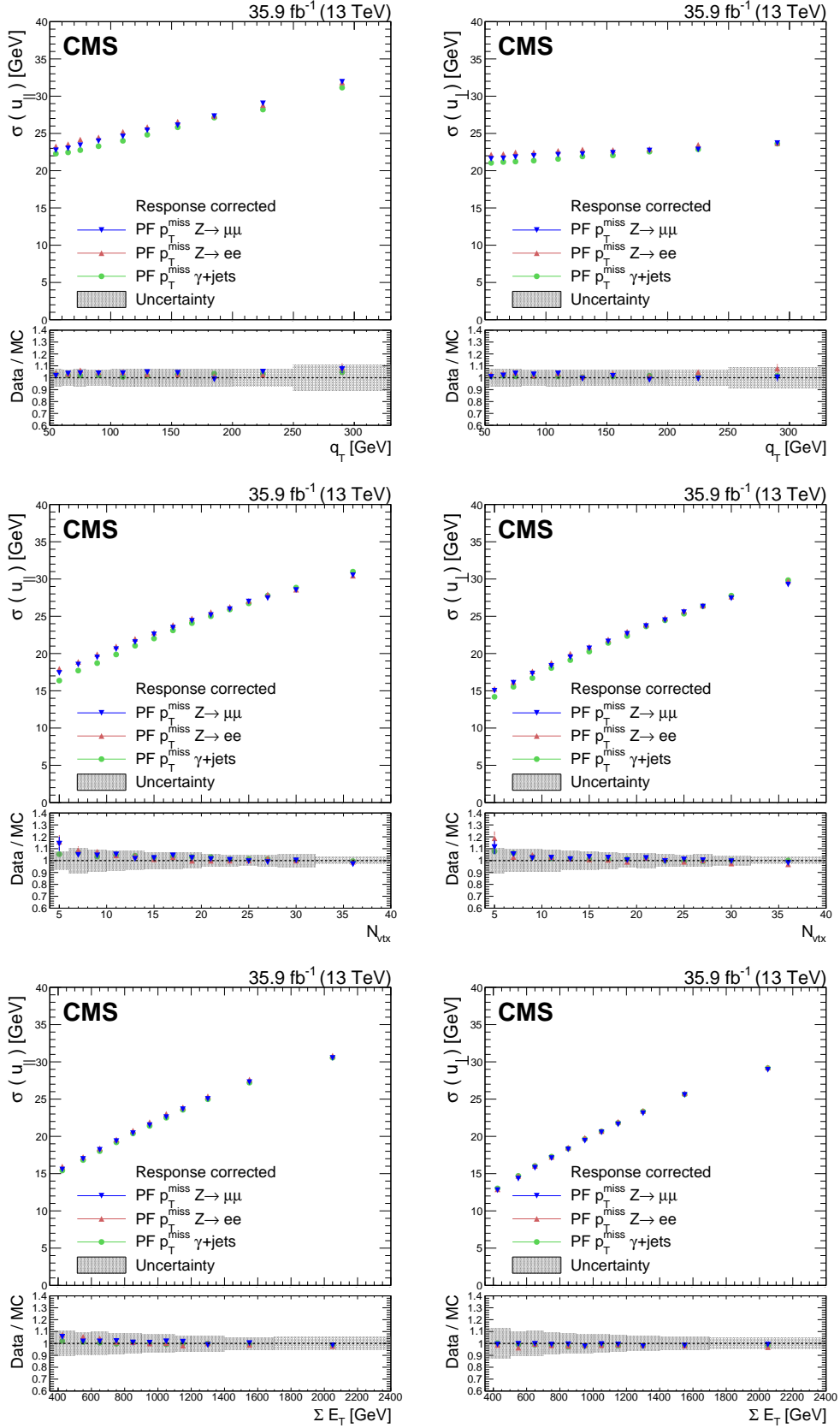
Figure 3-9. Resolution of the $u_\parallel$ and $u_\perp$ components of the hadronic recoil as a function of, q_T (upper row), the reconstructed vertices (middle row), and the scalar p_T sum of all PF candidates (lower row), in $Z \rightarrow \mu^+\mu^-$ $Z \rightarrow e^+e^-$ and γ+jets events. In each plot, the upper panel shows the resolution in data, whereas the lower panel shows the ratio of data to simulation. The band corresponds to the

Table 3-1. Parametrization results of the resolution curves for the $u_\parallel$ and $u_\perp$ components as a function of N_{vtx}. The parameter values for σ_c are obtained from data and simulation, and the values for σ_{PU} are obtained from data, along with a ratio R_{PU} of data and simulation. The uncertainties displayed for both components are obtained from the fit, and for simulation the JES, the JER, and UE uncertainties are added in quadrature.

Process	σ_c(data)[GeV]	σ_c(MC)[GeV]	σ_{PU}(data)[GeV]	$R_r = \sigma_{PU}$(data)$/\sigma_{PU}$(MC)
		$u_\parallel$ component		
$Z \to \mu^+\mu^-$	13.9 ± 0.07	11.9 ± 1.53	3.82 ± 0.01	0.95 ± 0.04
$Z \to e^+e^-$	14.6 ± 0.09	12.0 ± 1.09	3.80 ± 0.02	0.95 ± 0.03
γ+jets	12.2 ± 0.10	10.2 ± 1.98	3.97 ± 0.02	0.97 ± 0.05
		$u_\perp$ component		
$Z \to \mu^+\mu^-$	10.3 ± 0.08	8.58 ± 2.20	3.87 ± 0.01	0.97 ± 0.04
$Z \to e^+e^-$	10.7 ± 0.10	8.71 ± 1.76	3.89 ± 0.01	0.96 ± 0.03
γ+jets	9.04 ± 0.11	6.93 ± 2.70	3.94 ± 0.01	0.97 ± 0.04

3.5.3 Performance of PUPPI p_T^{miss} algorithm

The PUPPI p_T^{miss} distributions in the dilepton samples are shown in Fig. 3-10. The data distributions are modeled well by the simulation, in both the muon and the electron channels. Similar to the case of PF p_T^{miss}, the p_T^{miss} resolution in these events is dominated

Table 3-2. Parametrization results of the resolution curves for $u_\parallel$ and $u_\perp$ components as a function of the scalar p_T sum of all PF candidates. The parameter values for σ_0 are obtained from data and simulation, whereas the σ_s are obtained from data along with the ratio R_s, the ratio of data and simulation. The uncertainties displayed for both components are obtained from the fit, and for simulation the JES, the JER, and UE uncertainties are added in quadrature.

Process	σ_0(data)[GeV]	σ_0(MC)[GeV]	σ_s[GeV$^{1/2}$]	$R_s = \sigma_s$(data)$/\sigma_s$(MC)
		$u_\parallel$ component		
$Z \to \mu^+\mu^-$	1.98 ± 0.07	0.85 ± 2.45	0.64 ± 0.01	0.95 ± 0.11
$Z \to e^+e^-$	2.18 ± 0.09	0.19 ± 2.90	0.64 ± 0.01	0.92 ± 0.11
γ+jets	1.85 ± 0.09	0.94 ± 2.52	0.64 ± 0.01	0.96 ± 0.11
		$u_\perp$ component		
$Z \to \mu^+\mu^-$	-1.63 ± 0.06	-1.72 ± 2.53	0.68 ± 0.01	0.99 ± 0.11
$Z \to e^+e^-$	-1.42 ± 0.08	-1.98 ± 2.95	0.69 ± 0.01	0.96 ± 0.12
γ+jets	-1.16 ± 0.08	-1.31 ± 2.53	0.68 ± 0.01	0.98 ± 0.11

by the resolution of the hadronic activity, but the PUPPI-weighted PF candidates yield a much improved resolution for jets compared to the PF case. This is also reflected in the uncertainty shown in the figures, which includes the uncertainties due to jet energy scale and resolution, and the energy scale of the unclustered particles.
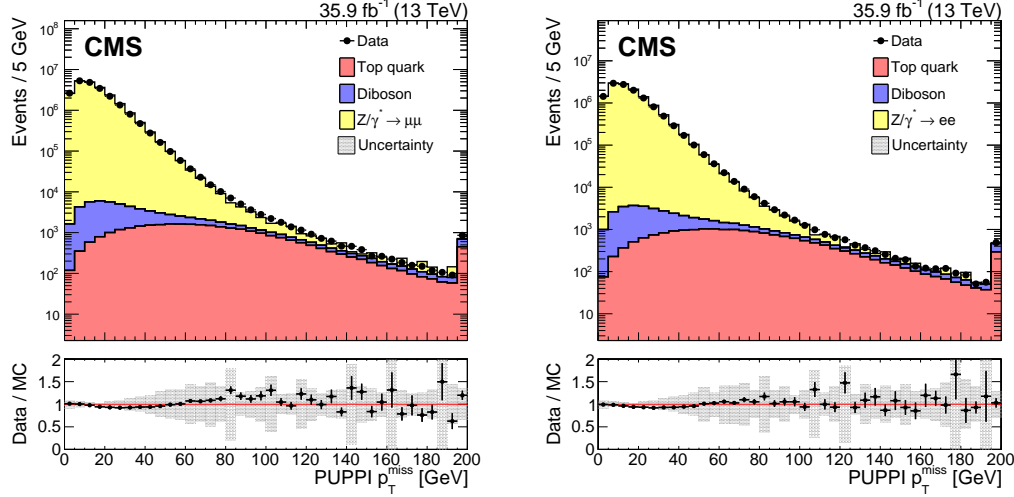

Figure 3-10. Upper panels: Distributions of PUPPI p_T^{miss} in $Z \to \mu^+\mu^-$ (left) and $Z \to e^+e^-$ (right) events. The last bin includes all events with $p_T^{miss} > 195\,\text{GeV}$. Lower panels: Data-to-simulation ratio. The band corresponds to the systematic uncertainties due to the JES, the JER, and variations in the UE added in quadrature, estimated from the $Z \to e^+e^-$ sample.

The distributions in $Z \to \mu^+\mu^-$ and $Z \to e^+e^-$ events of the vectorial sum $u_\parallel + q_T$ and of $u_\perp$ using PUPPI p_T^{miss}, are shown in Fig. 3-11. Following the same arguments as in the PF p_T^{miss} case, in events with no genuine p_T^{miss}, the vectorial sum of $u_\parallel$ and q_T is symmetric around zero, whereas for processes with genuine p_T^{miss} an asymmetric behavior is observed. The distribution of $u_\perp$ is symmetric around zero. Simulation describes data well for all distributions.

Figure 3-12 shows the PUPPI p_T^{miss} response as a function of q_T, extracted from data and simulation in $Z \to \mu^+\mu^-$ and $Z \to e^+e^-$ events. The response reaches unity for $Z \to \mu^+\mu^-$ events at a boson p_T of $150\,\text{GeV}$; while for PF p_T^{miss} the response is close to unity at $100\,\text{GeV}$. The slower rise of the response to unity is due to the removal of PF
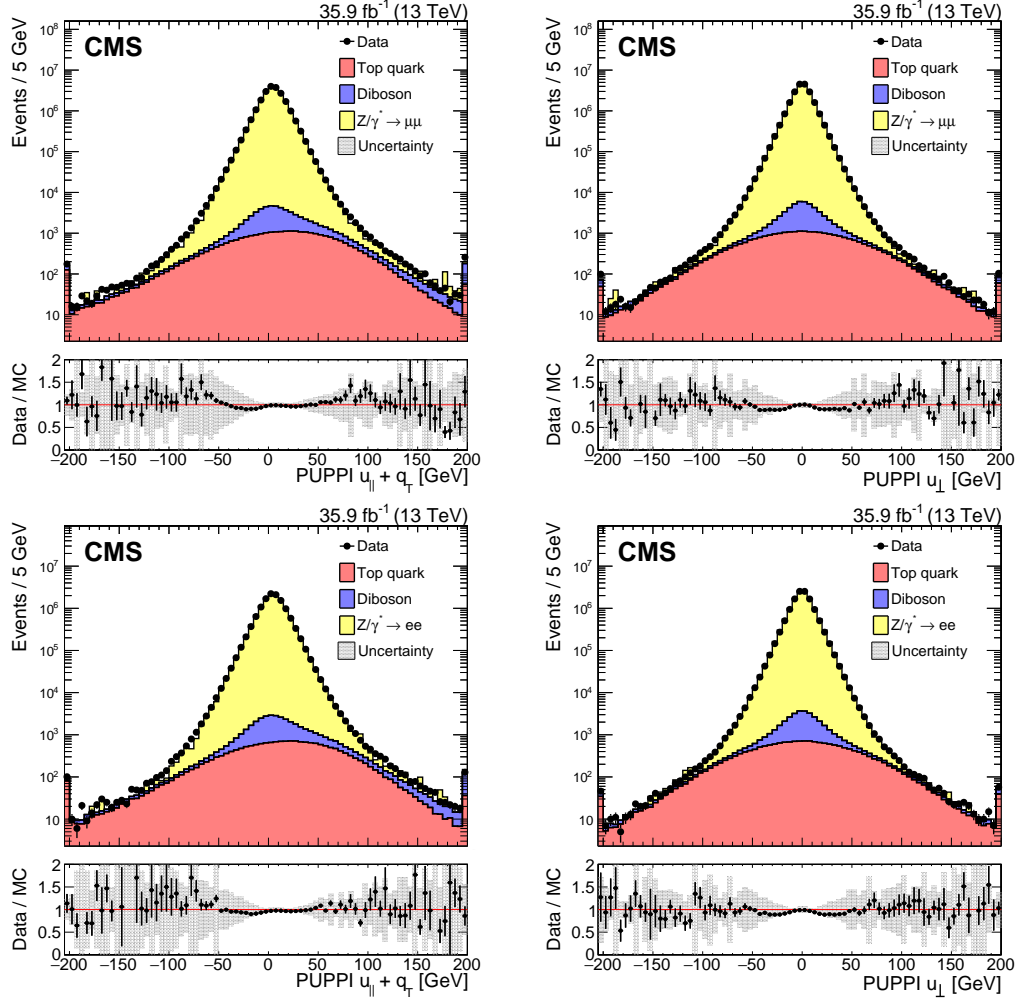
Figure 3-11. Upper panels: Distributions of the $u_\parallel + q_T$ and $u_\perp$ components of the hadronic recoil, in data (filled markers) and simulation (solid histograms), for the $Z \to \mu^+\mu^-$ (upper) and $Z \to e^+e^-$ (lower) events. The first and the last bins include all events below -195 and above +195, respectively. Lower panel: Data-to-simulation ratio. The band corresponds to the systematic uncertainties due to the JES, the JER, and variations in the UE added in quadrature, estimated from the $Z \to e^+e^-$ sample.

candidates that are wrongly associated with pileup interactions by the PUPPI algorithm. Similarly to PF p_T^{miss}, there is no response correction for the UE for PUPPI p_T^{miss}, which results in an underestimated response for low q_T. The response of p_T^{miss} is found to agree between the different samples within 2%.
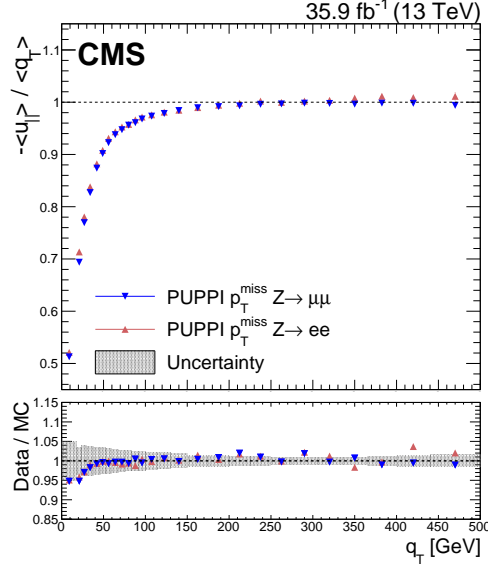

Figure 3-12. Upper panel: Response of PUPPI p_T^{miss}, defined as $-\langle u_\parallel \rangle / \langle q_T \rangle$, in data in $Z \to \mu^+\mu^-$ and $Z \to e^+e^-$ events. Lower panel: ratio of the PUPPI p_T^{miss} response in data and simulation. The band corresponds to the systematic uncertainties due to the JES, the JER, and variations in the UE added in quadrature, estimated from the $Z \to e^+e^-$ sample.

In Fig. 3-13, the results obtained for the case of PUPPI p_T^{miss} are overlayed with the ones obtained using PF p_T^{miss}. Compared to the case of PF p_T^{miss}, the resolutions show a much reduced dependence on the number of pileup interactions. The resolution in different samples is parametrized using Eq. (3–12), and the result of the parameterization are given in Table 3-3. Each additional pileup interaction is found to degrade the resolution of each component by up to 2 GeV. This resolution degradation corresponds to half of what is observed in the case of PF p_T^{miss}.
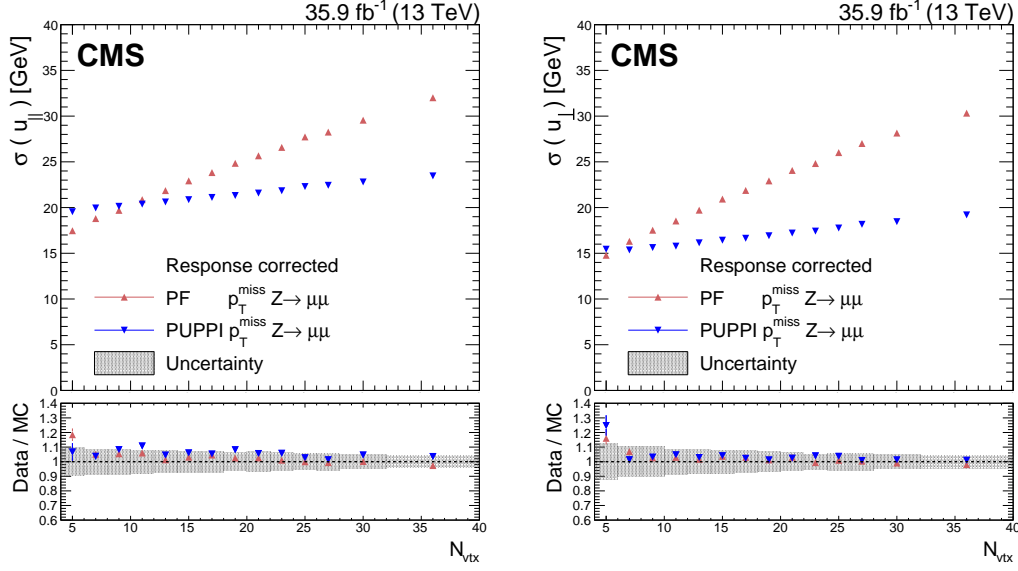
Figure 3-13. Upper panels: PUPPI and PF p_T^{miss} resolution of $u_\parallel$ (left) and $u_\perp$ (right) components of the hadronic recoil as a function of N_{vtx}, in $Z \to \mu^+\mu^-$ events. Lower panels: Data-to-simulation ratio. The systematic uncertainties due to the JES, the JER, and variations in the UE are added in quadrature and displayed with a band.

Table 3-3. Parameterization results of the resolution curves for PUPPI $u_\parallel$ and $u_\perp$ components as a function of N_{vtx}. The parameter values for σ_c are obtained from data and simulation, and the values for σ_{PU} are obtained from data, along with the ratio R_{PU} of data and simulation. The uncertainties displayed for both the components are obtained from the fit, and for simulation the JES, the JER,

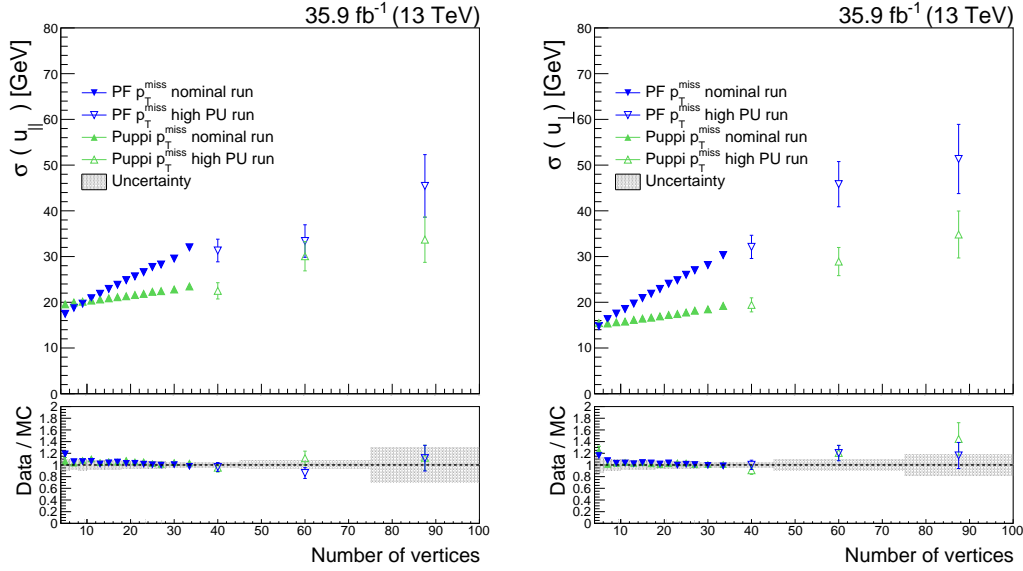| Process	σ_c(data)[GeV]	σ_c(MC)[GeV]	σ_{PU}(data)[GeV]	$R_{PU} = \sigma_{PU}(data)/\sigma_{PU}(MC)$
		$u_\parallel$ component		
$Z \to \mu^+\mu^-$	18.9 ± 0.05	17.5 ± 0.74	1.93 ± 0.02	0.97 ± 0.11
$Z \to e^+e^-$	18.9 ± 0.06	17.4 ± 0.80	1.94 ± 0.03	0.98 ± 0.12
		$u_\perp$ component		
$Z \to \mu^+\mu^-$	14.2 ± 0.04	13.6 ± 0.59	1.78 ± 0.01	0.97 ± 0.09
$Z \to e^+e^-$	14.3 ± 0.05	13.6 ± 0.59	1.80 ± 0.02	0.96 ± 0.09

Figure 3-14. Upper panels: PUPPI and PF p_T^{miss} resolution of $u_\parallel$ (left) and $u_\perp$ (right) components of the hadronic recoil as a function of N_{vtx}, in $Z \to \mu^+\mu^-$ events. The blue (green) markers correspond to the PF (PUPPI) p_T^{miss} reconstruction algorithm, with filled (open) markers for the nominal run (high pileup run). Lower panels: Data-to-simulation ratio. The systematic uncertainties due to the JES, the JER, and variations in the UE are added in quadrature and displayed with a band.

3.5.4   High pileup studies

By 2025, the LHC will be upgraded with a goal to increase the integrated luminosity by a factor of 10 beyond the original design. The new design, known as High Luminosity LHC (HL-LHC), will pose a major challenge due to the unprecedented increase of pileup expected (sim140) in the collision events per bunch crossing. In preparation, LHC delivered a special pp collisions data with conditions similar to the ones expected at the HL-LHC. The "high pileup" data set, as it will be referred in what follows, corresponds to an integrated luminosity of 5pb^{-1}. The bunch setup for the data was three isolated bunches of an average pileup of 70-100, and 2×48 bunch trains, corresponding to an average pileup of 35-50. Dedicated simulated samples were produced with similar conditions. Using the high pileup data set and the dedicated MC simulations, the performance of the PF $p_{\mathrm{T}}^{\mathrm{miss}}$ and Puppi $p_{\mathrm{T}}^{\mathrm{miss}}$ algorithms are studied in dimuon samples. The $p_{\mathrm{T}}^{\mathrm{miss}}$ resolution of the $u_{\parallel}$and $u_{\perp}$components of the hadronic recoil as a function of the number of reconstructed vertices is shown in Fig. 3-14. The results obtained from the high pileup data is overlayed with the ones obtained from the nominal data and found to be in agreement within the statistical uncertainties. Furthermore, the Puppi $p_{\mathrm{T}}^{\mathrm{miss}}$ is found to have more stable resolution across the full range.