

UNIVERSIDADE FEDERAL DO ABC
CENTRO DE CIÊNCIAS NATURAIS E HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA



Tese de Doutorado

**"Busca por ressonâncias pesadas na observação de jatos
hadrônicos e energia faltante no experimento CMS do
LHC"**

Estudante:

David Romero Abad

Orientador:

Eduardo de Moraes Gregores

Santo André

05 de janeiro de 2017

I dedicate this thesis to my family, in particular to my wife Margarita and my son Matias.

Resumo

O Large Hadron Collider (LHC) do CERN foi projetado e construído com o objetivo de proporcionar a observação de fenômenos na fronteira das altas energias. O Compact Muon Solenoide (CMS) é um dos quatro experimentos que observam o resultado das colisões hadrônicas de altas energias geradas pelo LHC, podendo medir com grande precisão diversos parâmetros do Modelo Padrão das interações fundamentais bem como propiciar o descobrimento de novos constituintes da matéria e suas interações. Em sua nova fase de operação, iniciada em junho de 2015 e prevista para durar até 2020, ele vem gerando colisões de prótons a 13 e 14 TeV, propiciando um ambiente de descobertas em regiões do espaço de fase ainda não acessíveis. O objetivo deste trabalho é investigar a possível existência de ressonâncias pesadas que tenham por sinal característico seu decaimento em bósons vetoriais massivos do Modelo Padrão, em particular em um par de bósons vetoriais VZ ($V=W,Z$). Tais ressonância são previstas de existir em muitas extensões do Modelo Padrão, tais como as que preveem a existência de Dimensões Extras espaciais ou Modelos Simplificados que descrevem novos bosons vetoriais de spin um. Estudaremos o indício da existência de tais ressonâncias analisando os sinais experimentais surgidos quando o bóson V decai em dois jatos hadrônicos e Z boson decai invisivelmente em dois neutrinos.

Palavras-chave: Física de Altas Energias; Física de Partículas; Colisores Hadrônicos; Física Além do Modelo Padrão.

Abstract

The Large Hadron Collider (LHC) at CERN was designed with the purpose to observe new phenomena on the high energies frontier. The Compact Muon Solenoid (CMS) is one of the four experiments that examine the outcome of hadronic collisions at high energies generated by LHC. It can measure with great accuracy various parameters of the Standard Model of fundamental interactions as well as facilitate the discovery of new constituents of matter and their interactions. In this new stage of operation, initiated in June 2015 and expected to last until 2020, the LHC generates collisions of protons at energies of 13 and 14 TeV, providing an environment of discoveries in regions of phase space still not accessible. The objective of this work is to investigate the existence of heavy resonances which have as characteristic signature, to decay in massive vector bosons of the Standard Model, particularly in a pair of vector bosons VZ ($V=W,Z$). These kind of resonances are predicted by many extensions of the Standard Model, such as Extra Dimensions or Simplified Models that describe new vector bosons of spin one. We will examine the evidence of such resonances by analyzing the experimental signatures that arises when the V boson decays into two jets and the Z boson decays invisibly into two neutrinos.

Key-words: High energy physics; Particle physics; Hadron colliders; Physics beyond standard model.

Aknowledgments

First of all, I would like to thank my advisor, Eduardo de Moraes Gregores, who gave me the opportunity to be involved in an area of research as apasionante as the experimental particle physics and to be part of the CMS collaboration. All his advice and recommendations have been very valuable during the whole process that involved the present thesis and for the future of my career.

I would also like to express my gratitude to the Universidade Federal do ABC (UFABC), for hosting me as a doctoral student and for the economic support to participate in events and research.

I would like to thank the entire SPRACE group for all the support that I received in infrastructure, equipment, discussions and advices. I would like to thank explicitly to: Sérgio Novaes, Sandra Padula, Pedro Mercadante, Sérgio Lietti, Ailton Shinoda, Márcio Costa, Rogério Iope, Sidney Santos, Angelo Santos, Chang-Seong Moon, Sudha Ahuja, Sunil Dogra, Thiago Tomei, César Bernardes and José Ruiz. I want to give particular thanks to José, César and Thiago, who were directly involved in this work and whose contribution was crucial for the development of this thesis.

I would like to thank the Núcleo de Computação Científica (NCC) of the Universidade Estadual Paulista (UNESP) for the Computational support, vital for the development of this work.

I would like to thank CAPES for providing me a scholarship during the whole period of doctoral studies.

Last but no least, I would like to thanks my family. I would lack space to list the reasons.

Contents

Abstract	II
Aknowledgments	IV
List of Figures	1
List of Tables	9
1 Introduction	11
2 Physics Motivation	15
2.1 The Standard Model of Fundamental Interactions	15
2.2 Beyond Standard Model	18
2.2.1 The Hierarchy Problem	18
2.2.2 Dark Matter	19
2.2.3 Extra Dimensions	20
2.3 Benchmark Models	21
2.3.1 Bulk Graviton Model	21
2.3.2 Heavy Vector Triplets	24
3 Experimental Apparatus	28
3.1 Large Hadron Collider (LHC)	28
3.2 CMS Detector	29
3.2.1 The Magnet	30
3.2.2 CMS Tracking system	33

3.2.3	Electromagnetic Calorimeter (ECAL)	35
3.2.4	Hadronic Calorimeter (HCAL)	37
3.2.5	The Muon System	39
3.3	CMS Data Acquisition and Triggering	41
3.4	Physics objects	43
3.4.1	Track reconstruction	43
3.4.2	Primary Vertex (PV)	44
3.4.3	Particle Flow (PF)	46
3.4.4	Jets	47
3.4.5	Missing Transverse Energy (\cancel{E}_T)	55
4	Data Analysis	61
4.1	Data Samples and Monte Carlo Simulation	62
4.1.1	Trigger and data samples	62
4.1.2	Simulated Samples	63
4.2	Object Reconstruction and Identification	66
4.2.1	Primary Vertices	66
4.2.2	Missing Transverse Energy	67
4.2.3	Jets	70
4.2.4	Hadronic V identification using jet substructure	74
4.2.5	Muons	75
4.2.6	Electrons	76
4.2.7	Taus	77
4.2.8	Photons	78
4.2.9	Transverse mass	78
4.3	Event Selection	79
4.4	Background Modeling	83
4.4.1	Alpha ratio method	84
4.5	Signal Modeling	94
4.6	Systematics Uncertainties	100
4.6.1	Systematic uncertainties in the background estimation	100
4.6.2	Integrated Luminosity	101
4.6.3	Jet Substructure Scale Factors	101

4.6.4	QCD scale	101
4.6.5	PDF (Parton distribution function)	101
4.6.6	Trigger SF	103
4.6.7	B-tagging	103
4.6.8	Pile-up	104
4.6.9	Jet Energy Corrections	105
4.6.10	Jet Energy Resolution	106
4.6.11	Leptons and photons ID	107
4.6.12	Jet mass calibration uncertainty	107
4.6.13	Jet mass resolution uncertainty	107
4.6.14	p_T extrapolation uncertainty	108
4.6.15	E_T^{miss} unclustered energy uncertainty	108
4.6.16	Cross section	109
4.6.17	Nuisance parameter impacts	109
4.7	Results	112
4.7.1	Crosscheck	115
4.7.2	Signal Significance	115
4.7.3	Competitiveness of the result	116
5	Conclusions	118
A	HLT description and efficiency	120
A.1	Trigger Paths	120
A.2	Trigger Efficiency	121
A.3	Crosscheck of trigger efficiency measurement	123
B	Recoil Correction	126
C	Statistics	131
C.1	Definition of Probability Distribution Function (PDF)	131
C.2	Nuisance Parameters and Systematic Uncertainties	132
C.3	Likelihood Function	132
C.4	Hypothesis Tests	133
C.5	Likelihood Ratio in the Search for a New Signal	134

C.6 Upper limits	135
C.7 Modified Frequentist Approach: The CL_S Method	136
Bibliography	138

List of Figures

1.1	Left: Feynman diagram for the production of a generic resonance X decaying into dibosons VZ ($V = Z, W$) and subsequently into quarks and neutrinos. Right: Schematic illustration of the boosted topology where the hadronic boson is reconstructed as a single jet.	12
2.1	Particles of the Standard Model	17
2.2	Radiative Corrections to the Higgs Mass.	18
2.3	Graviton probability function	22
2.4	Left: Production cross sections of the bulk graviton resonance as a function of the resonance mass for $\tilde{k} = 0.1$ ($\tilde{k} = k/M_{Pl}$). For comparison purposes the predicted cross sections at both the centre of mass energies of 8 TeV and 13 TeV are shown. Right: Branching Ratios for the two-body decays of the bulk graviton model.	24
2.5	Left: Production cross sections of the neutral and charged resonances as a function of the resonance mass in Model B. For comparison purposes the predicted cross sections at both the centre of mass energies of 8 TeV and 13 TeV are shown. Right: Branching Ratios for the two-body decays of the charged W' in the Model B.	27
3.1	The CERN accelerator complex.	29
3.2	Schematic view of the CMS detector	30
3.3	η and ϕ coordinates in the CMS detector.	30
3.4	General artistic view of the 5 modules composing the superconducting coil.	31

3.5	Schematic views of the CMS detector, with the numbering convention for azimuthal sectors (S), wheels (W), barrel yoke layers (L) and endcap disks (D). Left: transverse view at $z = 0$. Right: longitudinal view of one quarter of the detector.	32
3.6	Value of $ \vec{B} $ (left) and field lines (right) predicted on a longitudinal section of the CMS detector, at a central magnetic flux density of 3.8 T.	32
3.7	Left: Sktech of the CMS pixel detector. Right: Quarter of a slice of the CMS pixel detector by a plane which contains its axis of symmetry. . . .	34
3.8	Strip and pixel silicon detector	35
3.9	Layout of the CMS electromagnetic calorimeter showing the arrangement of crystal modules, supermodules and endcaps, with the preshower in front.	36
3.10	Longitudinal view of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters.	38
3.11	View of an HB wedge.	38
3.12	Muon Drift Tubes	39
3.13	Cathode Strip Chamber	40
3.14	Resistive Plate Chambers	40
3.15	Muons Stations	41
3.16	CMS Triggering System	42
3.17	Geometrical description of the closest approach point of a track (curved line) to the beam line: transverse (d_0) and longitudinal (z_0) impact parameters.	43
3.18	Track reconstruction efficiencies for single isolated muons passing high-purity quality requirements. Results are shown as a function of η (left), for $p_T = 1, 10, \text{ and } 100$ GeV. They are also shown as a function of p_T (right), for the barrel, transition, and endcap regions, which are defined by the η intervals of 0–0.9, 0.9–1.4 and 1.4–2.5, respectively.	45
3.19	Primary-vertex reconstruction efficiency as a function of the number of tracks in a cluster, measured in minimum-bias data and in MC simulation. .	46

3.20	A slice through the CMS detector. CMS consists of a central silicon Tracker (Pixel and Strips), an Electromagnetic Calorimeter, a Hadron Calorimeter, a Superconduction Solenoid Magnet, a massive iron return yoke instrumented with Muon Chambers. The depicted interactions present an ideal detector behavior for the different particles μ , e , charged and neutral hadrons, and γ .	48
3.21	pp-collision resulting in a collimated spray of particles, a jet.	48
3.22	CMS event display for a dijet reaction	49
3.23	Charged hadron subtraction.	51
3.24	Schematic picture of the factorized multi-level jet correction.	52
3.25	τ_{21} distribution characterising jet substructure in simulated samples of highly boosted and longitudinally polarized W bosons. Thick dashed lines represent the generator predictions without pileup interactions and without CMS detector simulation. The histograms are the expected distributions after full CMS simulation with pileup corresponding to an average number of 12 and 22 interactions.	54
3.26	Jet pruned mass distribution in simulated samples of highly boosted and longitudinally polarized W bosons. The red lines represent a resonance decaying in two W bosons (signal) and the black lines represent a W+jets process (background).	55
3.27	Invisible and visible particles	56
3.28	The E_T^{miss} distributions for events passing the dijet selection with the 2016 cleaning algorithms applied including the one based on jet identification requirements (filled markers), without the 2016 cleaning algorithms applied (open markers), and from simulation (filled histograms).	57
3.29	Vector sum of the reconstructed particles in the event.	58
4.1	Electroweak corrections for the W/Z boson in function of the boson p_T .	64
4.2	Left: Comparison between the PU profile in data (blue) and the Poisson density function in MC (green). Right: Number of primary vertices distribution after the reweight procedure in events that pass the final analysis selection.	66

4.3	Comparison between Raw PF E_T^{miss} and Type-1 PF E_T^{miss} corresponding to a signal sample of 1 TeV.	67
4.4	HBHENoiseFilter efficiency vs AK8 jet p_T in simulation for a signal sample of 1 TeV.	69
4.5	$Z \rightarrow \ell\ell$ event kinematics in the transverse plane. The transverse recoil vector \vec{u}_T is split into parallel and perpendicular components to the direction of the boson p_T	70
4.6	E_T^{miss} distribution before and after applying the recoil correction for W +jets and Z + jets samples.	70
4.7	Comparison between corrected and uncorrectd jets (JEC) for the jet mass distribution corresponding to a graviton signal sample of 1 TeV ($X \rightarrow ZZ$).	71
4.8	Jet hadronic energy fractions after the cleaning cuts.	72
4.9	ak4Jets multiplicity.	73
4.10	Derived efficiency maps for b-tagging in function of the p_T and $ \eta $ of the jet in MC samples (left) $t\bar{t}$ sample (right) Z +jets sample, for jets with flavor b.	73
4.11	Distribution of p_T (left) and τ_{21} (right) for the leading AK8 jet in events passing the final selection in the pruned jet mass signal region for data, SM backgrounds and for a bulk graviton signal sample (1.2 TeV resonance mass and $\tilde{k} = 0.5$).	75
4.12	Minimum ΔR between reconstructed and generated muons for a signal mass point of 1 TeV.	77
4.13	Signal efficiency for $G_{Bulk} \rightarrow ZZ$ and $W' \rightarrow WZ$ in the jet + E_T^{miss} final state for different mass points and different categories after the final analysis selection.	81
4.14	Top:(left) AK8 Jet transverse momentum. (right) Missing transverse energy. Center: (left) AK8 Jet pseudorapidity. (right) min $\Delta\phi$ distribution. Bottom: (left) Pruned jet mass. (right) Candidate transverse mass. The plots show the comparison between data(black dots) and simulation.	82
4.15	Top : Z/W +jets fits in the HP category. Center : Z/W +jets fits in the LP category. Bottom : Comparison between the Z +jets and W +jets fits.	87

4.16	Fit of the jet pruned mass in the HP category. Top: (left) Dominant MC background fit with the ErfExp function (right) Subdominant MC background fit with the Gaus2 function. Bottom: Distribution of the jet pruned mass in the HP categories. All selections are applied except the final m_{jet} signal window requirement. Data are shown as black markers. The contribution of the V+jets background is extrapolated from the sideband to the signal region.	88
4.17	Fit of the jet pruned mass in the LP category. Top: (left) Dominant MC background fit with the ErfExp function (right) Subdominant MC background fit with the Gaus2 function. Bottom: Distribution of the jet pruned mass in the LP categories. All selections are applied except the final m_{jet} signal window requirement. Data are shown as black markers. The contribution of the V+jets background is extrapolated from the sideband to the signal region.	89
4.18	Fit of the VZ candidate transverse mass in the HP category. Top: (left) Dominant MC background fit in the sideband region (right) Dominant MC background fit in the signal region. Center: (left) alpha transfer function (right) Data fit in the sideband region. Bottom: Final prediction on the background estimation in signal region.	92
4.19	Fit of the VZ candidate transverse mass in the LP category. Top: (left) Dominant MC background fit in the sideband region (right) Dominant MC background fit in the signal region. Center: (left) alpha transfer function (right) Data fit in the sideband region. Bottom: Final prediction on the background estimation in signal region.	93
4.20	Fit of the Bulk Graviton signal samples for different mass points (800-1400 GeV) in the HP category.	95
4.21	Fit of the Bulk Graviton signal samples for different mass points (1600-2000 GeV) in the HP category.	96
4.22	Fit of the Bulk Graviton signal samples for different mass points (800-1400 GeV) in the LP category.	97
4.23	Fit of the Bulk Graviton signal samples for different mass points (1600-2000 GeV) in the LP category.	98

4.24	Top: Fit of the W prime signal samples for different mass points in the HP category. Bottom: Fit of the W prime signal samples for different mass points in the LP category	99
4.25	Linear interpolation in the Crtsyal-Ball fit model with a step of 100 GeV for bulk graviton samples.	99
4.26	Fit of the jet pruned mass in data using different template functions. (left): High purity category; (right): Low purity category.	100
4.27	Systematic uncertainties due to the factorization and renormalization scales for different signal mass points.	102
4.28	Left : Transverse Mass of the candidate with the nominal value of the PDF. Also we show the scale up and down in one standard deviation for a signal of 2 TeV. Right: PDF systematic uncertainties for different signal mass point.	102
4.29	Systematic uncertainties due to the trigger SF for different signal mass points.	104
4.30	Left : Systematic uncertainties due to b-tagging uncertainty SF for different signal mass points. Right : Number of b-tag jets for different signal mass points.	104
4.31	Systematic uncertainties due to b-tagging uncertainty SF. Top : Nominal and scale up/down variations values in the transverse mass distribution for (Left) $t\bar{t}$ sample (right) W+jets sample .Bottom : (left) Z+jets sample (right) Number of b-tag jets for different backgrounds models.	105
4.32	Systematic uncertainties on the pile-up due to different minbias cross-section for a signal mass of 2 TeV.	106
4.33	Systematic uncertainties due to the JEC for different signal mass points.	106
4.34	Systematic uncertainties due to jet mass calibration. Left : Nominal and scale up/down variations values in the pruned jet mass distribution for a signal point of 1000 GeV. Right : Systematic uncertainties in (%) for different mass points.	107
4.35	Systematic uncertainties due to jet mass resolution. The figure shows the systematic uncertainties in (%) for different mass points.	108

4.36	Systematic uncertainties due to p_T extrapolation. The figure shows the systematic uncertainties in (%) for different mass points.	109
4.37	Impact of the systematic uncertainties in the analysis.	111
4.38	Observed and expected 95% CL upper limit on Bulk graviton production cross section times the branching fraction of $G_{bulk} \rightarrow ZZ$ assuming an integrated luminosity of 2.3 fb^{-1} . The limit is obtained with the Asymptotic CLs technique. Top: (left) Limits for the HP category (right) Limit for the LP category. Bottom: Limits for the combination of both categories.	113
4.39	Observed and expected 95% CL upper limit on the W' production cross section times the branching fraction of $G_{W'} \rightarrow WZ$ assuming an integrated luminosity of 2.3 fb^{-1} . The limit is obtained with the Asymptotic CLs technique. Top: (left) Limits for the HP category (right) Limits for the LP category. Bottom: Limits for the combination of both categories. .	114
4.40	Observed and expected 95% CL upper limit on Bulk graviton production cross section times the branching fraction of $G_{bulk} \rightarrow ZZ$ assuming an integrated luminosity of 2.3 fb^{-1} . In the figure we show the comparison between the full CLs method (blue) and the asymptotic limit (black). A good agreement is observed in the limit obtained using both methods. .	115
4.41	Signal significance in terms of p-values for the the Bulk graviton and W' models.	116
4.42	Observed and expected limits comparison bewteen this analysis and two similar searches with 2015 dataset.	117
A.1	High Level Trigger efficiency as function of the transverse missing energy without muons for the selected paths. The figure shows the turn-on curve for data and MC simulation and the Data/MC ratio.	122
A.2	High Level Trigger efficiency as function of the transverse missing energy for the selected paths. The figure shows the turn-on curve for data and MC simulation.	124
A.3	High level trigger efficiency as function of the transverse missing energy for the selected path in RS signal samples. The figure shows the turn-on curve for different mass points.	125

B.1	Kinematic and recoil properties in the $Z \rightarrow \mu\mu$ process in Data after applying the selection	127
B.2	Kinematic and recoil properties in the $Z \rightarrow \mu\mu$ process in MC after applying the selection	128
B.3	Fits on the parallel (u_1) components of the recoil in data with a double gaussian model.	129
B.4	Fits on the perpendicular (u_2) components of the recoil in data with a double gaussian model.	129
B.5	Fits on the parallel (u_1) components of the recoil in MC with a double gaussian model	130
B.6	Fits on the perpendicular (u_2) components of the recoil in MC with a double gaussian model	130

List of Tables

2.1	Fermions Generations	16
2.2	Fundamental Interactions	16
4.1	HLT Trigger path with their respective criteria.	62
4.2	Data samples.	62
4.3	Monte Carlo simulated signal samples for Bulk graviton ($\tilde{k} = 0.5$) and HVT(B) models.	64
4.4	Monte Carlo background samples.	65
4.5	NNLO QCD flat scale factors.	65
4.6	Jet energy resolution scaling factors and uncertainty.	72
4.7	Scale factors for the τ_{21} efficiency selection derived from data and simu- lation in a top-quark enriched sample.	75
4.8	Final analysis selection.	81
4.9	Background categories.	83
4.10	Jet mass regions.	83
4.11	Background Composition.	84
4.12	Analytic functions to fit the jet pruned mass distribution.	85
4.13	Expected and observed background yields in signal region.	86
4.14	Summary of the shapes used for fit the M_{VZ}^T spectra of each category. . .	90
4.15	Different mass points to fit the signal shape for Bulk graviton model. . .	94
4.16	Different mass points to fit the signal shape for W prime model.	95
4.17	Signal yields for different template functions.	100
4.18	V-tagging scale factors and their systematics uncertainties for each category.	101

4.19 Systematic uncertainties in the analysis.	110
A.1 Trigger efficiency for different E_T^{miss} (NoMu) values in Data and MC. . .	123
A.2 Trigger efficiency for different E_T^{miss} values in Data and MC.	125

Introduction

The standard model of particle physics (SM) [1–3] is a very successful framework that describes the elementary particles and its interactions as proven by many experiments [4–13]. However, there are some experimental and theoretical facts that strongly point to the conclusion that the SM is incomplete and that at some higher energy it must be embedded into some new theory. One open question for example is the *hierarchy problem* that leads naturally to physics beyond the standard model (BSM), possibly at the TeV scale [14–19]. There are many classes of BSMs that predict heavy resonances with masses of the order of a TeV and couple with SM particles. The channels where the resonances decay into fermions have much stronger limits compared to channels where the resonances decay into SM vector bosons and Higgs, both from Electroweak Precision Tests (EWPT) and direct searches [20]. One possible solution to the hierarchy problem is based on the original Randall-Sundrum (RS) model [21], [22], where gravity spreads into a small extra dimension. In this scenario, the existence of a spin-2 Kaluza-Klein graviton is predicted and according to the model, the decays of gravitons to pairs of photons and leptons are favored. Indeed, the branching fraction for the decay of a RS graviton into dibosons is very small, in particular around 7% in case of $G \rightarrow ZZ$. One interesting extension of the RS model is the Bulk graviton model [23–25], this scenario allow the SM fields to propagate in the extra dimension ("Bulk"). The most relevant difference that present this model in comparison with the RS is a much larger braching fraction for the graviton to decay in dibosons (WW, ZZ , and Higgs). Another possible interesting intrepretation is the Heavy Vector Triplet (HVT) model, which

predict spin-1 heavy resonances, such as heavy charged W' and neutral Z' [26]. The diboson final states are also common in composite Higgs models, where the Higgs bosons is a pseudo-Nambu-Goldstone boson of a broken global symmetry [27], [28].

In this document we describe a search for heavy resonances ($M_X^T \gtrsim 1$ TeV) decaying into a pair of SM vector bosons VZ ($V = Z, W$) with the V decaying hadronically and a Z decaying into neutrinos, as shown in Figure 1.1 (left). Since the mass of the exotic resonance is much higher than the masses of the bosons, the two bosons are produced with a high transversal momentum and consequently their decay products are created with a small angular separation. This is denominated *boosted topology*. In particular, the decay products of the hadronically decaying bosons cannot be resolved by the default jet algorithms, but are instead reconstructed as a single jet object (V-jet) as shown in Figure 1.1 (right).

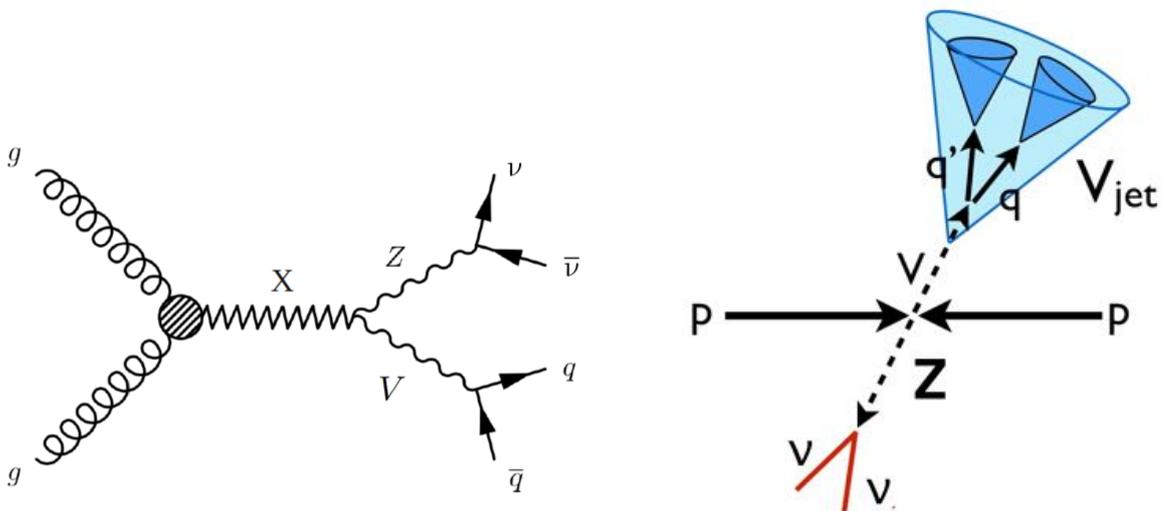


Figure 1.1: Left: Feynman diagram for the production of a generic resonance X decaying into dibosons VZ ($V = Z, W$) and subsequently into quarks and neutrinos. Right: Schematic illustration of the boosted topology where the hadronic boson is reconstructed as a single jet.

Many techniques developed theoretically and experimentally were tested in order to optimize the identification of the boosted bosons. In special, for the case of the hadronic decays, jet substructure techniques [29–31] have a fundamental role in these analyses, increasing the efficiency to identify boosted bosons and at the same time decreasing the fake-rate from quark/gluon-jets and the interferences from pile-up and underlying event [32–42]. This also helps to suppress the SM background, which mainly originates from the production of $V + jets$ and non-resonant VV events. The background

contributions are estimated using a data-driven technique in sidebands of the jet mass distribution of the reconstructed hadronic V boson candidates. The final state in a pair of bosons results in a huge spectrum of search channels because the bosons have many decay modes. Since 2011 LHC Runs with proton-proton collisions at 7 TeV, the CMS and ATLAS collaboration try to cover as much as possible these channels [43–52].

This analysis is based on proton-proton collision data at $\sqrt{s} = 13$ TeV collected by the CMS experiment at the CERN Large Hadron Collider (LHC) during 2015 and correspond to an integrated luminosity of 2.3 fb^{-1} . To assess the sensitivity of the search, to optimise the event selection, and for comparison with data, two specific benchmark models are used. The signal studied is the production of a narrow resonance with mass above 0.8 TeV, with the bulk graviton and the HVT charged resonance acting as benchmarks for the spin-2 and spin-1 hypotheses respectively. Narrow here refers to the assumption that the natural width of the resonance is much smaller than the experimental resolution.

The electroweak processes V +jets, where the Z decays in neutrinos and the W leptonically, are the dominant backgrounds reported in this search. Both represent around 80% of the total background. The W +jets background is reduced by applying a veto in the events where a lepton is identified. Certain number of W +jets events persist after the veto because the lepton from the W decay is not identified in those cases. Even after applying the veto, this background is considered important because, in comparison with the Z boson, the W boson production presents a cross section that is an order of magnitude larger. Subdominant backgrounds arise from the $t\bar{t}$ process, which is reduced by vetoing events in which jets originated from hadronization of bottom quarks are identified, from dibosons decays, and from QCD multijet events in which large artificial E_T^{miss} appears from jet energy mismeasurements and detector noise.

The experimental strategy is to reconstruct and identify the two bosons and to combine their information into a variable that can discriminate between signal and background and on which a statistical study can be performed. The signal of a new resonance X decaying to dibosons (VZ) is sought by comparing the transverse mass distribution observed in data and the data-driven background prediction from the standard model. Results are interpreted in terms of exclusion limits of the bulk graviton and the HVT

model (W'), under the assumption of a negligible width with respect to the experimental resolution (narrow-width approximation).

The document is organized as follows. A brief review of the Standard Model and the physics motivations behind some interesting extensions are presented in Chapter 2. Chapter 3 briefly describes the LHC machine and the CMS detector components. In addition, this chapter presents an introduction to the physics objects reconstruction (tracks, jets and missing transverse energy) and their properties, which will be very useful further in the analysis. Chapter 4 presents the data analysis for a search of a heavy resonance decaying into $VZ \rightarrow \nu\bar{\nu}q\bar{q}$ using the full CMS pp collision data collected in 2015. This chapter describe the data samples used in the analysis, the event selection, the signal and background modeling, the treatment of the systematics uncertainties, the statistical discrimination between the signal and background in data and the interpretation of the results in terms of upper limits. The compatibility between data and background only hypothesis is tested. It is obtained 95% C.L. limits on the new resonance production cross section times $X \rightarrow VZ$ branching ratio as a function of the resonance mass between 800 GeV and 2000 GeV. Chapter 5 presents the conclusion of this thesis.

Chapter 2

Physics Motivation

2.1 The Standard Model of Fundamental Interactions

The electroweak theory of the SM was proposed by Glashow, Weinberg and Salam [1–3] to describe the electromagnetic and weak interactions of quarks and leptons, which is based on the local gauge group $SU(2)_L \otimes U(1)_Y$. Combined with quantum chromodynamics (QCD), which is the theory of strong interactions between quarks and gluons, with local gauge group $SU(3)_C$, the model foresees a unified framework to detail these three forces of nature.

It is necessary to make a differentiation between different types of fundamental particles involved in the SM. The particles are divided into: bosons (particles of integer spin), responsible for transmitting the fundamental forces of the nature, and fermions (particles of half-integer spin) that are the constituents of matter. Since not all fermions have the same properties, they have been divided into two types: leptons and quarks. One of the differences is that quarks have fractional electric charge while the charge of the leptons are multiples of the electron charge. The quarks exhibit a very peculiar property called "confinement", which means that free quarks have not been observed. Quarks feel all interactions, but leptons are not affected by the strong force.

In particle physics, a generation is a division of elementary particles. Between generations, particles differ only in their mass. All interactions and quantum numbers are identical. There are three generations according to SM of particle physics. Each member of an higher generation has bigger mass than the corresponding particle of

the previous generation. This hierarchy of mass causes particles to decay from high to low generations, which explains why ordinary matter (atoms) is made of particles of the first generation. Every atom is then composed of first generation particles. The second and third generations of charged particles do not form normal matter and are only seen in extremely high-energy environments. The table below summarizes the main properties of fermions:

Fermions			
Generation	Fermions	Mass [MeV]	Charge ($Q/ e $)
1^a	u	2.3	$2/3$
	d	4.8	$-1/3$
	e	0.511	-1
	ν_e	$< 2 \times 10^{-6}$	0
2^a	c	1.275×10^3	$2/3$
	s	95	$-1/3$
	μ	105.66	-1
	ν_μ	< 0.19	0
3^a	t	173.21×10^3	$2/3$
	b	4.18×10^3	$-1/3$
	τ	1777	-1
	ν_τ	< 18.2	0

Table 2.1: Fermions Generations

Unlike leptons, quarks are confined within hadrons, and they are not seen as physical particles. The masses of the quarks can not be measured directly, but can be determined indirectly through their influence on hadronic properties.

The different interactions are described in the quantum language in terms of bosons exchange between the constituents fermions.

Types of Interactions				
Interaction	Gauge Group	Boson	Symbol	Relative Magnitude
Strong	$SU(3)$	gluons (8 types)	g	1
Electromagnetic	$U(1)$	photon	γ	10^{-2}
Weak	$SU(2)$	intermediate vector bosons	W^\pm, Z^0	10^{-7}
Gravitational	?	Graviton (hypothetical)	G	10^{-39}

Table 2.2: Fundamental Interactions

As the table 2.2 shows, there are four types of fundamental interactions. The strong interactions are responsible to bind the quarks inside the proton and the neutron, while maintaining the neutron and the proton in the nucleus. The force between quarks is mediated by massless particles called gluons. Electromagnetism is responsible for

binding electrons in the atom, the atoms in the molecules, and intermolecular forces in liquids and solids. These interactions are mediated by the exchange of photons. Weak interactions are typified by β nuclear decay processes, which involves the emission of an electron and a neutrino by a radioactive nucleus. The mediators of the weak force are bosons W^\pm and Z^0 , with masses of the order of 100 times the proton mass. Gravitational interactions act on all types of particles. As can be seen from the table 2.2, the relative magnitude of the gravitational interaction is very small, thus, for practical purposes it is not considered within the SM.

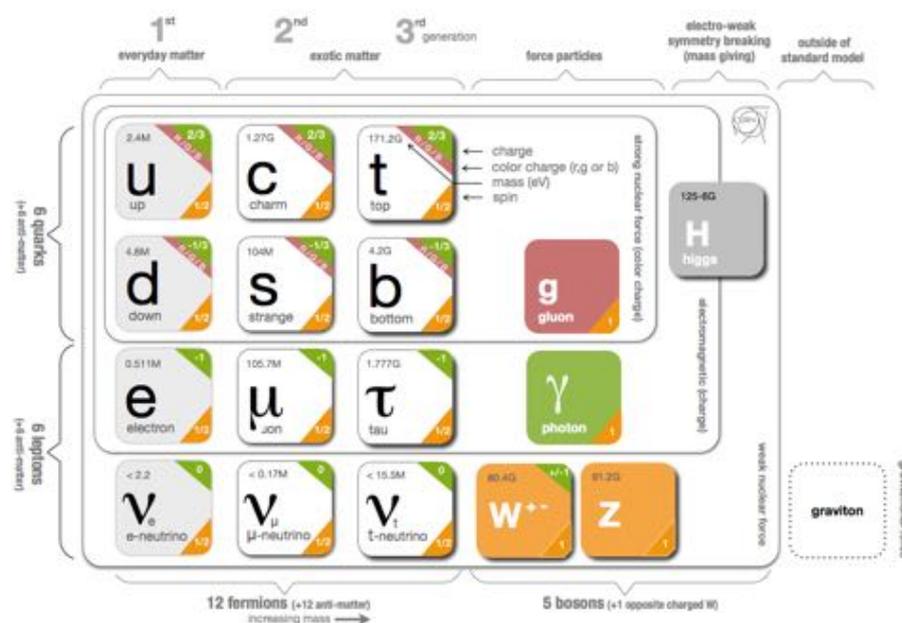


Figure 2.1: Particles of the Standard Model

There exist a very large list of experimental verification for the SM of particle physics, here we just summarize some of them for illustration purposes:

- Experimental indication of the existence of weak neutral currents (1973-CERN) [53]
- Discovery of the tau lepton (1975-SLAC) [54]
- The discovery of the W and Z bosons (1983-CERN) [7–9, 55]
- The discovery of the top quark (1995-FERMILAB) [10, 56]
- The discovery of the Higgs boson (2012-CERN) [4, 5]

For a detailed description of the standard model of particle physics, we refer the reader

to the references: [57–60].

2.2 Beyond Standard Model

Although the Standard Model accurately describes the fundamental interactions in nature and agrees with all the experimental data we have at our disposal today it is still incomplete. Perhaps it is only a part of a bigger picture that includes new physics. Some of the unanswered main questions are: Why is the weak scale so much smaller than the Planck scale? What is the origin of the difference between matter and antimatter, and is it related to the origin of the matter in the Universe? What is the nature of the astrophysical dark matter? How does one unify the fundamental interactions?

In this section we will give a little insight into some of these problems and some models which try to give an answer.

2.2.1 The Hierarchy Problem

When radiative corrections are applied to the Higgs mass, for example at one loop level (fig. 2.2), we need to integrate over the momentum of the virtual particles. In general we have to bound the integral by a cut-off (Λ) related with the next energy scale in the theory. If the next scale is gravity, Λ is the Planck scale $M_P \sim 10^{18}$ GeV. Thus, if the SM were valid up to the Planck scale, then the Higgs mass m_H , and therefore the minimum of the Higgs potential v , would be driven from the weak scale to the Planck scale by the radiative corrections (eqn. 2.1).

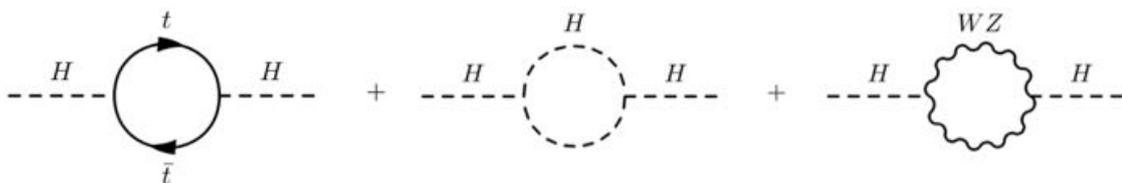


Figure 2.2: Radiative Corrections to the Higgs Mass.

$$m_H^2 = m_{H,0}^2 + \frac{3g^2}{32\pi^2} \frac{\Lambda^2}{m_W^2} \left[m_H^2 + 2m_W^2 + m_Z^2 - \frac{4}{3}m_t^2 \right] \quad (2.1)$$

To avoid this, one has to adjust the Higgs bare mass $m_{H,0}$ to one part in 10^{17} . This is quite unnatural, and is what we call the gauge **hierarchy problem**. In order to solve this unnatural fine tuning some theories beyond SM were proposed, for example, Supersymmetry (SUSY), Composite Higgs and Extra Dimensions. We will focus only in the last of these models.

2.2.2 Dark Matter

Dark matter is a hypothetical kind of matter that cannot be seen with telescopes but accounts for most of the matter in the universe. Dark matter neither emits nor absorbs light or any other electromagnetic radiation at any significant level. This means that it has no electric charge and can interact only via gravitational force, or weak force similar to neutrinos. The existence and properties of dark matter are inferred from many sources,

Velocity curves of spinning galaxies In 1970 an American astronomer, Vera Rubin, measured the speed of stars in rotating galaxies accurately enough to convince the scientific community. She observed that stars in spinning galaxies were all rotating at roughly the same velocity, no matter their distance to the galactic centre. This is in contradiction with Kepler's law that describes the rotation of planets around the Sun. This could only happen if huge amounts of invisible matter filled the entire galaxy and beyond.

Gravitational lensing We know that light moves in a straight line in free space. In the presence of a massive object such as a star or a galaxy, the space is deformed and light follows the curvature of the distorted space. Light coming from a distant galaxy will bend when passing near a massive clump of dark matter and the galaxy will appear shifted, as if coming from different places.

Cosmic microwave background Astrophysicists can infer how much dark matter exists by studying the cosmic microwave background. From the amount of radiation associated to each frequency, astrophysicists can calculate the quantity of dark matter contained in the Universe.

Experiments at the (LHC) may supply more direct evidence about dark matter. According to many theories, dark matter particles would be light enough to be produced at the LHC. If they were generated at the LHC, they would escape through the detectors leaving no signal. However, they would transport energy and momentum, so one could infer their existence from the amount of energy and momentum "missing" after a collision. Dark matter candidates arise frequently in theories that suggest physics beyond the Standard Model, such as Supersymmetry and Extra Dimensions.

2.2.3 Extra Dimensions

Why is gravity so much weaker than the other fundamental forces? One possibility is that we don't feel the full effect of gravity because part of it spreads to extra dimensions. If extra dimensions exist, they could explain why gravity is weaker than the other forces of nature.

How could we test for extra dimensions? Some theorists suggest that a particle called the "graviton" is associated with gravity. If gravitons exist, it should be possible to create them at the LHC, but they would rapidly disappear into extra dimensions. A graviton might escape our detectors, leaving an empty zone that we notice as an imbalance in momentum and energy in the event. We would need to carefully study the properties of the missing object to work out whether it is a graviton escaping to another dimension or something else. This method of searching for missing energy in events is also used to look for dark matter or supersymmetric particles.

2.3 Benchmark Models

Despite that this search is trying to be as general as possible, it is important to have some benchmark models (Theoretical models that predict heavy resonances decaying into dibosons) in order to compare with our expectations from SM and the observed events from LHC. Two well-motivated benchmark models for spin-2 hypothesis (Bulk graviton) and spin-1 hypothesis (W' HVT model B) will be discussed in the next sections.

2.3.1 Bulk Graviton Model

Lisa Randall and Raman Sundrum proposed a model where there is only one warped extra dimension which is compactified on the S^1/Z_2 orbifold [21, 22]. Two 4D branes (the Planck brane and the TeV brane) are separated by the fifth extra dimension with size r_c (fig. 2.3). Even though the extra dimension is curved, the brane itself remains static and flat, that is, it preserves 4D Lorentz invariance. This means that the induced metric at every point along the extra dimension has to be the ordinary flat 4D Minkowski metric, and the components of the 5D metric depend only on the fifth coordinate y . The ansatz for the most general metric satisfying these properties is given by:

$$ds^2 = e^{-A(y)} dx^\mu dx^\nu \eta_{\mu\nu} - dy^2 \quad (2.2)$$

Where $\eta_{\mu\nu} = \text{diag}(-1, 1, 1, 1)$ and the amount of curvature (warping) along the extra dimension depends on the function $e^{-A(y)}$, which is therefore called the warp-factor. This type of geometry is called "non-factorizable" because the metric of the 4D subspace is y -dependent. In the simplest version of the RS model it is assumed that the SM fields live on the so-called TeV brane while gravity lives everywhere. Solving the 5D Einstein's equations provides a unique solution for these quantities and also determines that $A(y) = k|y|$, where k is a dimensionful parameter. A basic assumption of this model is that there are no large mass hierarchies present, so that we expect that $k \sim M_*$, the 5D fundamental or Planck scale. In fact, once we solve Einstein's equations and plug the solutions back into the original action and integrate over y we find that:

$$M_{Pl}^2 = \frac{M_*^3}{k} \left(1 - e^{-2\pi k r_c}\right) \quad (2.3)$$

The warp factor $e^{-\pi kr_c}$ will be a very small quantity which implies that M_{Pl} , M_* and k have essentially comparable magnitudes following from the assumption that no hierarchies exist. If we calculate the Ricci curvature invariant for this 5D space, we find that it is constant, $R_5 = -20k^2$ and thus k is a measure of the constant curvature of this space. A space with constant negative curvature is called an Anti-DeSitter space and so this 5D version is called AdS_5 .

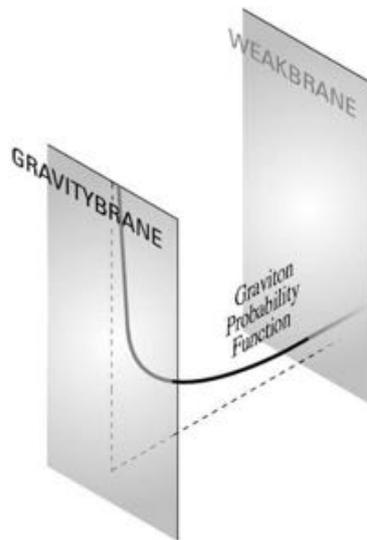


Figure 2.3: Graviton probability function

It will be assumed that all dimensionful parameters in the action will have their mass scale set by $M_* \sim M_{Pl} \sim k$ so that there is no fine-tuning. However, the warp factor rescales them as one moves about in y so that, in particular, all masses will appear to be of order the TeV scale on the SM brane. This means that if there is some mass parameter, m , in the action which is order M_{Pl} , we on TeV brane will measure it to be reduced by the warp factor, i.e., $me^{-\pi kr_c}$. Note that if $kr_c \sim 11$ (a small hierarchy) this exponential suppression reduces a mass of order 10^{18} GeV to only 1 TeV. Thus the ratio of the weak scale to M_{Pl} is explained through an exponential factor and no large ratios appear anywhere else in the model. It has been shown by Goldberger and Wise [61] that values of $kr_c \sim 11$ are indeed natural and can be provided by a stable configuration. Hence we have obtained a true solution to the hierarchy problem. If we consider the

action for the Higgs field on the TeV brane:

$$S = \int d^4x dy \sqrt{-g} \left(g^{\mu\nu} \partial_\mu H^\dagger \partial_\nu H - \left(H^2 - v_0^2 \right)^2 \right) \delta(y - \pi r_c) \quad (2.4)$$

From this we see that the vev that we observe on the SM brane is not v_0 but

$$v = v_0 e^{-\pi k r_c} \quad (2.5)$$

which is of order the TeV scale. Even though gravitons are spin-2, it turns out that their masses and wave functions are identical to the case of a scalar field in the RS bulk which is far simpler to analyze. If we solve the Klein-Gordon equation, but now in the case of curved space, after a separation of variables via the KK decomposition the solutions are linear combination of J_2, Y_2 Bessel functions and the mass of the KK states are given by:

$$m_n = x_n k e^{-\pi r_c} \quad (2.6)$$

where x_n are roots of $J_1(x_n) = 0$. Here $x_n = 0, 3.8317, 7.0155, 10.173, \dots$ etc. Since $k e^{-\pi k r_c}$ is of the order of a few hundred GeV at most, we see that the KK graviton masses are of similar magnitude with comparable, but unequal, spacing, i.e., the KK gravitons have approximately weak/TeV scale masses. We thus have weak scale graviton KKs with weak scale couplings that should be produced as spin-2 resonances at colliders.

Different models with warped extra dimensions allow the SM fields to propagate in the ED. In these models, as a consequence of the localization of SM particles near the Planck or the TeV brane, decays to diphotons and dileptons are suppressed by a factor proportional to the volume of the extra dimension. This scenario is more compatible with electroweak precision tests and limits on flavor-changing neutral current processes than the original RS1. The different couplings of the graviton to the SM fields result in two distinctive effects: the branching fraction to SM vector-boson pairs can become dominant for certain values of the model parameters, and a very strong enhancement in the longitudinal polarization of the vector bosons is predicted. Because of the aforementioned suppression of photon and fermion couplings, the total production cross section is also smaller with respect to RS1 gravitons. The Bulk graviton model proposed by Agashe–Davoudiasl–Perez–Soni [23] predict that kind of behaviour.

Figure 2.4 shows the production cross section and the branching fraction for a heavy resonance of spin 2 in the bulk graviton model.

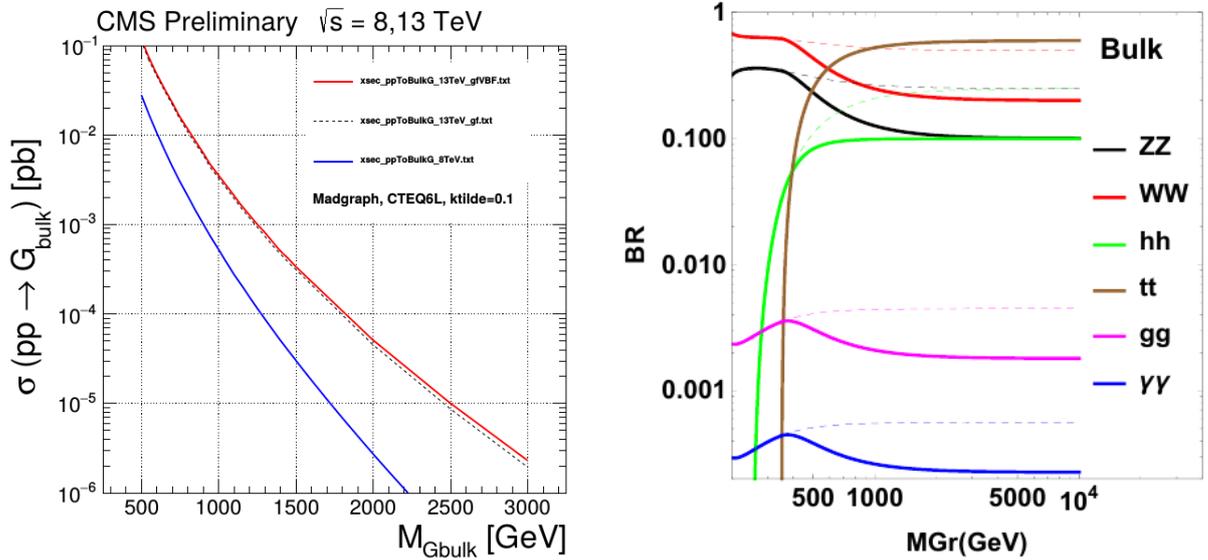


Figure 2.4: Left: Production cross sections of the bulk graviton resonance as a function of the resonance mass for $\tilde{k} = 0.1$ ($\tilde{k} = k/M_{Pl}$). For comparison purposes the predicted cross sections at both the centre of mass energies of 8 TeV and 13 TeV are shown. Right: Branching Ratios for the two-body decays of the bulk graviton model. [62, 63]

2.3.2 Heavy Vector Triplets

The paper from Pappadopulo, Thamm, Torre, and Wulzer [26] defines a model-independent strategy to study narrow resonances. The method is based on a simplified phenomenological Lagrangian, which recreates a large class of explicit models. The approach is to focus on direct experimental manifestation of new physics, through the production of reasonably narrow new particles. The model adopts the "Simplified Model" strategy; the resonant searches are typically not sensitive to all details and free parameters of the underlying model, but only to those that control the mass of the resonance and the interaction involved in its production and decay. Therefore, it uses a simplified description of the resonance, defined by the phenomenological Lagrangian where only the relevant coupling and mass parameters are retained. As the Simplified Model is not supposed to be a complete theory, it is constructed to describe only the on-shell resonance production and decay. Many different new physics effects, not included in the Simplified Model, might change the prediction. For this reason only around the

peak the Simplified Model prediction is trustable.

The paper is focused in the simple but well-motivated example of electroweak-charged spin-1 resonances which are a common prediction of many New Physics scenarios. In the simplified framework, in addition to the SM fields and interactions, the model consider a real vector V_μ^a ($a = 1, 2, 3$) in the adjoint representation of $SU(2)_L$ with vanishing hypercharge. This field describes one charged and one neutral heavy spin-1 particle with the eigenstates defined by:

$$V_\mu^\pm = \frac{V_\mu^1 \mp iV_\mu^2}{\sqrt{2}}, \quad V_\mu^0 = V_\mu^3 \quad (2.7)$$

The dynamics of the new vector is given by a phenomenological Lagrangian:

$$\begin{aligned} \mathcal{L}_V = & -\frac{1}{4}D_{[\mu}V_{\nu]}^a D^{[\mu}V^{\nu]a} + \frac{m_V^2}{2}V_\mu^a V^{\mu a} \\ & + ig_V c_H V_\mu^a H^\dagger \tau^a \overleftrightarrow{D}^\mu H + \frac{g^2}{g_V} c_F V_\mu^a J_F^{\mu a} \\ & + \frac{g_V}{2} c_{VVV} \epsilon_{abc} V_\mu^a V_\nu^b D^{[\mu}V^{\nu]c} + g_V^2 c_{VVHH} V_\mu^a V^{\mu a} H^\dagger H \\ & - \frac{g}{2} c_{VWV} \epsilon_{abc} W^{\mu\nu a} V_\mu^b V_\nu^c \end{aligned} \quad (2.8)$$

with:

$$D_{[\mu}V_{\nu]}^a = D_\mu V_\nu^a - D_\nu V_\mu^a, \quad \text{and} \quad D_\mu V_\nu^a = \partial_\mu V_\nu^a + g\epsilon^{abc} W_\mu^b V_\nu^c \quad (2.9)$$

where g denotes the $SU(2)_L$ gauge coupling. The V_μ^a fields are not mass eigenstates and they mix with W_μ^a after the electroweak symmetry breaking (EWSB). The second line of the equation above contains the interactions of V with the Higgs current and with the SM left-handed fermions currents ($J_F^{\mu a}$).

$$J_F^{\mu a} = \sum_f \bar{f}_L \gamma^\mu \tau^a f_L \quad (2.10)$$

where $\tau^a = \sigma^a/2$. In the Higgs current term, the coupling constant c_H leads to vertices involving the physical Higgs and the three unphysical Goldstone bosons. According to the Equivalence Theorem [64], the Goldstones represent the longitudinally polarized SM vector bosons W and Z in the high-energy regime. The parameter c_H controls the

V interactions with the SM vectors and with the Higgs, and in particular its decay into bosonic channels. By othe hand, the term c_F describes the direct interaction with fermions, which is responsible for the resonance production by Drell-Yang mechanism and for its fermionic decays. The parameter g_V represent the typical strength of the of V interactions, ranging from $g_V \sim g \sim 1$ for a typical weakly-coupled case up to $g_V \simeq 4\pi$ in the extreme strong limit. In the third and fourth lines none of the terms contain vertices of one V with light SM fields, therefore they do not contribute directly to V decays and single production processes which are the only ones relevant for LHC phenomenology. After the EWSB and diagonalization of the mass matrix that mix the heavy vector triplets with the SM gauge bosons one obtains:

$$M_V^2 = m_V^2 + g_V^2 C_{VVHH} \hat{v}^2 \quad (2.11)$$

In this equation \hat{v} denotes the Higgs field Vaccum Expectation Value (VEV), defined by $\langle H^\dagger H \rangle = \hat{v}^2/2$, which can differ significantly from the EWSB scale $v = 246$ GeV. The goal of the model is to describe new vector bosons with masses at or above TeV scale, but keeping the SM masses $m_{W,Z} \sim 100$ GeV. Therefore the model require a hierarchy in the mass spectrum:

$$\frac{M_V}{m_{W,Z}} \lesssim 10^{-1} \ll 1 \quad (2.12)$$

The partial widths for diboson decays are:

$$\Gamma_{V_0 \rightarrow W_L^+ W_L^-} \simeq \Gamma_{V_\pm \rightarrow W_L^\pm Z_L} \simeq \frac{g_V^2 c_H^2 M_V}{192\pi} \left[1 + \mathcal{O}(\zeta^2) \right] \quad (2.13)$$

Two different models A and B were intoduced in the paper, inspired by weakly coupled extensions of the SM gauge group and strongly coupled scenarios of EWSB, i.e. Composite Higgs models. They can be considered in different regions of g_V , Model A: relatively small ($g_V \lesssim 3$) and Model B: relatively large ($g_V \gtrsim 3$). In our case we used the model B with $g_V = 3$ as a benchmark, because as it can observe from equation 2.8, a relatively large value of g_V favors the interaction between V and vector bosons while suppress the interaction between V and fermions. Very large values of the coupling ($g_V \gg 3$) leads to extremely broad resonance, with $\Gamma/M \gg 0.1$, for which the experimental searches

for a narrow resonance are no longer motivated. Figure 2.5 shows the production cross section and the branching fraction for a heavy resonance of spin one in the model B of the HVT framework.

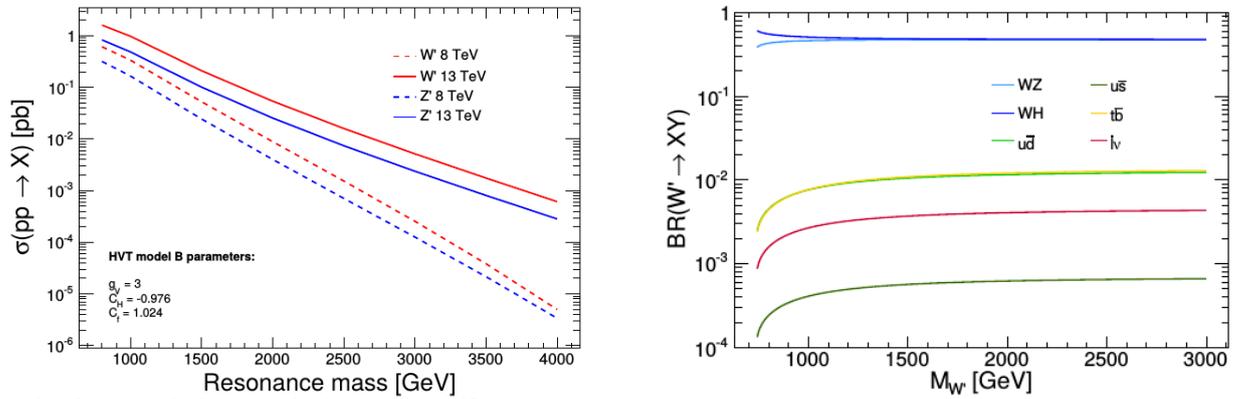


Figure 2.5: Left: Production cross sections of the neutral and charged resonances as a function of the resonance mass in Model B. For comparison purposes the predicted cross sections at both the centre of mass energies of 8 TeV and 13 TeV are shown. Right: Branching Ratios for the two-body decays of the charged W' in the Model B. [26, 62]

Chapter 3

Experimental Apparatus

3.1 Large Hadron Collider (LHC)

The Large Hadron Collider is the largest and most powerful particle accelerator ever built. It boosts protons, to produce two beams travelling in opposite directions, which collide at four points where the two rings of the machine intersect. The design energy per proton beam is of 7 TeV. The protons of the LHC circulate around the ring in well defined bunches. In the LHC, under nominal operating conditions, each proton beam has 2808 bunches, with each bunch containing about 10^{11} protons. They measure a few centimetres long and a millimetre wide when they are far from a collision point. As they approach the collision points, they are squeezed to about $16 \mu\text{m}$ to allow for a greater chance of proton-proton collisions. The LHC uses a bunch spacing of 25 ns (or about 7 m), which corresponds to a frequency of 40 MHz [65]. Nowadays, each proton beam flying around the LHC has an energy of 6.5 TeV, so when two protons collide the collision energy is 13 TeV. There are seven experiments installed at the LHC (Fig. 3.1); The biggest experiments consist in two general-purpose detectors to investigate the largest range of physics possible : The Compact Muon Solenoid (CMS) and A Toroidal LHC Apparatus (ATLAS), and two specialized for focussing on specific phenomena: A Large Ion Collider Experiment (ALICE) and Large Hadron Collider beauty experiment(LHCb). The smaller experiments on the LHC are the TOTAL Elastic and diffractive cross section Measurement (TOTEM), Large Hadron Collider forward experiment (LHCf) and the Monopole and Exotics Detector at the LHC (MOEDAL).

The first two experiments are focused on "forward particles", protons or heavy ions that brush past each other rather than meeting head on when the beams collide and the last experiment searches for a hypothetical particle called the magnetic monopole. TOTEM will be installed close to the CMS interaction point, LHCf will be installed near ATLAS and MOEDAL near LHCb.

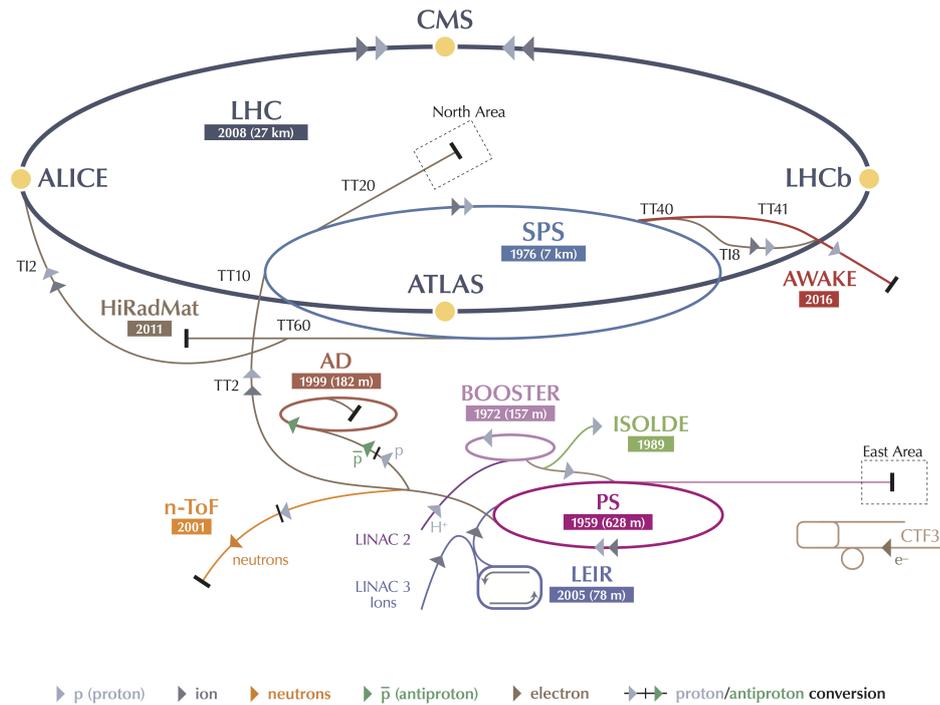


Figure 3.1: The CERN accelerator complex.

3.2 CMS Detector

The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the superconducting solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap sections. Muons are measured in gas-ionization detectors embedded in the steel flux-return yoke outside the solenoid. Extensive forward calorimetry complements the coverage provided by the barrel and endcap detectors. A right-handed coordinate system is used with its origin at the nominal interaction point (IP). The x-axis points to the center of the LHC ring, the y-axis is vertical and points upward, and the z-axis is parallel to the counterclock-wise beam direction. The

azimuthal angle ϕ is measured with respect to the x-axis in the xy-plane and the polar angle θ is defined with respect to the z-axis, while the pseudorapidity is defined as $\eta = -\ln[\tan(\theta/2)]$. Figure 3.20 shows a schematic view of the CMS detector and figure 3.3 shows the η and ϕ coordinates in the CMS detector.

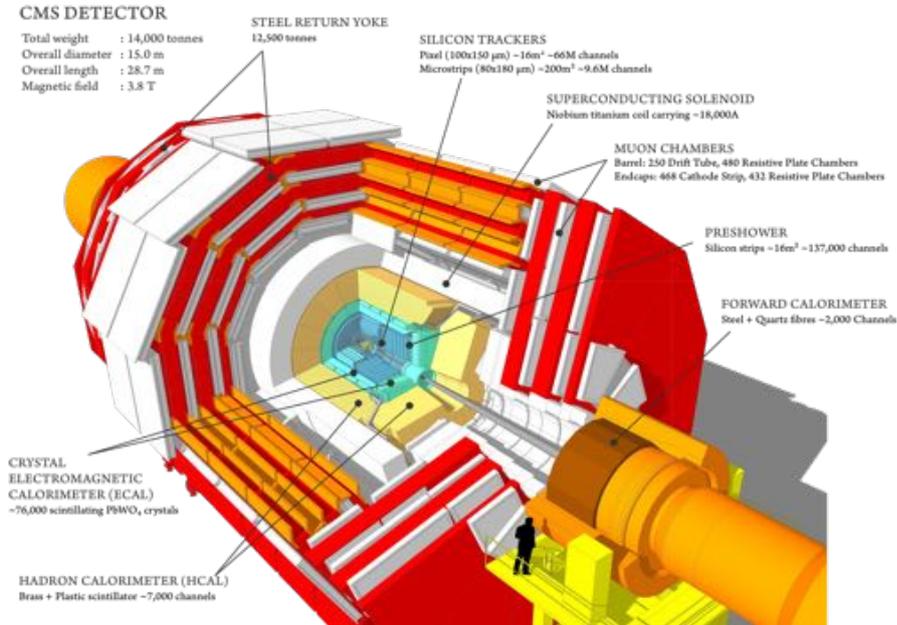


Figure 3.2: Schematic view of the CMS detector .

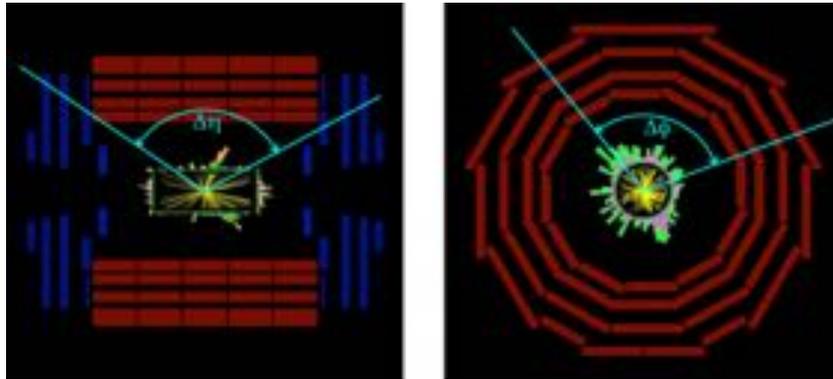


Figure 3.3: η and ϕ coordinates in the CMS detector.

3.2.1 The Magnet

The CMS detector contains a 3.8 T superconducting solenoid (which produce an axial field) with a free bore of a diameter of 6 m and a length of 12.5 m, enclosed inside a 12 000 t yoke made of common structural steel. The inner diameter of the coil is large enough to set up the tracking system and the full calorimetry. Due to the number of ampere-turns

required for generating a field of 3.8 T, the winding in the coil is composed of 5 modules, with four layers of conductor each. The coil is indirectly cooled by saturated helium at 4.5 K circulating in the thermosiphon mode through a network of pipes welded to the external mandrels [66]. Figure 3.4 shows an artistic view of the superconducting solenoid. The yoke is composed of five three-layered dodecagonal barrel wheels and three endcap disks at each end. In the barrel region the innermost yoke layer is 295 mm thick and each of the two outermost ones is 630 mm thick. The yoke contributes to only 8% of the central magnetic flux density; its main role is to increase the field homogeneity in the tracker volume and to reduce the stray field by returning the magnetic flux of the solenoid. The demand for good momentum resolution, without making tight requests on the spatial resolution of the muon chambers, influence to the choice of a high solenoidal magnetic field. Since the magnet is the main component of CMS in terms of size, weight and structural rigidity, it is used as the principal structural element to support all barrel detector components. Figure 3.5 shows a schematic view of the CMS detector and their magnet components.

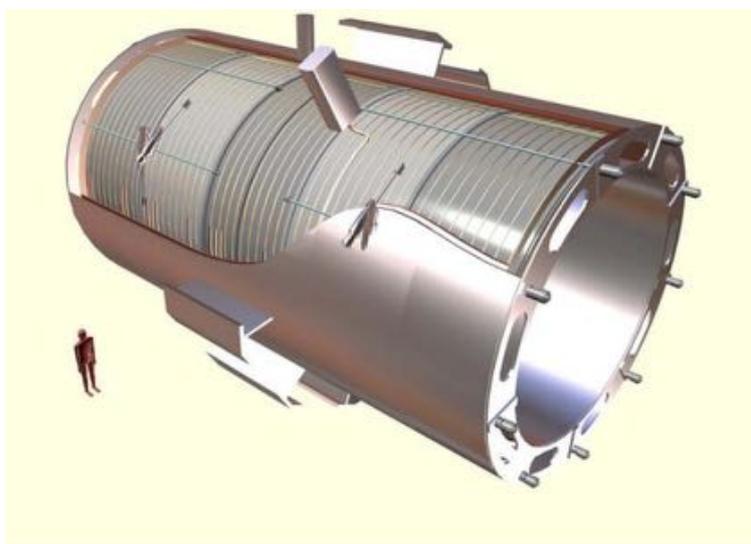


Figure 3.4: General artistic view of the 5 modules composing the superconducting coil.

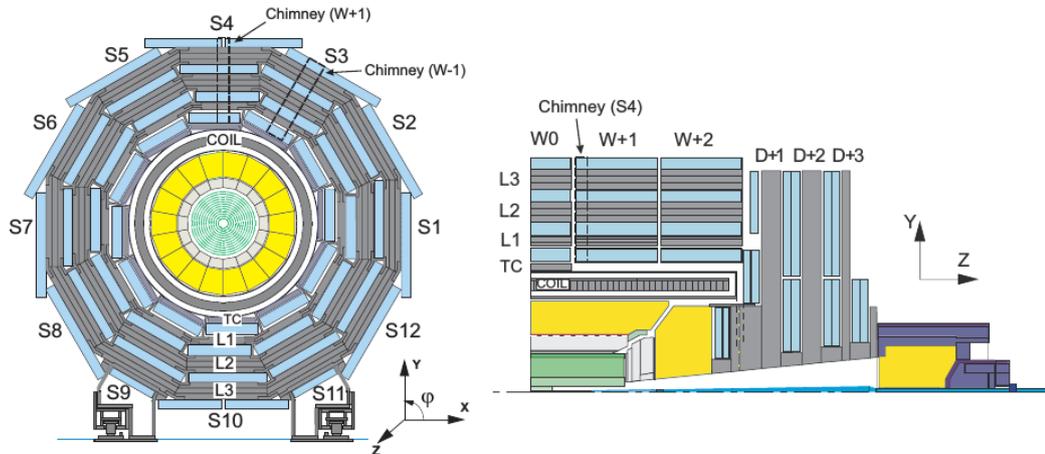


Figure 3.5: Schematic views of the CMS detector, with the numbering convention for azimuthal sectors (S), wheels (W), barrel yoke layers (L) and endcap disks (D). Left: transverse view at $z = 0$. Right: longitudinal view of one quarter of the detector. .

Simulation and reconstruction of events in the CMS detector require knowledge of the magnetic field in the entire detector, both in the inner tracking region and in the complex configuration of the steel return yoke. Figure 3.6 shows the predicted magnetic flux density on a longitudinal section of the CMS detector.

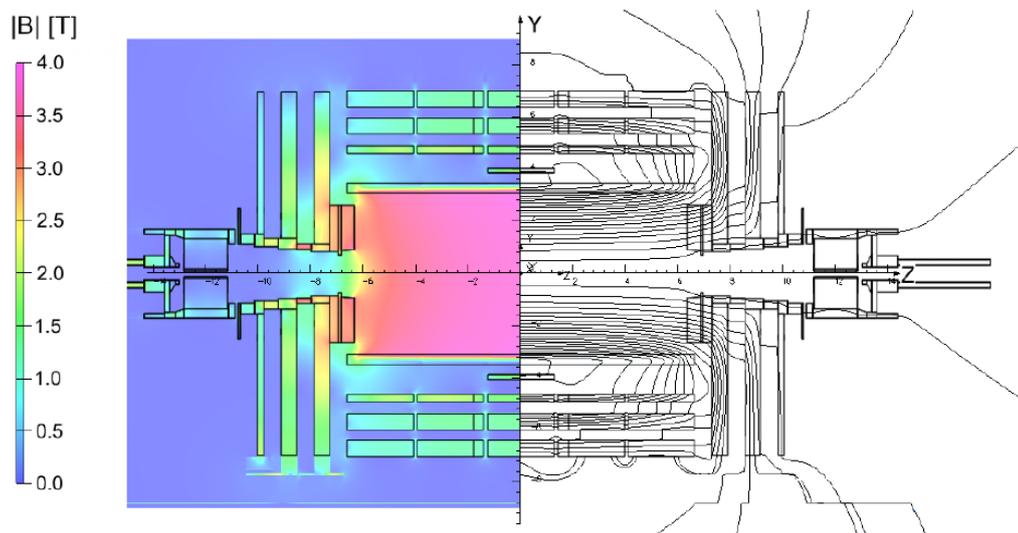


Figure 3.6: Value of $|\vec{B}|$ (left) and field lines (right) predicted on a longitudinal section of the CMS detector, at a central magnetic flux density of 3.8 T. .

3.2.2 CMS Tracking system

The main idea behind "tracking" is to provide a precise and efficient measurement of the trajectories of charged particles emerging from the LHC collisions. The entire tracking system is enclosed in a huge solenoid magnet, which produces an approximately uniform magnetic field of 3.8 T pointing along the direction on the LHC beam. As the charged particles travel from the center of the detector, their trajectories are bended. Along its path, each charged particle leaves hits in the detecting material. In a process called track reconstruction, CMS software connects the hits and produces a track. From the information of the charge of the particle, the intensity of the magnetic field and the radius of the path, the software determine the transverse momentum of the particle. The tracking system is divide into two different subsystems, the Silicon Pixels and the Silicon Strips [67, 68].

Silicon pixels

The pixel tracker allows the reconstruction of charged particle trajectories in the nearest region to the interaction point. In addition, the pixel detector is essential for the reconstruction of secondary vertices from b and tau decays, and forming seed tracks for the outer track reconstruction and high level triggering. The CMS pixel detector consists of about 65 million of pixels, each of $100 \times 150 \mu\text{m}^2$, spread over three cylindrical layers with mean radii of 4.4, 7.3, and 10.2 cm with respect to the beam and around 53 cm in length. The system is completed by endcaps which consist of 2 disks in each extremity, extending the acceptance of the tracker up to a pseudorapidity of $|\eta| < 2.5$. Figure 3.7(left) shows a perspective view of the CMS pixel system. Figure 3.7(right) shows a quarter of a slice of a transverse view of the pixel detector. The centre of the detector is at the left-bottom corner of the drawing, in the interaction region. The horizontal axis is parallel to the LHC beams. The vertical axis points along the radius. Various pseudo-rapidity values are shown at the ends of the black dashed lines. When a particle passes through the silicon detector, it ionizes the material, producing an electron-hole pairs. These charge carriers are divided in opposite directions by an electric field, and pulled into "contacts". Then, the charge built up on those contacts produces a current that flows into the electronics.

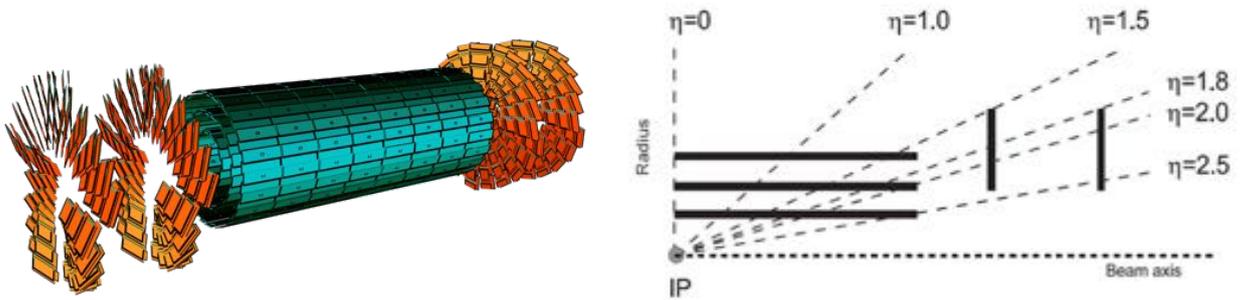


Figure 3.7: Left: Sketch of the CMS pixel detector. Right: Quarter of a slice of the CMS pixel detector by a plane which contains its axis of symmetry.

For high p_T tracks, the transverse impact parameter resolution reaches $10 \mu\text{m}$. The nominal momentum resolution is 0.7% (5%) at 1 GeV (1 TeV) and the primary vertex resolution is $60 \mu\text{m}$ ($500 \mu\text{m}$) for vertices with many tracks (few tracks with low p_T) [69].

Silicon Strips

The Silicon Strip Tracker (SST) consists of four main subsystems, shown in Figure 3.8: the four-layer Tracker Inner Barrel (TIB), the six-layer Tracker Outer Barrel (TOB), the three-disk Tracker Inner Disks (TID), and the nine-disk Tracker End Caps (TEC). Each TID disk is made of three rings of modules, while TEC disks have seven rings. The whole SST has a diameter of 2.4 m and a length of 5.5 m, being the largest silicon detector ever built with an active area of 198m^2 . Its acceptance ranges over a region in pseudo-rapidity $|\eta| < 2.5$. This component of the tracker consists of 15 148 detector modules and comprises 9.3 million detector channels. Each detector module consists of a carbon or graphite fibre frame, which supports the silicon sensor and the associated front-end readout electronics.

The physical principle behind the strip detector is the same than the pixel detector. As a charged particle crosses the detector, it interacts with the electrons from the material producing a small pulse of current during a very short time. This small amount of charge is amplified by electronic components, resulting in "hits", which are used for the path reconstruction.

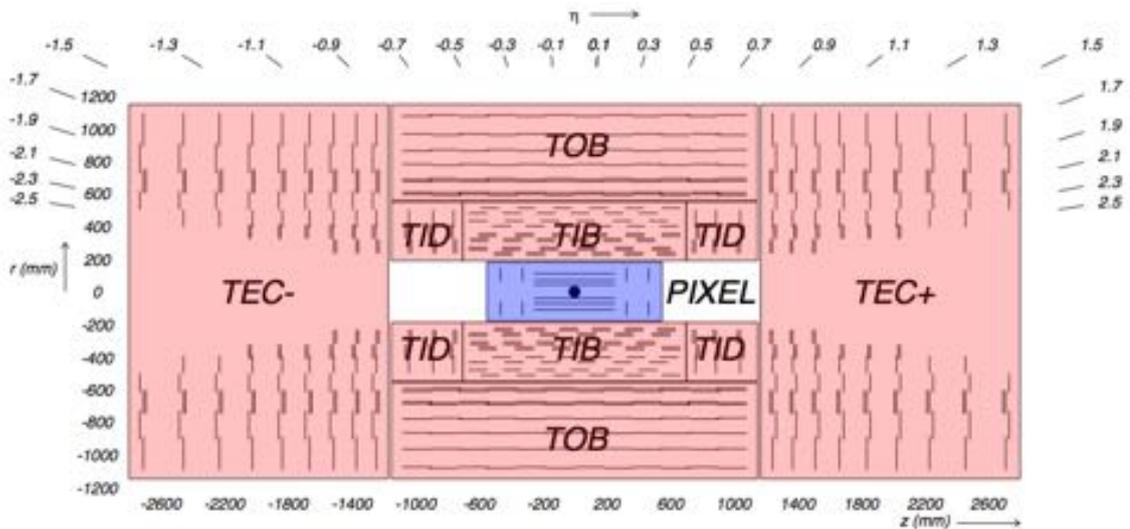


Figure 3.8: Strip and pixel silicon detector .

3.2.3 Electromagnetic Calorimeter (ECAL)

The electromagnetic calorimeter of CMS (ECAL) is an hermetic and homogeneous calorimeter. It is divided in three sections: barrel, preshower and two endcap sections. The ECAL barrel (EB) covers the pseudo-rapidity region $|\eta| < 1.48$ and is constructed from 61200 lead tungstate crystals. The crystals are grouped into units, called supermodules, of 1700 crystals. There are 36 supermodules in the barrel. The ECAL Preshower (ES) is a sampling calorimeter with two layers: lead radiators initiate electromagnetic showers from incoming photons/electrons whilst silicon strip sensors placed after each radiator measure the deposited energy and the transverse shower profiles. The total thickness of the Preshower is 20 cm. The principal aim of the CMS Preshower detector is to identify neutral pions in the endcaps within a fiducial region $1.653 < |\eta| < 2.6$. It also helps the identification of electrons against minimum ionizing particles, and improves the position determination of electrons and photons with high granularity. The ECAL endcaps (EE) cover the pseudo-rapidity region $1.48 < \eta < 3.0$. Each endcap is made from two 'Dees' and 7244 crystals. The crystals are grouped into modules of 25 crystals, known as supercrystals. The inner and outer boundaries of the endcaps are made more circular by the addition of smaller units known as partial supercrystals [67, 70]. Figure 3.9 shows a schematic view of the CMS electromagnetic calorimeter. Photodetectors are used in the ECAL in order to collect the light emitted by the crystals. The configuration of

the magnetic field and the expected level of radiation led to different choices: avalanche photodiodes in the barrel and vacuum phototriodes in the endcaps. The characteristics of the lead tungstate (PbWO_4) crystals make them an appropriate choice for operation at LHC. The high density (8.28 g/cm^3), short radiation length ($X_0 = 0.89 \text{ cm}$) and small Molière radius (2.2 cm) result in a fine granularity and a compact calorimeter. The scintillation decay time is very fast with 80% of the scintillation light collected within 25ns (in the LHC bunches of protons collide every 25ns).

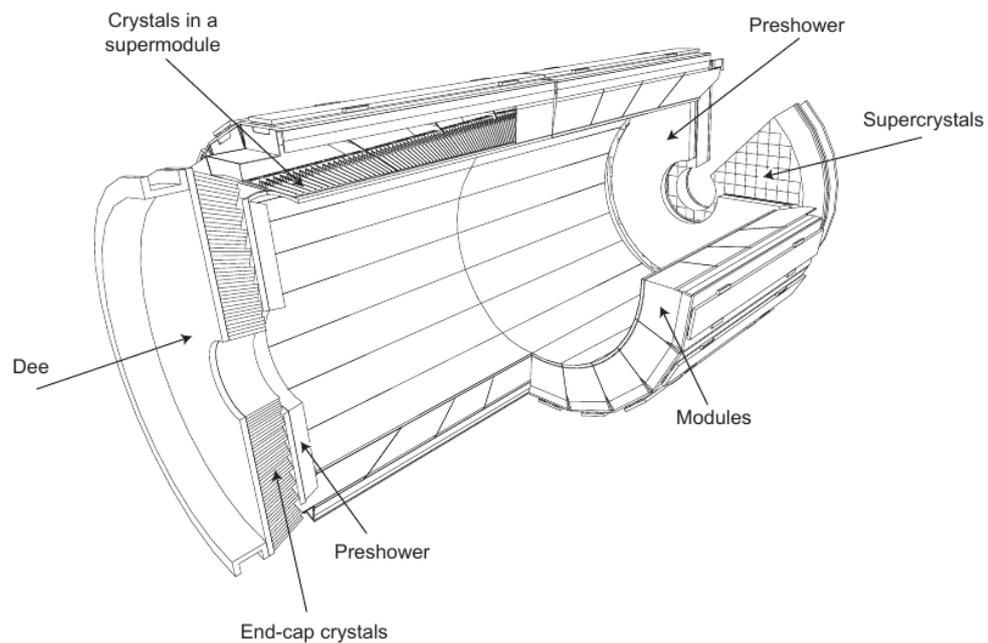


Figure 3.9: Layout of the CMS electromagnetic calorimeter showing the arrangement of crystal modules, supermodules and endcaps, with the preshower in front.

The energy of the particles are measured in the ECAL in the following way. As the particle (electron or photon) go through the crystals (which act as both absorber and the scintillator) it interact with the material producing a shower of secondary particles (via pair-production and bremsstrahlung processes). The amount of secondary particles is proportional to the energy of the incoming particle and the amount of light produced by the scintillator is proportional to the number of secondary particles, resulting in the amount of light proportional to the energy of the initial particle. The material of the crystals is dense enough to stop a shower initiated by an electron or a photon.

A typical energy resolution for the ECAL is:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{12\%}{E}\right)^2 + (0.3\%)^2 \quad (3.1)$$

where E is the energy in GeV.

3.2.4 Hadronic Calorimeter (HCAL)

The HCAL is a sampling calorimeter which determine the position, energy and arrival time of a particle using alternating layers of "absorber" and fluorescent "scintillator" materials that produce a rapid light pulse when the particle passes through. Special optic fibres collect up this light and deliver it into readout boxes (RBX) where photodetectors (hybrid photodiodes (HPDs)) amplify the signal (each readout box contains 4 HPDs). When the amount of light in a given region is summed up over many layers of tiles in depth, called a "tower", this total amount of light is a measure of a particle's energy [67, 71].

The hadron calorimeter barrel is radially restricted between the outer extent of the electromagnetic calorimeter ($R = 1.77$ m) and the inner extent of the magnet coil ($R = 2.95$ m). This constrains the total amount of material which can be put in to absorb the hadronic shower. Therefore, an outer hadron calorimeter is placed outside the solenoid complementing the barrel calorimeter. Beyond $|\eta| = 3$, the forward hadron calorimeters placed at 11.2 m from the interaction point extend the pseudorapidity coverage down to $|\eta| = 5.2$.

The HCAL is organized into barrel (formed by two sections : Hadron Barrel(HB) in the region $|\eta| < 1.4$ and Hadron Outer (HO) in the region $|\eta| < 1.26$), Hadron Endcap (HE) [$1.3 < |\eta| < 3.0$] and Hadron Forward (HF) [$2.9 < |\eta| < 5.0$] sections.(Fig. 3.10)



Figure 3.10: Longitudinal view of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters.

The HB is a sampling calorimeter covering the pseudorapidity range $|\eta| < 1.4$, resulting in 2304 towers with a segmentation $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. The HB consists of 36 identical azimuthal wedges which form the two half-barrels (HB+ and HB-).

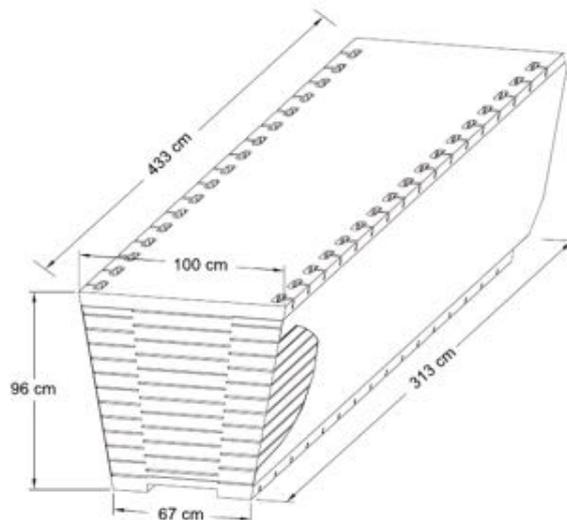


Figure 3.11: View of an HB wedge.

The wedges (Fig. 3.11) are constructed out of flat brass absorber plates aligned parallel to the beam axis. The HB baseline active material is 3.7 mm thick Kuraray SCSN81 plastic scintillator, chosen for its long term stability and moderate radiation hardness. The granularity of the HCAL is 25 times coarser than that of the ECAL, which would not allow charged and neutral hadrons to be spatially separated in jets with a transverse momentum much above 100 GeV/c. The hadron energy resolution in the combined ECAL-HCAL system is, however, of the order of 10% at 100 GeV. This resolution allows

neutral hadrons to be detected as an energy excess on top of the energy deposited by the charged hadrons pointing to the same calorimeter cells.

3.2.5 The Muon System

Muons are an unmistakable signature of most of the physics LHC is designed to explore. The ability to trigger on and reconstruct muons at the highest luminosities is central to the concept of CMS. Muons can penetrate several metres of iron without interacting. Unlike most particles they are not stopped by any of the calorimeter detectors ($\tau \approx 2.2 \mu\text{s}$). Therefore, chambers to detect muons are placed at the very edge of the experiment where they are the only particles likely to register a signal.

The CMS muon system is designed to have the capability of reconstructing the momentum and charge of muons over the the entire kinematic range of the LHC. The Muon system is a class of tracking detector, and is divided in two main regions: the Barrel ($|\eta| < 1.2$) and the endcap ($1.2 < |\eta| < 2.4$). There are three types of detectors in the Muon System:

Drift Tubes (DT) The drift tube (DT) chamber (Fig. 3.12) system measures muon positions in the barrel part of the detector. The chamber volume is filled with a Ar(85 %)/CO₂(15 %) gas mixture, kept at atmospheric pressure. When a muon or any charged particle passes through the volume, it knocks electrons off the atoms of the gas. By registering where along the wire electrons hit, as well as by calculating the muon's original distance away from the wire, DTs give two coordinates for the muon's position.

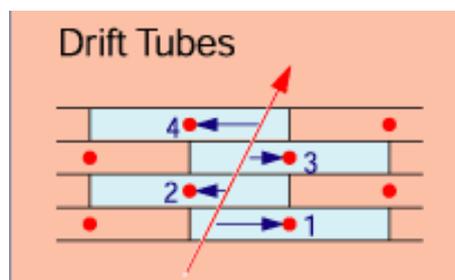


Figure 3.12: Muon Drift Tubes

Cathode Strip Chamber (CSC) Cathode strip chambers (CSC) are used in the endcap disks where the magnetic field is inhomogeneous and particle rates are high. CSCs consist of arrays of positively charged "anode" wires crossed with negatively charged copper "cathode" strips within a gas volume. When muons pass through, they knock electrons off the gas atoms, which flock to the anode wires creating an avalanche of electrons. Positive ions move away from the wire and towards the copper cathode, also inducing a charge pulse in the strips, at right angles to the wire direction.

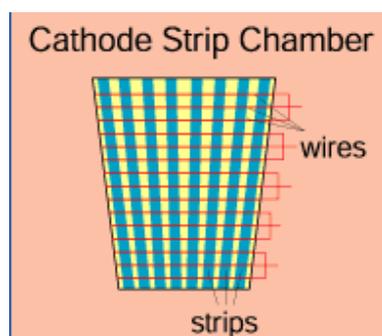


Figure 3.13: Cathode Strip Chamber

Resistive Plate Chambers (RPC) Resistive plate chambers (RPC) are fast gaseous detectors that provide a muon trigger system parallel with those of the DTs and CSCs. RPCs consist of two parallel plates, a positively-charged anode and a negatively-charged cathode, both made of a very high resistivity plastic material and separated by a gas volume. When a muon passes through the chamber, electrons are knocked out of gas atoms. These electrons in turn hit other atoms causing an avalanche of electrons. The electrodes are transparent to the signal (the electrons), which are instead picked up by external metallic strips after a small but precise time delay. The pattern of hit strips gives a quick measure of the muon momentum, which is then used by the trigger to make immediate decisions about whether the data are worth keeping. RPCs combine a good spatial resolution with a time resolution of just one nanosecond (one billionth of a second).



Figure 3.14: Resistive Plate Chambers

In total there are 1400 muon chambers: 250 drift tubes (DTs) and 540 cathode strip chambers (CSCs) track the particles' positions and provide a trigger, while 610 resistive plate chambers (RPCs) form a redundant trigger system, which quickly decides to keep the acquired muon data. The Barrel Detector consists of 4 concentric "stations" (Fig. 3.15) of 250 chambers inside the magnet return yoke of CMS, which is in turn divided into 5 wheels. Each wheel is divided into 12 sectors, each covering a 30° azimuthal angle. The 2 innermost stations, named MS1 and MS2, consist of "sandwiches" made of a DT chamber placed between 2 RPCs. The 2 outermost stations, MS3 and MS4, consist of packages of a DT chamber coupled to a layer made of 1, 2, or 4 RPCs, depending on the sector and station, placed on the innermost side of the station.

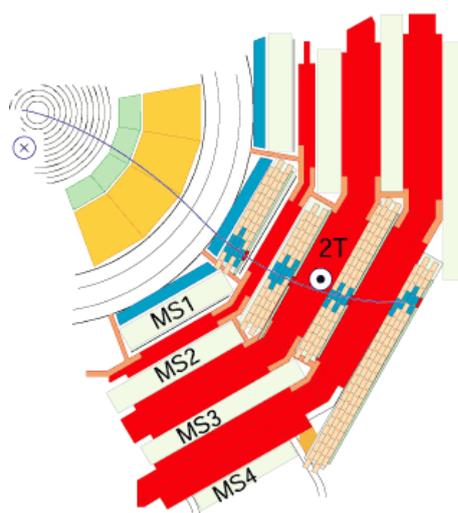


Figure 3.15: Muons Stations

3.3 CMS Data Acquisition and Triggering

When CMS performs at its peak, about one billion proton-proton interactions will take place every second inside the detector. There is no way that data from all these events could be read out, and even if they could, most would be less likely to reveal new phenomena. We therefore need a "trigger" that can select the potentially interesting events, and reduce the rate to just a few hundred "events" per second, which can be read out and stored on computer disk for subsequent analysis. This task is performed by the trigger system, which is the start of the physics event selection process. The rate is reduced in two steps called Level-1 (L1) Trigger and High-Level Trigger (HLT),

respectively. The Level-1 Trigger consists of custom-designed, largely programmable electronics, whereas the HLT is a software system implemented in a filter farm of about one thousand commercial processors. The rate reduction capability is designed to be at least a factor of 10^6 for the combined L1 Trigger and HLT.

Level 1 of the trigger is an extremely fast and wholly automatic process that looks for simple signs of interesting physics, e.g. particles with a large amount of energy or in unusual combinations. The Level 1 trigger select the best 100,000 events each second from the billion available. For the next test, the HLT assimilate and synchronise information from different parts of the detector to recreate the entire event and send it to a farm of more than 1000 standard computers.

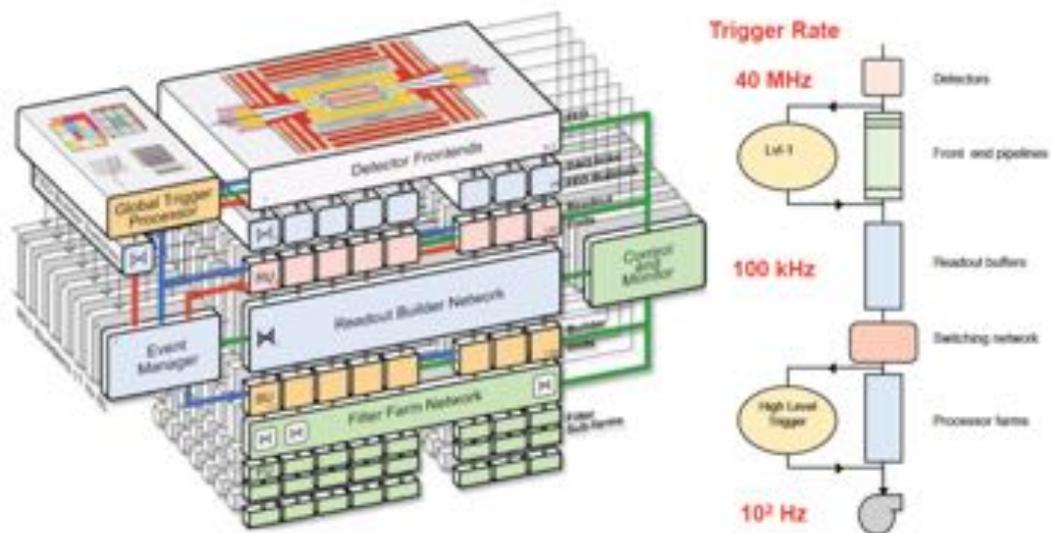


Figure 3.16: CMS Triggering System

3.4 Physics objects

3.4.1 Track reconstruction

Under nominal conditions of the LHC at $\sqrt{s} = 13$ TeV a typical instantaneous luminosity around $10^{34} \text{cm}^{-2} \text{s}^{-1}$ is expected, with the proton bunches intersecting at intervals of 25 ns. Therefore, the CMS tracker will be crossed by about 1000 charged particles at each bunch crossing, producing an average of more than twenty proton-proton (pp) collisions. These multiple interactions are known as *pileup* [72], to which prior or later bunch crossings can also contribute because of the finite time resolution of the detector.

The first step in processing the data prior to track reconstruction is the efficient detection of *hits*, which represent the positions in the sensors of the tracker where charged particles passed through. Track reconstruction is the process of using the hits on the pixel (section 3.2.2) and strip tracker (section 3.2.2) to estimate the momentum and the position parameters (the longitudinal z_0 and transverse d_0 distances relative to the beam axis) of the charged particles responsible for the hits (tracks). Figure 3.17 shows the position parameters of a track.

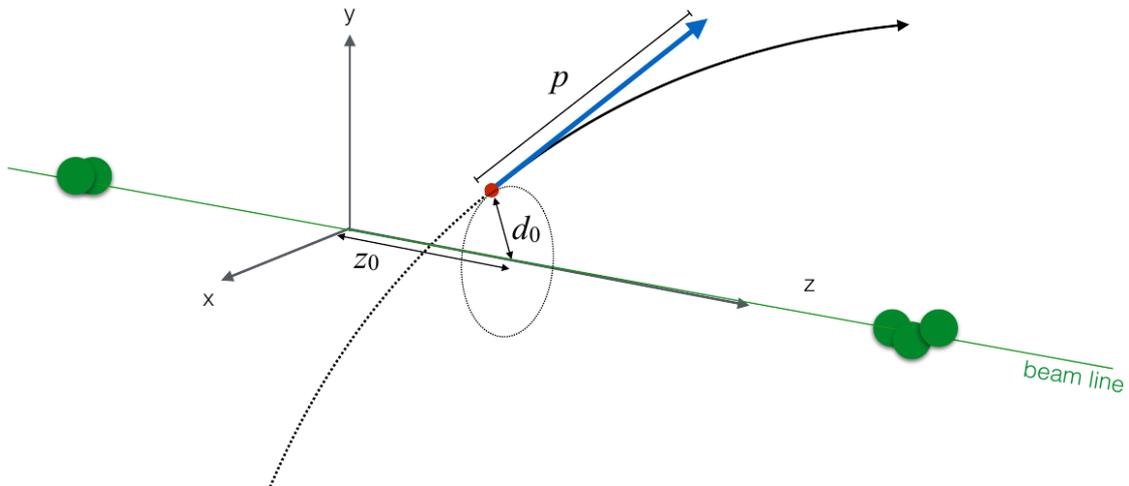


Figure 3.17: Geometrical description of the closest approach point of a track (curved line) to the beam line: transverse (d_0) and longitudinal (z_0) impact parameters.

Reconstructing the trajectories of charged particles is a computationally challenging task. Hence, CMS developed a software, which support pattern recognition and track fitting in the same framework, called *Combinatorial Track Finder* (CTF) [73]. This software

is based on an adaptation of the *Combinatorial Kalman filter* method [74–76], which in turn is an extension of the *Kalman filter* method [77].

The collection of reconstructed tracks is produced by multiple passes (iterations) of the CTF track reconstruction sequence, in a process called *iterative tracking*. The idea behind iterative tracking is to search in the initial iterations for the easiest tracks to find (e.g., of relatively large p_T , and produced near the interaction region). After each iteration, the hits related with tracks are removed in a search for more difficult classes of tracks (e.g., low- p_T , or greatly displaced tracks). Each iteration process contains four steps:

- Seed generation provides initial track candidates found using only a few (2 or 3) hits. A seed defines the initial estimate of the trajectory parameters and their uncertainties
- Track finding is based on a Kalman filter. It extrapolates the seed trajectories along the expected flight path of a charged particle, searching for additional hits that can be assigned to the track candidate.
- The track-fitting module is used to provide the best possible estimate of the parameters of each trajectory by means of a Kalman filter and smoother.
- Track selection sets quality flags, and discards tracks that fail certain specified criteria

Muons are reconstructed better than any other charged particle in the tracker, as they mainly interact with the silicon detector through ionization of the medium and, unlike electrons, their energy loss through bremsstrahlung is negligible. Muons therefore tend to cross the entire volume of the tracking system, producing detectable hits in several sensitive layers of the detector. For isolated muons with $1 < p_T < 100$ GeV, the tracking efficiency is $>99\%$ over the full η -range of tracker acceptance, and does not depend on p_T as shown in Fig.3.18. The fake rate is completely negligible.

3.4.2 Primary Vertex (PV)

The objective of primary-vertex reconstruction [73] is to determine the location, and the related uncertainty, of all proton-proton (pp) interaction vertices in each event,

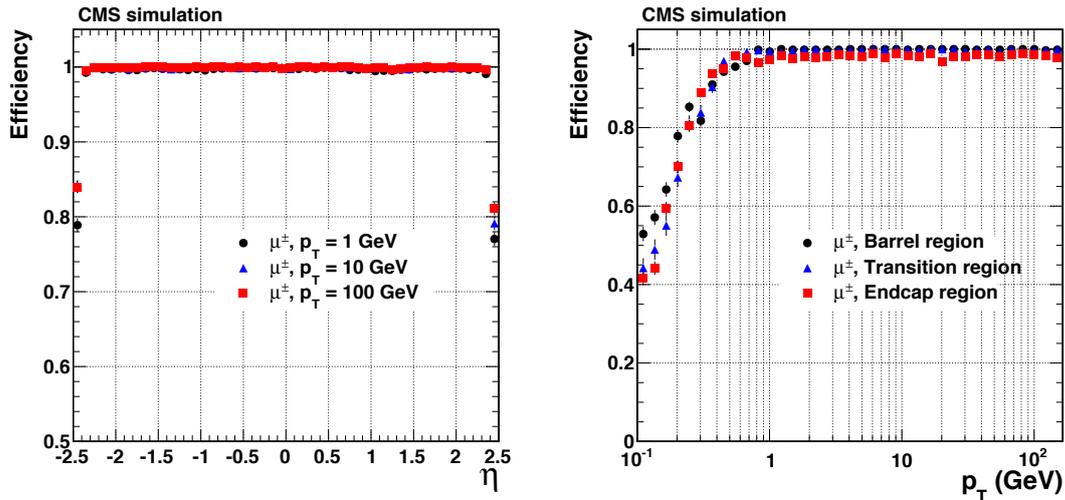


Figure 3.18: Track reconstruction efficiencies for single isolated muons passing high-purity quality requirements. Results are shown as a function of η (left), for $p_T = 1, 10,$ and 100 GeV. They are also shown as a function of p_T (right), for the barrel, transition, and endcap regions, which are defined by the η intervals of $0-0.9, 0.9-1.4$ and $1.4-2.5$, respectively. [73]

including vertices originated from the primary event or from additional pp interactions (pileup), using the available reconstructed tracks. The method consists of three steps:

- Selection of the tracks
- Clustering of the tracks that appear to originate from the same interaction vertex
- Fitting for the position of each vertex using its associated tracks

Tracks are selected if they are produced rapidly in the primary interaction region. The selected tracks are clustered based on information from their z -coordinates and their point of closest approach to the centre of the beam spot. Track clustering is performed using a deterministic annealing (DA) algorithm [78], finding the global minimum for a problem with many degrees of freedom. After identifying candidate vertices based on the DA clustering, the candidates with at least two tracks are then fitted using an adaptive vertex fitter [79] to compute the best estimate of vertex parameters (position, covariance matrix, number of degrees of freedom for the vertex, and weights of the tracks used in the vertex). In the adaptive vertex fit, each track in the vertex have a weight between 0 and 1, which indicate the likelihood that it certainly belongs to the vertex. Tracks that are compatible with the position of the reconstructed vertex have a weight close to 1, while tracks that lie more than a few standard deviations from the

vertex have small weights. The number of degrees of freedom in the fit is defined as

$$n_{\text{dof}} = -3 + 2 \sum_{i=1}^{\text{\#tracks}} w_i \quad (3.2)$$

where w_i is the weight of the i th track, and the sum runs over all tracks associated with the vertex. The value of n_{dof} is therefore strongly correlated with the number of tracks compatible with arising from the interaction region. For this reason, n_{dof} can be also used to select true proton-proton interactions.

The primary-vertex efficiency is estimated to be close to 100% when more than two tracks are used to reconstruct the vertex. Figure 3.19 shows the efficiency of the primary-vertex reconstruction as a function of the number of tracks clustered in z . The effect of pileup on the efficiency is checked using simulated minimum-bias events, with and without added pileup, and the loss of efficiency is found to be $< 0.1\%$ for the pileup with a mean value of 8 [80].

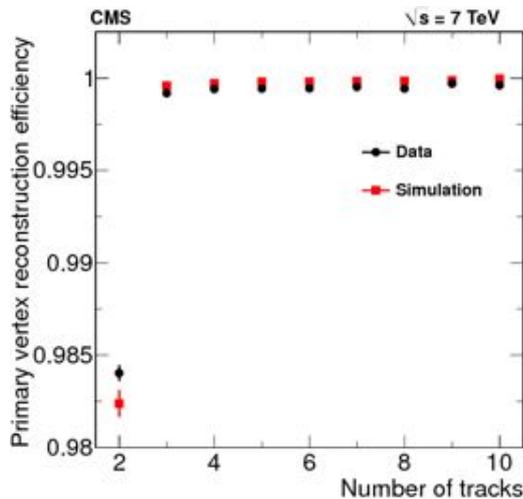


Figure 3.19: Primary-vertex reconstruction efficiency as a function of the number of tracks in a cluster, measured in minimum-bias data and in MC simulation. [73].

3.4.3 Particle Flow (PF)

The global event reconstruction (also called particle-flow event reconstruction) reside in the reconstruction and identification of each single particle with an optimized combination of all subdetector information. In this process, the identification of the particle type

(photon, electron, muon, and hadrons) plays an important role in the determination of the particle direction and energy.

- **Photons** (*e.g.* coming from π^0 decays or from electron bremsstrahlung) are identified as ECAL energy clusters not linked to the extrapolation of any charged particle trajectory to the ECAL. The energy of photons is directly obtained from the ECAL measurement.
- **Electrons** (*e.g.* coming from photon conversions in the tracker material or from b-hadron semileptonic decays) are identified as a primary charged particle track and potentially many ECAL energy clusters corresponding to this track extrapolation to the ECAL and to possible bremsstrahlung photons emitted along the way through the tracker material. The energy of electrons is determined from a combination of the track momentum at the main interaction vertex, the corresponding ECAL cluster energy, and the energy sum of all bremsstrahlung photons attached to the track.
- **Muons** (*e.g.* from b-hadron semileptonic decays) are identified as a track in the central tracker consistent with either a track or several hits in the muon system, associated with an energy deficit in the calorimeters. The energy of muons is obtained from the corresponding track momentum.
- **Charged hadrons** are identified as charged particle tracks neither identified as electrons, nor as muons. The energy of charged hadrons is determined from a combination of the track momentum and the corresponding ECAL and HCAL energy, corrected for the response function of the calorimeters to hadronic showers.
- **Neutral hadrons** are identified as HCAL energy clusters not linked to any charged hadron trajectory, or as ECAL and HCAL energy excesses with respect to the expected charged hadron energy deposit. The energy of neutral hadrons is obtained from the corresponding corrected ECAL and HCAL energy.

3.4.4 Jets

Jets are the experimental signatures of quarks and gluons produced in high-energy processes such as hard scattering of partons in proton-proton collisions. The LHC

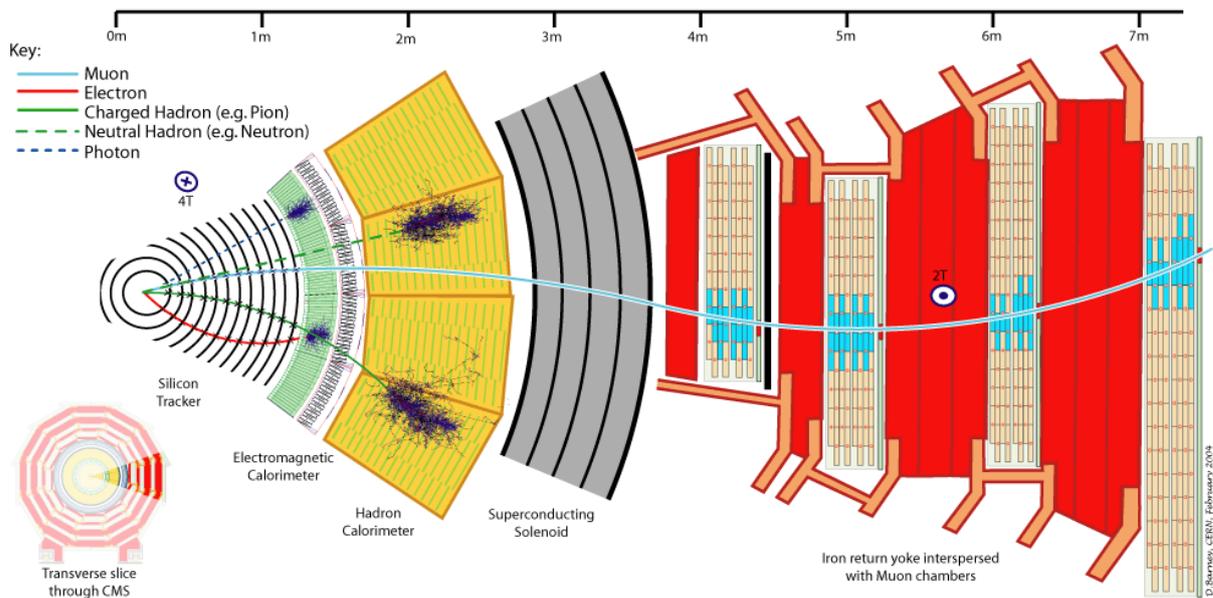


Figure 3.20: A slice through the CMS detector. CMS consists of a central silicon Tracker (Pixel and Strips), an Electromagnetic Calorimeter, a Hadron Calorimeter, a Superconduction Solenoid Magnet, a massive iron return yoke instrumented with Muon Chambers. The depicted interactions present an ideal detector behavior for the different particles μ , e , charged and neutral hadrons, and γ .

collides protons containing colored partons: quarks, antiquarks and gluons. Almost immediately after being produced, a quark or gluon fragments and hadronises, leading to a collimated spray of energetic hadrons: a jet (Fig. 3.21).

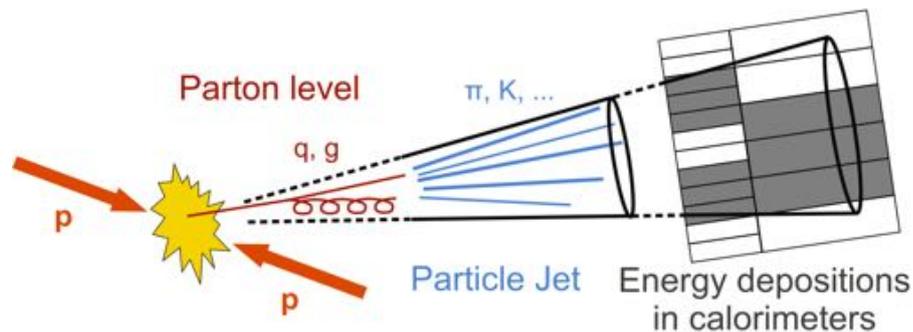


Figure 3.21: pp-collision resulting in a collimated spray of particles, a jet.

Jets are obvious structures when one looks at an event display, and by measuring their energy and direction one can get close to the idea of the original parton (Fig. 3.22).

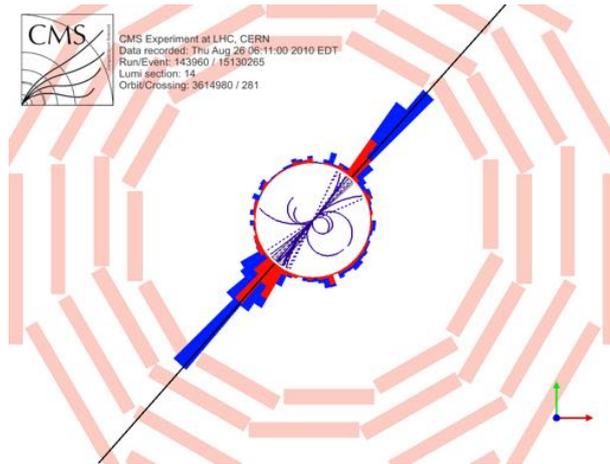


Figure 3.22: CMS event display for a dijet reaction

Jet clustering algorithms provide a set of rules for grouping particles into jets. They usually involve one or more parameters that indicate how close two particles must be for them to belong to the same jet. Additionally they are always associated with a recombination scheme, which indicates what momentum to assign to the combination of two particles. Taken together, a jet algorithm with its parameters and a recombination scheme form a jet definition. There are many types of jet algorithms in the market, but in Run-2 CMS will mainly use the anti- k_t algorithm [81]. The method introduce two kind of distances. First, the distance d_{ij} between entities i and j , defined by:

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{\Delta R^2} \quad (3.3)$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and ΔR is the geometrical distance $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. Here k_{ti} , y_i and ϕ_i are the transverse momentum, the rapidity and the azimuth of the particle i respectively. The distance d_{iB} between the entity i and the beam (B) is defined by:

$$d_{iB} = k_{ti}^{2p} \quad (3.4)$$

The algorithm compute both distances d_{ij} and d_{iB} for the entity i . If the smallest is d_{ij} , it combine (sum of four momenta) the two particles i and j , update the distances and then proceed to find the next smallest. If the smallest is d_{iB} , the algorithm remove the particle i , and call it a jet. The process is repeated until all particles are clustered into a

jet. The parameter p governs the relative power of energy vs geometrical scales to distinguish the three algorithms: $p = 1$ (k_t algorithm [82]), $p = 0$ (Cambridge-Aachen algorithm [83]) and $p = -1$ (anti- k_t algorithm).

Particle Flow Jet (PFJet)

The Particle-Flow (PF) jets are reconstructed by clustering the four-momentum vectors of particle-flow candidates. The particle-flow algorithm combines the information from all relevant CMS sub-detectors to identify and reconstruct all visible particles in the event, namely muons, electrons, photons, charged hadrons, and neutral hadrons. The energy of charged hadrons is determined from a combination of the track momentum and the corresponding ECAL and HCAL energy, calibrated for the non-linear response of the calorimeters. The energy of neutral hadrons is obtained from the corresponding calibrated ECAL and HCAL energy. The PF jet momentum and spatial resolutions are greatly improved with respect to calorimeter jets, as the use of the tracking detectors and of the high granularity of ECAL allows resolution and measurement of charged hadrons and photons inside a jet, which together constitute $\sim 85\%$ of the jet energy.

Charged hadron subtraction (CHS)

Contamination to the jet from pileup degrades the ability to reconstruct the jet observables. Previous pile up corrections applied in Run I help to correct the four-vector of the jet but not the jet structure observables. One new approach is use tracking information which takes advantage of the fact that a large fraction of the pileup vertices are separated in space from the vertex of interest. Therefore, charged particles from pileup vertices can be removed from the jets, in a process called "charged hadron subtraction" (CHS) (Fig. 3.23). Charged Hadron Subtraction (CHS) is a technique used to reduce the effect of "in-time pileup" on reconstructed physics objects. In this approach, charged hadrons unambiguously associated to pileup vertices are removed from the event and the remaining PF candidates are allowed to cluster to form jets.

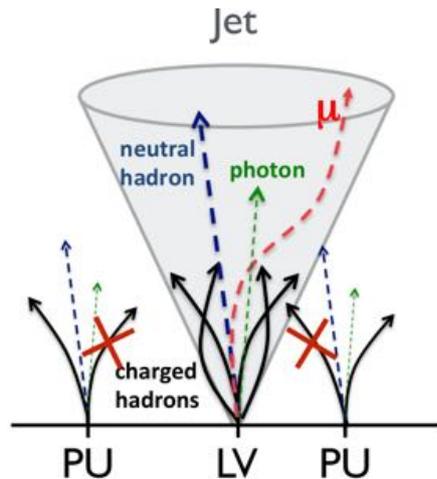


Figure 3.23: Charged hadron subtraction.

Jet Energy Correction (JEC)

Jet energy corrections need to be applied to account for the non-linear and non-uniform response of the CMS calorimeters [84]. They associate, on average, the p_T of a reconstructed jet to the p_T of the corresponding particle jet. Jet energy measured in the detector is typically different from the corresponding particle jet energy. The latter is obtained in the simulation by clustering, with the same jet algorithm, the stable particles produced during the hadronization process that follows the hard interaction. The main cause for this energy mismatch is the non-uniform and non-linear response of the CMS calorimeters. Furthermore, electronics noise and additional pp interactions in the same bunch crossing (event pile-up) can lead to extra unwanted energy. The purpose of the jet energy correction is to relate, on average, the energy measured in the detector to the energy of the corresponding particle jet.

CMS uses a factorized multi-level jet correction, shown schematically in Fig. 3.24, in which the correction must be applied in the following fixed sequence:

1. **L1:Offset**: Required correction for pile-up and electronic noise.
2. **L2:Relative (η)**: Required correction for variations in jet response with pseudorapidity relative to a control region.
3. **L3:Absolute (p_T)**: Required correction to particle level versus jet p_T in the control region.

4. **EMF:** Optional correction for variations in jet response with electromagnetic energy fraction.
5. **Flavor:** Optional correction to particle level for different types of jets (light quark, c, b, gluon).
6. **Underlying Event:** Optional correction for underlying event energy due to soft interactions involving spectator partons.
7. **Parton:** Optional correction to parton level.

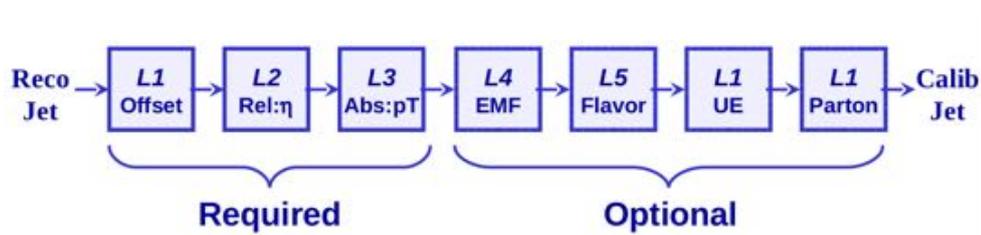


Figure 3.24: Schematic picture of the factorized multi-level jet correction.

Factorization facilitates the use of data-driven corrections, breaking the correction into pieces that are naturally measured in collider data. Combined correction brings back the jet to the particle level.

V-tagging

Generally, V tagging methods ($V=Z,W$ weak vector boson) have depended largely on leptonic decay channels. Hadronic signatures deal with the relatively poorer reconstruction of jets and large multijet backgrounds from QCD processes at hadron colliders. Several recent developments have improved the tagging of hadronically decaying weak vector bosons. Many of these advances have resulted from the analysis of the internal components of a jet, i.e. its substructure.

A more effective identification of hadronic V decays allows many analyses to profit from the substantially larger branching fraction of hadronic channels. This, in turn, may provide significant gains in searches for new physics.

Unresolved jets

For highly boosted weak vector bosons, the hadronic decay products can be merged into a single jet. For distance parameter $R = 0.8$, this occurs for boson p_T above 200 GeV [85]. The radiation profile of the individual hard partons within a merged jet must be explicitly resolved for an accurate calculation of the boson mass. This contrasts with the resolved scenario, for which the boson mass can be determined simply from the properties of the individually reconstructed jets. A new class of observables has been developed for disentangling the radiation profiles of proximate partons [29–31, 86]. Jet mass is the most natural discriminator between jets originating from V decays and those originating from single partons. Jet grooming techniques improve mass resolution by reducing the effects from pileup and underlying event. The following grooming algorithms were used:

N-subjettiness: This method [36] uses the distribution of jet constituents relative to the jet axis in order to quantify how well the jet can be divided into N subjets. The computation is done by reclustering the jet using the k_t -algorithm until N protojets are left. The direction of the remaining jets are then used to compute the "N-subjettiness" as

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \times \min(\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k}) \quad (3.5)$$

with the normalization factor d_0 :

$$d_0 = \sum_k p_{T,k} \times R_0 \quad (3.6)$$

and R_0 is the clustering parameter of the original jet, $p_{T,k}$ is the p_T of the k -th jet constituent and $\Delta R_{n,k}$ is its distance to the n -th subjet. In particular, the ratio of the "2-subjettiness" to the "1-subjettiness" ($\tau_2/\tau_1 = \tau_{21}$) has excellent capability at separating jets originating from boosted vector bosons from jets originating from quarks and gluons.

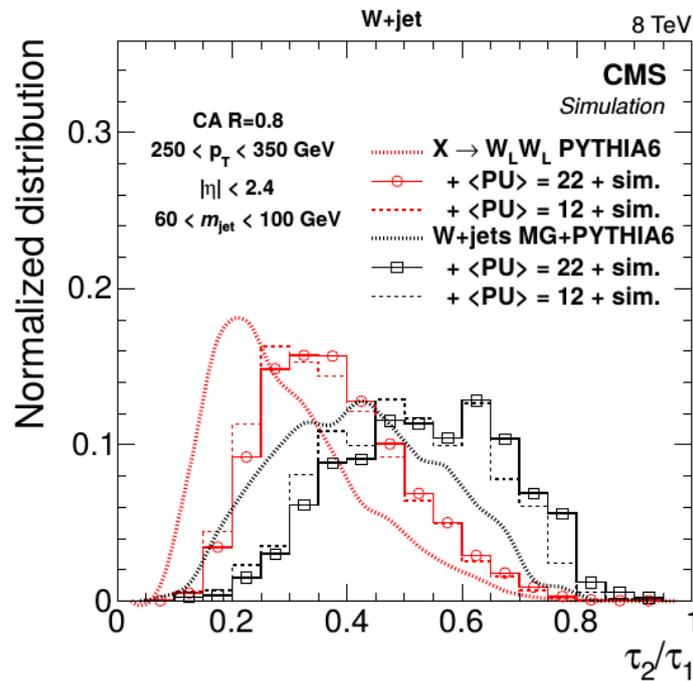


Figure 3.25: τ_{21} distribution characterising jet substructure in simulated samples of highly boosted and longitudinally polarized W bosons. Thick dashed lines represent the generator predictions without pileup interactions and without CMS detector simulation. The histograms are the expected distributions after full CMS simulation with pileup corresponding to an average number of 12 and 22 interactions. [86].

Pruning: This method [33] attempts to isolate subjet showers by removing soft, large angle particles from each subjet. Pruning will remove the uncorrelated contributions from underlying events and pile up that make significant contributions to the jet mass. The mass of the resulting pruned jet is small if we start with a QCD jet, and near the particle mass if we start with a jet containing the decay products of a heavy particle.

$$\Delta R_{ij} > r_{\text{cut}} \times \frac{2m}{p_T}, \quad z = \min \left(\frac{p_{T_i}, p_{T_j}}{p_{T_{i+j}}} \right) < z_{\text{cut}} \quad (3.7)$$

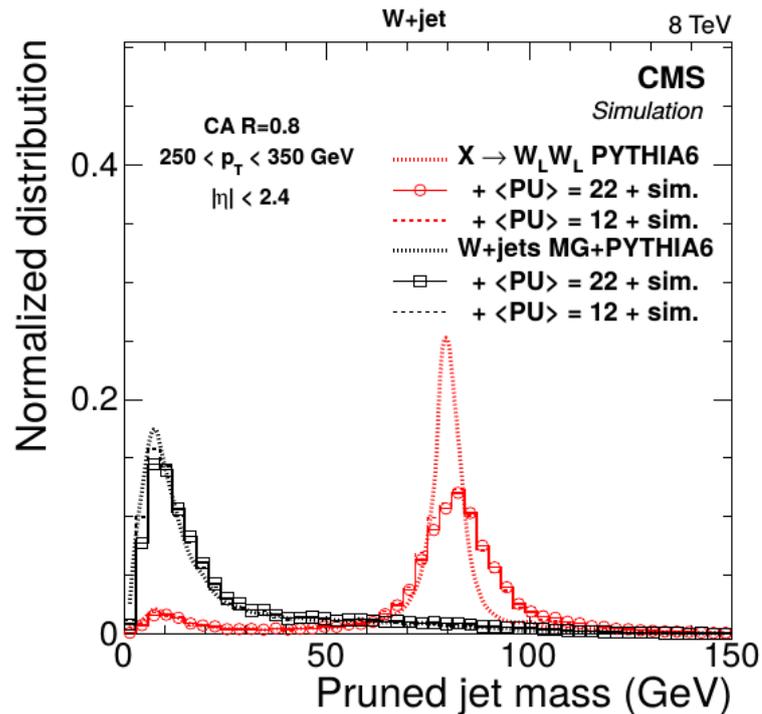


Figure 3.26: Jet pruned mass distribution in simulated samples of highly boosted and longitudinally polarized W bosons. The red lines represent a resonance decaying in two W bosons (signal) and the black lines represent a W+jets process (background) [86].

3.4.5 Missing Transverse Energy (\cancel{E}_T)

Missing transverse momentum plays a critical role in many physics analyses at the LHC. It is a key variable in many searches for physics beyond the standard model, such as extra dimensions and supersymmetry as well as for collider dark matter searches. Some neutral particles which interact weakly with matter (*i.e.* neutrinos) leave the detector without producing any direct response in the detector components. The presence of such particles (also called invisible particles, Fig. 3.27) must be implied from the imbalance of total momentum considering that the detector is hermetic. The vector momentum imbalance in the plane perpendicular to the beam direction is particularly useful in pp and $p\bar{p}$ colliders, and is known as missing transverse momentum, here denoted $\vec{\cancel{E}}_T$. Its magnitude is called missing transverse energy, and is denoted \cancel{E}_T (MET).

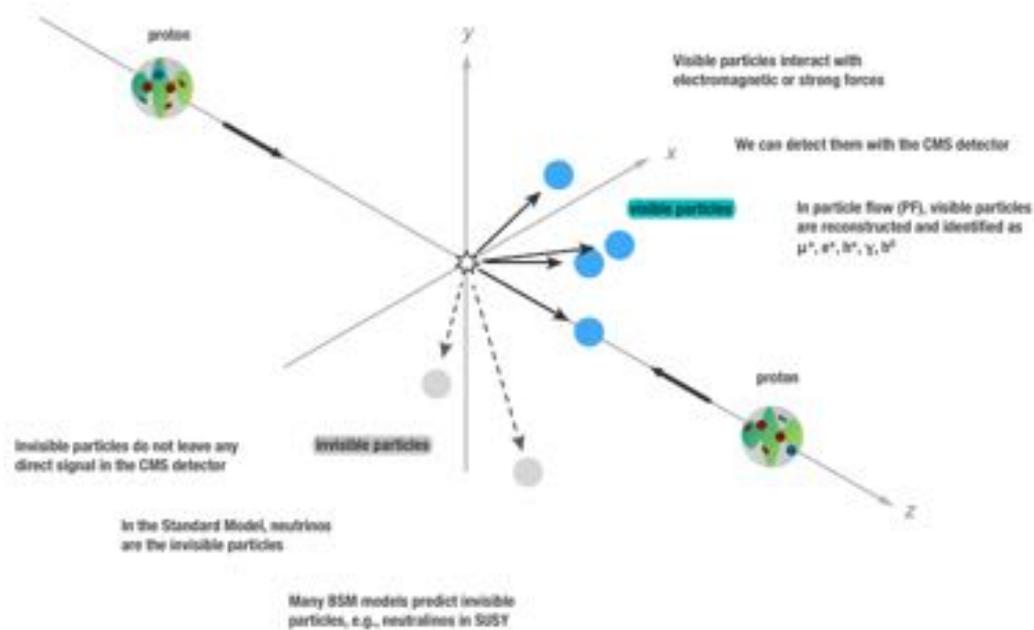


Figure 3.27: Invisible and visible particles

In general \vec{E}_T is calculated as the negative of the vector sum of the components of momentum transverse to the beam axis of all final-state particles reconstructed in the detector. CMS has developed three distinct algorithms to reconstruct \vec{E}_T . (a) Calo E_T based on calorimeter energies and calorimeter tower geometry, (b) TC E_T calculated by replacing the calorimeter tower energies matched to charged hadrons with their corresponding charged-track momenta, and (c) PF E_T calculated using a complete particle-flow technique. In this work we will focus on PF E_T .

The E_T reconstruction is sensitive to detector malfunctions and various reconstruction effects resulting in particle momentum mismeasurements and particle misidentification. Precise calibration of all physics objects is crucial for the E_T reconstruction, and E_T is particularly sensitive to multiple proton-proton interactions in the same, earlier, and later bunch crossings (pileup interactions). Thus, it is essential to study reconstruction in detail with data.

In Run-I, identifying and removing the causes of large fake E_T was the major challenge in E_T reconstruction at CMS. By the time Run-I ended in early 2013, it was developed a matured set of E_T filters to reject such large fake E_T . These filters are used in the event selections of many physics analyses. Below is a list of the main filters:

- CSC tight beam halo filter

- HBHE noise filter and with isolated noise rejection
- ECAL dead cell trigger primitive (TP) filter
- Bad EE Supercrystal filter

Some filters are used online in HLT. After the \cancel{E}_T filters are applied, the agreement of the \cancel{E}_T spectrum with simulation, significantly improves (Fig. 3.28)

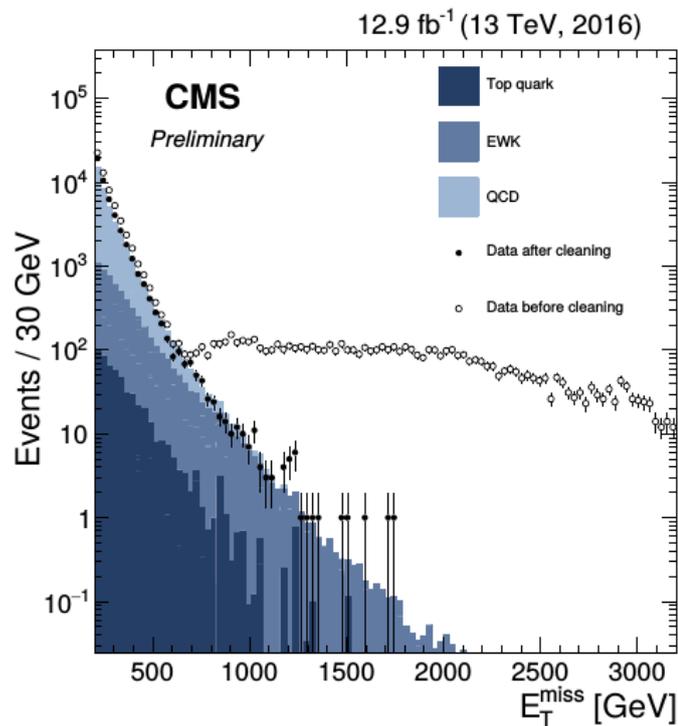


Figure 3.28: The E_T^{miss} distributions for events passing the dijet selection with the 2016 cleaning algorithms applied including the one based on jet identification requirements (filled markers), without the 2016 cleaning algorithms applied (open markers), and from simulation (filled histograms) [87].

Particle Flow MET (PF \cancel{E}_T)

The particle flow technique aims to reconstruct a complete, unique list of particles in each event using an optimal combination of information across all CMS subdetector systems. Particles which are reconstructed and identified include muons, electrons (with associated bremsstrahlung photons), photons (unconverted and converted), and charged and neutral hadrons. The PF \cancel{E}_T hereafter called MET is then simply the negative

vector sum of all such reconstructed particles in the event (Fig. 3.29).

$$\vec{E}_T = - \sum_{i \in \text{vis.}} \vec{p}_{T_i} \quad (3.8)$$

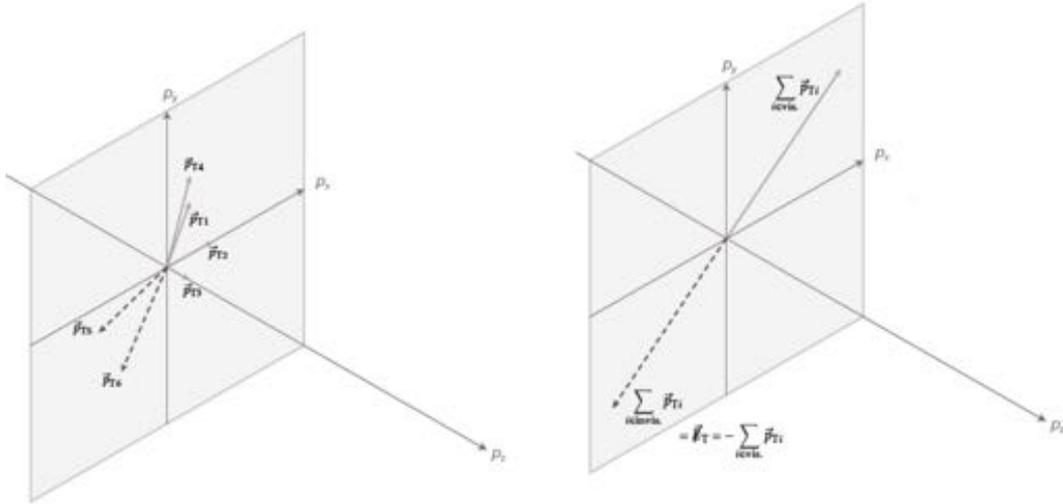


Figure 3.29: Vector sum of the reconstructed particles in the event.

The Type-I Correction

Raw MET is the negative of the vector sum of all reconstructed particles. The raw MET is systematically different from true MET, i.e., the transverse momentum carried by invisible particles, for many reasons including the non-compensating nature of the calorimeters and detector misalignment. To make MET a better estimate of true MET, corrections can be applied. The Type-I correction is the most popular MET correction in CMS. This correction is a propagation of the jet energy corrections (JEC) to MET. The Type-I correction replaces the vector sum of transverse momenta of particles which can be clustered as jets with the vector sum of the transverse momenta of the jets to which JEC is applied.

$$\vec{E}_T^{\text{raw}} = - \sum_{i \in \text{all}} \vec{p}_{T_i} \quad (3.9)$$

The particles can be classified into two disjoint sets: either clustered as jets or unclustered

$$\vec{E}_T^{\text{raw}} = - \sum_{i \in \text{jets}} \vec{p}_{T_i} - \sum_{i \in \text{uncl.}} \vec{p}_{T_i} \quad (3.10)$$

The vector sum of p_T of all particles clustered as jets is the same as the vector sum of p_T of all jets.

$$\sum_{i \in \text{jets}} \vec{p}_{T_{\text{jet}}}^{\text{raw}} = \sum_{i \in \text{jets}} \vec{p}_{T_i} \quad (3.11)$$

$$\vec{E}_T^{\text{raw}} = - \sum_{\text{jet}} \vec{p}_{T_{\text{jet}}}^{\text{raw}} - \sum_{i \in \text{uncl.}} \vec{p}_{T_i} \quad (3.12)$$

The Type-I correction replaces the raw jet p_T with the corrected jet p_T

$$\vec{C}_T^{\text{Type-I}} = \sum_{\text{jet}} \vec{p}_{T_{\text{jet}}}^{\text{raw}} - \sum_{\text{jet}} \vec{p}_{T_{\text{jet}}}^{\text{JEC}} \quad (3.13)$$

The Type-I correction is a vector term that can be added to raw MET

$$\vec{E}_T^{\text{Type-I}} = \vec{E}_T^{\text{raw}} + \vec{C}_T^{\text{Type-I}} \quad (3.14)$$

The Type-I corrected MET can be written as

$$\vec{E}_T^{\text{Type-I}} = - \sum_{\text{jet}} \vec{p}_{T_{\text{jet}}}^{\text{JEC}} - \sum_{i \in \text{uncl.}} \vec{p}_{T_i} \quad (3.15)$$

Transverse Mass

In this analysis we perform a search in the JET + MET final state as we will discuss in detail in the next chapter. The strategy is search for an excess related with the mass of the resonance. For that reason is necessary introduce a variable associated with this magnitude. Since the invisible particles are not directly detected in the experiment, it is difficult to reconstruct the mass of the resonance. Consider a single heavy particle of mass M which decays in a JET (labeled particle 1) and MET (labeled particle 2). The

mass of the parent particle can be constrained with the quantity M_T defined by:

$$\begin{aligned} M_T^2 &= (E_T(1) + E_T(2))^2 - (\vec{p}_T(1) + \vec{p}_T(2))^2 \\ &= E_T(1)^2 + E_T(2)^2 + 2E_T(1)E_T(2) - \vec{p}_T(1)^2 - \vec{p}_T(2)^2 - 2\vec{p}_T(1) \cdot \vec{p}_T(2) \end{aligned} \quad (3.16)$$

Considering that:

$$E_T^2 = m^2 + \vec{p}_T^2, \text{ and } m(1), m(2) \approx 0 \quad (3.17)$$

We obtain:

$$M_T^2 = 2 |\vec{p}_T(1)| |\vec{p}_T(2)| (1 - \cos \phi) \quad (3.18)$$

Remember that: $p_T(1) = p_T^{\text{jet}}$, and $E_T = p_T(2)$. Finally we get:

$$M_T = \sqrt{2 p_T^{\text{jet}} E_T [1 - \cos \Delta\phi(\text{jet}, E_T)]} \quad (3.19)$$

Chapter 4

Data Analysis

In this chapter we detail the analysis strategy used to search for a new heavy resonance decaying into dibosons VZ ($V=Z,W$) in the $\text{Jet} + E_{\text{T}}^{\text{miss}}$ final state. The experimental strategy is to reconstruct and identify the two bosons (VZ) and to combine their information into a variable (M_{VZ}^T) that can discriminate between signal and background and on which a statistical study can be performed. Therefore, the V-jet tagging procedure is essential in order to correctly identify the jet that arise from the V decay in the boosted regime. The signal of a new resonance X decaying to dibosons (VZ) is sought via a bump search in the in V-jet + $E_{\text{T}}^{\text{miss}}$ transverse mass. In order to perform a reliable comparison between data and simulation a data-driven strategy is adopt to estimate the dominant backgrounds. For the subdominant backgrounds the estimation is taken directly from simulation. In this analysis, we modeling the signal using the Bulk graviton and the HBT model B (W') as benchmarks models in the narrow-width approximation regime. This chapter is organized as follows. The trigger, the data and the simulation samples used by this study are described in section 4.1. The reconstruction and identification of the physics objects introduced in the previuos chapter are detailed in section 4.2. The event selection and the analysis efficiency is discussed in section 4.3. The modeling of the background and signal M_{VZ}^T distribution are studied in section 4.4 and section 4.5. The systematic uncertainties are summarized in section 4.6. Finally, section 4.7 shows the limit on the production cross section times branching ratio and the compatibility between data and the background-only hypothesis, quantified by the p-value.

4.1 Data Samples and Monte Carlo Simulation

4.1.1 Trigger and data samples

The data employed for this search were collected by the CMS experiment at $\sqrt{s} = 13$ TeV, and correspond to a total integrated luminosity of 2.3 fb^{-1} . Signal event candidates are recorded online using a trigger designed to select H_T^{miss} and E_T^{miss} with a lower threshold of 90 GeV in each case. The H_T^{miss} is computed as the magnitude of a vectorial sum of the transverse momenta of all jets with p_T greater than 20 GeV. To reject events arising from atypical detector performance, supplementary selection requirements are set on the jets used in the H_T^{miss} calculation. The E_T^{miss} is defined as the magnitude of the negative vectorial sum of the transverse momentum of all the particles identified at the trigger level. Identified muons are removed from the event before the E_T^{miss} and H_T^{miss} are calculated.¹ And additional support trigger was used in combination with the main signal trigger. This trigger select events that contains E_T^{miss} with a lower threshold of 170 GeV. Selected events are required to have $E_T^{\text{miss}} > 250$ GeV to guarantee a trigger efficiency greater than 98% for all events used in the analysis. The trigger paths are reported in the Table 4.1. For additional information about the trigger paths and the efficiency calculation, we refer the reader to the Appendix A.

Table 4.1: HLT Trigger path with their respective criteria.

Trigger path	Criteria
PFMETNoMu90_JetIdCleaned_PFMHTNoMu90_IDTight (unprescaled)	$E_T^{\text{miss}} > 90 \text{ GeV}, H_T^{\text{miss}} > 90 \text{ GeV}$
PFMET170_* (unprescaled)	$E_T^{\text{miss}} > 170 \text{ GeV}$

We use about 2.3 fb^{-1} of data collected during the Run2015C and Run2015D era (Table 4.2) and reconstructed with the CMSSW 76X release. We employ only lumisections that have been declared good for analysis by the central certification team.

Table 4.2: Data samples.

Sample	Number of events
MET/Run2015D	17996789
MET/Run2015C	106269

¹In CMS, the development of the triggers is quite generic, thought to be used by several analyzes. Removal of the muons from the indicated objects allow to use the same trigger to select $Z \rightarrow \mu + \mu$ events.

4.1.2 Simulated Samples

The analysis makes use of various simulated event samples for modeling the SM background and signal processes. Two benchmark models, the bulk graviton (spin-2) and the HVT (spin-1), are used to illustrate typical signal behavior. Simulated signal samples of bulk graviton resonances decaying to dibosons (ZZ) and subsequently to quarks and neutrinos were generated at leading order (LO) with the MADGRAPH5 [88] program interfaced with PYTHIA8 [89, 90] for the parton showering and hadronization, considering a coupling constant $\tilde{k} = 0.5$. For the HVT model inclusive $W' \rightarrow WZ$ samples are generated at LO with MADGRAPH5 and followed with PYTHIA8, for showering and fragmentation, with a relative resonance width of 0.1%. For both models we considered defined values of the resonances mass in the range $0.8 \leq m_X \leq 2$ TeV in steps of 100 GeV. We restrict the analysis to scenarios where the natural width of the resonance is sufficiently small to be neglected when compared to the detector resolution. This makes our modelling of the detector effects on the signal shape independent of the actual model used for generating the events. The signal samples used in the analysis are shown in Table 4.3.

Simulated samples were produced for the Z+jets, W+jets, $t\bar{t}$, dibosons and QCD multijet processes in order to describe the contribution expected from SM backgrounds. The main components of the total background are represented by Z + jets ($Z \rightarrow \nu\bar{\nu}$) and W+jets ($W \rightarrow \ell\nu$) production. These as well as the QCD multijets sample, are simulated with MADGRAPH5 in LO mode and matched to PYTHIA8 using the CUETP8M1 tune for hadronization and fragmentation. Double counting by the matrix element calculation and parton showering is resolved by using the MLM matching prescription [91]. The SM background contribution from $t\bar{t}$ events was modeled at next-to-leading order (NLO) with the aMC@NLO program [92], interfaced with PYTHIA8. Inclusive non-resonant dibosons simulated samples (WW/WZ/ZZ) were generated at LO with PYTHIA8. All the background samples used in this analysis are listed in Table 4.4.

The V+jets simulated samples were rescaled using next-to-next-to-leading order (NNLO) QCD correction in the cross sections (Table 4.5) and NLO QCD electroweak (EW) correction in the V boson p_T domain (Fig. 4.1) [93].

Minimum bias events were included during the production of the simulated samples to

account for contributions from additional proton-proton collisions (pileup), with the number of reconstructed primary vertices matching those in data. The simulation is corrected from perceptible differences between data and simulation in the trigger and identification/isolation efficiency of leptons (electrons, muons, taus), photons and jets originating from hadronization of bottom quarks (b-jets).

In all the simulated samples, the events were generated using the NNPDF 3.0 [94] set of parton distribution functions. The simulation of the detector response was modeled with the GEANT4 package [95].

Table 4.3: Monte Carlo simulated signal samples for Bulk graviton ($\tilde{k} = 0.5$) and HVT(B) models.

Sample	Cross Section (pb)	Nevents
BulkGravToZZToZhadZinv_narrow_M-800_13TeV	0.065	100000
BulkGravToZZToZhadZinv_narrow_M-1000_13TeV	0.017	99200
BulkGravToZZToZhadZinv_narrow_M-1200_13TeV	0.006	95800
BulkGravToZZToZhadZinv_narrow_M-1400_13TeV	0.0023	100000
BulkGravToZZToZhadZinv_narrow_M-1600_13TeV	0.002048662	100000
BulkGravToZZToZhadZinv_narrow_M-1800_13TeV	0.00065516	98000
BulkGravToZZToZhadZinv_narrow_M-2000_13TeV	0.000324162	99200
WprimeToWZ_width0p1_M-800_TuneCUETP8M1_13TeV	0.6807755239	49200
WprimeToWZ_width0p1_M-1200_TuneCUETP8M1_13TeV8	0.2503529114	50000
WprimeToWZ_width0p1_M-2000_TuneCUETP8M1_13TeV	0.1389695229	49200

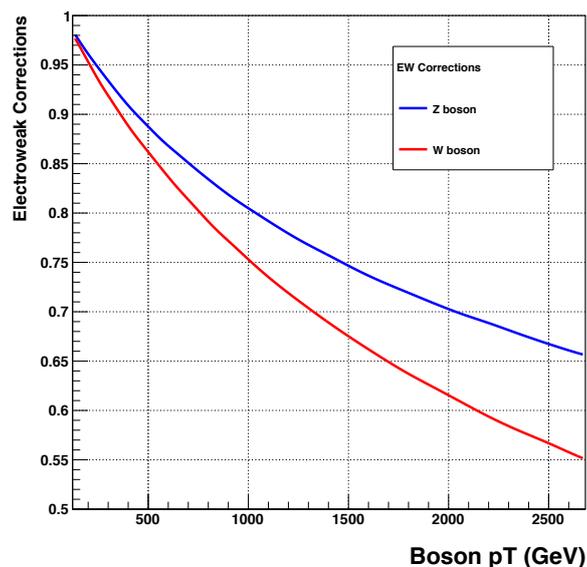


Figure 4.1: Electroweak corrections for the W/Z boson in function of the boson p_T .

Table 4.4: Monte Carlo background samples.

Sample	Cross Section (pb)	N_{events}
$Z(\rightarrow \nu\bar{\nu})+\text{jets}, 100 < H_T < 200 \text{ GeV}$	280.35	5240199
$Z(\rightarrow \nu\bar{\nu})+\text{jets}, 200 < H_T < 400 \text{ GeV}$	77.67	5135542
$Z(\rightarrow \nu\bar{\nu})+\text{jets}, 400 < H_T < 600 \text{ GeV}$	10.73	954435
$Z(\rightarrow \nu\bar{\nu})+\text{jets}, H_T > 600 \text{ GeV}$	4.116	1033818
$W(\rightarrow \ell\nu)+\text{jets}, 100 < H_T < 200 \text{ GeV}$	1345	10205377
$W(\rightarrow \ell\nu)+\text{jets}, 200 < H_T < 400 \text{ GeV}$	359.7	4949568
$W(\rightarrow \ell\nu)+\text{jets}, 400 < H_T < 600 \text{ GeV}$	48.91	1943664
$W(\rightarrow \ell\nu)+\text{jets}, H_T > 600 \text{ GeV}$	18.77	1041358
QCD multijets, $100 < H_T < 200 \text{ GeV}$	27990000	82095800
QCD multijets, $200 < H_T < 300 \text{ GeV}$	1712000	18784379
QCD multijets, $300 < H_T < 500 \text{ GeV}$	347700	16909004
QCD multijets, $500 < H_T < 700 \text{ GeV}$	32100	19665695
QCD multijets, $700 < H_T < 1000 \text{ GeV}$	6831	15547962
QCD multijets, $1000 < H_T < 1500 \text{ GeV}$	1207	5049267
QCD multijets, $1500 < H_T < 2000 \text{ GeV}$	119.9	3939077
QCD multijets, $H_T > 2000 \text{ GeV}$	25.24	1981228
$t\bar{t}$	831.76	38475776
$t\bar{t}$ (Extension)	831.76	196937036
WW	118.7	988418
WZ	47.13	985600
ZZ	16.523	996944

Table 4.5: NNLO QCD flat scale factors.

Sample	Scale factor
Z + jets	1.23
W + jets	1.21

4.2 Object Reconstruction and Identification

4.2.1 Primary Vertices

In general, many primary vertices are reconstructed in an event because of the pileup contributions. In order to identify the vertex related to the main proton-proton collision in the analyzed event, we required that:

- Number of degrees of freedom: $n_{\text{dof}} > 4$
- longitudinal coordinate: $|z_0| < 24$ cm
- Transverse position: $d_0 < 2$ cm

If more than one vertex pass the previous conditions, we choose the one with the highest sum of transverse momenta $\sum p_T^2$ of the *tracks* associated to it. Due to disagreements in the number of primary vertices distribution between the MC samples and data a reweight procedure was implemented assuming a total inelastic cross section of $\sigma = 69\text{mb}$. Figure 4.2 (Left) shows the comparison for the PU in data and MC (Right) the distribution of the number of primary vertices after apply the reweight procedure.

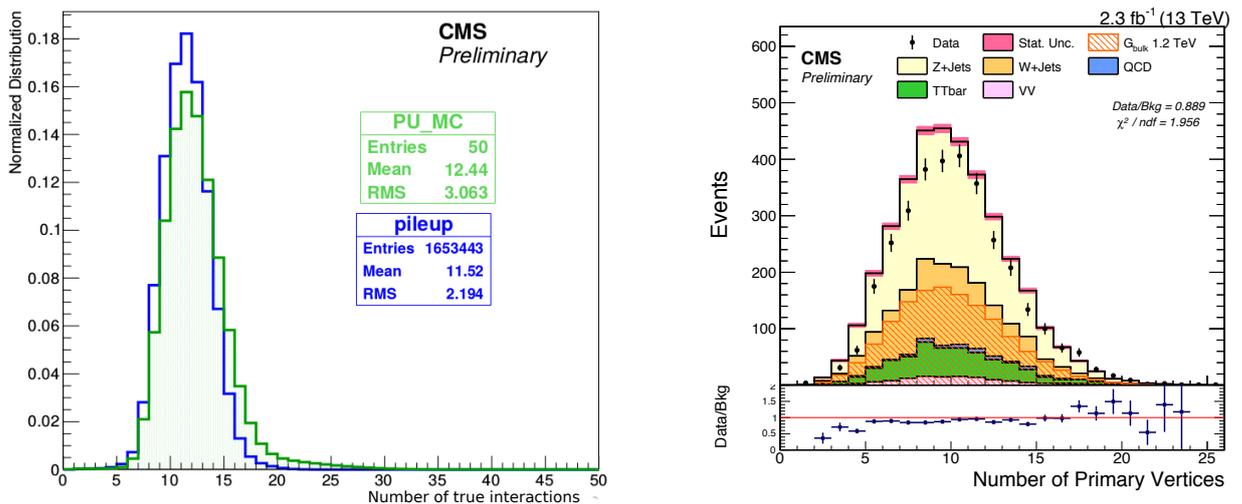


Figure 4.2: Left: Comparison between the PU profile in data (blue) and the Poisson density function in MC (green). Right: Number of primary vertices distribution after the reweight procedure in events that pass the final analysis selection.

4.2.2 Missing Transverse Energy

The raw missing transverse energy vector is computed as the negative vector sum of the transverse momenta of all particles reconstructed in the event, with magnitude denoted by E_T^{miss} . Corrections to the momenta of jets reconstructed in the event are further propagated to the E_T^{miss} (Type-1 corrections) [87]. Figure 4.3 show a comparison between the raw PF E_T^{miss} and the Type-1 PF E_T^{miss} .

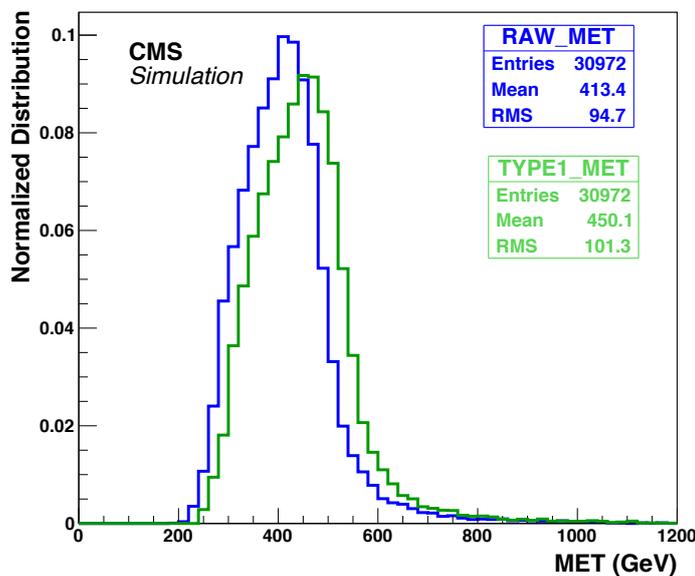


Figure 4.3: Comparison between Raw PF E_T^{miss} and Type-1 PF E_T^{miss} corresponding to a signal sample of 1 TeV.

A set of dedicated quality filters are applied in data and simulation to remove events with a large misreconstructed E_T^{miss} originated from detector noise and beam backgrounds [87]:

- **HBHENoiseFilter:** CMS has been observed anomalous signals in the Hadron Barrel (HB) and in the Hadron Endcap (HE) of the Hadronic calorimeter (HCAL). The fake energy can arise due to electronic noise in the hybrid photodiodes (HPDs) and in the readout boxes (RBX). These sources have been studied extensively in the data collected in the 2010 to 2012 LHC running period [96, 97]. The HBHENoiseFilter is used to remove events if many rechits are flagged.
- **HBHENoiseIsoFilter:** Similar than HBHENoiseFilter but with isolated noise rejection. The isolation-based noise filter utilizes a topological algorithm where energy

deposits in HCAL and ECAL as well as the tracking measurements are combined and compared in order to identify isolated instances of noise activity in HBHE.

- **CSCTightHalo2015Filter:** As the proton bunches circle the LHC, proton interactions with the residual gas particles can occur, producing showers of secondary particles which can interact with the CMS detector. This phenomenon is known as beam halo and can be a formidable machine-induced-background if the rates are high enough. Some observables are particularly vulnerable to beam halo backgrounds, such as E_T^{miss} . The CSCTightHalo2015Filter filter is used in order to reject events that present the above features.
- **EcalDeadCellTriggerPrimitiveFilter:** ECAL endcap (EE) supercrystals have been identified and masked which give anomalously high energies (TeV regime) due to producing high amplitude anomalous pulses, in several channels at once. Furthermore in a limited number of ECAL towers, the crystal by crystal precision readout is not available. For the cases where the precision readout is not available, the trigger primitive (TP) information is available, and can be used to estimate the magnitude of the visible energy “lost”. However, the TP information saturates above 127.5 GeV. If the TP value is found to be close to the saturation value for the “lost” channels, the event is filtered out.
- **eeBadScFilter:** In the ECAL, much of the electronics noise and spurious signals from particle interactions with the photodetectors is removed at the reconstruction based on ECAL information only (local reconstruction) with topological and timing based selection. The remaining effects that lead to high- E_T^{miss} signatures such as anomalously high energy supercrystals (cluster of clusters extended in ϕ to recover the spread energy), and the lack of information for channels due to not properly functioning precision readout are removed through dedicated event filters.

Figure 4.4 shows the efficiency for the events that pass the HBHENoiseFilter for signal sample of 1 TeV. Events containing a minimum E_T^{miss} of 250 GeV are required in the analysis in order to settle in the plateau of the trigger turn-on curve (Fig. A.1). Further corrections are applied to the E_T^{miss} in the V+jets simulated samples, based on the hadronic recoil information derived from Z+jets ($Z \rightarrow \ell\ell$) events in data and simulation.

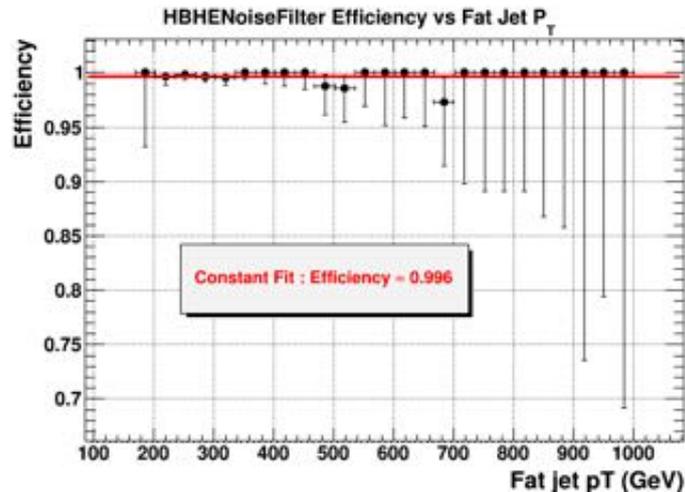


Figure 4.4: HBHENoiseFilter efficiency vs AK8 jet p_T in simulation for a signal sample of 1 TeV.

Recoil Correction

We use a data-driven method to model the V boson recoil (response and resolution) in order to improve the description of the missing energy in V+jets MC events. The term “recoil” here means the hadronic activity that balances the p_T of the boson. To derive the recoil correction we use a Z+jets ($Z \rightarrow \mu\mu$) process in data and simulation, fitting the response and resolution of the recoil as a function of Z p_T . The advantage of using the Z+jets process is that the Z boson can be selected without significant background and the p_T can be accurately reconstructed in data from the two final state leptons. The transverse recoil vector is defined as:

$$\vec{u}_T = -\vec{E}_T^{\text{miss}} - \sum_i \vec{\ell}_i \quad (4.1)$$

where $\vec{\ell}_i$ is the momentum of the lepton in which the Z decays. Figure 4.5 shows the kinematics of the process in the transverse plane. The method parametrizes the recoil in the parallel and perpendicular directions of the boson p_T , fitting these variables with a double gaussian model in different bins of the Z p_T . From the fits one can extract the mean and the σ of the gaussians, using different polynomial functions to fit these values and extract the response and resolution curves. After this process some scale factors are derived and applied to the V+jets simulated samples. Complementary information about the recoil method is reported in the Appendix B.

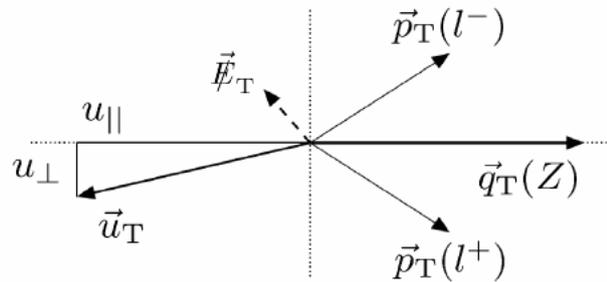


Figure 4.5: $Z \rightarrow \ell\ell$ event kinematics in the transverse plane. The transverse recoil vector \vec{u}_T is split into parallel and perpendicular components to the direction of the boson p_T .

The transverse missing energy distribution before and after applying the recoil corrections is shown in Figure 4.6 .

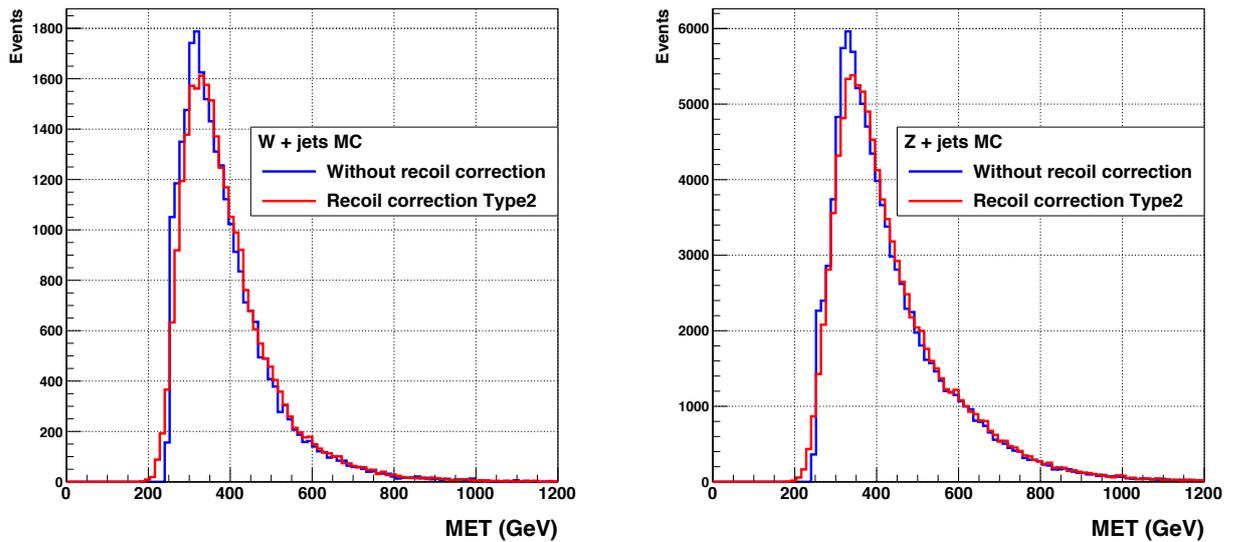


Figure 4.6: E_T^{miss} distribution before and after applying the recoil correction for W +jets and Z + jets samples.

4.2.3 Jets

Jets are reconstructed using the PF technique. Charged hadrons not originating from the primary vertex are discarded in a process called “charged hadron subtraction” (CHS). The resulting list of particles are used as input to the anti- k_T jet clustering algorithm with a distance parameter R , implemented in the FastJet package. It is applied to the

jets a technique based on jet areas that provides jet-by-jet corrections for pileup and underlying-event effects. Jet energies are further corrected using p_T and η dependent correction factors. These corrections are derived from MC simulation and are supplemented by residual corrections from dijet and photon+jet events in data. Figure 4.7 shows the comparison of the jet mass distributions before and after apply the JEC. As it can be observed the corrected distribution shows a peak close to the Z mass (91 GeV).

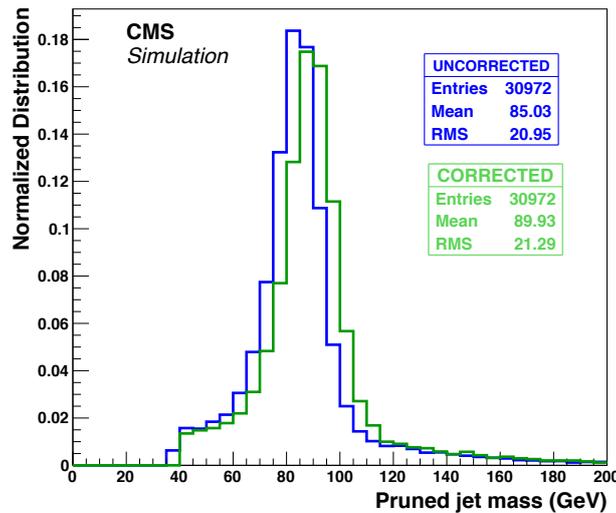


Figure 4.7: Comparison between corrected and uncorrected jets (JEC) for the jet mass distribution corresponding to a graviton signal sample of 1 TeV ($X \rightarrow ZZ$).

Loose jet identification criteria are applied to remove spurious jet-like features associated with calorimeter noise. To suppress additional instrumental and beam-related backgrounds, events are rejected if less than 10% of the energy of the highest p_T jet (leading jet) is carried by charged hadrons, or if more than 80% of this energy is carried by neutral hadrons. Figure 4.8 shows the charged hadron fraction (CHF) and neutral hadron fraction (NHF) distributions for events obtained with the full analysis selection after the cleaning cuts were applied.

In addition, jet energy resolution smearing factors reported in Table 4.6 were applied in the simulation aiming to reduce the difference between Data and MC.

To identify the hadronic decays of boosted V bosons, jets are clustered using the anti- k_T algorithm with a distance parameter $R = 0.8$ namely “AK8 jets”. The leading AK8 jet is required to be inside the tracker acceptance ($|\eta| < 2.4$) and to have $p_T > 200$ GeV. The AK8 jet with the highest p_T is associated to the $V \rightarrow q\bar{q}'$ candidate where the two

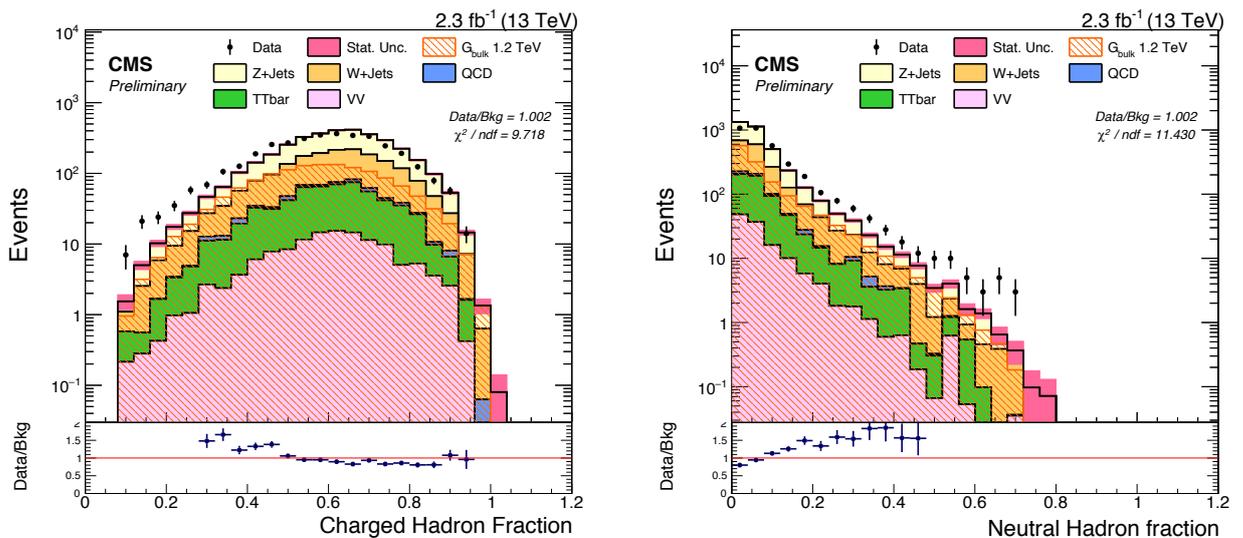


Figure 4.8: Jet hadronic energy fractions after the cleaning cuts.

Table 4.6: Jet energy resolution scaling factors and uncertainty.

$ \eta $ region	Data/MC SF
0.0-0.5	1.095 ± 0.018
0.5-0.8	1.120 ± 0.028
0.8-1.1	1.097 ± 0.017
1.1-1.3	1.103 ± 0.033
1.3-1.7	1.118 ± 0.014
1.7-1.9	1.100 ± 0.033
1.9-2.1	1.162 ± 0.044
2.1-2.3	1.160 ± 0.048
2.3-2.5	1.161 ± 0.060
2.5-2.8	1.209 ± 0.059
2.8-3.0	1.564 ± 0.321
3.0-3.2	1.384 ± 0.033
3.2-5.0	1.216 ± 0.050

quarks are merged to the same V-jet. In addition, another collection of jets clustered with the anti- k_T algorithm with a distance parameter $R = 0.4$, called “AK4 jets”, is used primarily for vetoing the presence of b-jets. The AK4 jets are required to have p_T larger than 30 GeV and $|\eta| < 2.4$. Figure 4.9 shows the ak4jets multiplicity in events that pass the final analysis selection.

To identify b-jets, the medium working point of the inclusive combined secondary vertex b-tagging algorithm (CSV) is applied to the reconstructed AK4 jets. We also required the b-jets to be spatially separated from the AK8 jets by at least $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.8$, where $\Delta\eta$ and $\Delta\phi$ are differences between the b-jet and the AK8 jet directions

4.2.4 Hadronic V identification using jet substructure

The decay of heavy resonances in dibosons $X \rightarrow VZ$ produce objects with very high p_T . When the boost of the V-boson is large enough i.e. $p_T > 200$ GeV, the final state hadrons from the decay $V \rightarrow q\bar{q}'$ merge into a single jet. In those cases, the traditional techniques relying on resolved jets are no longer applicable. However, jet substructure methods can be used to identify those jets arising from decays of W, Z or H bosons. In this analysis, a ‘‘V-tagging’’ technique based on two jet substructure methods, pruning and N-subjettiness, is used in order to discriminate between jets arising from V-decays and those from QCD backgrounds.

The leading AK8 jet in the event is considered a V-jet candidate if its pruned mass (m_{jet}), computed from the sum of the four-momenta of the constituents surviving the pruning, falls in the range $65 \leq m_{\text{jet}} \leq 105$ GeV. Jets coming from hadronic V decays in signal events are characterized by lower values of τ_{21} compared to the SM background. To optimize the analysis sensitivity, we distinguish two samples of data:

- High Purity (HP) : $\tau_{21} < 0.45$
- Low Purity (LP) : $0.45 < \tau_{21} < 0.75$

Events with $\tau_{21} > 0.75$ are rejected due to very low signal efficiency. The combined use of the pruned mass and the τ_{21} allows us to do the V-tagging of a jet with different degrees of purity. The p_T and τ_{21} distributions of the highest AK8 jet in the event for data, SM backgrounds, and a bulk graviton signal sample are shown in Fig. 4.11 after applying a $65 \leq m_{\text{jet}} \leq 105$ GeV requirement.

The discrepancies between data and simulation in the jet substructure variables m_{jet} and τ_{21} are corrected using scale factors for V-tagging efficiency. A sample of high- p_T W bosons, which decay hadronically and are reconstructed as a single AK8 jet, was studied in $t\bar{t}$ and single top-quark events. Scale factors for the τ_{21} selection efficiency are extracted following the method described in Ref. [98]. A simultaneous fit to the jet mass distributions, in both data and simulation, before and after the m_{jet} and τ_{21} requirements, is performed to separate the W-signal from the combinatorial components in the top-enriched sample. The fit results are used to extract data-to-simulation efficiency scale factors to identify an isolated hadronic W boson. The scale factors are reported in Table

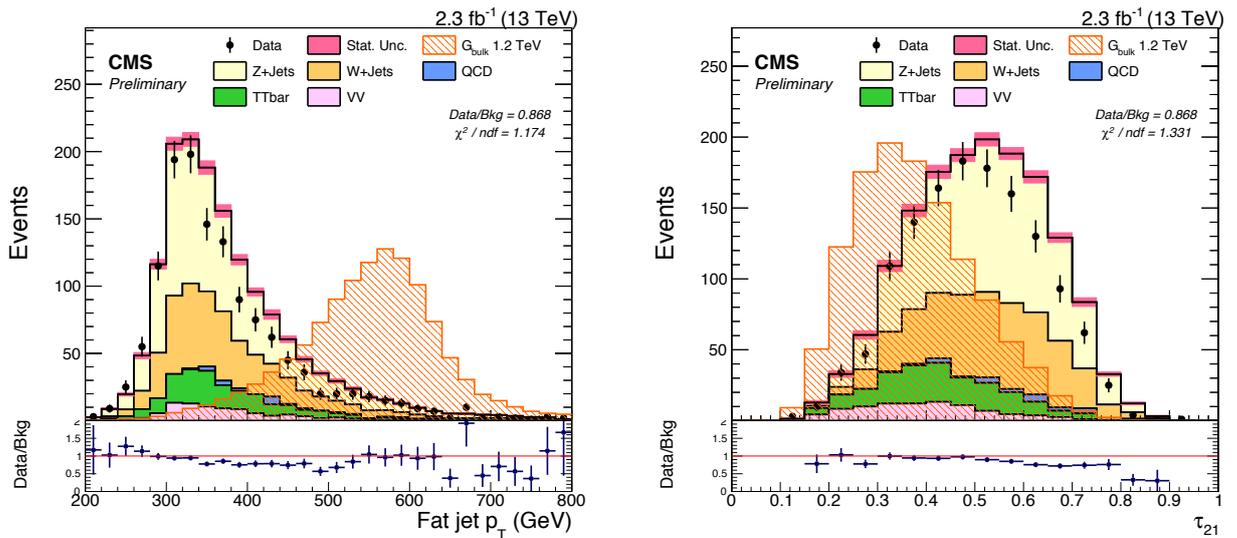


Figure 4.11: Distribution of p_T (left) and τ_{21} (right) for the leading AK8 jet in events passing the final selection in the pruned jet mass signal region for data, SM backgrounds and for a bulk graviton signal sample (1.2 TeV resonance mass and $\tilde{k} = 0.5$).

4.7 and are used to correct the total signal efficiency. The uncertainty on those scale factors is then assigned as systematic uncertainty of the method.

Table 4.7: Scale factors for the τ_{21} efficiency selection derived from data and simulation in a top-quark enriched sample.

τ_{21} Selection	Efficiency scale factor
$\tau_{21} < 0.45$	0.942 ± 0.063
$0.45 < \tau_{21} < 0.75$	1.268 ± 0.332

4.2.5 Muons

Muons candidates are reconstructed by associating track measurements in the inner tracker and in the muon system. A set of requirements based on the impact parameter of the track and on the number of hits recorded in the silicon tracker and in the muon chambers have to be fulfilled in order to identify loose muons [99, 100]. The resulting muons are required to have $p_T > 10$ GeV and $|\eta| < 2.4$. To reject nonprompt or misidentified leptons, requirements are imposed on the isolation criteria, based on the sum of deposited energies. The relative isolation parameter (RelIso) is defined as the contributions from the total transverse momentum of all charged hadrons, the transverse energies (E_T) of all photons, and the E_T of all neutral hadrons reconstructed

by the PF algorithm within a cone of radius $\Delta R < 0.4$ centered on the muon track direction, divided by the muon track p_T . Identified muons with RelIso values below 0.25 are considered isolated and used in the analysis.

In order to reduce electroweak backgrounds (W+jets), we reweight the events that contain identified muons. Some scale factors (SF) were derived centrally to improve the agreement between data and MC due to identification and Isolation of the muons. The value of the SFs depends on the kinematic properties of the muons (p_T and η). We use the SFs to define a weight called *muonWeight* as:

$$\text{muonWeight} = 1 - \text{SF} \quad (4.2)$$

The *muonWeight* will be applied only in the events that contain muons. The value of the *muonWeight* will ranging between 0 and 1.

An association between the reconstructed and the generated muons was performed using a geometrical matching with a cone of $\Delta R(\text{recoMuon}, \text{genMuon}) = 0.1$. To resolve ambiguities in the matching, among all the possible combinations, we choose the minimum ΔR in the calculation. In Figure 4.12 we show the minimum ΔR between the reconstructed and generated muons for a MC signal sample of 1 TeV after applying the final selection of the analysis. The application of the SF is based on the kinematic information (p_T, η) of the reconstructed muons that pass the MC matching, taking into account the generator level information of the muons in this process.

4.2.6 Electrons

Electrons candidates are reconstructed by associating a charged particle track with an ECAL supercluster, including energy depositions from final-state radiation [101]. The resulting electron candidates are required to have $p_T > 10$ GeV, $|\eta| < 2.4$, and to satisfy identification criteria designed to remove photon conversions, jets misidentified as electrons, and electrons from semileptonic decays of bottom and charm quarks. Identified electrons with RelIso values below 0.1 are considered isolated and used in the analysis. Events with identified electrons are vetoed to reduce W+jets background. SF's were derived centrally to improve the agreement between data and MC due to

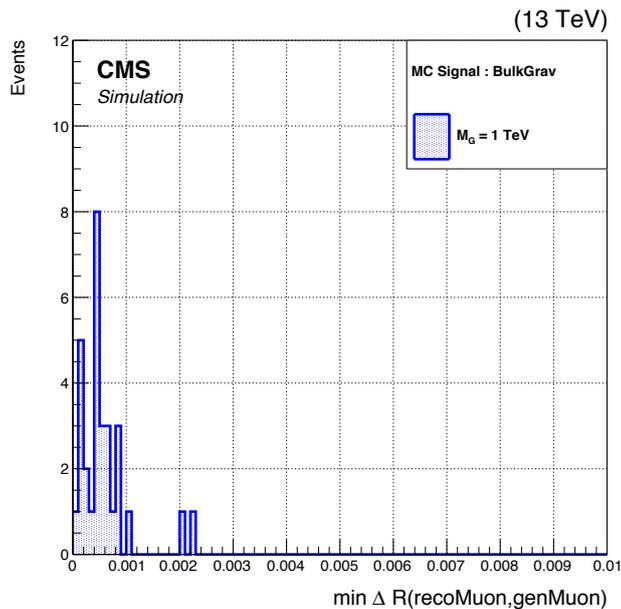


Figure 4.12: Minimum ΔR between reconstructed and generated muons for a signal mass point of 1 TeV.

identification and Isolation of the electrons. A MC matching was implemented in the reconstructed electrons aiming to apply the SF, taking into account the generator level information of the electrons.

4.2.7 Taus

The hadronic tau (τ_h) decays are reconstructed and identified [102] using the hadrons-plus-strips (HPS) algorithm. The algorithm is designed to reconstruct individual decay modes of the tau lepton, taking advantage of the PF algorithm. It also discriminates τ_h decays from quark and gluon jets, and from electrons and muons. In addition, tau-isolation requirements complement the identification process. Identified taus are required to have $p_T > 20$ GeV and $|\eta| < 2.3$ to be used in the analysis. Events with identified taus are removed in order to diminish W +jets background. We consider for the tau ID efficiency, the ratio between data and MC equal to 1, with an uncertainty of 6%. This recommendation is valid for all isolation discriminators, p_T and η range, in Run-2 analyses.

4.2.8 Photons

Photons candidates are reconstructed by clustering spatially correlated energy deposits in the ECAL [103]. Photon identification is based on two main categories of observables: shower-shape and isolation variables which help to discriminate among signal photons and photons that arise from neutral mesons ($\pi^0 \rightarrow \gamma\gamma$) produced in jets or electrons misidentified as photons. In order to reduce $Z + \gamma$ +jets and $W + \gamma$ +jets backgrounds we reject events that contain photons. Photons candidates are required to have a minimum p_T of 15 GeV and $|\eta| < 2.5$. With the objective to identify photons we use a loose cut-based working point identification. SFs were derived centrally to improve the agreement between data and MC due to identification and Isolation of the photons. We applied those scale factor over the MC samples to reweight the event.

4.2.9 Transverse mass

The hadronic V-boson candidate ($V \rightarrow q\bar{q}'$) is reconstructed from the AK8 jet with the highest p_T . Due to the invisible decay of the Z boson ($Z \rightarrow \nu\bar{\nu}$), the reconstruction of the resonance mass is not directly viable and its total momentum can be constrained only in the plane transverse to the beam direction. Therefore, we will use as the final discriminant a quantity called “transverse mass”, defined by

$$M_{VZ}^T = \sqrt{2p_T E_T^{\text{miss}} (1 - \cos(\Delta\phi))}$$

where p_T is the transverse momentum of the AK8 Jet and $\Delta\phi$ is the angle between the jet p_T and the E_T^{miss} vector.

4.3 Event Selection

The data used in this analysis was collected using a trigger designed to record events that contain large values of H_T^{miss} and E_T^{miss} . The details of the trigger were explained in the chapter 4.1. Offline, all events are required to have at least one primary vertex reconstructed within a 24 cm window along the beam axis, with a transverse distance from the nominal pp interaction region of less than 2 cm. In the presence of more than one vertex passing these requirements, the primary-event vertex is chosen to be the one with the highest total p_T^2 , summed over all the associated tracks.

In the signal region we selected events that contain highly energetic AK8 jets and large missing transverse energy. An event falls into this category if the AK8 jets have $p_T > 200$ GeV, $|\eta| < 2.4$, $65 \leq m_{\text{jet}} \leq 105$ GeV and $\tau_{21} < 0.75$. For the missing energy, we require $E_T^{\text{miss}} > 250$ GeV. If the selected event contains more than one AK8 jet with the previous requirements, the jet with the largest p_T is chosen in order to calculate the transverse mass variable. Also, a lower threshold of 600 GeV in the transverse mass is required in the selected events. Below this value, a new resonance would not be massive enough to produce a large fraction of boosted V bosons, resulting in a low value of the selection efficiency.

Background from leptonic W boson decays is reduced through rejection of events with isolated leptons (electrons, muons, taus) identified with loose selection criteria. Events containing b-jets are vetoed in the analysis in order to suppress $t\bar{t}$ background. Events containing identified photons are reject to reduce $V + \gamma + \text{jets}$ backgrounds. To suppress QCD multijet background in which large E_T^{miss} could arise, the minimum azimuthal angle between the E_T^{miss} direction and the AK4 jets with p_T greater than 30 GeV is required to be greater than 0.5. Table 4.8 summarize the final selection chosen in the analysis.

In order to set object disambiguation between possible VZ candidates, we use as a selection criteria the candidate with the highest transverse momentum in the event.

The signal efficiency in each category (ε_{cat}) is defined as the ratio between the number of signal events after the whole analysis selection in a chosen category ($N_{\text{sel}}^{\text{cat}}$) over the total number of generated signal events (N_{gen}).

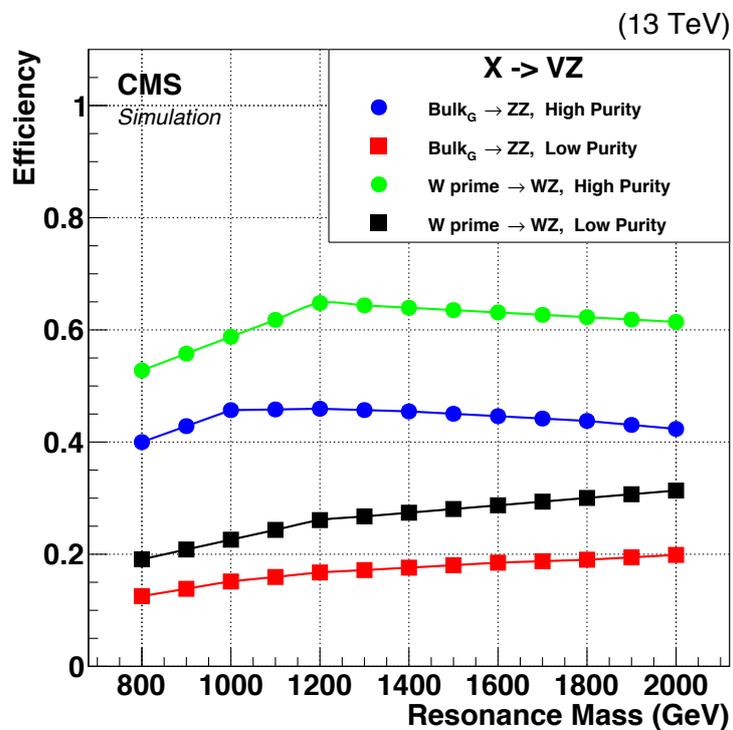
$$\varepsilon_{\text{cat}} = \frac{N_{\text{sel}}^{\text{cat}}}{N_{\text{gen}}}, \quad \text{cat} = \text{HP,LP} \quad (4.3)$$

The signal efficiencies are evaluated for the full analysis selection inside the VZ-enriched ($65 \leq m_{\text{jet}} \leq 105$ GeV) mass windows, and are shown in Figure 4.13 for the bulk graviton and W' signal models for several mass points and different categories. The high purity (HP) efficiency drops at high values of the resonance mass due to the inefficiency of the N-subjettines and jet mass selection; this is partially recovered in the low purity category (LP).

Figure 4.14 shows a set of comparison plots between data and simulation for interesting variables in the analysis. The final analysis selection except for a loose requirement in the mass of the jet $40 \leq m_{\text{jet}} \leq 220$ GeV was used to produce the plots in order to observe regions that will be employed for background estimation (sideband regions). In the figure, the simulation is composed by different backgrounds: Z+jets(light yellow), W+jets(orange), $t\bar{t}$ (green), QCD multijets(light blue) and diboson(light pink). The statistical uncertainty is shown in dark pink. In addition a Bulk graviton signal sample of 1.2 TeV is shown by the orange hatch region. The figure also show the ratio of the histograms (data/simulation) and the χ^2/ndf value as a figure of merit. In general a good agreement between data and simulation is observed in the control plots.

Table 4.8: Final analysis selection.

Selection	Value
High Level Trigger	PFMETNoMu90_JetIDCleaned_PFMHTNoMu90_IDTight OR HLT_PFMET170_*
E_T^{miss}	Type-I PF MET $p_T > 250$ GeV
Jets	AK8Jets $p_T > 200$ GeV, $ \eta < 2.4$ $65 < m_{\text{jet}}^{\text{pruned}} < 105$ GeV CHF > 0.1 , NHF < 0.8
	AK4Jets PFJetID Loose $p_T > 30$ GeV, $ \eta < 2.4$ CHF > 0.1 , NHF < 0.8 $\min\Delta\phi(\text{AK4Jet}, E_T^{\text{miss}}) > 0.5$ b-tag Veto : CSV > 0.8
Leptons (electrons, muons, taus)	Veto
Photons	Veto
VZ candidate transverse mass	$M_{VZ}^T > 600$ GeV

Figure 4.13: Signal efficiency for $G_{Bulk} \rightarrow ZZ$ and $W' \rightarrow WZ$ in the jet + E_T^{miss} final state for different mass points and different categories after the final analysis selection.

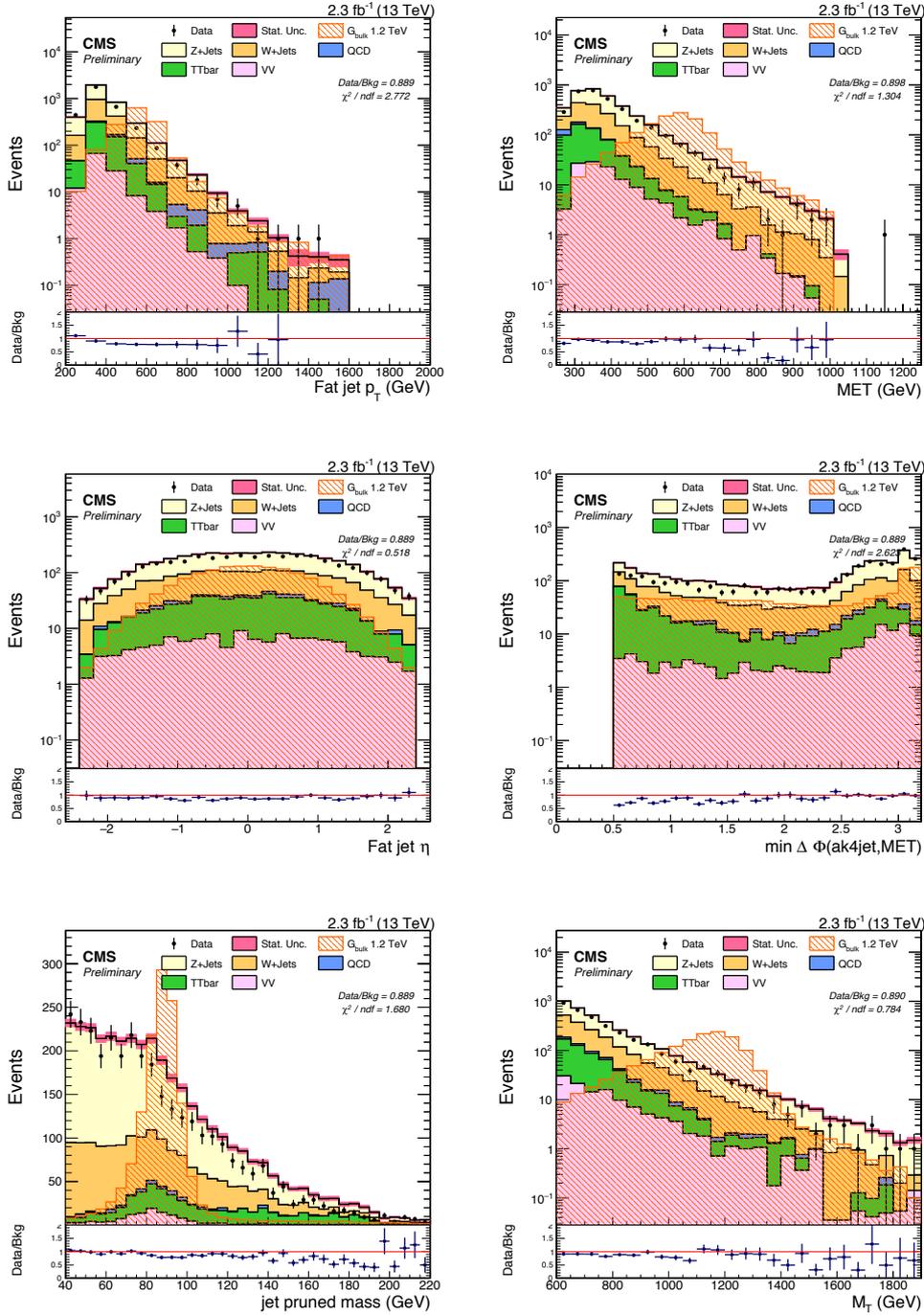


Figure 4.14: Top:(left) AK8 Jet transverse momentum. (right) Missing transverse energy. Center: (left) AK8 Jet pseudorapidity. (right) min $\Delta\phi$ distribution. Bottom: (left) Pruned jet mass. (right) Candidate transverse mass. The plots show the comparison between data(black dots) and simulation.

4.4 Background Modeling

To perform an accurate prediction of the SM backgrounds we used a data-driven strategy called *alpha ratio method* that will be discussed in the section 4.4.1. The dominant background contribution in the analysis originates from the V+jets processes (they represent between the 70-86% of the total background). The subdominant contributions come from $t\bar{t}$, diboson (WW/WZ/ZZ), and QCD multijets. The subdominant contribution yields and the transverse mass shapes (M_{VZ}^T) are primarily taken from simulation. Table 4.9 summarizes the background categories.

Table 4.9: Background categories.

Category	Backgrounds
Dominants	W+jets($W \rightarrow \ell\nu$), Z+jets ($Z \rightarrow \nu\nu$)
Subdominants	Multijets, $t\bar{t}$, dibosons (WW/WZ/ZZ)

To determine the dominant V+jets background in the signal region ($m_{\text{jet}} \in [65, 105]$ GeV), a signal-free sideband region is defined in the mass of the hadronic V candidate by the interval $m_{\text{jet}} \in [40, 65] \cup [135, 220]$ GeV. The region $m_{\text{jet}} \in [105, 135]$ GeV is not used in order to avoid any bias in the M_{VZ}^T shape due to possible contributions from new resonances in the HZ final state, in which the Higgs bosons would decay to a pair of b-quarks. Table 4.10 summarizes the different jet mass regions used for the background modeling.

Table 4.10: Jet mass regions.

Region	Interval (GeV)
Lower sideband (LSB)	[40, 65]
Lower signal (SR)	[65, 105]
Upper signal (Higgs)	[105, 135]
Upper sideband (USB)	[135, 220]

The background composition in percentages in each category is reported in table 4.11.

In general, we experience low statistics in data, particularly in the tail of transverse mass distribution, which is the main observable in order to determine any possible signal of new physics. For those cases unbinned maximum likelihood (ML) fits are preferred due to its robustness (statistically more powerful) and to avoid the information loss and

Table 4.11: Background Composition.

Backgrounds	High purity ($\tau_{21} < 0.45$)	Low purity ($0.45 < \tau_{21} < 0.75$)	No purity ($\tau_{21} < 0.75$)
V+jets	70%	86%	80%
$t\bar{t}$	18%	8%	11%
QCD+VV	12%	6%	9%

arbitrariness of the binning procedure [104]. Therefore, in the analysis we performed all the fits using the unbinned ML method.

4.4.1 Alpha ratio method

The alpha ratio method is based on the extrapolation of the background shape and yield from the jet pruned mass sideband to the signal region. The method relies on the assumption that the correlation between M_{VZ}^T and the jet pruned mass for the dominant background in data is reasonably well reproduced in simulation. The advantage of this approach is that most systematic uncertainties cancel out in the ratio. The total background prediction (normalisation and shape) as a function of the reconstructed resonance mass, M_{VZ}^T , is obtained separately for each purity category according to the formula:

$$\begin{aligned}
N_{\text{total}}^{\text{signal}}(M_{VZ}^T) &= N_{\text{DB}}^{\text{signal}}(M_{VZ}^T) + N_{\text{SB}}^{\text{signal}}(M_{VZ}^T) \\
&= \underbrace{N_{\text{data}}^{\text{sideband}}(M_{VZ}^T) \times (1 - R_0(M_{VZ}^T)) \times \alpha^{\text{MC}}(M_{VZ}^T)}_{\text{Shape}} \times \underbrace{F_{\text{DB}}}_{\text{Norm}} \\
&\quad + N_{\text{SB}}^{\text{signal}}(M_{VZ}^T)
\end{aligned} \tag{4.4}$$

where

- $N_{\text{total}}^{\text{signal}}(M_{VZ}^T)$ is the total background prediction in the signal region;
- $N_{\text{DB}}^{\text{signal}}(M_{VZ}^T)$ is the dominant background prediction in the signal region (W/Z + jets);
- $N_{\text{SB}}^{\text{signal}}(M_{VZ}^T)$ is the background prediction in signal region for the sum of the subdominant backgrounds ($t\bar{t}$, multijets, dibosons);

- $N_{\text{data}}^{\text{sideband}}(M_{VZ}^T)$ is the M_{VZ}^T distribution in data for the sideband region;
- $R_0(M_{VZ}^T) = N_{\text{SB}}^{\text{sideband}}(M_{VZ}^T) / N_{\text{data}}^{\text{sideband}}(M_{VZ}^T)$ is the fraction of subdominant backgrounds in the sideband region with respect to the total background in the sideband region (i.e. data). The $1 - R_0(M_{VZ}^T)$ multiplicative term represents the subtraction of the subdominant contribution from the data in the sideband region;
- $\alpha^{\text{MC}}(M_{VZ}^T) = N_{\text{DB}}^{\text{signal}}(M_{VZ}^T) / N_{\text{DB}}^{\text{sideband}}(M_{VZ}^T)$ is the dominant background prediction in the signal region divided by the one in the sideband region, calculated from MC. It represents the transfer function (from sideband to signal region) used to correct the data in the sideband region and extract the dominant background's M_{VZ}^T shape.
- F_{DB} is an overall scale factor used to set the normalisation of the dominant background prediction.

In summary, the method is divided in two parts, the shape and the normalization prediction for the dominant backgrounds. We will detail each in the following sections. The alpha ratio method is the standard strategy for background estimation used by CMS in similar searches [105–107].

Background normalization

The m_{jet} distribution is modeled with analytic functions on simulation, considering separately the dominant and subdominant backgrounds. The overall dominant background normalization in the signal region is determined from a fit to the m_{jet} distribution in the sideband region, after fixing both the shape and the normalization of the subdominant backgrounds. Table 4.12 show the functional forms employed to describe the jet pruned mass distribution.

Table 4.12: Analytic functions to fit the jet pruned mass distribution.

Name	Description	Function
ErfExp	An exponential times an error function	$e^a \cdot \frac{1 + \text{Erf}((x-b)/w)}{2}$
Gaus2	The addition of two gaussians	$f_0 e^{-(x-a)^2/2s^2} + (1 - f_0) e^{-(x-b)^2/2s^2}$

Table 4.13: Expected and observed background yields in signal region.

Category	Background	Expected	Observed	Syst. Uncer.
HP	Dominant + Sub-dominant	544 ± 40	507	20.5 %
LP	Dominant + Sub-dominant	820 ± 53	806	15.5%

To verify that the $W + \text{jets}$ and $Z + \text{jets}$ samples have the same shape and that we can treat them together as a dominant background, we performed a cross-check, fitting them separately as is shown in Figure 4.15. For the normalization procedure we did the following:

- We fit the dominant backgrounds with the **ErfExp** model in the total jet pruned mass distribution, as it can be seen in the top-left side of the figures 4.16, 4.17.
- We fit the subdominant backgrounds with the **Gaus2** model in the total jet pruned mass distribution, as it can be seen in the top-right side of the figures 4.16, 4.17.
- We used an extended model adding the probability density functions (PDFs) of the dominant and subdominants backgrounds. For the dominant backgrounds we let both the parameters of the normalization and shape float and for the subdominant backgrounds we fix the parameters of the normalization and shape. Then we fit the extended model to the data in sidebands.
- Using the results of the fits, we predicted the shape and the normalization (F_{DB}) of the dominant backgrounds in signal region. Figures 4.16, 4.17, bottom side, show the prediction in comparison with data. In addition, we can extract the estimation of the total yields in signal region as is reported in table 4.13.

The characterization of the jet pruned mass with the analytic functions was evaluated in simulation and compared against alternative functions. To consider the mis-modelling of the pruned jet mass, an uncertainty was included as a systematic error that ranges between 15-20 %. The expected and observed number of events in the signal region is given in Table 4.13; the result is reported for the HP and LP categories.

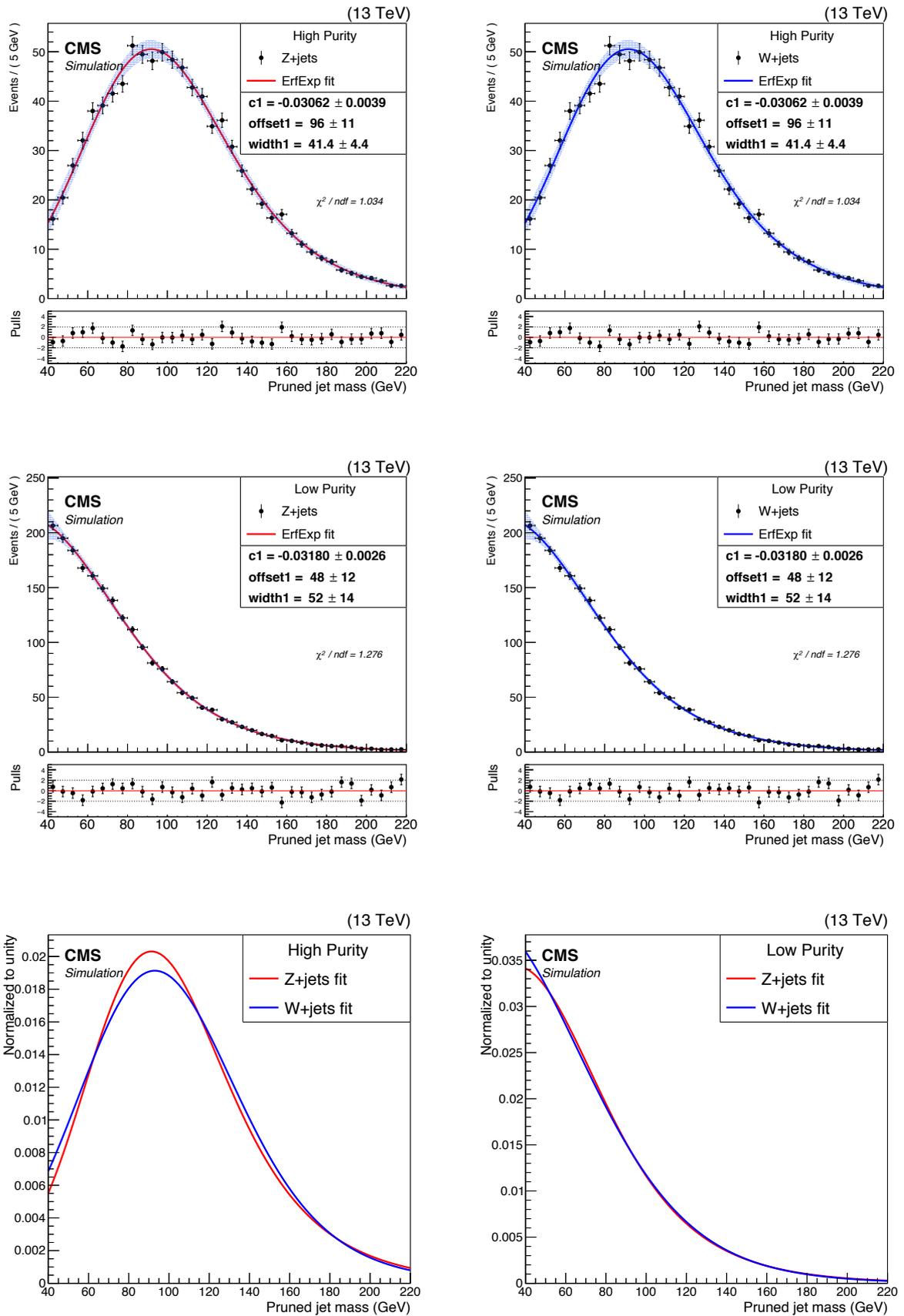


Figure 4.15: Top : Z/W +jets fits in the HP category. Center : Z/W +jets fits in the LP category. Bottom : Comparison between the Z +jets and W +jets fits.

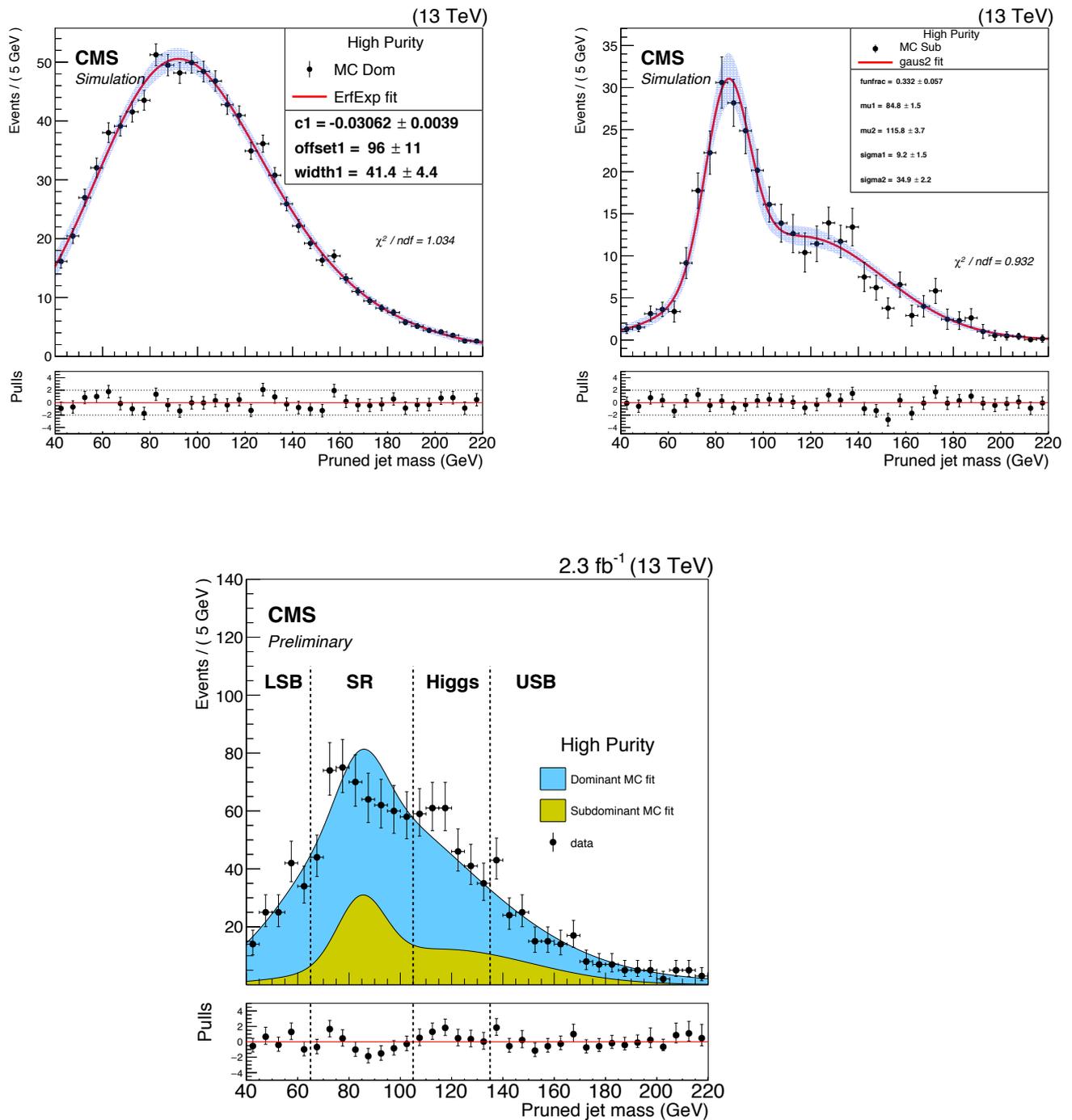


Figure 4.16: Fit of the jet pruned mass in the HP category. Top: (left) Dominant MC background fit with the ErfExp function (right) Subdominant MC background fit with the Gaus2 function. Bottom: Distribution of the jet pruned mass in the HP categories. All selections are applied except the final m_{jet} signal window requirement. Data are shown as black markers. The contribution of the V+jets background is extrapolated from the sideband to the signal region.

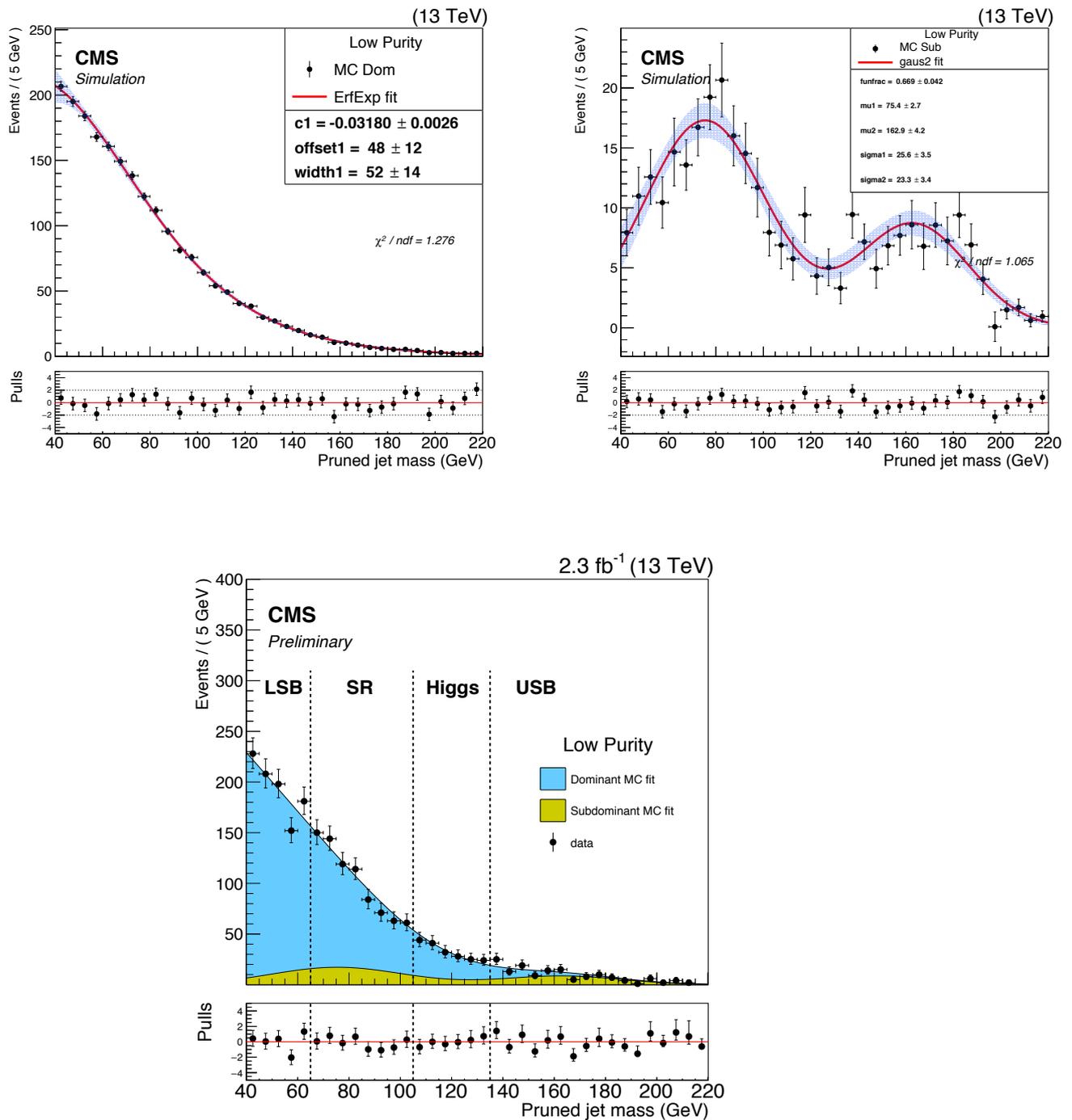


Figure 4.17: Fit of the jet pruned mass in the LP category. Top: (left) Dominant MC background fit with the ErfExp function (right) Subdominant MC background fit with the Gaus2 function. Bottom: Distribution of the jet pruned mass in the LP categories. All selections are applied except the final m_{jet} signal window requirement. Data are shown as black markers. The contribution of the V+jets background is extrapolated from the sideband to the signal region.

Background shape

Evidence of a new resonance decaying into diboson (VZ) would appear as a localized excess in the transverse mass distribution. Because of this, it is essential to perform a good estimation of the background shape in the M_{VZ}^T variable. Due to the correlation between the jet pruned mass and the transverse mass variables, the selections on the m_{jet} reported in table 4.10 define the sideband and the signal regions in the M_{VZ}^T distributions. The analytic model selected to estimate the background shape of the transverse mass distribution is the **ExpTail**:

$$F_{\text{ExpTail}}(x) = e^{-x/(a+bx)} \quad (4.5)$$

The choice of the functional forms for the VZ transverse mass for different categories is shown in table 4.14.

Table 4.14: Summary of the shapes used for fit the M_{VZ}^T spectra of each category.

Category	Fit Function	Regions	Sample
HPLP	ExpTail	signal and sideband	Data and MC

In order to predict the shape of the dominant backgrounds we followed these steps:

- We fit the subdominant backgrounds with an **ExpTail** model in sideband and signal region. After the fit we fixed all the parameters of the model.
- We subtracted the subdominant backgrounds from data in the sideband region.
- We performed a simultaneous fit using the **ExpTail** model for the dominant background in the signal and sideband region and for the data in sideband region. After the fit we fix all the parameters of the model. Figures 4.18, 4.19 show the results of the fits. The top side shows the dominant backgrounds fits in sideband and signal region and the right-center side show the data fit in sideband region.
- We extract the alpha transfer function using the fits of the dominant backgrounds in sideband and signal region. $\alpha^{MC}(M_{VZ}^T) = N_{\text{DB}}^{\text{signal}}(M_{VZ}^T) / N_{\text{DB}}^{\text{sideband}}(M_{VZ}^T)$. Figures 4.18, 4.19 show the alpha transfer function for high and low purity categories

respectively.

- We estimate the shape of the dominant backgrounds in signal region just correcting the M_{VZ}^T distribution of the data in sideband region with the transfer function. Figures 4.18, 4.19 show the final prediction in comparison with data. The solid curve represents the background estimation provided by the data-driven method. The hatched band includes both statistical and systematics uncertainties. The data are shown as black markers. The bottom panels show the corresponding pull distributions, quantifying the agreement between the background-only hypothesis and the data. The pull distribution is defined as the difference between the data and the background prediction, divided by the error on data. The error bars on the points represent the error on data. In those plots the normalization factor was already applied.

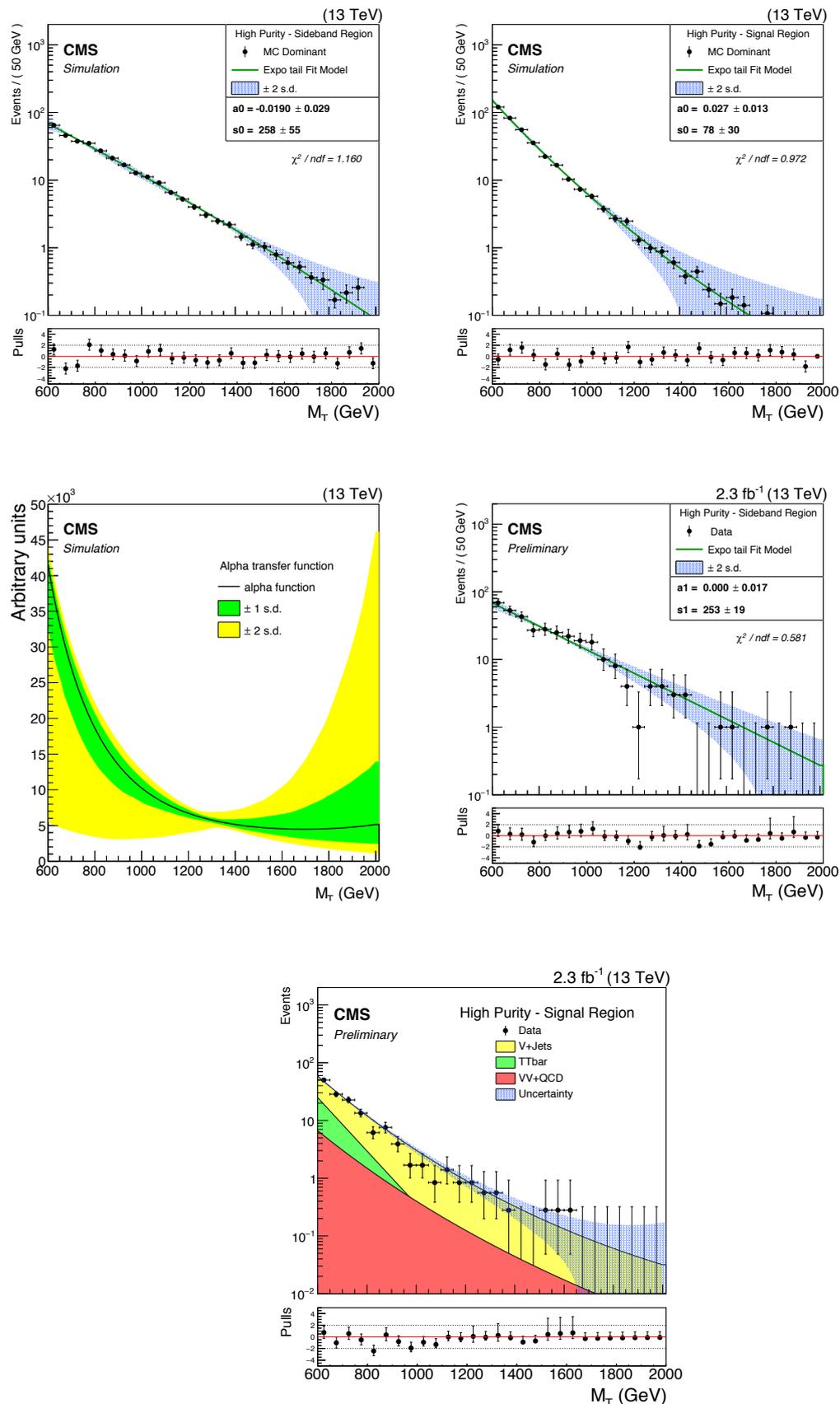


Figure 4.18: Fit of the VZ candidate transverse mass in the HP category. Top: (left) Dominant MC background fit in the sideband region (right) Dominant MC background fit in the signal region. Center: (left) alpha transfer function (right) Data fit in the sideband region. Bottom: Final prediction on the background estimation in signal region.

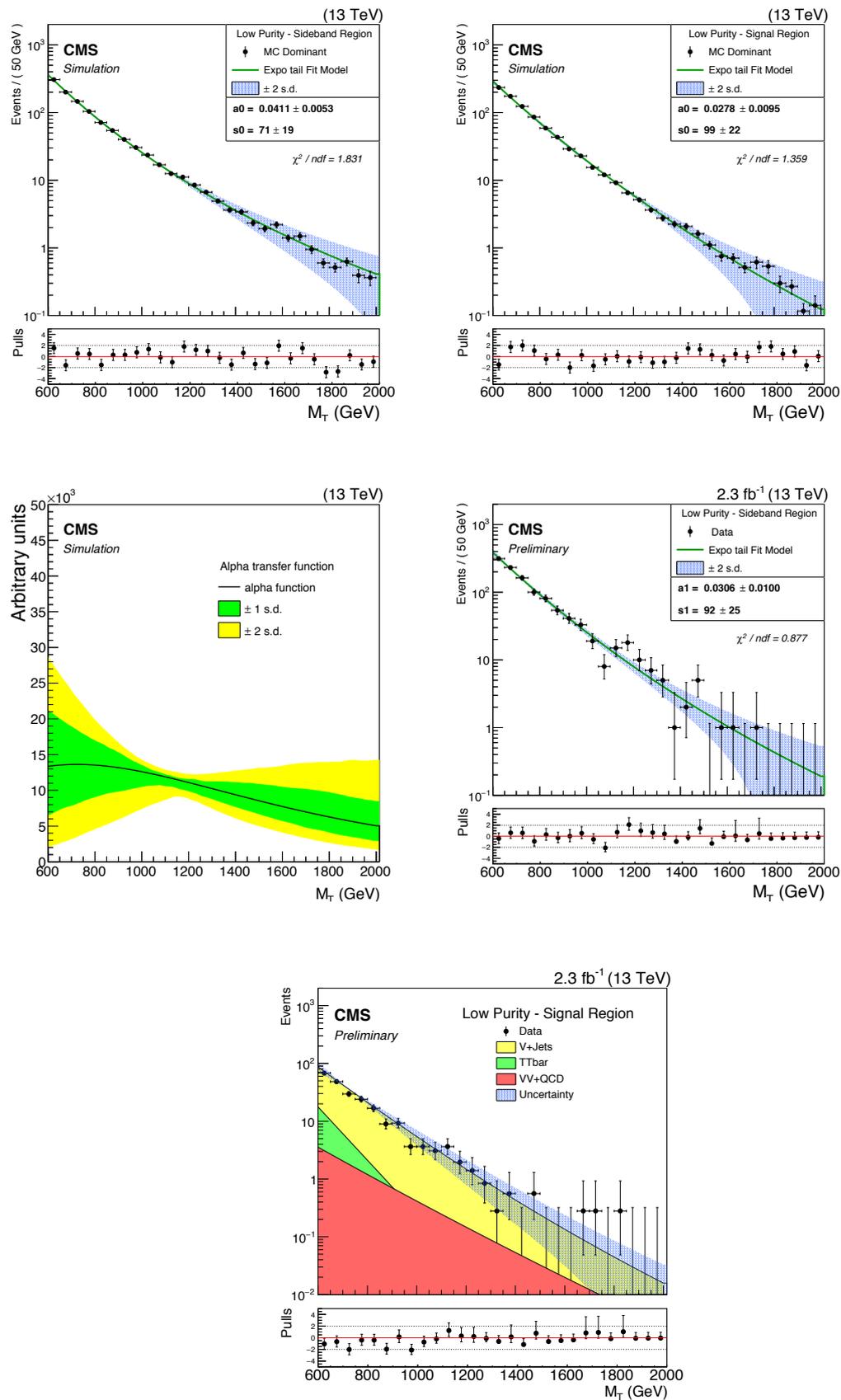


Figure 4.19: Fit of the VZ candidate transverse mass in the LP category. Top: (left) Dominant MC background fit in the sideband region (right) Dominant MC background fit in the signal region. Center: (left) alpha transfer function (right) Data fit in the sideband region. Bottom: Final prediction on the background estimation in signal region.

4.5 Signal Modeling

The shape of the reconstructed signal mass is extracted from the bulk graviton and W' samples to model the peak of the resonance. The natural width of the resonance is considered to be sufficiently small to be neglected when compared to the detector resolution. In the final analysis of the M_{VZ}^T spectrum, the discovery potential and the exclusion power depend both on an accurate description of the signal shape. We adopt an analytical description of the signal shape, choosing a single Crystal-Ball function (i.e. a Gaussian core with power-law low-end tail) to describe the CMS detector resolution. The typical width of the Gaussian core is about 5%-7% of the nominal mass. The analytical description of the signal shape allows us to probe mass points for which there is no generated sample by interpolating the shape parameter. No appreciable differences have been observed in the M_{VZ}^T signal shape between the low-purity and the high-purity categories. Table 4.15 and table 4.16 show the mass points and the fit range for each signal hypothesis. The transverse mass fits in the signal sample are shown in the figures 4.20,4.22, and 4.24. Figure 4.25 shows the simulated M_{VZ}^T distribution for resonance masses from 800 to 2000 GeV after the interpolation process. The different distributions are normalized to the corresponding efficiencies.

Table 4.15: Different mass points to fit the signal shape for Bulk graviton model.

Mass point (GeV)	Mean (GeV)	Fit window (GeV)
800	763	[620,900]
1000	930	[670,1190]
1200	1100	[720,1424]
1400	1283	[850,1657]
1600	1461	[1035, 1887]
1800	1640	[1168, 2112]
2000	1819	[1295, 2342]

Table 4.16: Different mass points to fit the signal shape for W prime model.

Mass point (GeV)	Mean (GeV)	Fit window (GeV)
800	760	[600,1000]
1200	1092	[560,1500]
2000	1685	[700, 2100]

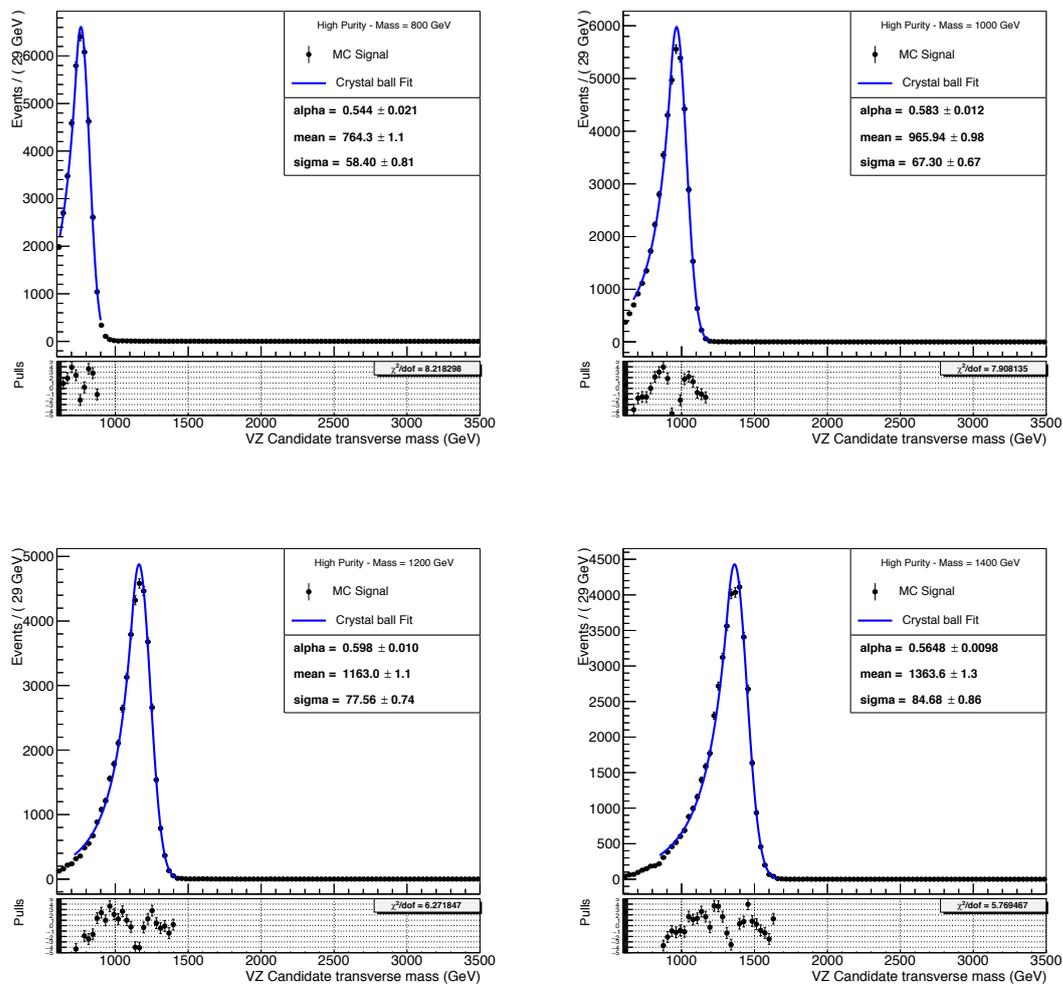


Figure 4.20: Fit of the Bulk Graviton signal samples for different mass points (800-1400 GeV) in the HP category.

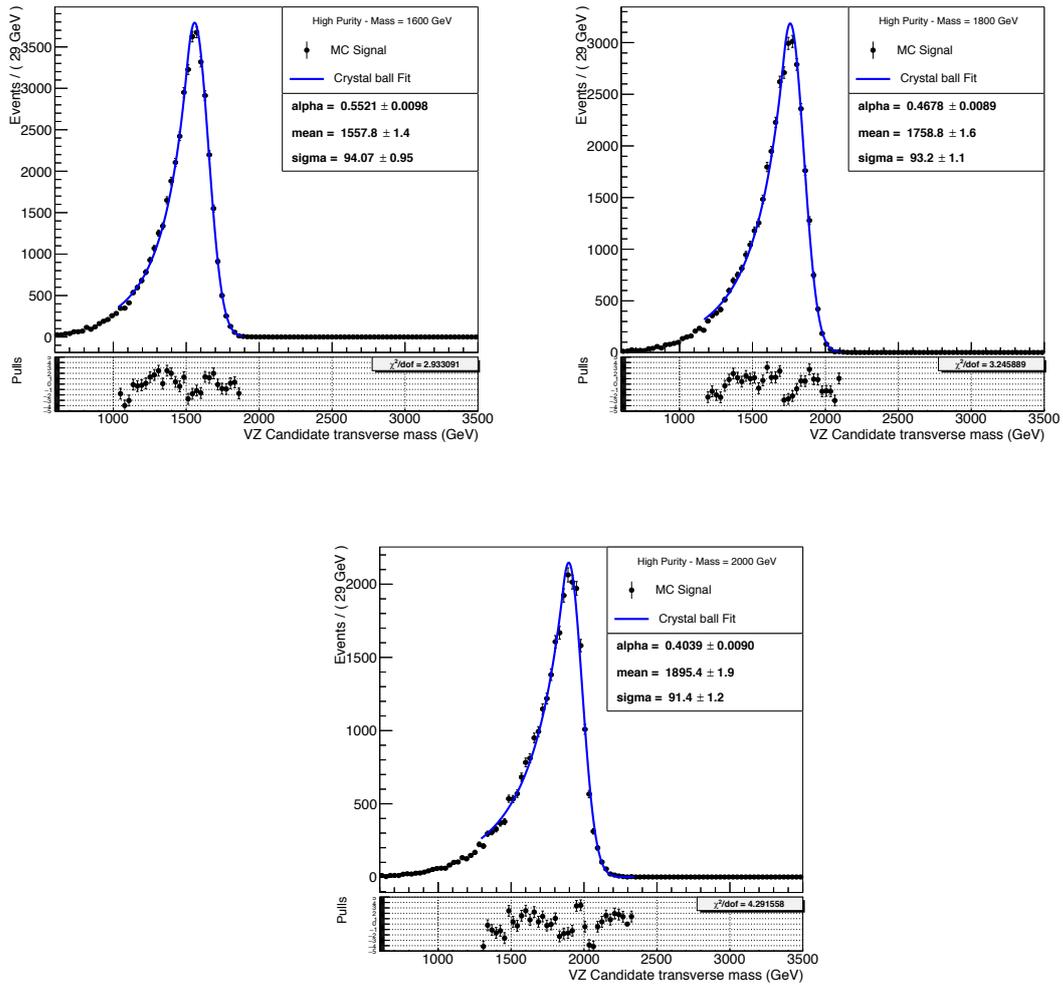


Figure 4.21: Fit of the Bulk Graviton signal samples for different mass points (1600-2000 GeV) in the HP category.

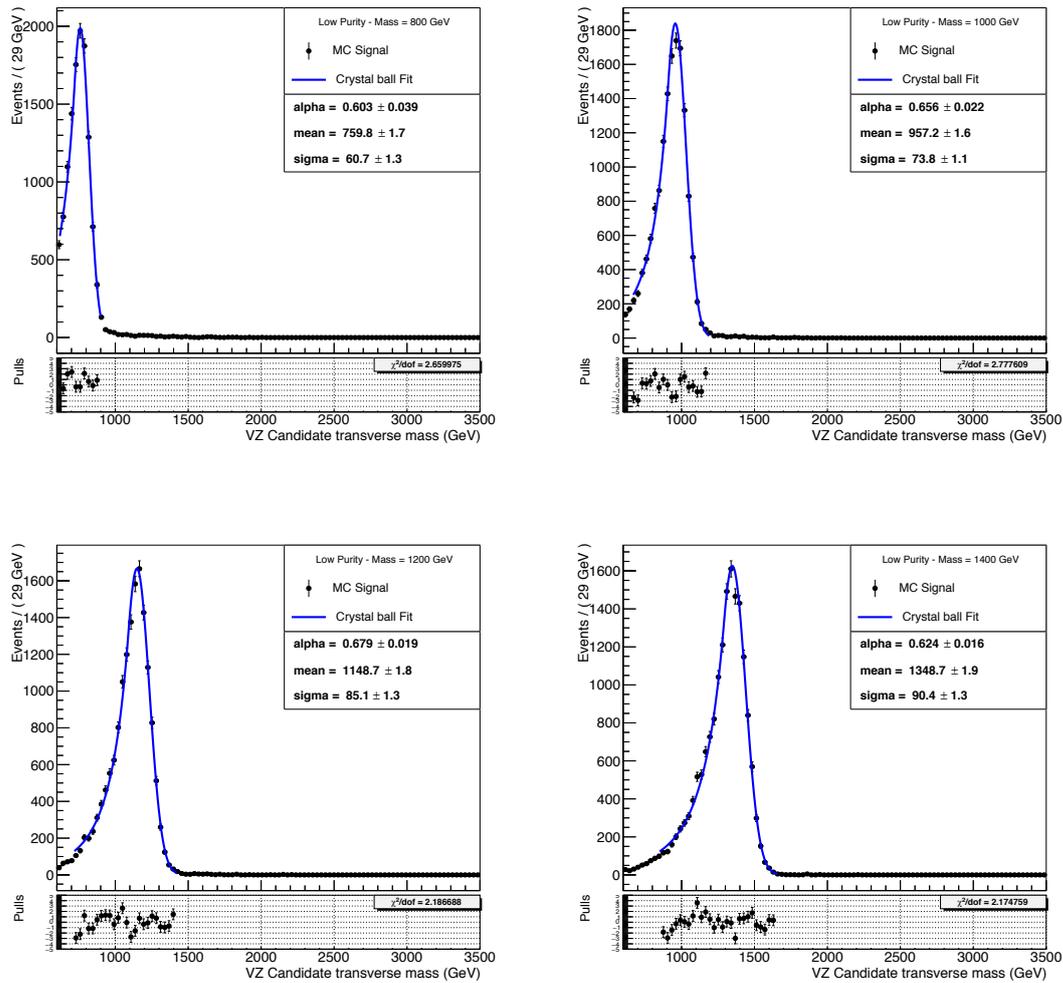


Figure 4.22: Fit of the Bulk Graviton signal samples for different mass points (800-1400 GeV) in the LP category.

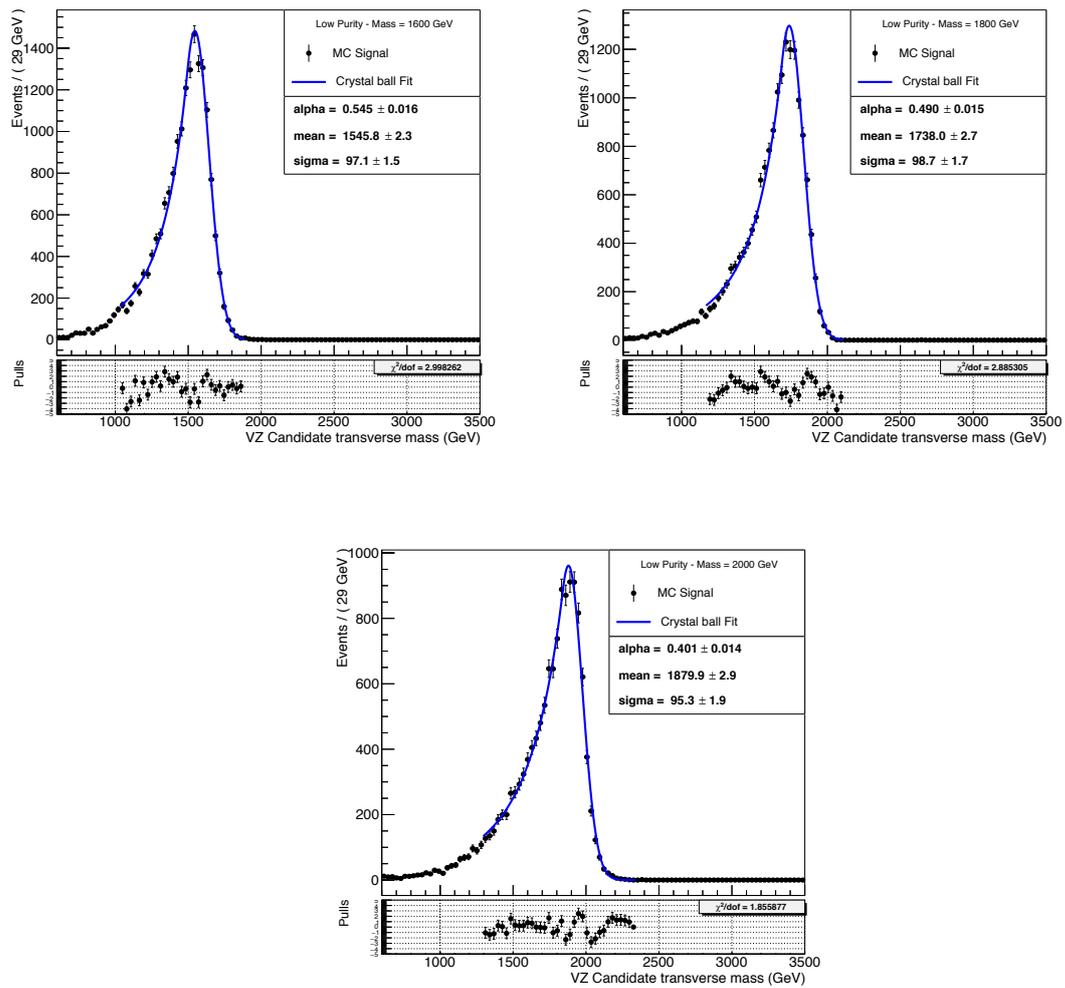


Figure 4.23: Fit of the Bulk Graviton signal samples for different mass points (1600-2000 GeV) in the LP category.

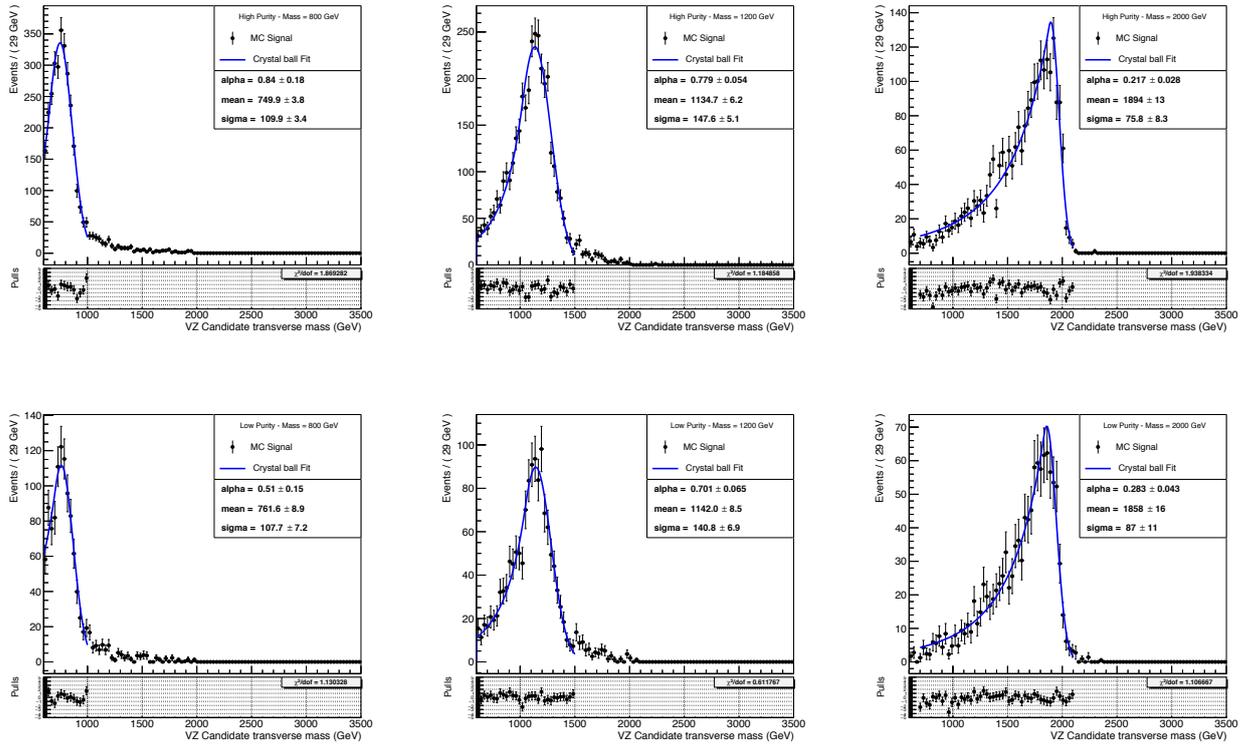


Figure 4.24: Top: Fit of the W prime signal samples for different mass points in the HP category. Bottom: Fit of the W prime signal samples for different mass points in the LP category

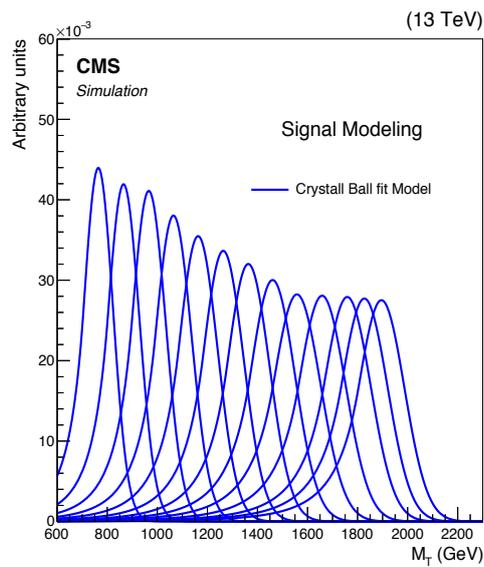


Figure 4.25: Linear interpolation in the Crtsyal-Ball fit model with a step of 100 GeV for bulk graviton samples.

4.6 Systematics Uncertainties

The systematic uncertainties from different sources are listed in table 4.19. In this chapter we describe in detail how they were determined for background estimation and signal prediction.

4.6.1 Systematic uncertainties in the background estimation

To estimate the systematic uncertainties for the normalisation of the background prediction we performed a fit in data with different template functions. Figure 4.26 shows the pruned jet mass distribution for different categories and the fits using different models. Based on these fits, we can estimate the yields in the signal region and assess a systematic uncertainty (δ) for the background normalisation, as it is shown in Table 4.17.

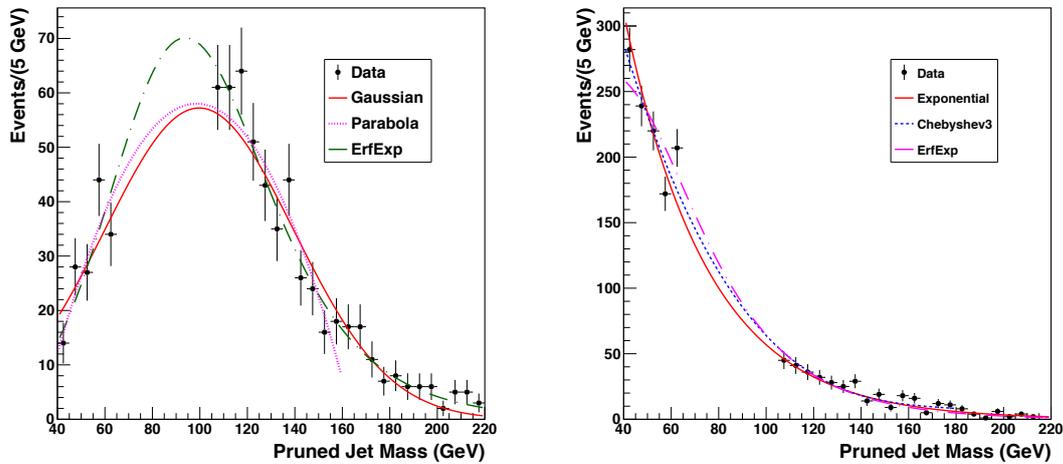


Figure 4.26: Fit of the jet pruned mass in data using different template functions. (left): High purity category; (right): Low purity category.

Table 4.17: Signal yields for different template functions.

Category	Gaussian	Parabola	ErfExp	Exponential	Chebyshev3	$(l-s)/s$	$(l-s)/l$	δ
HP	412	430	507	-	-	23%	18%	20.5%
LP	-	-	732	817	858	17%	14%	15.5%

4.6.2 Integrated Luminosity

For the luminosity, we considered a systematic uncertainty of 2.7%. This is related with the uncertainty on the number of signal events passing the final selection and is fully correlated in all the categories.

4.6.3 Jet Substructure Scale Factors

The V-tagging efficiency scale factors for 2015 data and simulation were derived in [108]. These scale factors were applied to the signal yields and their uncertainty which were anti-correlated taken as systematic.

Table 4.18: V-tagging scale factors and their systematics uncertainties for each category.

Category	Scale factor	Systematic Uncertainty
HP	0.942	1.067/0.933
LP	1.268	0.74/1.26

4.6.4 QCD scale

The impact of the systematic uncertainties due to the factorization and renormalization scales on the signal efficiency were evaluated using the weight values provided in the MC samples. The scale uncertainties were studied by varying the renormalization and factorization scales independently by a factor 1/2 and 2. The associated systematic uncertainty due to the shift of the peak varies between 0.2-0.5% for signal and 12% for the subdominant backgrounds. Figure 4.27 shows the QCD scale systematic uncertainties for different signal mass points.

4.6.5 PDF (Parton distribution function)

The PDF Systematic uncertainties for the NNPDF3.0 set was calculated using the PDF4LHC prescription [109]. The standard deviation was obtained as the RSM value

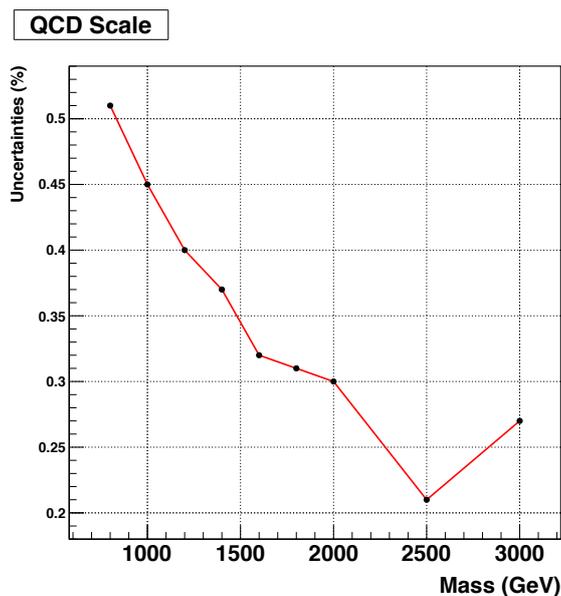


Figure 4.27: Systematic uncertainties due to the factorization and renormalization scales for different signal mass points.

of the weights per event. The evaluation was performed by raising and lowering the respective uncertainty by one standard deviation. The associated systematic uncertainty due to the shift of the signal peak varies between 8% and 18% for signal samples and 17% for subdominant backgrounds. Figure 4.28 shows the PDF systematic uncertainties for different signal mass points.

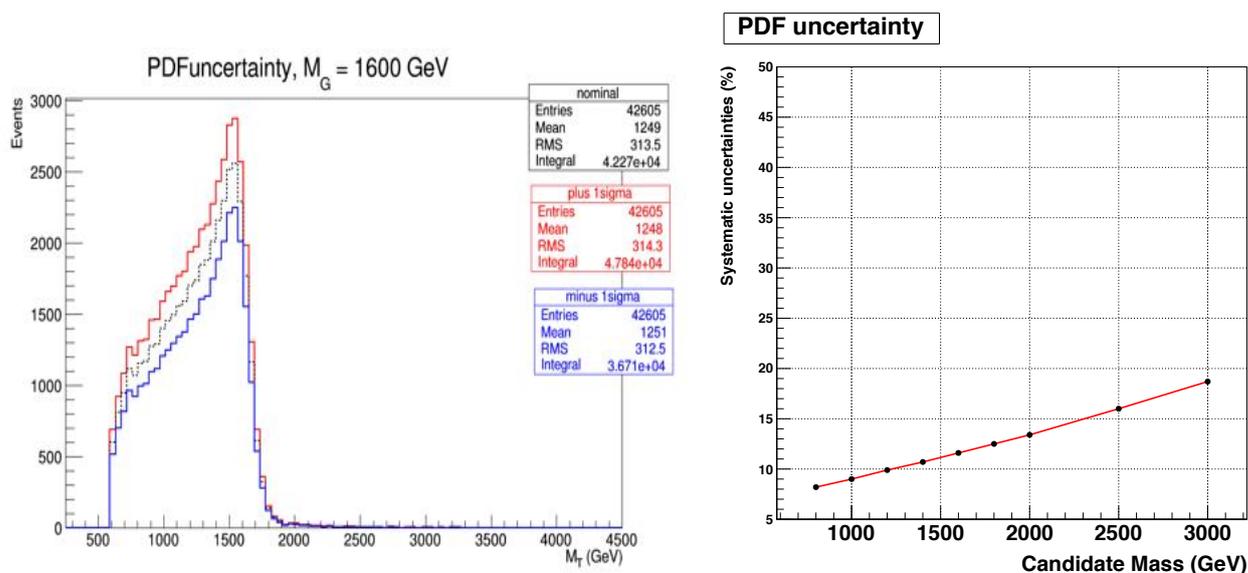


Figure 4.28: Left : Transverse Mass of the candidate with the nominal value of the PDF. Also we show the scale up and down in one standard deviation for a signal of 2 TeV. Right: PDF systematic uncertainties for different signal mass point.

4.6.6 Trigger SF

We estimate the uncertainty of the trigger SF from the ratio between the efficiency measured in Data and the efficiency measured in MC:

$$\text{ratio} = \frac{\text{efficiency}_{\text{Data}}}{\text{efficiency}_{\text{MC}}} \quad (4.6)$$

The uncertainty in the ratio is given by:

$$\Delta\text{ratio} = \text{ratio} \times \sqrt{\left(\frac{\Delta y}{y}\right)_{\text{eff-Data}}^2 + \left(\frac{\Delta y}{y}\right)_{\text{eff-MC}}^2} \quad (4.7)$$

where y give the value of the efficiency and Δy it uncertainty. We consider the SF as a weight in the event and then we define some variations:

$$\text{triggerWeight} = \text{ratio}$$

$$\text{triggerWeightup} = \text{ratio} + \Delta\text{ratio}$$

$$\text{triggerWeightdown} = \text{ratio} - \Delta\text{ratio} \quad (4.8)$$

$$(4.9)$$

Then we will estimate the impact in the transverse mass in signal after applying the trigger weights (up and down), as it can be observed in the figure 4.29. The associated systematic uncertainty due to the trigger SF is around 2.5%.

4.6.7 B-tagging

The effect of the b-tagging uncertainty is evaluated by varying the Combined Secondary Vertex (CSV) scale factors (up and down) for the respective flavor. The associated systematic uncertainty due to the shift of the signal peak varies between 0.02% and 0.04%. The figure 4.30 show the evaluation of the uncertainty for different mass points.

In the Figure 4.31 we show the evaluation for the backgrounds : Z+jets, W+jets and TTbar. In case of the TT bar the systematic uncertainty is 1 %, and for Z/W + jets 0.05 %.

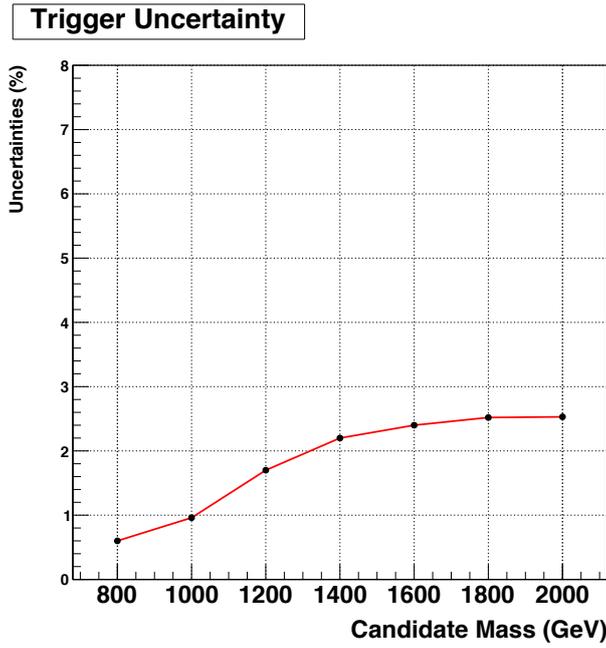


Figure 4.29: Systematic uncertainties due to the trigger SF for different signal mass points.

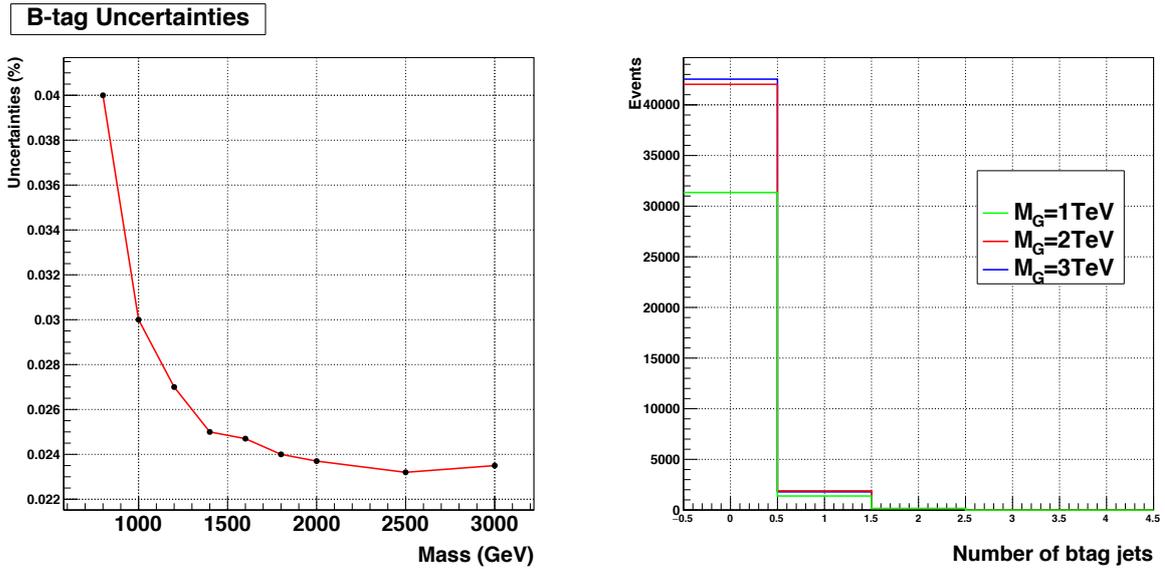


Figure 4.30: Left : Systematic uncertainties due to b-tagging uncertainty SF for different signal mass points. Right : Number of b-tag jets for different signal mass points.

4.6.8 Pile-up

For the PU uncertainties we test over different minbias cross-section (72mb, 69mb, 66mb) to obtain the pileup weights. In Figure 4.32 we show different scenarios for a signal mass of 2 TeV. In this case the uncertainty is of 0.02% for signal and 2.3% for the subdominant backgrounds.

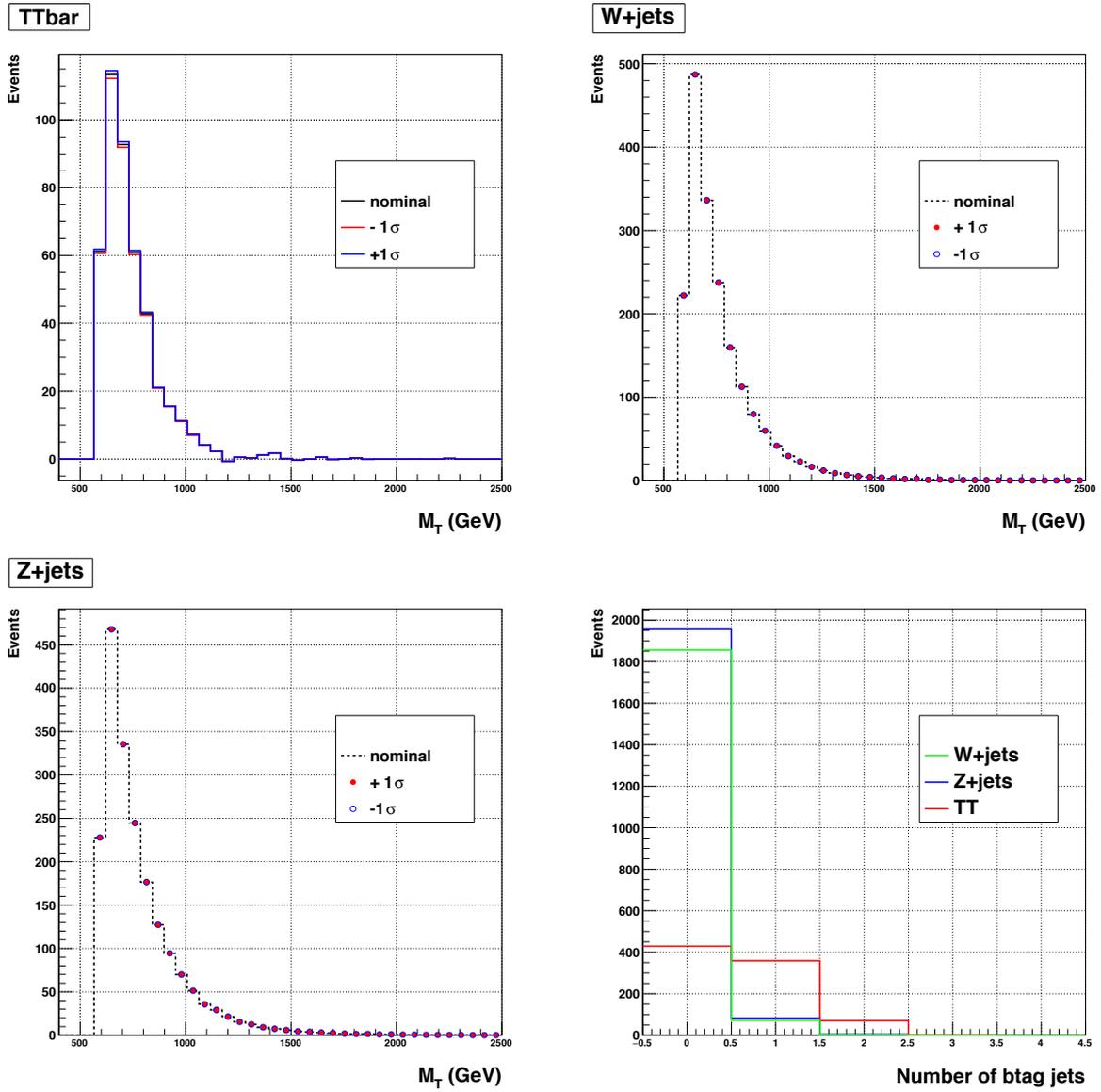


Figure 4.31: Systematic uncertainties due to b-tagging uncertainty SF. Top : Nominal and scale up/down variations values in the transverse mass distribution for (Left) $t\bar{t}$ sample (right) W+jets sample. Bottom : (left) Z+jets sample (right) Number of b-tag jets for different backgrounds models.

4.6.9 Jet Energy Corrections

The impact of the JEC uncertainties were evaluated scaling up and down the jet p_T . We considered the propagation of the JEC to the missing energy, so that the variation was applied in both jets and E_T^{miss} (up,up) and (down,down). The associated systematic uncertainty due to the shift of the signal peak varies between 0.09% and 1.76%. For the subdominant backgrounds is 3%. Figure 4.33 show the JEC systematic uncertainties for different signal mass points.

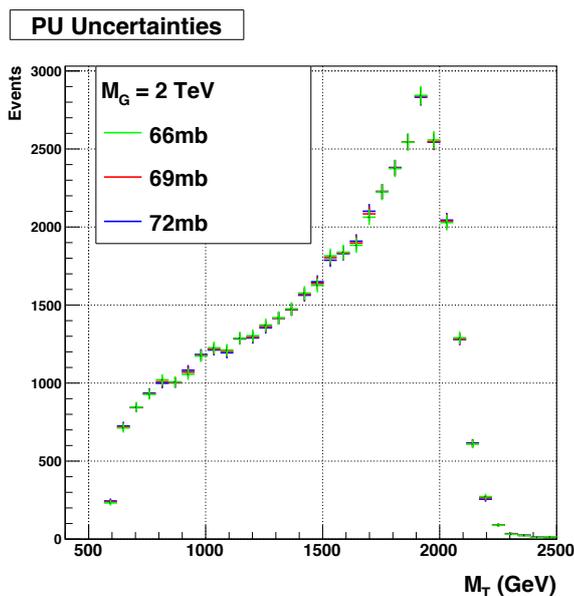


Figure 4.32: Systematic uncertainties on the pile-up due to different minbias cross-section for a signal mass of 2 TeV.

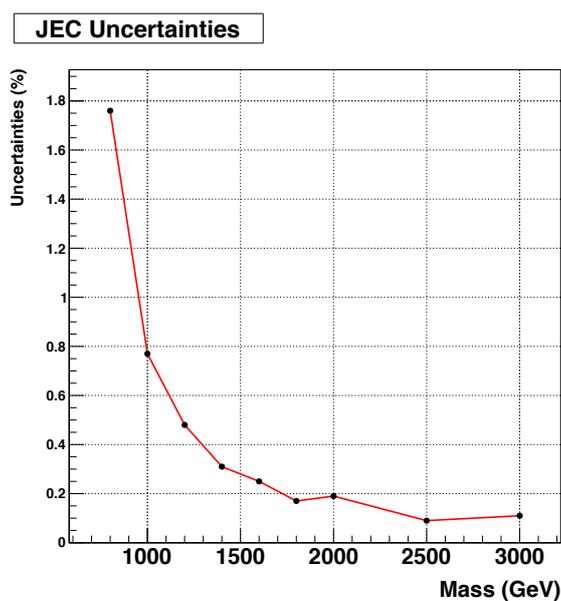


Figure 4.33: Systematic uncertainties due to the JEC for different signal mass points.

4.6.10 Jet Energy Resolution

We smeared the jets using Jet energy resolution scaling factors and the uncertainties were tested by varying its value by one standard deviation (up and down). We found that the systematic uncertainties for the Jet resolution are small, ranging from 0.009% to 0.06%. For the subdominant backgrounds we did not observe any significant variation.

4.6.11 Leptons and photons ID

The systematic uncertainties due to scale factors for leptons and photons are very small in the analysis. In the case of a signal mass of 1 TeV the propagation of the uncertainty in $\pm 1 \sigma$ gives no significant variation in the efficiency. For a mass of 3 TeV the uncertainty is of 0.02%. We consider these systematics negligible.

4.6.12 Jet mass calibration uncertainty

As we used a W/Z-tagger we needed to consider the uncertainty in the pruned jet mass calibration. We varied the jet pruned mass calibration to calculate the impact on the signal selection efficiency. The shift of the signal peak varied between 0.13% and 0.22%. The figure 4.34 show the systematic uncertainties due to the jet mass calibration for different mass points.

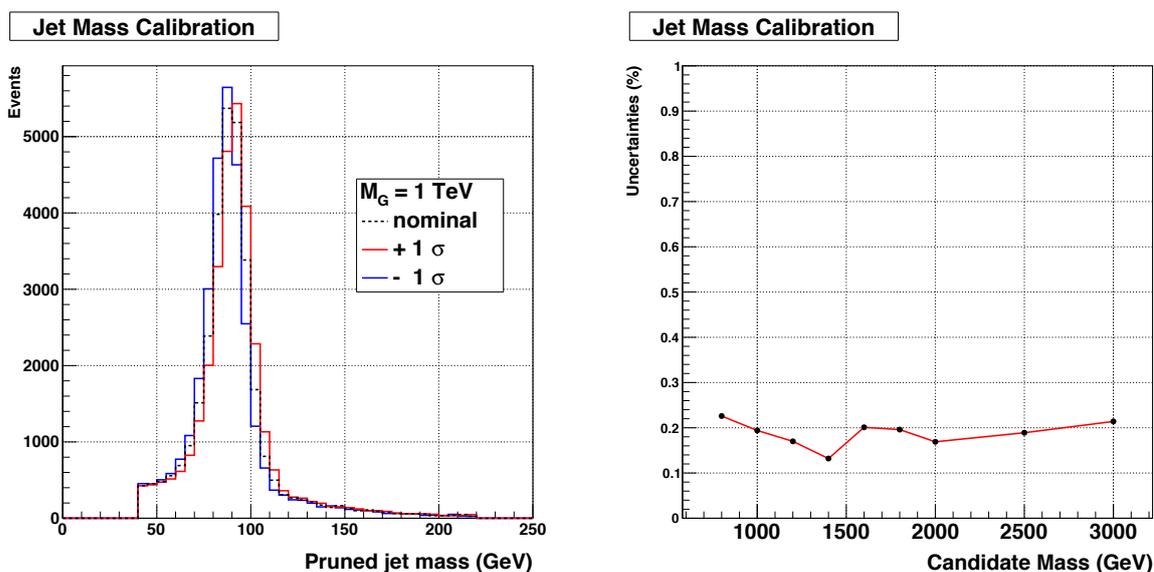


Figure 4.34: Systematic uncertainties due to jet mass calibration. Left : Nominal and scale up/down variations values in the pruned jet mass distribution for a signal point of 1000 GeV. Right : Systematic uncertainties in (%) for different mass points.

4.6.13 Jet mass resolution uncertainty

We considered the systematic uncertainties due to the pruned jet mass resolution. In that sense we varied the jet mass resolution from the central value in order to calculate

the impact on the signal selection efficiency. Note that in our case the scale factor is one. The shift of the signal peak varies between 0.93% and 1.50%. Figure 4.35 shows the systematic uncertainties due to the jet mass resolution for different mass points.

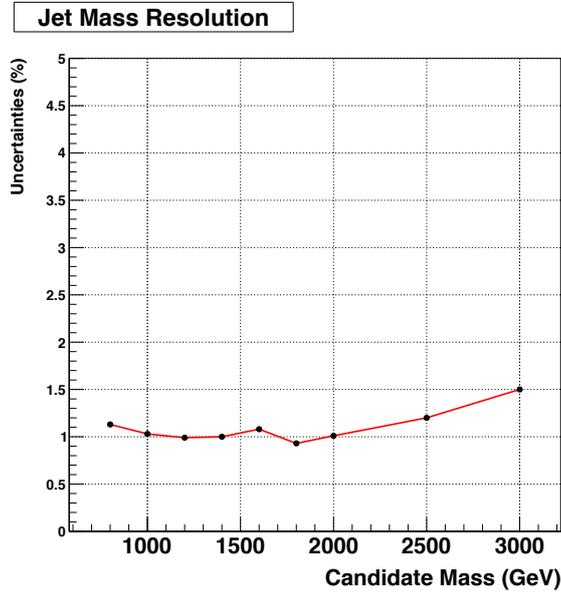


Figure 4.35: Systematic uncertainties due to jet mass resolution. The figure shows the systematic uncertainties in (%) for different mass points.

4.6.14 p_T extrapolation uncertainty

The impact of the extrapolation uncertainties on the τ_{21} selection due to propagation to higher momenta is taken into account. Extrapolation uncertainties for a W/Z p_T of interest can be estimated from the double ratio of the selection efficiency from Pythia and Herwig samples with W/Z p_T of 200 GeV and signal with W/Z p_T of interest. We use the formula : $0.059 * \ln(p_T/200)$ GeV to estimate this uncertainty [108]. As an approximation we will consider the p_T of the jets as half the value of the mass of the resonance. Figure 4.36 shows the systematic uncertainties due to the p_T extrapolation for different mass points.

4.6.15 E_T^{miss} unclustered energy uncertainty

We considered the impact of the unclustered energy uncertainties in the analysis. We used the Run2 style, varying each particle type by his own resolution. With this purpose

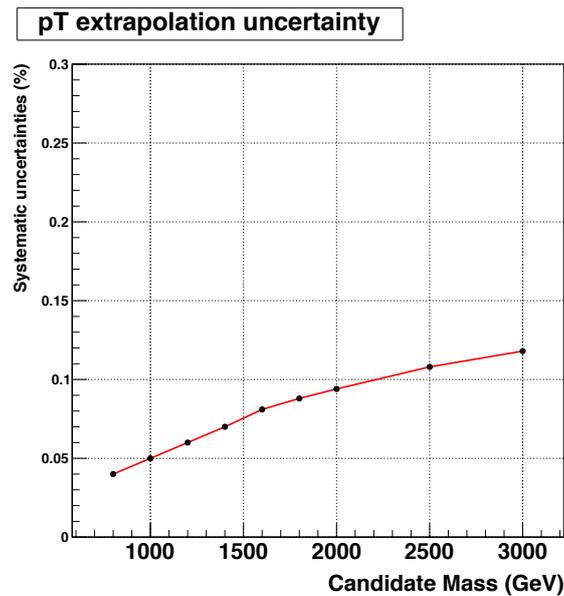


Figure 4.36: Systematic uncertainties due to p_T extrapolation. The figure shows the systematic uncertainties in (%) for different mass points.

in 76X we needed to rerun the E_T^{miss} from the miniAOD and used the resolution files provided by the JME POG. Then we scaled up and down the E_T^{miss} . The associated systematic uncertainty due to the shift of the signal peak varies between 0.02% and 0.13% for signal and is around 3.6% for the subdominant backgrounds.

4.6.16 Cross section

Systematics in the normalization of the subdominant backgrounds of 10%, 20% and 50% were included for the top, diboson and QCD backgrounds respectively to account for the uncertainty in their production cross-sections.

We summarize the systematic uncertainties obtained in this section in the table 4.19. All the uncertainties with values less or equal than 0.5 % were considered to be negligibles in the analysis.

4.6.17 Nuisance parameter impacts

The impact of a nuisance parameter (NP) θ (systematic uncertainties) on a parameter of interest (POI) μ (signal strength) is defined as the shift $\Delta\mu$ that is induced as θ is fixed

Table 4.19: Systematic uncertainties in the analysis.

Source	Signal	Dominant	Subdominant
Luminosity	2.7 %	-	-
Boosted V-Tagging	7% (HP), 26% (LP)	-	7% (HP), 26% (LP)
Background estimation	-	20.5 % (HP), 15.5% (LP)	-
Leptons and Photons ID	0.02 %	-	-
JEC (Jet and E_T^{miss})	0.09-1.76 %	-	3%
JER	0.009-0.006 %	-	-
Factorization and renormalization scales	0.21-0.5 %	-	12 %
PDF	8-18 %	-	17%
Trigger SF	2.5 %	-	-
b-tagging efficiency SF	0.02-0.04%	0.05 %	1 %
Pile up	0.02%	-	2.3 %
Jet mass Calibration	0.13-0.22 %	-	-
Jet mass Resolution	0.93-1.50 %	-	-
p_T extrapolation uncertainty	0.04-0.11 % (HP)	-	-
E_T^{miss} unclustered energy	0.02-0.13 %	-	3.6 %
Cross section	-	-	30 %

and brought to its $+1\sigma$ and -1σ post-fit values, with all other parameters profiled as normal. This is effectively a measure of the correlation between the NP and the POI, and is useful for determining which NPs have the largest effect on the POI uncertainty. The direction of the $+1\sigma$ and -1σ impacts on the POI indicates whether the parameter is correlated or anti-correlated with it. Figure 4.37 shows the impact of the systematic uncertainties for a bulk graviton signal of 1.2 TeV.

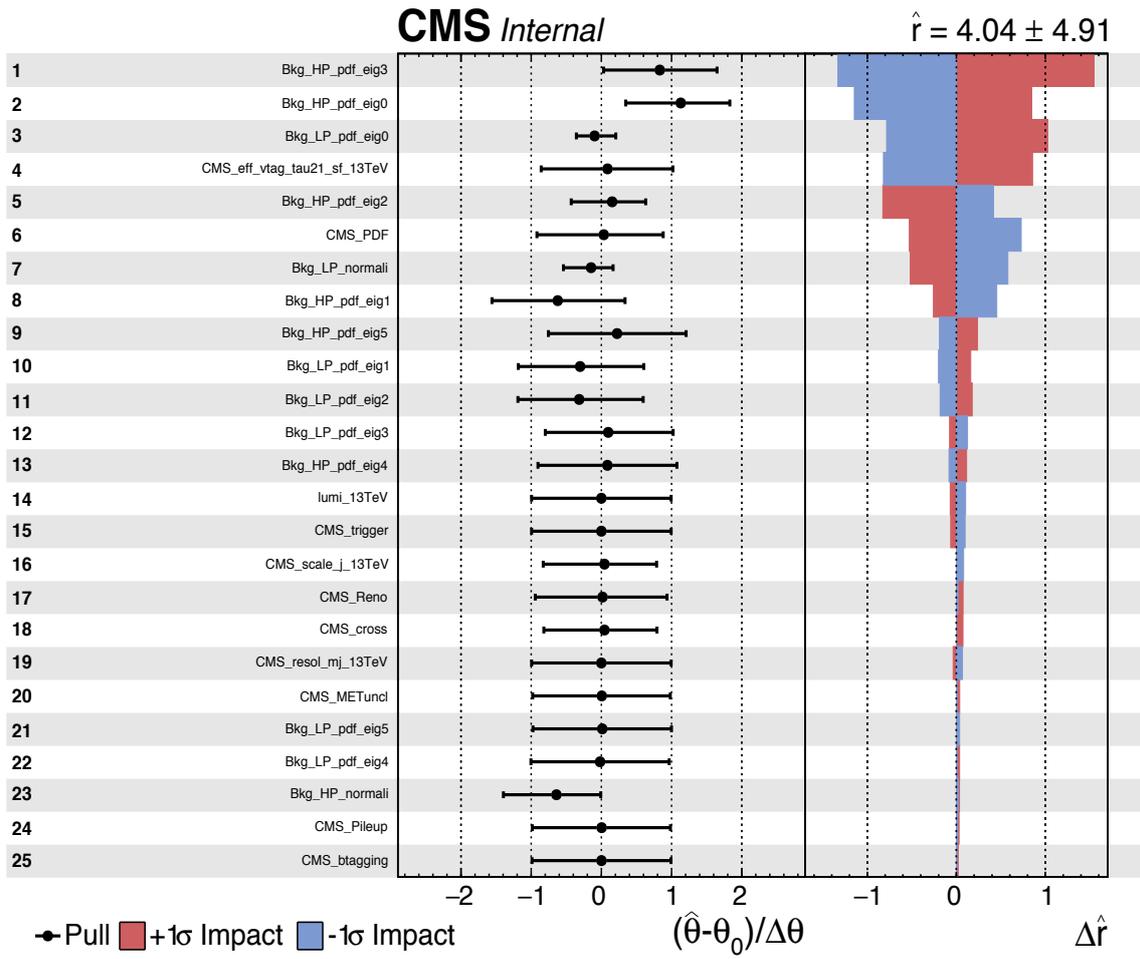


Figure 4.37: Impact of the systematic uncertainties in the analysis.

4.7 Results

In this section, we use the signal shapes M_{VZ}^T presented in Section 4.5, the signal efficiency presented in Section 4.3, the background parameterization introduced in Section 4.4, and the systematic uncertainties discussed in Section 4.6 to perform a search for new resonances in the M_{VZ}^T spectrum of data selected with the criteria summarized in Section 4.3.

Employing the probability density functions (PDFs) obtained from the fits and the respective yields for background and signal, we used an hypothesis test (H_0 : only background and H_1 : signal plus background), to set upper bounds for the signal strength μ ($\mu < \mu^{up}$) in the modified frequentist approach (CLs method) [110–112] (Appendex C.7). The "test statistic" is a profile likelihood ratio [113] with asymptotic approximations [112, 114]. The limits are computed using an unbinned shape analysis. Systematic uncertainties are treated as nuisance parameters, and profiled in the statistical interpretation using log-normal probability distribution functions [115]. The signal strength is defined by the ratio of the number of observed signal events over the number of expected signal events. The translation to the cross section limits are obtained using $\sigma_{Obs} = \mu\sigma_{TH}$.

Upper limits at 95% confidence level (CL) are obtained on the production cross section of a new resonance decaying to the ZZ and WZ final state using the bulk graviton model and the W' model to compute the signal efficiencies respectively, for each of the two categories in the analysis, high-purity and low-purity, under the narrow-width approximation assuming an integrated luminosity of 2.3 fb^{-1} . The resulting limits for each category and their combination are shown in the figures 4.38 and 4.39.

The limits show the cross-sections that are excluded at 95 % confidence level for each resonance mass hypothesis. The continuous black line represent the observed limit (signal + background hypothesis) and the dotted line show the expected limit (background only hypothesis). The green and yellow bands represent the 65% and 95% ranges of expectation, which are denoted by $\pm 1\sigma$ and $\pm 2\sigma$ deviations from the expected limits. For the Bulk graviton hypothesis, the achieved sensitivity is not enough to exclude this model. The theoretical prediction for the HVT model B with coupling constant $g_V = 3$ allow exclusion of masses below 2000 GeV.

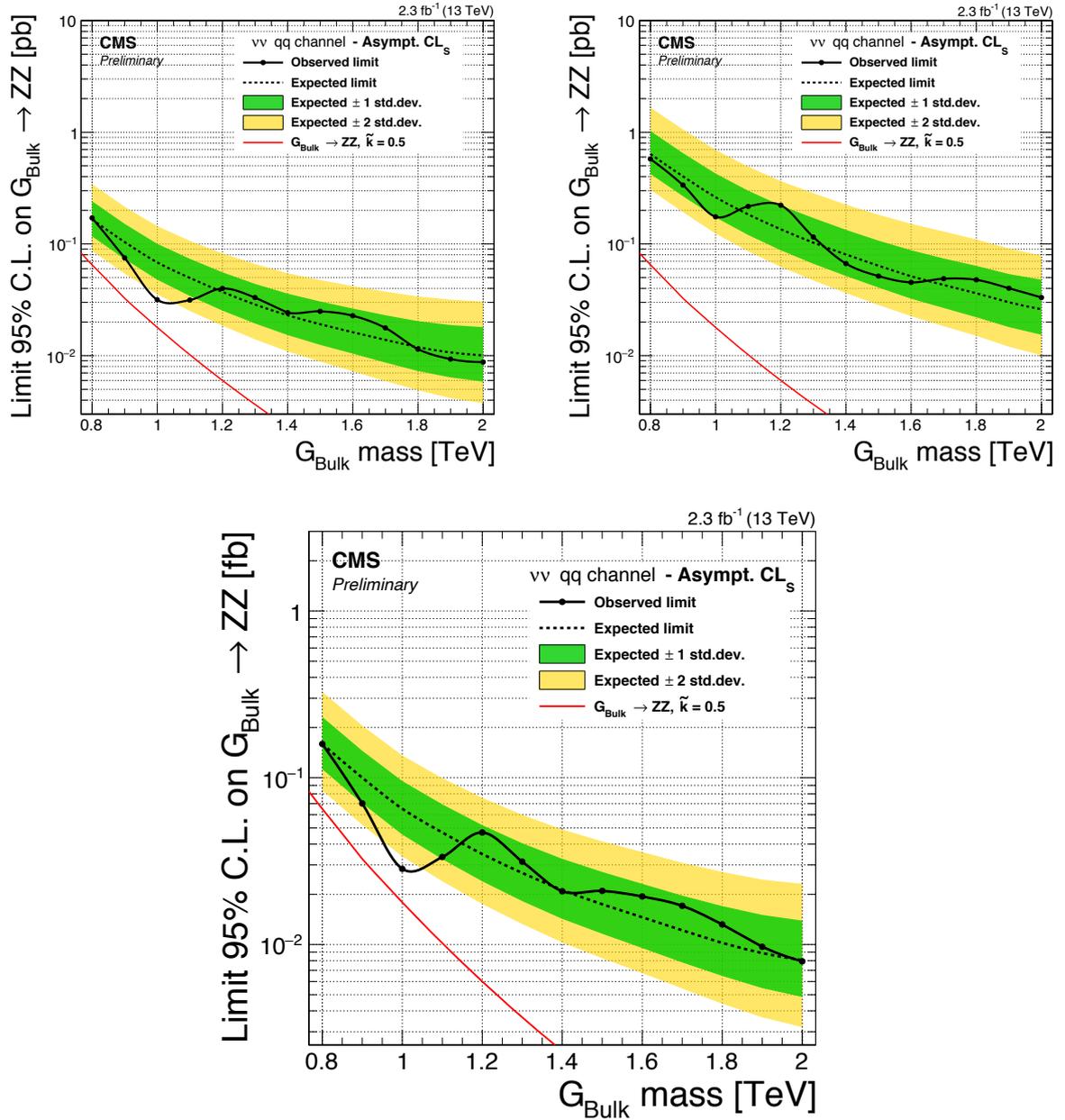


Figure 4.38: Observed and expected 95% CL upper limit on Bulk graviton production cross section times the branching fraction of $G_{\text{bulk}} \rightarrow ZZ$ assuming an integrated luminosity of 2.3 fb^{-1} . The limit is obtained with the Asymptotic CLs technique. Top: (left) Limits for the HP category (right) Limit for the LP category. Bottom: Limits for the combination of both categories.

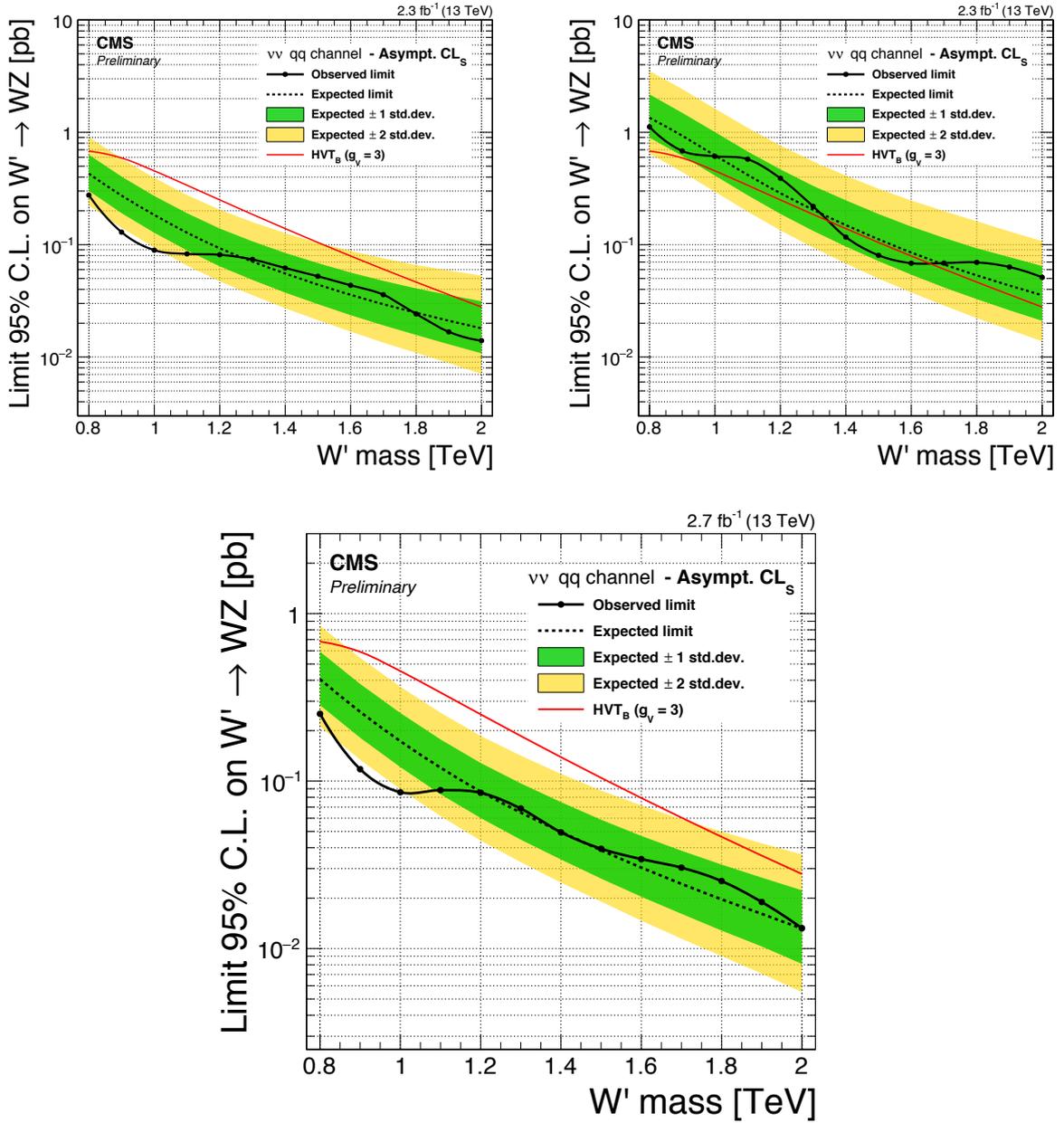


Figure 4.39: Observed and expected 95% CL upper limit on the W' production cross section times the branching fraction of $G_{W'} \rightarrow WZ$ assuming an integrated luminosity of 2.3 fb^{-1} . The limit is obtained with the Asymptotic CL_s technique. Top: (left) Limits for the HP category (right) Limits for the LP category. Bottom: Limits for the combination of both categories.

4.7.1 Crosscheck

As a crosscheck of the asymptotic limit we evaluated the fully frequentist CLs limit with the *The HybridNew method* [116], which is the current recommended method by the LHC Higgs Combination Group. Figure 4.40 shows the comparison between the asymptotic and full CLs limit for the bulk graviton model in the HP and LP categories.

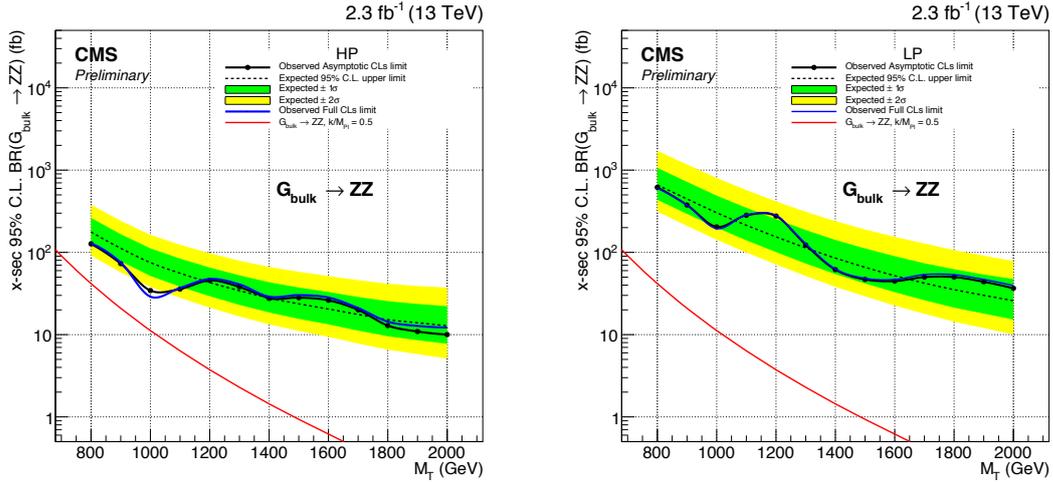


Figure 4.40: Observed and expected 95% CL upper limit on Bulk graviton production cross section times the branching fraction of $G_{bulk} \rightarrow ZZ$ assuming an integrated luminosity of 2.3 fb^{-1} . In the figure we show the comparison between the full CLs method (blue) and the asymptotic limit (black). A good agreement is observed in the limit obtained using both methods.

4.7.2 Signal Significance

Given an observed data sample, claiming the discovery of a new signal requires to determine if the sample is sufficiently inconsistent with the hypothesis that only background is present in the data. A test statistic can be used to measure how consistent or inconsistent the observation is with the hypothesis of the presence of background only. A quantitative measurement of the inconsistency with the background-only hypothesis is given by the *significance*, defined from the probability p (p-value) that the considered test statistics (profile-likelihood ratio [114]) assumes a value greater or equal to the observed one in the case of pure background fluctuation. By convention one claims the “observation” of the signal under investigation if the significance is at least 3σ , which corresponds to a probability of background fluctuation (p-value) of 1.35×10^{-3} . One

claims the “evidence of” the signal (discovery) in case the significance is at least 5σ , corresponding to a p-value of 2.87×10^{-7} [112]. Figure 4.41 shows the signal significance for the Bulk graviton and W' model B hypothesis for masses between 0.8 and 2 TeV. As it can be observed in the analysis, the signal significances for some mass values are less than 1σ which correspond to a probability of background fluctuation (p-value) of 1.59×10^{-1} .

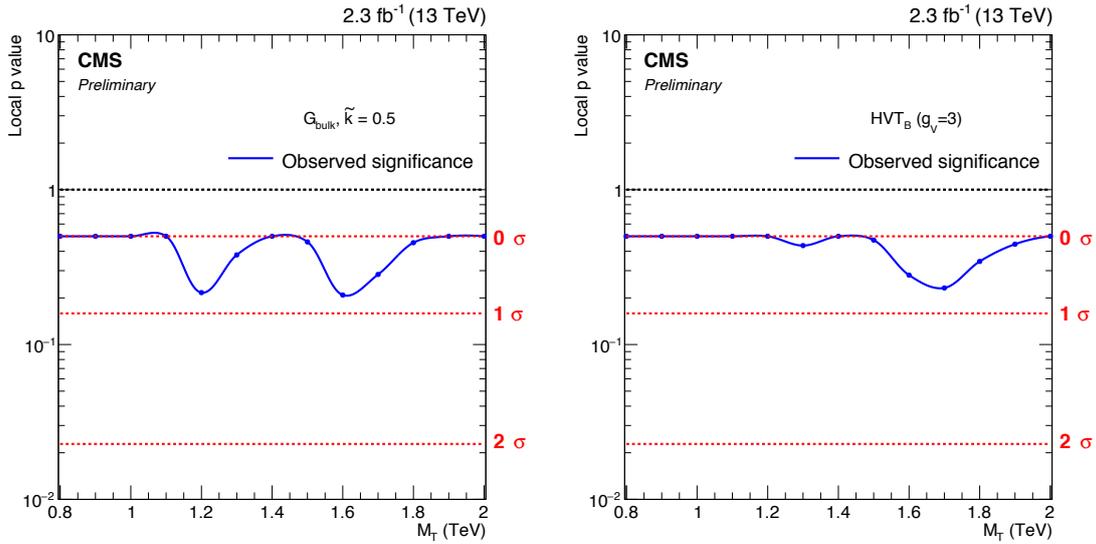


Figure 4.41: Signal significance in terms of p-values for the the Bulk graviton and W' models.

4.7.3 Competitiveness of the result

The results of this analysis are competitive with all other results with 2015 dataset in CMS. Figure 4.42 shows the observed and expected limits in the 10-100 fb range, for masses between 0.8-2 TeV, for three analyses: this one, B2G-16-010 [117] and EXO-15-002 [118]. The outcome of this thesis provides a benchmark setup for the jet+MET channel against which the analysis with the full 2016 dataset can be compared.

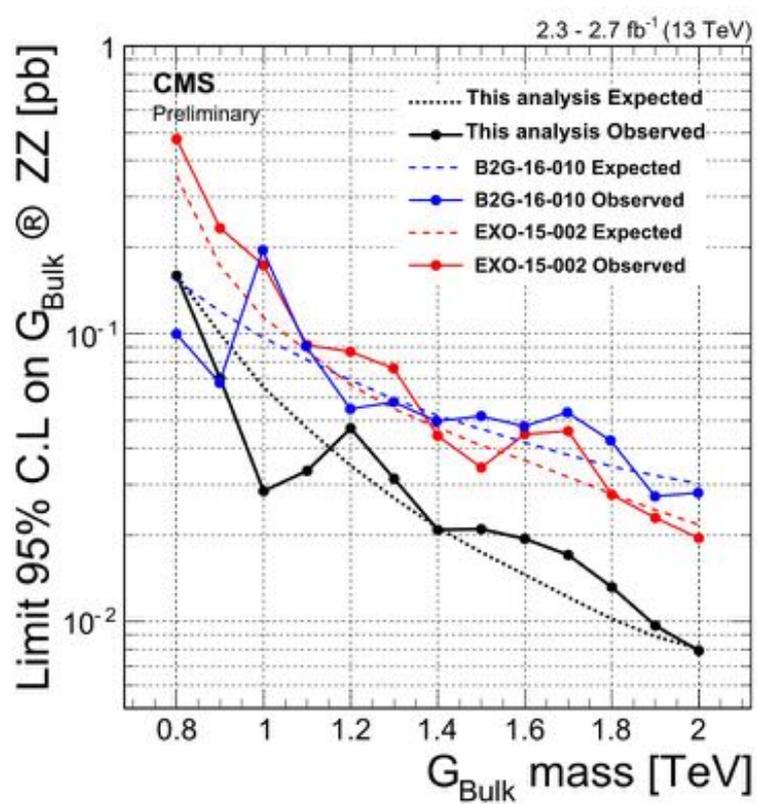


Figure 4.42: Observed and expected limits comparison between this analysis and two similar searches with 2015 dataset.

Chapter 5

Conclusions

Searches for new heavy resonances are one of the main areas of the CMS and ATLAS physics programmes at the Large Hadron Collider (LHC) at CERN [119, 120]. One interesting scenario is the case in which the resonance decays in a pair of vector bosons [121]. Diboson resonances are predicted in several extensions of the SM such as models with vectorial heavy resonances, extra dimensions and composite Higgs. These models favor the decay of the resonance in diboson while reduce the branching fractions into fermions [17, 26]. The search for high-mass resonances decaying into vector bosons benefits greatly from the increase in centre-of-mass energy of proton–proton collisions at the Large Hadron Collider (LHC) from 8 TeV to 13 TeV.

CMS and ATLAS have searched for heavy diboson resonances in various final states during the Run 1 and Run 2 periods. Both collaborations have dedicated and experienced groups working on the diboson signatures such as: $X \rightarrow VV \rightarrow \text{jet}+\text{jet}$, $X \rightarrow VW \rightarrow \text{jet}+\ell + E_T$, $X \rightarrow VZ \rightarrow \text{jet}+\ell\ell$ and $X \rightarrow VZ \rightarrow \text{jet}+\nu\nu$ [122–126].

In this thesis, we presented a search for high-mass resonances X decaying into $VZ \rightarrow q\bar{q}'\nu\bar{\nu}$ (where V can be a Z or a W) with a final state composed of a highly energetic jet and large missing transverse energy. Results are based on an integrated luminosity of 2.3 fb^{-1} of proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$ delivered by the LHC and collected with the CMS detector at CERN. Jet substructure were used to separate the signal from the standard model background when the boost of the V causes the hadronization products of the two quarks to merge into one single fat jet. No significant excess of events in data above the predicted background has been observed. In the absence of evidence

on the existence of a new resonance, we set 95% CLs upper limits on bulk graviton and HVT(W') production cross section time branching fraction ($X \rightarrow VZ$) for resonance masses between 0.8 and 2.0 TeV, interpreting the result in different BSM contexts.

A similar search with the same final state is currently being performed by CMS but using data collected during the 2016 year with an integrated luminosity of 12.9 fb^{-1} .

Appendix A

HLT description and efficiency

A.1 Trigger Paths

In the analysis we search for a signal with one energetic jet emerging from a boosted boson (Z or W) and large amount of missing transverse energy (E_T^{miss}). Therefore, we employ a data sample that was collected using triggers prepared to select events that have E_T^{miss} and a quantity associated to jets (H_T^{miss}). The main signal trigger path chosen is:

- `PFMETNoMu90_JetIdCleaned_PFMHTNoMu90_IDTight`

The E_T^{miss} (MET) is defined as the negative vectorial sum of the transverse momentum of all the particles in the event and the H_T^{miss} (MHT) is defined as the negative vectorial sum of the transverse momentum of all the jets in the event. Both quantities were calculated using particle flow (PF) objects adding back the vector momenta of all PF muons (NoMu). The main trigger path present thresholds of 90 GeV in E_T^{miss} and H_T^{miss} . In order to prevent collecting noise events online that would be discarded offline, some additional requirements are included in the trigger path. They rely on the online application of the jet energy corrections for jets with transverse momentum bigger than 20 GeV, a selection on neutral hadron energy fraction ($\text{NHF} < 0.9$) and tight jet identification conditions in the calculation of the H_T^{miss} .

In addition, an inclusive E_T^{miss} trigger path is employed (in an OR) with the main signal trigger path:

- HLT_PFMET170_*

This path acts as a support trigger to gain acceptance at high E_T^{miss} in order to recover inefficiency of the main trigger path. The missing transverse momentum threshold applied in the online selection is 170 GeV, where the E_T^{miss} is calculated using the PF algorithm. The signal HLT paths are seeded at level 1 (in an OR) by L1ETM50, L1ETM60 and L1ETM70.

A.2 Trigger Efficiency

In order to measure the trigger efficiency in data we used the reference trigger method with an unbiased data sample of single muon events, collected with the trigger path HLT_IsoMu20_v* and grouped in the Primary Dataset SingleMu. For the measurement of the efficiency in MC an inclusive W + jets sample was used.

For the offline selection we imposed the following conditions in Data and MC:

- We required at least one muon with the following properties:
 - $p_T > 10 \text{ GeV}$, $|\eta| < 2.5$
 - Tight muon ID
 - Relative isolation with PU correction less than 0.15
- We required at least one ak4jet with the following properties:
 - $p_T > 100 \text{ GeV}$, $|\eta| < 2.5$
 - Loose jet ID
 - Cleaning cuts : CHF > 0.1, NHF < 0.8
 - $\Delta R(\text{jet}, \text{leptons}) > 0.5$
- General requirements:
 - At least one good offline primary vertex
 - Events passing the E_T^{miss} filters
 - $\min \Delta\phi(\text{AK4Jet}, E_T^{\text{miss}}) > 0.4$

$$- \frac{|\text{CALO MET} - \text{MET}|}{\text{MET}} < 0.5$$

The performance of the trigger efficiency is measured in Data and MC, using the definition:

$$\text{efficiency} = \frac{\text{passed}(\text{HLT_IsoMu20} \ \&\& \ (\text{OR of Signal Triggers}) \ \&\& \ \text{Offline Selection})}{\text{passed}(\text{HLT_IsoMu20} \ \&\& \ \text{Offline Selection})} \quad (\text{A.1})$$

The numerator will register the events that pass the HLT_IsoMu20_v* and (PFMETNoMu90_JetIdCleaned_PFMHTNoMu90_IDTight OR HLT_PFMET170_*) and the offline selection. While the denominator will register only the events that pass HLT_IsoMu20_v* and the offline selection.

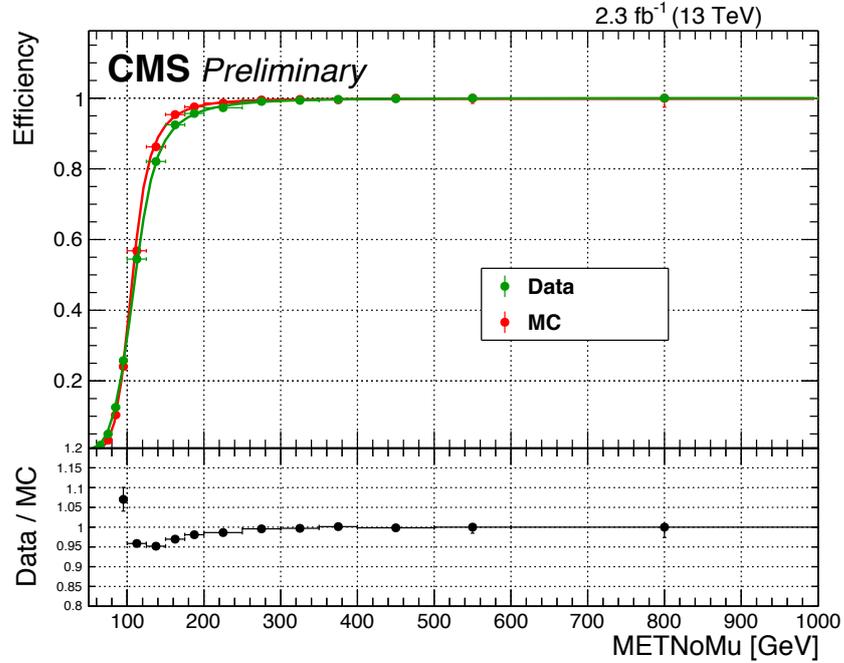


Figure A.1: High Level Trigger efficiency as function of the transverse missing energy without muons for the selected paths. The figure shows the turn-on curve for data and MC simulation and the Data/MC ratio.

For events passing the selection $E_T^{\text{miss}}(\text{NoMu}) > 250$ GeV, the trigger presents an efficiency of around 99%.

In the analysis, the trigger decision is applied on both data and MC and residual Data/MC scale factors obtained from the Data/MC ratio are imposed to the MC with the resulting uncertainty considered later on as systematics.

Table A.1: Trigger efficiency for different E_T^{miss} (NoMu) values in Data and MC.

E_T^{miss} (NoMu) value (GeV)	Trigger efficiency in MC (%)	Trigger efficiency in Data (%)
200	98	96.5
250	99.2	98.5
300	99.5	99.2
350	99.6	99.5
400	99.7	99.7
500	99.79	99.8
1000	99.8	100

A.3 Crosscheck of trigger efficiency measurement

As a crosscheck of the previous efficiency measurement we used a SingleElectron data sample for data and a W +jets sample for MC with the following reference trigger paths:

- HLT_Ele32_eta2p1_WPTight_Gsf_v*
- HLT_Ele105_CaloIdVT_GsfTrkIdT_v*

For the offline selection we imposed the following conditions in Data and MC:

- We required at least one electron with the following properties:
 - $p_T > 40$ GeV, $|\eta| < 2.5$
 - Tight electron ID
- We require at least one ak4jet with the following properties:
 - $p_T > 100$ GeV, $|\eta| < 2.5$
 - Loose jet ID
 - Cleaning cuts : CHF > 0.1 , NHF < 0.8
 - ΔR (jet,leptons) > 0.5

- General requirements:
 - At least one good offline primary vertex
 - Events pass the E_T^{miss} filters
 - $\min \Delta\phi (\text{AK4Jet}, E_T^{\text{miss}}) > 0.4$
 - $\frac{|\text{CALO MET} - \text{MET}|}{\text{MET}} < 0.5$

The performance of the trigger efficiency is measured in Data and MC using a similar definition as the one given in (A.1). In the figure A.2 we show the turn-on curve for the high level trigger efficiency as function of the transverse missing energy for data and W+jets events.

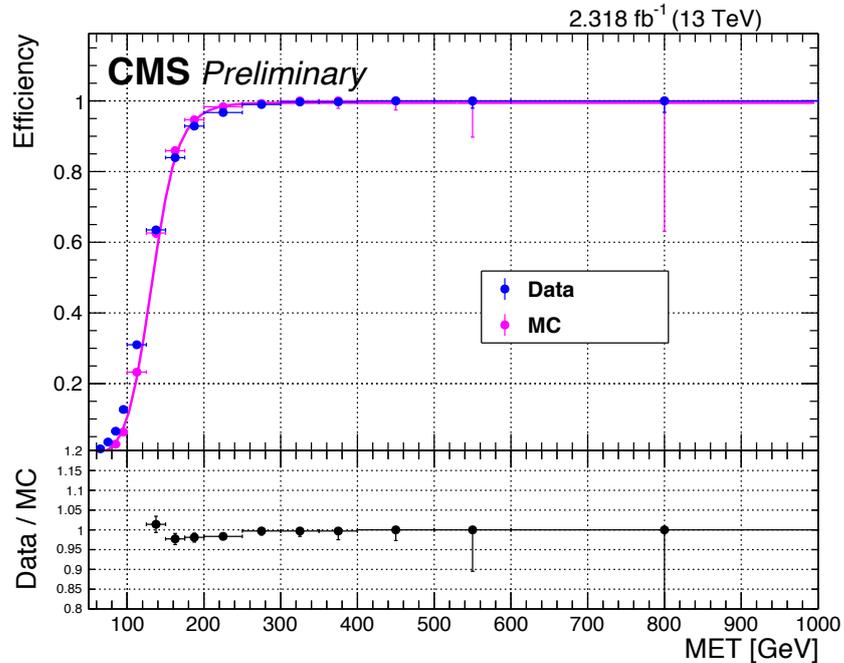


Figure A.2: High Level Trigger efficiency as function of the transverse missing energy for the selected paths. The figure shows the turn-on curve for data and MC simulation.

For events passing the selection $E_T^{\text{miss}} > 250$ GeV, the trigger is 99% efficient. As can be observed from the figures A.1, A.2 and the tables A.2, A.1, we obtained similar results for the trigger efficiency using the SingleMu or the SingleElectron dataset. In addition, the figure A.3 shows the trigger turn on curves for MC signal samples, where the analysis selection was applied.

Table A.2: Trigger efficiency for different E_T^{miss} values in Data and MC.

E_T^{miss} value (GeV)	Trigger efficiency in MC (%)	Trigger efficiency in Data (%)
200	96.5	94.5
250	99.1	98.5
300	99.4	99.4
350	99.46	99.7
400	99.47	99.7
500	99.47	99.7
1000	99.47	99.7

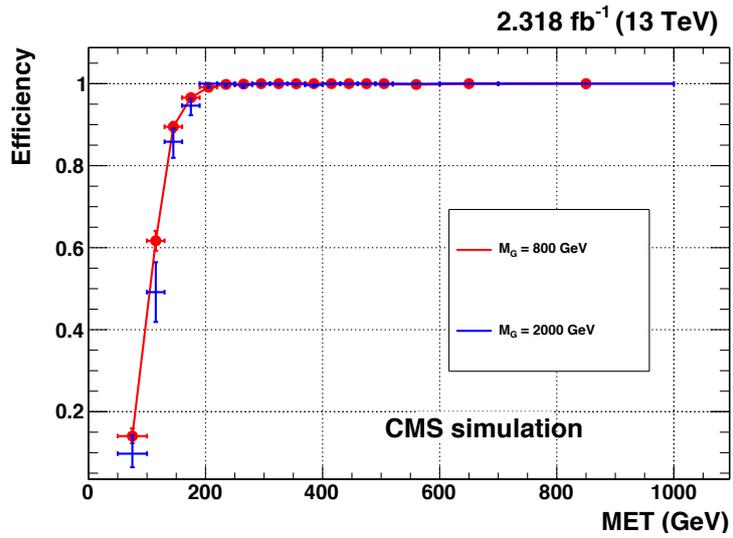


Figure A.3: High level trigger efficiency as function of the transverse missing energy for the selected path in RS signal samples. The figure shows the turn-on curve for different mass points.

Appendix B

Recoil Correction

To derive this kind of correction, we used a singleMuon sample for data and a DYJet-sToLL for MC. For the $Z \rightarrow \mu\mu$ process we used the following selection:

In Data:

- Trigger : HLT_IsoMu20_v* OR HLT_IsoTkMu20_v*
- Good offline primary vertex
- $p_T > 25 \text{ GeV}$, $|\eta| < 2.4$
- Tight muon ID
- Relative isolation with $\Delta\beta$ correction of less than 0.15
- $\Delta R(\text{muon}, \text{muon}) > 0.5$
- $60 < m_Z < 120 \text{ GeV}$
- Opposite charge muons

In MC we applied the same selection, but in addition:

- PU reweighting
- Muon Id/Iso and trigger scale factors
- JEC

Figure B.1 shows the main kinematic properties and recoil variables after the selection.

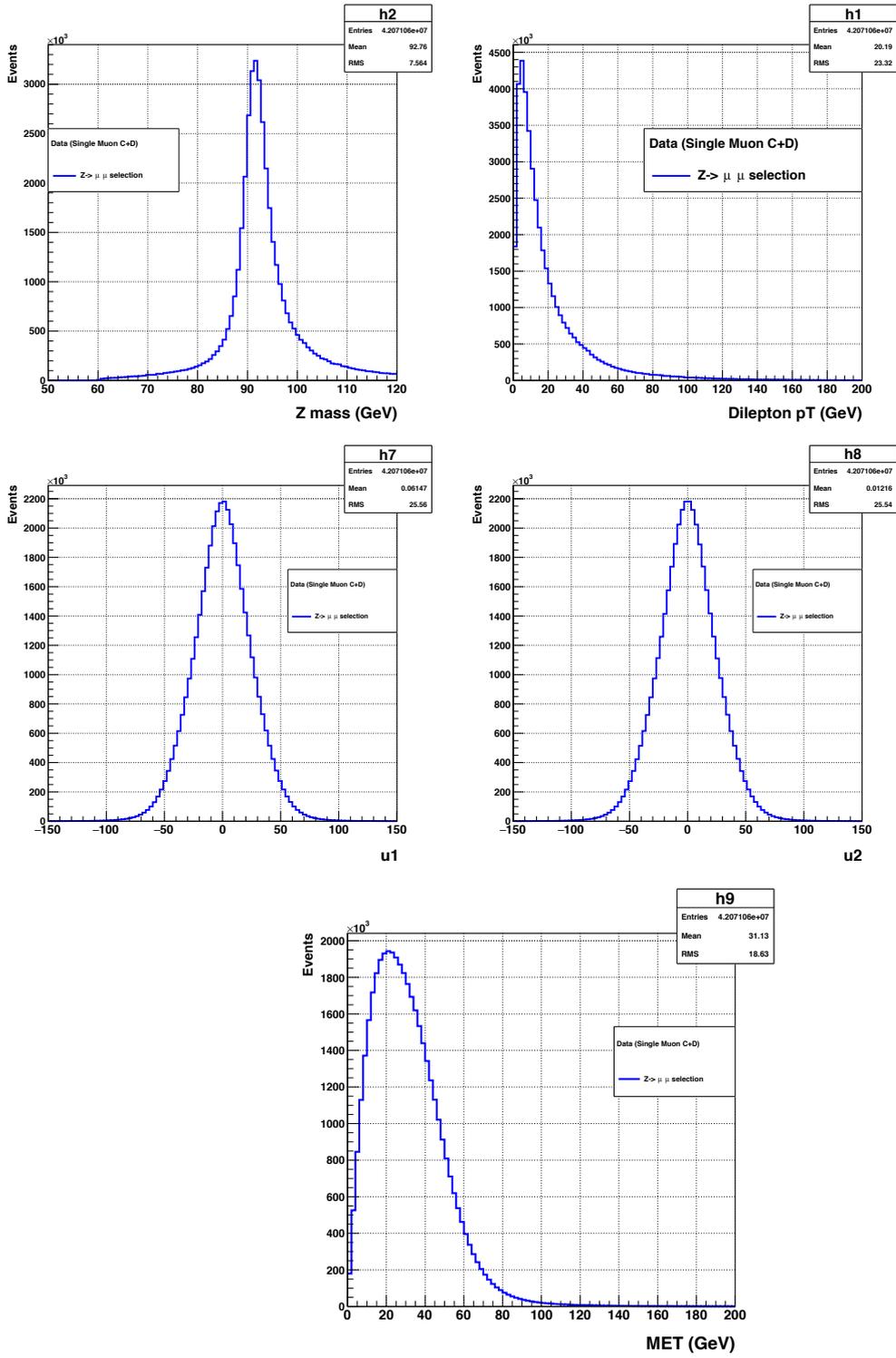


Figure B.1: Kinematic and recoil properties in the $Z \rightarrow \mu\mu$ process in Data after applying the selection .

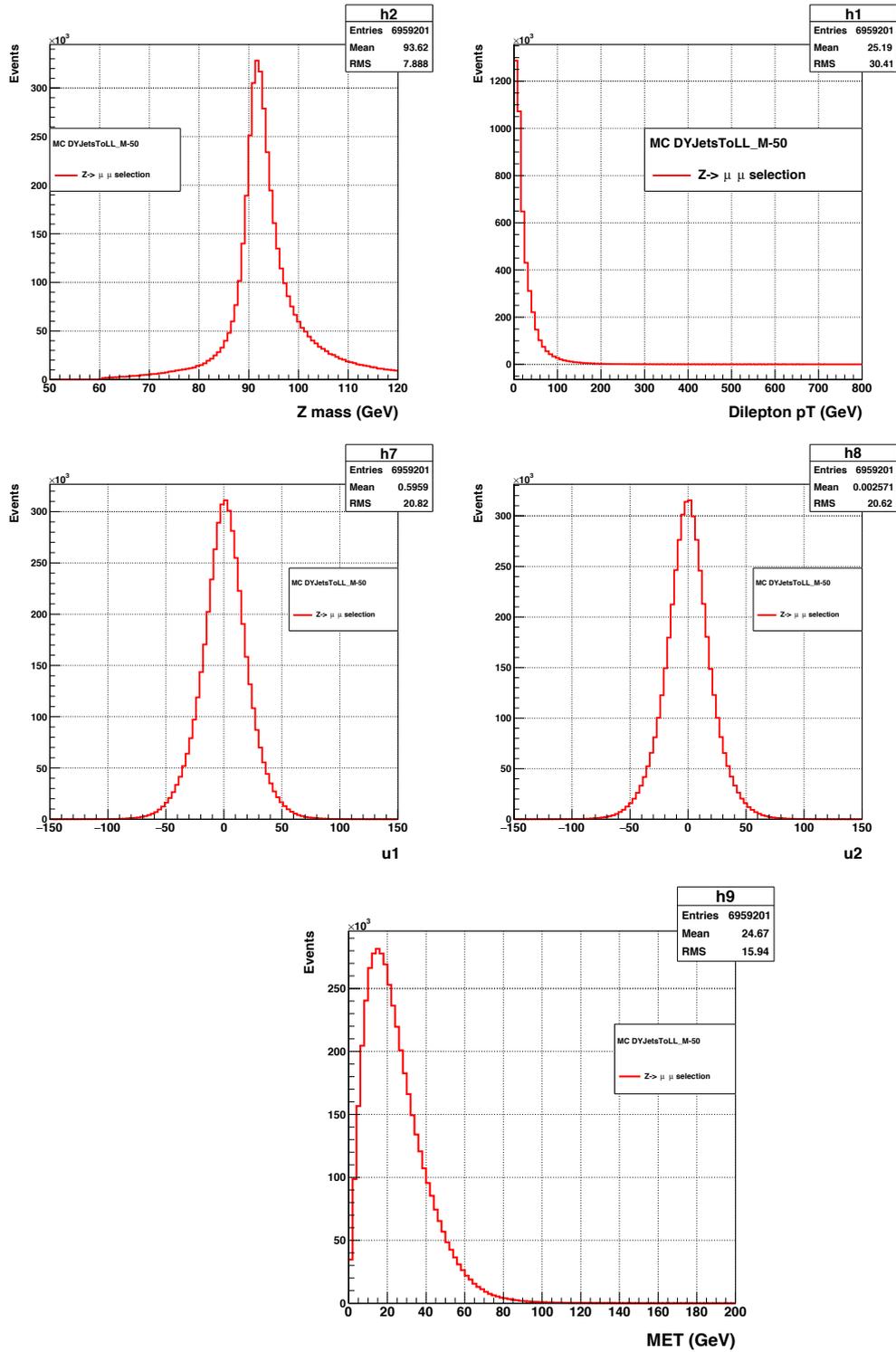


Figure B.2: Kinematic and recoil properties in the $Z \rightarrow \mu\mu$ process in MC after applying the selection .

Figures B.3, B.4 B.5, B.6 show the fit of the recoil in the parallel (u_1) and perpendicular (u_2) directions of the boson p_T with a double gaussian model in different bins of the Z p_T .

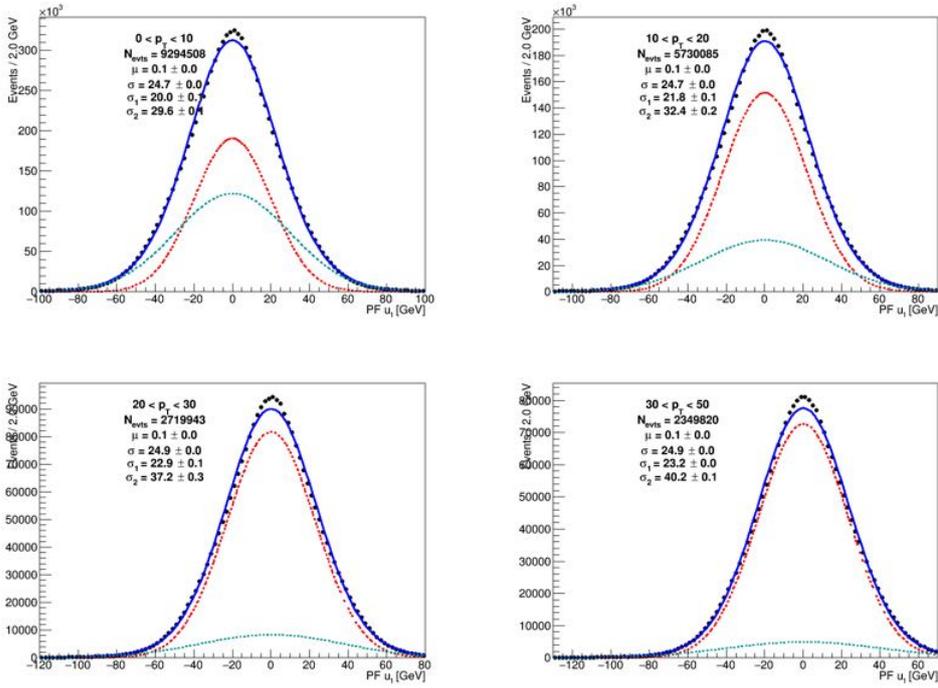


Figure B.3: Fits on the parallel (u_1) components of the recoil in data with a double gaussian model.

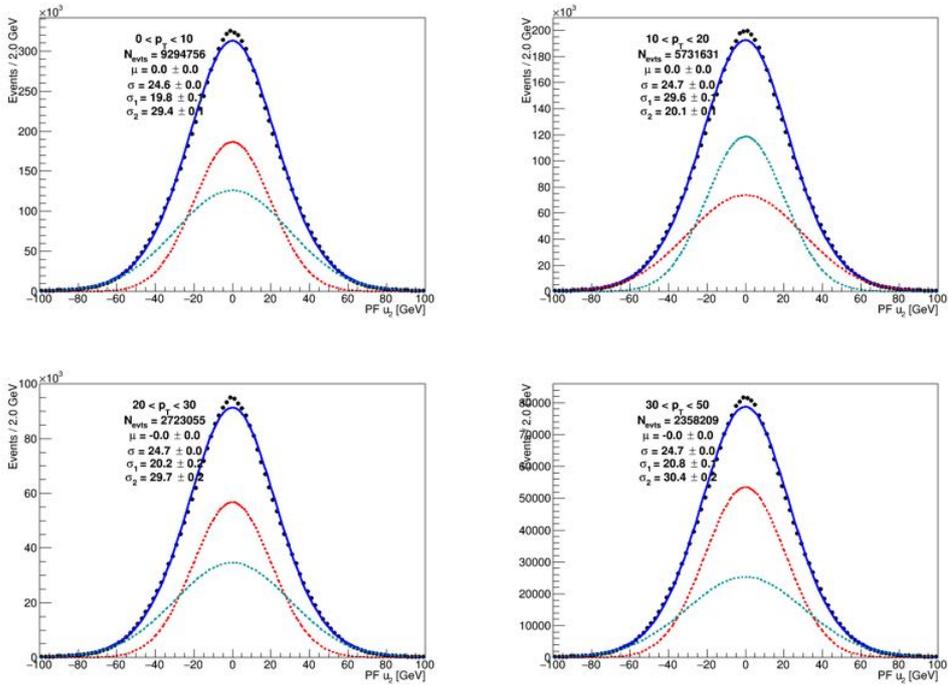


Figure B.4: Fits on the perpendicular (u_2) components of the recoil in data with a double gaussian model.

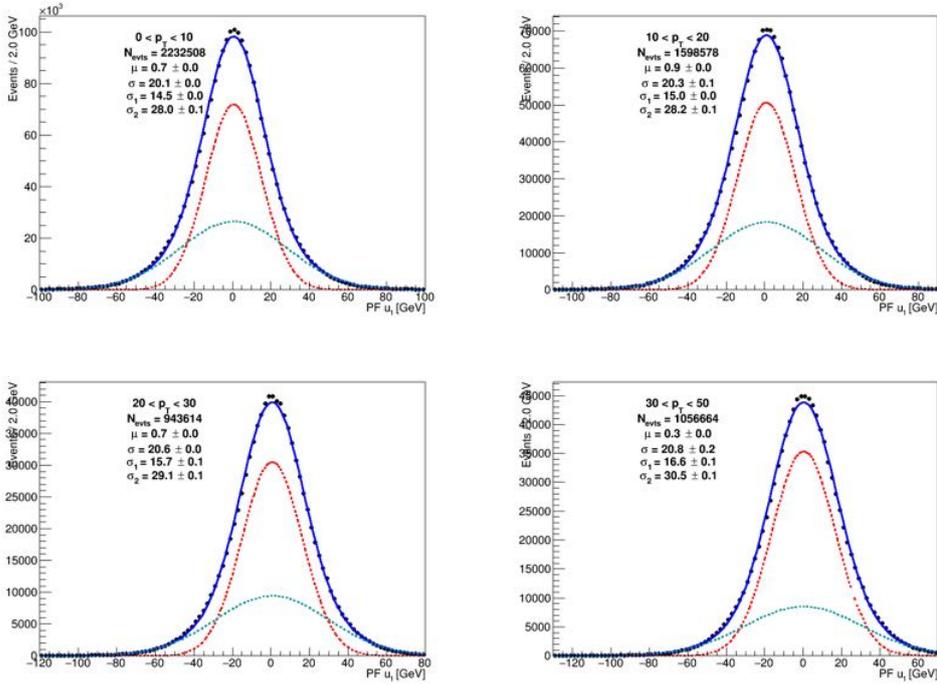


Figure B.5: Fits on the parallel (u_1) components of the recoil in MC with a double gaussian model

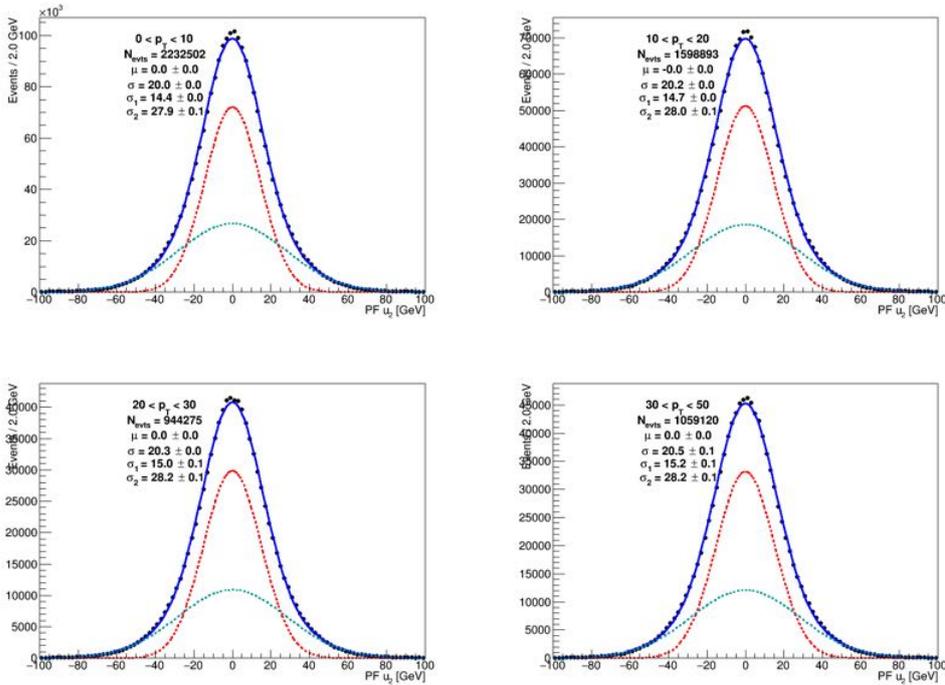


Figure B.6: Fits on the perpendicular (u_2) components of the recoil in MC with a double gaussian model

Statistics

C.1 Definition of Probability Distribution Function (PDF)

Consider a sample space¹:

$$\vec{x} = (x_1, \dots, x_n) \in \Omega \subseteq \mathbb{R}^n \quad (\text{C.1})$$

Each random extraction (an experiment, in our case of interest) will lead to an outcome (measurement) of one point \vec{x} in the sample space Ω . We can associate to any point \vec{x} a probability density $f(\vec{x}) = f(x_1, \dots, x_n)$ which is a real value greater or equal to zero. The probability of an event A , where $A \subseteq \Omega$ (the probability that $\vec{x} \in A$) is given by:

$$P(A) = \int_A f(x_1, \dots, x_n) d^n x \quad (\text{C.2})$$

The function f is called probability distribution function (PDF) and have the following property:

$$\int_{\Omega} f(x_1, \dots, x_n) d^n x = 1 \quad (\text{C.3})$$

¹The main reference used in this appendix is [112].

C.2 Nuisance Parameters and Systematic Uncertainties

In order to define the PDF describing a data model, in many cases it is necessary to introduce parameters that are not of direct interest to our problem. For instance, when determining (“fitting”) the yield of a signal peak, it is sometimes needed to determine from data other parameters, like the experimental resolution that gives the peak width, detector efficiencies that are needed to determine the signal production yield from the measured signal yield, parameters to define the shapes and amounts of possible backgrounds, and so on. Those parameters are often referred to as nuisance parameters. In some cases, nuisance parameters can not be determined from the same data sample used to measure the parameters of interest and their estimate should be taken from other measurements. The uncertainty on their determination, external to the considered fit problem, will reflect into uncertainties on the estimate of parameters of interest. Uncertainties due to the propagation of imperfect knowledge of nuisance parameters that can not be constrained from the same data sample used for the main fit of the parameters of interest gives rise to systematic uncertainties, while uncertainties purely related to the fit are referred to as statistical uncertainties.

C.3 Likelihood Function

Given a sample (x_1, \dots, x_n) of n random variables whose PDF is known and depends on m parameters, $(\theta_1, \dots, \theta_m)$, the likelihood function is defined as the probability density at the point (x_1, \dots, x_n) given a fixed set of values of the $(\theta_1, \dots, \theta_m)$.

In our particular case, we define the likelihood function as the PDF that characterizes our set of experimental observables, evaluated at the values of those observables that corresponds to our data sample, for given values of the unknown parameters. If we measure the values x_1, \dots, x_n of n random variables and our PDF model depends on m unknown parameters $\theta_1, \dots, \theta_m$ we define the likelihood function L as:

$$L(x_1, \dots, x_n; \theta_1, \dots, \theta_m) = f(x_1, \dots, x_n; \theta_1, \dots, \theta_m) \quad (\text{C.4})$$

where f is the (joint) PDF of the random variables x_1, \dots, x_n . If we have N repeated

measurements each consisting of the n values of the random variables (x_1, \dots, x_n) we can consider the probability density corresponding to the total sample:

$$\vec{x} = \left\{ (x_1^1, \dots, x_n^1), \dots, (x_1^N, \dots, x_n^N) \right\} \quad (\text{C.5})$$

If we assume that the events are independent of each other, the likelihood function of the sample consisting of the N events can be written as the product of the PDFs corresponding to the measurement of each single event, i.e.:

$$L(\vec{x}; \vec{\theta}) = \prod_{i=1}^N f(x_1^i, \dots, x_n^i; \theta_1, \dots, \theta_m) \quad (\text{C.6})$$

C.4 Hypothesis Tests

A hypothesis test is a statistical procedure that is designed to test a claim. One example could be to determine whether a sample of events is composed of background only or contains a mixture of background plus signal events. In statistical literature when two hypotheses are present, one is called the null hypothesis (H_0), and the other is called the alternative hypothesis (H_1). So, the first step in a hypothesis testing is to state the null and alternative hypothesis.

The next step is to define a "test statistic". In general, a test statistic is selected or defined in such a way as to quantify, within observed data, behaviours that would distinguish the null from the alternative hypothesis

A test statistic that ensures the optimal performance is provided by the Neyman–Pearson lemma [127]. According to this lemma, such test statistic is defined as the ratio of the likelihood functions evaluated for the observed data sample \vec{x} under the two hypotheses H_1 and H_0 :

$$\lambda(\vec{x}) = \frac{L(\vec{x}|H_1)}{L(\vec{x}|H_0)} \quad (\text{C.7})$$

C.5 Likelihood Ratio in the Search for a New Signal

Two hypotheses H_1 and H_0 are represented as two possible sets of values Θ_1 and Θ_0 of the parameters $\vec{\theta} = (\theta_1, \dots, \theta_N)$ that characterize the PDFs. Usually we want to use the number of events N as information in the likelihood definition, hence we multiply it by a Poissonian factor corresponding to the probability to observe a number of events N :

$$L(\vec{x}_1, \dots, \vec{x}_N; \theta_1, \dots, \theta_m) = \frac{e^{-\nu(\vec{\theta})} \nu(\vec{\theta})^N}{N!} \prod_{i=1}^N f(\vec{x}_i; \vec{\theta}) \quad (\text{C.8})$$

In the Poissonian term the expected number of events ν may also depend on the parameter $\vec{\theta}$: $\nu = \nu(\vec{\theta})$. We want to discriminate between two hypotheses, which are the presence of only background events in our sample, i.e.: $\nu = b$, against the presence of both signal and background, i.e.: $\nu = \mu s + b$. Here we have introduced the multiplier μ , called *signal strength*, assuming that the expected signal yield from theory is s . We consider all possible values of the expected signal yield, given by μs , by varying μ while keeping s constant at the value predicted from theory.

The hypothesis H_0 corresponding to the presence of background only is equivalent to $\mu = 0$, while the hypothesis H_1 corresponding to the presence of signal plus background allows any non-null positive value of μ .

The PDF $f(\vec{x}_i; \vec{\theta})$ can be written as superposition of two components, one PDF for the signal and another for the background, weighted by the expected signal and background fractions, respectively:

$$f(\vec{x}_i; \vec{\theta}) = \frac{\mu s}{\mu s + b} f_s(\vec{x}_i; \vec{\theta}) + \frac{b}{\mu s + b} f_b(\vec{x}_i; \vec{\theta}) \quad (\text{C.9})$$

In this case the likelihood is then:

$$L_{s+b}(\vec{x}_1, \dots, \vec{x}_M; \mu; \vec{\theta}) = \frac{e^{-(\mu s(\vec{\theta}) + b(\vec{\theta}))}}{N!} \prod_{i=1}^N \left(\mu s f_s(\vec{x}_i; \vec{\theta}) + b f_b(\vec{x}_i; \vec{\theta}) \right) \quad (\text{C.10})$$

Under the hypothesis H_0 ($\mu = 0$) the likelihood function can be written as:

$$L_b(\vec{x}_1, \dots, \vec{x}_M; \vec{\theta}) = \frac{e^{-b(\vec{\theta})}}{N!} \prod_{i=1}^N \left(b f_b(\vec{x}_i; \vec{\theta}) \right) \quad (\text{C.11})$$

The likelihood ratio:

$$\lambda(\mu, \vec{\theta}) = \frac{L_{s+b}(\vec{x}_1, \dots, \vec{x}_M; \mu; \vec{\theta})}{L_b(\vec{x}_1, \dots, \vec{x}_M; \vec{\theta})} \quad (\text{C.12})$$

$$= e^{-(\mu s(\vec{\theta}))} \prod_{i=1}^N \left(\frac{\mu s f_s(\vec{x}_i; \vec{\theta})}{b f_b(\vec{x}_i; \vec{\theta})} + 1 \right) \quad (\text{C.13})$$

Taking the negative logarithm of the likelihood function, we have:

$$-\ln \lambda(\mu, \theta) = \mu s(\vec{\theta}) - \sum_{i=1}^N \ln \left(\frac{\mu s f_s(\vec{x}_i; \vec{\theta})}{b f_b(\vec{x}_i; \vec{\theta})} + 1 \right) \quad (\text{C.14})$$

C.6 Upper limits

The goal of many experiments is to search for new physics phenomena. If an experiment provides a convincing measurement of a new signal, the result should be published claiming a discovery. If the outcome is not sufficiently convincing, the publication can anyway quote an upper limit to the “intensity” of the new signal, which usually allow to exclude parameter sets of a new theory.

In the frequentist approach, the achieved significance level is used to claim a discovery. The significance measures the probability that, in case of presence of background only, a statistical fluctuation in data might have produced by chance the observed features that are interpreted as a new signal.

In case no convincing new signal is observed, in many cases it is interesting to quote as result of the search for the new phenomena, the upper limit on the expected yield of the hypothetical new signal (limits on the signal strength: μ). From upper limits of the signal yield it is often possible to indirectly derive limits on the properties of the new signal.

In the frequentist approach the procedure to set an upper limit is a special case of determination of the confidence interval for the unknown parameter (signal strength: μ). Instead of estimating the parameter (μ) by a single value, an interval that is likely to include the parameter is given. Thus, confidence intervals are used to indicate the reliability of an estimate. How likely the interval contains the parameter is determined by

the confidence level. Increasing the desired confidence level will widen the confidence interval.

In order to determine an upper limit instead of a central interval, the choice of the interval with the desired confidence level (95 %, for example) should be fully asymmetric, becoming $\mu \in [0, \mu^{\text{UP}})$. When the outcome of an experiment is an upper limit, one usually quotes:

$$\mu < \mu^{\text{UP}} \quad \text{at 95 \% confidence level (C.L.)} \quad (\text{C.15})$$

C.7 Modified Frequentist Approach: The CL_S Method

The so-called modified frequentist approach will be illustrated using the test statistic adopted in the original proposal, which is the ratio of the likelihood functions evaluated under two different hypotheses: the presence of signal plus background (H_1 , corresponding to the likelihood function L_{s+b}), and the presence of background only (H_0 , corresponding to the likelihood function L_b).

In order to quote an upper limit using the frequentist approach, the distribution of the test statistics λ (or equivalently $-2 \ln \lambda$) in the hypothesis of signal plus background ($s + b$) has to be known, and the p-value corresponding to the observed value $\lambda = \hat{\lambda}$ (denoted below as CL_{s+b}) has to be determined.

The proposed modification to the purely frequentist approach consists of finding two p-values corresponding to both the $s + b$ and b hypotheses (here the set of parameters $\vec{\theta}$ also includes μ):

$$CL_{s+b}(\vec{\theta}) = P_{s+b}(\lambda(\vec{\theta}) \leq \hat{\lambda}), \quad (\text{C.16})$$

$$CL_b(\vec{\theta}) = P_b(\lambda(\vec{\theta}) \leq \hat{\lambda}) \quad (\text{C.17})$$

From those two probabilities, the following quantity can be derived:

$$CL_s(\vec{\theta}) = \frac{CL_{s+b}(\vec{\theta})}{CL_b(\vec{\theta})} \quad (\text{C.18})$$

Upper limits are determined excluding the range of the parameters of interest (e.g.: the signal strength μ) for which $CL_s(\vec{\theta})$ is lower than the conventional exclusion confidence level, typically 95 %. For this reason, the modified frequentist approach is often referred to as the CL_s method.

Bibliography

- [1] S. L. Glashow, Nucl. Phys. **22**, 579 (1961).
- [2] S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967).
- [3] A. Salam, Eighth Nobel Symposium, edited by Almqvist and Wiksell p. 367 (1968).
- [4] G. Aad et al. (ATLAS Collaboration), Phys. Lett. B **716**, 1 (2012).
- [5] S. Chatrchyan et al. (CMS Collaboration), Phys. Lett. B **716**, 30 (2012).
- [6] A. Benvenuti et al., Phys. Rev. Lett. **34**, 419 (1975).
- [7] G. Arnison et al. (UA1 Collaboration), Phys. Lett. B **122**, 103 (1983).
- [8] G. Arnison et al. (UA1 Collaboration), Phys. Lett. B **126**, 398 (1983).
- [9] P. Bagnaia et al. (UA2 Collaboration), Phys. Lett. B **129**, 130 (1983).
- [10] F. Abe et al. (CDF Collaboration), Phys. Rev. Lett. **74**, 2626 (1995).
- [11] DELPHI, OPAL, ALEPH, L3, Nucl. Instrum. Meth. **A378**, 101 (1996).
- [12] S. W. SLD Collaboration, Proceedings of the 24th Annual SLAC Summer Institute on Particle Physics: The Strong Interaction, From Hadrons to Partons, edited by L. Dixon, J. Chan, L. DePorcel (Springfield, NTIS) p. 405 (1997).
- [13] C. Caso et al., Eur. Phys. J. C **3**, 1 (1998).
- [14] N. Arkani-Hamed, S. Dimopoulos, and G. Dvali, Phys. Lett. B **429**, 263 (1998).

- [15] A. H. Chamseddine, R. Arnowitt, and P. Nath, *Phys. Rev. Lett.* **49**, 970 (1982).
- [16] R. Barbieri, S. Ferrara, and C. A. Savoy, *Phys. Lett. B* **119**, 343 (1982).
- [17] K. Agashe, R. Contino, and A. Pomarol, *Nucl. Phys. B* **719**, 165 (2005).
- [18] R. Contino, L. Da Rold, and A. Pomarol, *Phys. Rev. D* **75**, 055014 (2007).
- [19] L. Randall and R. Sundrum, *Phys. Rev. Lett.* **83**, 3370 (1999).
- [20] J. L. Hewett, F. J. Petriello, and T. G. Rizzo, *JHEP* **09**, 030 (2002).
- [21] L. Randall and R. Sundrum, *Phys. Rev. Lett.* **83**, 3370 (1999).
- [22] L. Randall and R. Sundrum, *Phys. Rev. Lett.* **83**, 4690 (1999).
- [23] K. Agashe, H. Davoudiasl, G. Perez, and A. Soni, *Phys. Rev.* **D76**, 036006 (2007).
- [24] A. L. Fitzpatrick, J. Kaplan, L. Randall, and L.-T. Wang, *JHEP* **09**, 013 (2007).
- [25] O. Antipin, D. Atwood, and A. Soni, *Phys. Lett. B* **666**, 155 (2008).
- [26] D. Pappadopulo, A. Thamm, R. Torre, and A. Wulzer, *JHEP* **09**, 060 (2014).
- [27] B. Bellazzini, C. Csáki, and J. Serra, *Eur. Phys. J.* **C74**, 2766 (2014).
- [28] R. Contino, D. Marzocca, D. Pappadopulo, and R. Rattazzi, *JHEP* **10**, 081 (2011).
- [29] S. D. Ellis, J. Huston, K. Hatakeyama, P. Loch, and M. Tonnesmann, *Prog. Part. Nucl. Phys.* **60**, 484 (2008).
- [30] J. Thaler and K. Van Tilburg, *JHEP* **02**, 093 (2012).
- [31] S. Chatrchyan et al. (CMS Collaboration), *JHEP* **05**, 090 (2013).
- [32] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, *Phys. Rev.* **D80**, 051501 (2009).
- [33] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, *Phys. Rev. D* **81**, 094023 (2010).
- [34] D. Krohn, J. Thaler, and L.-T. Wang, *JHEP* **1002**, 084 (2010).
- [35] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, *Phys.Rev.Lett.* **100**, 242001 (2008).

- [36] J. Thaler and K. Van Tilburg, JHEP **1103**, 015 (2011).
- [37] J. Butterworth, B. Cox, and J. R. Forshaw, Phys. Rev. D **65**, 096014 (2002).
- [38] G. Aad et al. (ATLAS Collaboration), Phys. Rev. D **86**, 072006 (2012).
- [39] L. G. Almeida, S. J. Lee, G. Perez, G. F. Sterman, I. Sung, et al., Phys. Rev. D **79**, 074017 (2009).
- [40] S. D. Ellis, A. Hornig, T. S. Roy, D. Krohn, and M. D. Schwartz, Phys. Rev. Lett. **108**, 182003 (2012).
- [41] A. J. Larkoski, G. P. Salam, and J. Thaler, JHEP **1306**, 108 (2013).
- [42] D. Krohn, M. D. Schwartz, T. Lin, and W. J. Waalewijn, Phys. Rev. Lett. **110**, 212001 (2013).
- [43] G. Aad et al. (ATLAS Collaboration), Eur. Phys. J. **C75**, 69 (2015).
- [44] G. Aad et al. (ATLAS Collaboration), JHEP **12**, 055 (2015).
- [45] G. Aad et al. (ATLAS Collaboration), Phys. Lett. **B755**, 285 (2016).
- [46] S. Chatrchyan et al. (CMS Collaboration), Phys. Lett. **B718**, 1208 (2013).
- [47] S. Chatrchyan et al. (CMS Collaboration), JHEP **02**, 036 (2013).
- [48] S. Chatrchyan et al. (CMS Collaboration), Phys. Lett. **B723**, 280 (2013).
- [49] V. Khachatryan et al. (CMS Collaboration), JHEP **08**, 173 (2014).
- [50] V. Khachatryan et al. (CMS Collaboration), JHEP **08**, 174 (2014).
- [51] CMS Collaboration, CMS-PAS-EXO-14-010 (2015).
- [52] V. Khachatryan et al. (CMS Collaboration), JHEP **02**, 145 (2016).
- [53] F. Hasert et al., Physics Letters B **46**, 138 (1973).
- [54] M. L. Perl et al., Phys. Rev. Lett. **35**, 1489 (1975).
- [55] M. Banner et al., Physics Letters B **122**, 476 (1983).
- [56] S. Abachi et al. (D0 Collaboration), Phys. Rev. Lett. **74**, 2632 (1995).

- [57] S. F. Novaes, in *Particles and fields. Proceedings, 10th Jorge Andre Swieca Summer School, Sao Paulo, Brazil, February 6-12, 1999* (1999), pp. 5–102.
- [58] C. Quigg, *Gauge theories of the strong, weak, and electromagnetic interactions*, Frontiers in physics (Benjamin/Cummings Pub. Co., Advanced Book Program, 1983), ISBN 9780805360202.
- [59] M. Thomson, *Modern Particle Physics*, Modern Particle Physics (Cambridge University Press, 2013), ISBN 9781107034266.
- [60] T. Cheng and L. Li, *Gauge Theory of Elementary Particle Physics*, Oxford science publications (Clarendon Press, 1984), ISBN 9780198519614.
- [61] W. D. Goldberger and M. B. Wise, *Phys. Lett.* **B475**, 275 (2000).
- [62] CMS Collaboration, CMS-AN-15-196 (2015).
- [63] A. Oliveira, *Gravity particles from Warped Extra Dimensions, predictions for LHC* (2014), [arXiv:1404.0102](https://arxiv.org/abs/1404.0102).
- [64] Chanowitz, Michael S. and Gaillard, Mary K., *Nucl. Phys.* **B261**, 379 (1985).
- [65] O. S. Bruning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, CERN-2004-003-V1, CERN-2004-003, CERN-2004-003-V-1 (2004).
- [66] CMS Collaboration, CERN-LHCC-97-10 (1997).
- [67] S. Chatrchyan et al. (CMS Collaboration), *JINST* **3**, S08004 (2008).
- [68] V. Karimäki, M. Mannelli, P. Siegrist, H. Breuker, A. Caner, R. Castaldi, K. Freudenreich, G. Hall, R. Horisberger, M. Huhtinen, et al. (CMS Collaboration), *The CMS tracker system project: Technical Design Report*, Technical Design Report CMS (CERN, Geneva, 1997).
- [69] V. Khachatryan et al. (CMS Collaboration), *Eur. Phys. J. C* **70**, 1165 (2010).
- [70] CMS Collaboration, *The CMS electromagnetic calorimeter project: Technical Design Report*, Technical Design Report CMS (CERN, Geneva, 1997).
- [71] CMS Collaboration, CERN-LHCC-97-31 (1997).

- [72] CMS collaboration, *Journal of Physics: Conference Series* **404**, 012045 (2012).
- [73] CMS Collaboration, *JINST* **9**, P10009 (2014).
- [74] P. Billoir, *Comput. Phys. Commun.* **57**, 390 (1989).
- [75] P. Billoir and S. Qian, *Nucl. Instrum. Meth.* **A294**, 219 (1990).
- [76] R. Mankel, *Nucl. Instrum. Meth.* **A395**, 169 (1997).
- [77] R. Fruhwirth, *Nucl. Instrum. Meth.* **A262**, 444 (1987).
- [78] K. Rose, *Proceedings of the IEEE* **86**, 2210 (1998).
- [79] R. Frühwirth, W. Waltenberger, and P. Vanlaer, *Tech. Rep.*, CERN (2007).
- [80] CMS Collaboration, CMS-PAS-TRK-10-005 (2010).
- [81] M. Cacciari, G. P. Salam, and G. Soyez, *JHEP* **04**, 063 (2008).
- [82] S. Catani, Y. Dokshitzer, M. Seymour, and B. Webber, *Nucl. Phys. B* **406**, 187 (1993).
- [83] Y. Dokshitzer, G. Leder, S. Moretti, and B. Webber, *JHEP* **08**, 001 (1997).
- [84] CMS Collaboration, *JINST* **6**, P11002 (2011).
- [85] CMS Collaboration, CMS-PAS-JME-14-002 (2014).
- [86] CMS Collaboration, CMS-PAS-JME-13-006 (2013).
- [87] CMS Collaboration, CMS-PAS-JME-16-004 (2016).
- [88] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *JHEP* **07**, 079 (2014).
- [89] T. Sjostrand, S. Mrenna, and P. Z. Skands, *JHEP* **05**, 026 (2006).
- [90] T. Sjostrand, S. Mrenna, and P. Z. Skands, *Comput. Phys. Commun.* **178**, 852 (2008).
- [91] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, *JHEP* **07**, 001 (2003).

- [92] S. Frixione and B. R. Webber, *JHEP* **06**, 029 (2002).
- [93] S. Kallweit, J. M. Lindert, P. Maierhofer, S. Pozzorini, and M. Schönherr, *JHEP* **04**, 021 (2016).
- [94] R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali, *Nucl. Phys.* **B849**, 296 (2011).
- [95] S. Agostinelli et al. (GEANT4), *Nucl. Instrum. Meth.* **A506**, 250 (2003).
- [96] CMS Collaboration, *JINST* **6**, P09001 (2011).
- [97] CMS Collaboration, *JINST* **10**, P02006 (2015).
- [98] V. Khachatryan et al. (CMS Collaboration), *JHEP* **12**, 017 (2014).
- [99] CMS Collaboration, *JINST* **7**, P10002 (2012).
- [100] CMS Collaboration, CMS-PAS-PFT-10-003 (2010).
- [101] CMS Collaboration, *JINST* **10**, P06005 (2015).
- [102] CMS Collaboration, *JINST* **11**, P01019 (2016).
- [103] CMS Collaboration, *JINST* **10**, P08010 (2015).
- [104] W. Verkerke and D. P. Kirkby, *eConf* **C0303241** (2003).
- [105] CMS Collaboration, CMS-PAS-EXO-12-021 (2013).
- [106] CMS Collaboration, CMS-PAS-EXO-12-022 (2013).
- [107] CMS Collaboration, CMS-PAS-EXO-15-002 (2015).
- [108] CMS Collaboration, CMS AN-16-215 (2016).
- [109] J. Butterworth et al., *J. Phys.* **G43**, 023001 (2016).
- [110] A. L. Read, *J. Phys.* **G28**, 2693 (2002).
- [111] T. Junk, *Nucl. Instrum. Meth.* **A434**, 435 (1999).
- [112] Luca Lista, *Statistical Methods for Data Analysis in Particle Physics* (Springer, 2016).

- [113] ATLAS and CMS Collaboration, ATL-PHYS-PUB-2011-011, ATL-COM-PHYS-2011-818, CMS-NOTE-2011-005 (2011).
- [114] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Eur. Phys. J.* **C71**, 1554 (2011).
- [115] Frederick James, *Statistical Methods in Experimental Physics* (World Scientific, 2010).
- [116] CMS Collaboration, *Documentation of the RooStats -based statistics tools for Higgs PAG* (2016), URL <https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideHiggsAnalysisCombinedLimit>.
- [117] CMS Collaboration, CMS-PAS-B2G-16-010 (2016).
- [118] CMS Collaboration, CMS-PAS-EXO-15-002 (2015).
- [119] ATLAS and CMS Collaborations, *PoS DIS2014*, 114 (2014).
- [120] CMS Collaboration, *J. Phys. Conf. Ser.* **623** (2015).
- [121] F. Santanastasio, *Nucl. Part. Phys. Proc.* **273–275**, 649 (2016).
- [122] V. Khachatryan et al. (CMS Collaboration), *JHEP* **08**, 173 (2014).
- [123] V. Khachatryan et al. (CMS Collaboration), *JHEP* **08**, 174 (2014).
- [124] CMS Collaboration, CMS-PAS-EXO-11-061 (2012).
- [125] G. Aad et al. (ATLAS Collaboration), *JHEP* **12**, 055 (2015).
- [126] G. Aad et al. (ATLAS Collaboration), *EPJC* **75**, 263 (2015).
- [127] J. Neyman and E. S. Pearson, *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* **231**, 289 (1933).