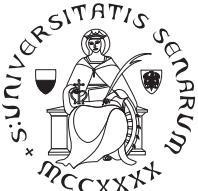


UNIVERSITA' di SIENA

DIPARTIMENTO DI SCIENZE FISICHE, DELLA TERRA E DELL' AMBIENTE
SEZIONE DI FISICA



UNIVERSITÀ
DI SIENA
1240

Search for new resonances in p-p collisions using fully leptonic W^+W^- decays with the CMS detector

Tesi di Dottorato di Ricerca in Fisica Sperimentale
In partial fulfillment of the requirements fo the Ph.D.
Thesis in Experimental Physics

Candidato:
Dr. Lorenzo Russo

Supervisor:
Prof. Vitaliano Ciulli

dedica a

This work is partially funded by MIUR Italy, under contract 2012Z23ERZ of PRIN 2012, “H-TEAM: Trigger, Elettronica Avanzata e Metodi innovativi per misure di precisione nel settore dell’ Higgs ad LHC”.

Abstract

A major breakthrough in modern experimental particle physics came with the discovery of the Higgs boson by the LHC experiments ATLAS and CMS in 2012. The discovered particle is compatible with the Standard Model (SM) Higgs mechanism predictions: the only unknown parameter, the boson's mass, has been measured to be close to 125 GeV. Nevertheless, in order to determine whether the SM Higgs sector is complete, precise measurements of the Higgs boson coupling strengths, CP structure and transverse momentum are required. A complementary and important strategy is the search for additional heavy scalars, that would prove the presence of beyond-the-SM (BSM) physics in the form of a non-minimal Higgs sector. The existence of such a sibling Higgs boson, denoted X , is motivated in many BSM scenarios, so the search for additional scalar resonances in the full mass range accessible at the LHC remains one of the main objectives of the experimental community.

The search for a high mass Higgs boson has been performed using Run-I and early Run-II data in many different decay channels and an upper limit on its cross section has been determined. With the full 2016 data collected by CMS experiment at 13 TeV, approximately 36 fb^{-1} , it is now possible to set a very tight upper limit on the high mass Higgs boson cross section. One of the most sensitive decay channels, for Higgs masses above 200 GeV, is the decay to a pair of W bosons. The fully leptonic, $2\ell 2\nu$, final state (with $\ell = e$ or μ) is considered in this analysis.

The fully leptonic channel has a clear signature due to the presence of the two isolated leptons and moderate missing-transverse-energy (MET) that provides indirect evidence of the neutrinos presence. The search is performed in a wide range of masses from 200 GeV up to 3 TeV. The events are categories in four different categories, optimised for the gluon-gluon fusion and for the vector-boson fusion (VBF) production mechanisms.

The signal is interpreted in terms of the electroweak (EW) singlet model including a detailed simulation of the on interference between X signal, SM Higgs boson H and WW backgrounds.

In addition in the search for new particles, the Two-Higgs-doublet model (2HDM) is a well motivated extension of the SM. A final interpretation of the EW singlet model results in term of 2HDM model is also performed.

Contents

1 The Standard Model, the Higgs Boson and New Scalar Particles	1
1.1 Phenomenology of the Standard Model	1
1.2 The Higgs Boson	6
1.3 New Scalar Particles	10
2 The CMS experiment at LHC	23
2.1 The Large Hadron Collider	23
2.2 The Compact Muon Solenoid experiment	24
2.3 Data recoiled and future plans	33
3 Monte Carlo Generators	37
3.1 Hard process	39
3.2 Parton shower	41
3.3 Multiple Interaction	46
3.4 Hadronization	46
3.5 Hadronic Decays and Electromagnetic Radiation.	49
3.6 Jets Reconstruction	49
3.7 Main Monte Carlo generators	49
3.8 Monte Carlo sample in High Mass Analysis	51
4 Event Reconstruction	57
4.1 The Particle Flow	57
4.2 Lepton reconstruction and identification	58
4.3 Jet reconstruction and identification	61
4.4 b-jet identification	64
4.5 The Missing Transverse Energy	68
4.6 Fake Lepton Background Estimation	70
5 High mass resonances searching	73
5.1 The X to WW search: summary	73
5.2 Discriminating variable	74
5.3 Signal interpretation: EW singlet, 2HDM and MSSM	74
5.4 Study of the Interference effects	78
5.5 Main Background processes	78
5.6 Data sample and Triggers	81
5.7 Opposite Flavor final state	85
5.8 Same Flavor final state	98

5.9	Systematic uncertainties	108
6	Results and Interpretation	113
6.1	Statistical interpretation	113
6.2	Limits and results	115
A	Boost-invariant variables	119
B	Electron Efficiencies from Tag and Probe Method	121

Chapter 1

The Standard Model, the Higgs Boson and New Scalar Particles

1.1 Phenomenology of the Standard Model

The Standard Model (SM) of particle physics [1] is a description of the nature which best explain the fundamental structure of matter and the fundamental forces which govern all known phenomena. The SM gives a quantitative description of three of the four interactions in nature: electromagnetism, weak interactions and strong nuclear force. Developed in the early 1970s by Glashow [2], Weinberg [3] and Salam [4], it has successfully explained almost all experimental results and precisely predicted a wide variety of phenomena. It is a renormalizable quantum field theory, compatible with special relativity.

General Picture

The main constituents of the SM are shown in Fig. 1.1. These are the particle composing the ordinary matter and responsible of the forces. The particles involved are characterized by the spin, the mass, and the quantum numbers determining their interactions. The quarks are subject to all the three forces and, in particular, are the only fermions to possess a “colour” charge, which is responsible of the strong nuclear force, as described by Quantum Chromo Dynamics (QCD). Because of the QCD colour confinement properties, quarks do not exist as free states but can be experimentally observed only as bound states. The proton and neutron, are composed by three quarks (called baryons). The particle composed by a quark-antiquark are called “meson”. Quark flavour is conserved in electromagnetic and strong interactions but not in weak ones, as quark mass eigenstates do not correspond to the weak interaction eigenstates. Their mixing is described by the Cabibbo–Kobayashi–Maskawa (CKM) matrix. The leptons have no colour charge and are subject only to the electromagnetic and weak forces. The charged leptons of the three families are respectively denoted as the electron (e), muon (μ) and tau lepton (τ). The only stable lepton is the electron. To each lepton corresponds a neutrino. The mass of neutrino is unknown but their flavour oscillations prove a non-zero mass [5]. The gluons, the W^\pm bosons, the Z boson and the photon (γ) are boson that compose

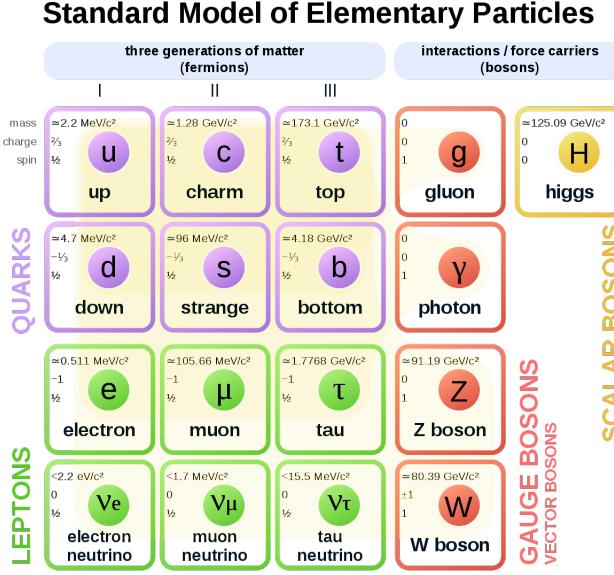


Figure 1.1. Main constituents of the Standard Model.

the SM gauge sector. The gluons are the mediators of the strong interactions. They are massless, electrically neutral and carry color quantum number and they can interact with themselves. The W^\pm and Z bosons are the mediator of the weak interactions. Their mass is ~ 81 GeV and ~ 90 GeV respectively. These particles are unstable and decay in other particles. Finally the photon is massless, chargeless, non self-interacting and mediates the electromagnetic interactions.

To summarize, the SM Lagrangian may be written as the sum of three parts:

$$\mathcal{L}_{SM} = \mathcal{L}_{QCD} + \mathcal{L}_{EWK} + \mathcal{L}_H , \quad (1.1)$$

where \mathcal{L}_{QCD} is the quantum chromodynamics Lagrangian that describes the interactions of quarks and gluons, the \mathcal{L}_{EWK} is the the electroweak Lagrangian that describes the interactions of the fermions with the Z and W^\pm bosons. The \mathcal{L}_H is the Higgs part of the Lagrangian: the Lagrangian symmetries seems to forbid the introduction of mass terms without spoiling its gauge invariance. Higgs' proposal solves this problem by spontaneously breaking the Lagrangian symmetry (Sec. 1.2).

Quantum Chromodynamics

Quantum Chromodynamics describes (QCD) the interactions of quarks and gluons, mediated by the strong force through the colour charge. The QCD Lagrangian has the local gauge invariance under the $SU(3)_C$ group and it is given by [6],

$$\begin{aligned} \mathcal{L}_{QCD} = & \sum_q [\bar{\psi}_{q,a}(i\gamma^\mu \partial_\mu \delta_{ab} - g_s \gamma^\mu t_{ab}^C A_\mu^C - m_q \delta_{ab}) \psi_{q,b} \\ & - \frac{1}{4} F_{\mu\nu}^A F^{A\mu\nu}] , \end{aligned} \quad (1.2)$$

where repeated indices are summed over. The γ^μ are the Dirac γ -matrices. The

$\psi_{q,a}$ are quark-field spinors for a quark of flavor q and mass m_q , with a color-index a that runs from $a = 1$ to $N_c = R$ 3, i.e. quarks come in three “colors”. Quarks are said to be in the fundamental representation of the $SU(3)$ color group. The A_μ^C correspond to the gluon fields, with C running from the eight kinds of gluon. The t_{ab}^C correspond to eight 3×3 matrices and are the generators of the $SU(3)_C$ group. The quantity g_S is the QCD coupling constant. Finally, the field tensor $F_{\mu\nu}^A$ is given by,

$$F_{\mu\nu}^A = \partial_\mu A_\nu^A - \partial_\nu A_\mu^A - g_S f_{ABC} A_\mu^B A_\nu^C, \quad (1.3)$$

$$[t^A, t^B] = i f_{ABC} t^C, \quad (1.4)$$

where the f_{ABC} are the structure constants of the $SU(3)_C$ group. Neither quarks nor gluons are observed as free particles. Hadrons are color-singlet (i.e. color-neutral) combinations of quarks, anti-quarks, and gluons. The fundamental parameters of QCD are the coupling g_S (or $\alpha_S = \frac{g_S^2}{4\pi}$) and the quark masses m_q . If the quark masses are fixed, there is only one free parameter in the QCD Lagrangian, that is α_S . This constant is not a physical observable but rather a quantity defined in the context of perturbation theory, which enters prediction for experimental observable.

Electroweak Model

The electroweak interactions is based on the gauge group $SU(2)_L \otimes U(1)_Y$. The $SU(2)_L$ group refers to the weak isospin charge (I), and $U(1)_Y$ to the weak hypercharge (Y). Left-handed (L) fermions are paired in $I = 1/2$ isospin doublets, whereas right-handed (R) fermions in $I = 0$ singlets. The presence of these local gauge symmetries introduces four vector bosons: three for the $SU(2)$ group, the W_i fields ($i = 1, 2, 3$), and one for $U(1)$, the B field. This gives rise to a quantum field theory, invariant under local gauge symmetries, whose Lagrangian is expressed as:

$$\mathcal{L}_{EW} = \sum_f \bar{\psi} i\gamma^\mu \mathcal{D}_\mu \psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{1}{4} \vec{E}_{\mu\nu} \cdot \vec{E}^{\mu\nu}, \quad (1.5)$$

where the sum is extended over all the fermions f and where covariant derivatives which preserve the local gauge invariance have the following form:

$$\begin{aligned} \mathcal{D}_\mu &= \partial_\mu + ig \vec{W}_\mu \cdot \frac{\vec{\tau}}{2} + i \frac{g'}{2} Y B_\mu, \\ F_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu, \\ E_{\mu\nu}^\alpha &= \partial_\mu W_\nu^\alpha \partial_\nu W_\mu^\alpha - g \epsilon^{\alpha\beta\gamma} W_\mu^\beta W_\nu^\gamma, \end{aligned} \quad (1.6)$$

where $\vec{\tau}$ indicates the three Pauli matrices, g and g' are the coupling constants which correspond respectively to $SU(2)_L$ and $U(1)_Y$. The physical fields are obtained as linear combinations of these fields:

$$\begin{aligned}
A_\mu &= \sin \theta_W W_\mu^3 + \cos \theta_W B_\mu, \\
Z_\mu &= \cos \theta_W W_\mu^3 - \sin \theta_W B_\mu, \\
W_\mu^\pm &= \frac{W_\mu^1 \mp W_\mu^2}{\sqrt{2}}.
\end{aligned} \tag{1.7}$$

The above equations represent two neutral particles (the photon, described by the A_μ field, and the Z boson) and two charged particles (the W^+ and W^- bosons). We have further introduced the angle θ_W , which is known as the weak mixing angle or Weinberg angle. Up to here, the theory is necessarily incomplete: all particles it describes are massless, contradicting experimental evidence. The Lagrangian symmetries, on the other hand, seem to forbid the introduction of mass terms without spoiling its gauge invariance. Higgs' proposal solves this problem by spontaneously breaking the Lagrangian symmetry.

Experimental evidence

The experimental study of the Standard Model has made a quantum leap in the last 30 years. First of all the theory predicted the existence of W^\pm and Z bosons in the mass range from 60 to 93 GeV [7]. In 1976 Rubbia, Cline and McIntyre proposed the transformation of an existing high-energy proton accelerator into a proton–antiproton collider as a quick and relatively cheap way to achieve collisions above threshold for W and Z production. This proposal was adopted at CERN Super Proton Synchrotron (SPS) collider and the first proton–antiproton collisions were collected in 1981. In the following years the W and Z bosons have been observed by UA1 and UA2 experiments with a mass of 80 GeV and 91 GeV respectively, Fig. 1.2. With the electron-positron colliders at a center-of-mass energy equal to Z mass, precise measurements of the fundamental parameters of electroweak theory could be made. In 1989, two e^+e^- colliders operation started: the Stanford Linear Collider (SLC) at SLAC, and the circular Large Electron Positron collider (LEP) at CERN. Precision electroweak tests covering the measurements at the Z pole have been conducted by SLC and LEP experiments. W boson properties were also measured at the LEP collider, which reach 209 GeV center-of-mass energy, well above the threshold for W^+W^- pair production, Fig. 1.3. The results of measurements confirmed the SM prediction. The next missing SM piece was top quark. It was a necessary component of the SM of electroweak interactions, but there was no consistent theoretical guidance as to what its mass should be. The only way to observe a top quark with such a high mass was at the collider with the highest-energy, the Tevatron antiproton-proton collider at Fermilab. The existence of the top quark was firmly established in 1995 with simultaneous announcements by both the CDF [8] and the DO [9] experiments with results that demonstrated a mass of around 174 GeV, Fig.1.4. After the top quark discovery, the only missing part to the SM was the Higgs boson particle that has been observed at ATLAS and CMS experiment at LHC proton-proton collider at CERN in 2012 (see Sec.1.2).

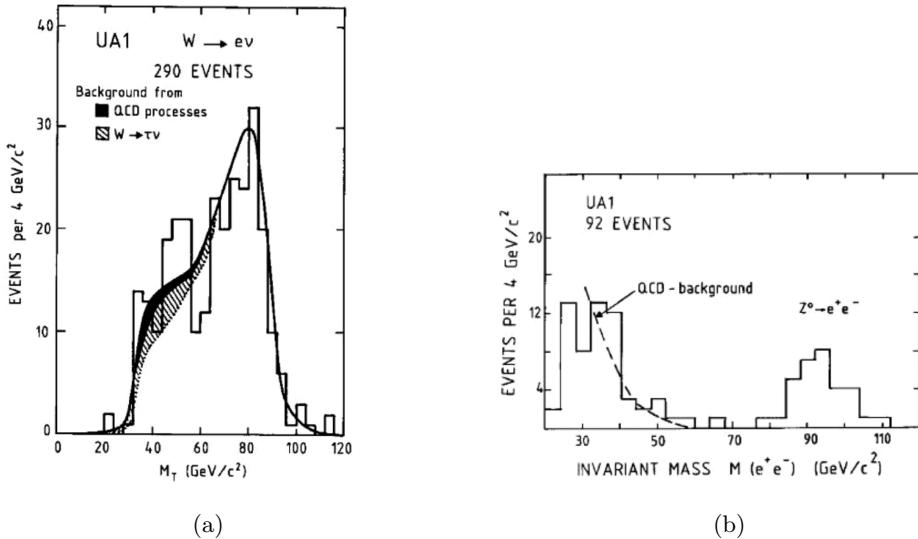


Figure 1.2. (a) Transverse mass distribution for all $W \rightarrow e\nu$ events recorded by UA1 between 1982 and 1985. (b) Invariant mass distribution of all e^+e^- pairs recorded by UA1 between 1982 and 1985

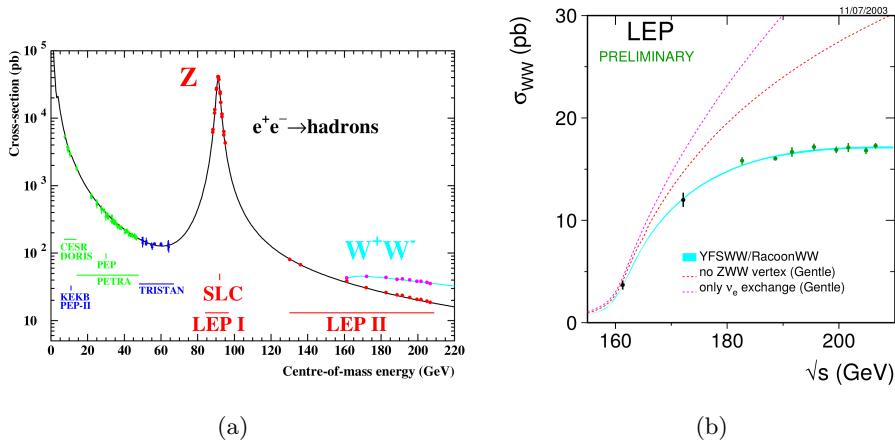


Figure 1.3. (a) The cross-section for the production of hadrons in e^+e^- annihilations. The measurements are shown as dots with error bars. The solid line shows the prediction of the SM. (b) The measured W-pair production cross section compared to the SM and alternative theories not including trilinear gauge couplings.

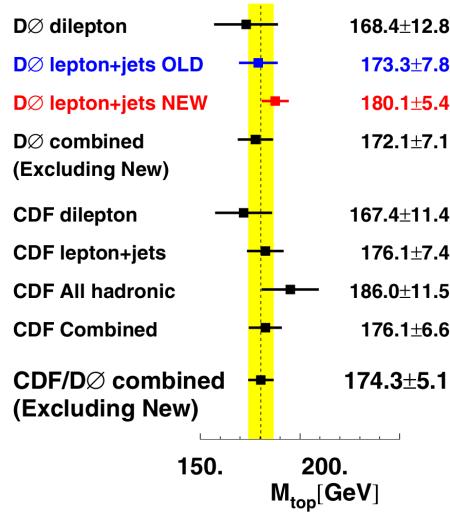


Figure 1.4. Top quark mass measurements.

1.2 The Higgs Boson

A Lagrangian is said to have a symmetry when it is invariant under a group of transformations. However the fact that the weak gauge bosons have a mass different from zero indicates that the gauge symmetry is broken. Also the fermion masses can not be included without violating gauge symmetry in the \mathcal{L}_{QCD} . The mass terms can be introduced with the Spontaneous Symmetry Breaking Mechanism, adding \mathcal{L}_H , that gives mass to the weak bosons and fermions and leaves the photon massless. This mechanism has been proposed in 1964 independently by Higgs [10] and Brout and Englert [11]. With the spontaneous symmetry breaking mechanism, a new particle which couples to the massive fermions and to the boson emerges. This particle is called Higgs boson and its mass is a free parameter of the theory. In 2012, 48 years after this hypothesis was formulated, the Higgs boson has been observed by ATLAS [12] and CMS [13] experiments at LHC.

The Brout–Englert–Higgs mechanism

The symmetry of SM Lagrangian, \mathcal{L}_{SM} , is $SU(2)_L \otimes U(1)_Y \otimes SU(3)_C$, where L , Y and C refers to isospin, hypercharge and color quantum numbers, respectively. To break the symmetry, a scalar field Φ (Higgs field) is introduced. The field is an isospin doublet ($SU(2)_L$):

$$\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \Phi_1 + i\Phi_2 \\ \Phi_3 + i\Phi_4 \end{pmatrix}$$

where Φ_j with $j = 1, 2, 3, 4$ are real fields used to manifest the complexity of Φ^+ and Φ^0 . The simplest Lagrangian of a self-interacting scalar field is,

$$\mathcal{L}_H = (D^\mu \Phi^\dagger)(D_\mu \Phi) - \mu^2 \Phi^\dagger \Phi - \lambda (\Phi^\dagger \Phi)^2 , \quad (1.8)$$

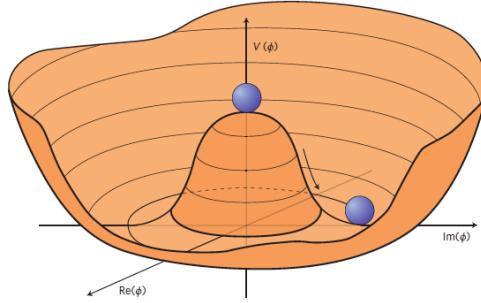


Figure 1.5. Higgs fields potential with two degree of freedom.

where λ needs to be positive for the potential to be bounded from below and μ^2 is a mass term for the Φ field. The ground state (vacuum) of the theory is defined as the state where the energy density is at a minimum. If the μ parameter is chosen so that $\mu^2 < 0$, the symmetry of the potential may be broken, and the minimum value is,

$$v \equiv \sqrt{\frac{-\mu^2}{\lambda}} = \Phi^\dagger \Phi. \quad (1.9)$$

The choice for the sign of the parameters μ^2 and λ gives to the potential $V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2$ the “mexican hat” shape, as illustrated in Fig. 1.5. In the perturbation theory, the Φ fields is expanded around the minimum, that is chosen among the set of states which satisfy Eq. 1.9. All this states break the rotational symmetry of the Lagrangian. The Φ field is expressed as,

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \quad (1.10)$$

The $h(x)$ field gives a particle with mass equal to,

$$m_H = \sqrt{2}\mu = v\sqrt{2\lambda}, \quad (1.11)$$

that is a free parameter of the theory. However some theoretical constrains could be imposed. The value of v parameter in Eq. 1.11 could be determinate by the Fermi constant, G_F , as,

$$v = \frac{2m_W}{g} = (\sqrt{2}G_F)^{-1/2}, \quad (1.12)$$

with the actual G_F measurements obtained with the muon lifetime [14], a value of ~ 250 GeV is obtained for v . However the model is not predictive on the value of the λ parameter. Nevertheless, additional theoretical arguments place approximate upper and lower bounds on m_H [6]. The lower bound on the Higgs mass is given by the vacuum state stability, that leads to requiring λ to be positive at all energies. The upper edge is imposed by the Planck scale. So a Higgs boson in a range of $130 < m_H < 180$ GeV is consistent with the theory.

The Higgs boson at LHC

The Higgs boson has been searched in several experiments (Fig. 1.6) located at different colliders (LEP, SLC, Tevatron) without a clear evidence of such particle.

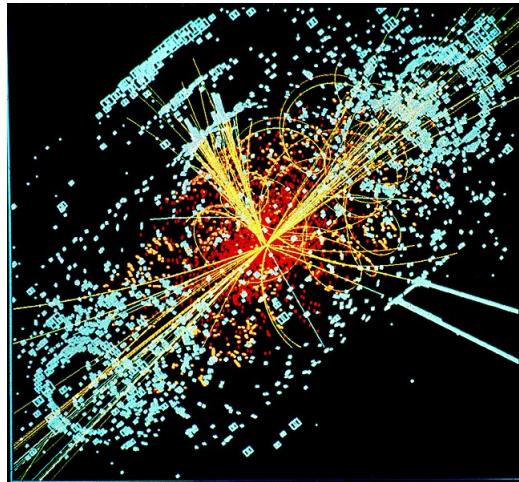
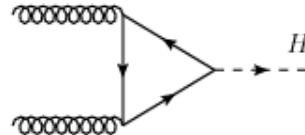


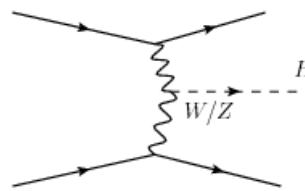
Figure 1.6. An example of simulated data modeled for the CMS particle detector on the Large Hadron Collider (LHC) at CERN.

Therefore, a proton proton collider, located at CERN, that has been designed and built to Higgs boson researches specifically. In a proton collision the Higgs boson could be produced in different ways. The production modes are the gluon gluon fusion, the vector boson fusion (VBF), the vector-boson associated production and the top-quark associated production. Below are described the details of the different mechanisms:

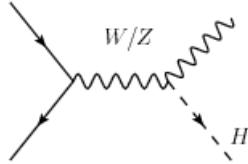
- *Gluon gluon fusion ($gg \rightarrow H$):* this is the main Higgs boson production mode. Here a couple of gluons interact via a heavy quark loop and give rise to a Higgs boson.



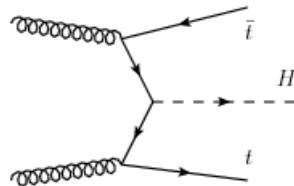
- *Vector Boson Fusion (VBF) ($qq \rightarrow qqH$):* each incoming quarks emits a virtual W or Z boson that interact generating a Higgs boson. The quarks after emitting the vector bosons proceed in the forward direction and represent the peculiar signature of this production mode.



- *Vector boson associated production* (or *Higgsstrahlung*): the Higgs boson is emitted from a W^\pm or a Z boson which has been produced by a pair quark-antiquark.



- *Top-quark associated production*: a pair of top quarks, originated from the splitting of two incoming gluons, interacts to yields to a Higgs boson.



The Higgs boson cross section of the different production modes, depend on the center of mass energy of the collider and by the supposed mass of the particle. The cross section increases with the energy of the colliding particles (\sqrt{s}) and decreases with the mass of the boson, m_H . The second behaviour is shown in Fig. 1.10. The Higgs boson is an unstable particle and decays in a variety of different finale states, Fig 1.11. For a Higgs boson of mass $m_H \sim 125$ GeV (that is the mass measured by ATLAS and CMS) the channel with the largest branching ratio is in $\bar{b}b$ quark, followed by the W^+W^- , $\tau^+\tau^-$ and ZZ channels.

Experimental Results

The discovery or exclusion of the SM Higgs boson was one of the primary scientific goals of LHC. On the 4th of July 2012, as the observation of a new boson by the ATLAS [12] and the CMS [13] collaborations has been announced, as an excess of events near 125 GeV was reported by both experiments, Fig. 1.7. Latest results from both the experiments confirm that the properties of the discovered particle are consistent with the hypothesis of being the Higgs boson predicted by the Standard Model. The measurements of Higgs boson couplings via exclusive production modes and decay channels, of its spin-parity, and of its differential production cross section, need to be thoroughly investigated to verify that they correspond precisely to the SM predictions. It is being done using the new data, collected from 2015 at LHC with increased center-mass-energy of 13 TeV. Precision measurements of the Higgs boson's properties are important to understand why the Higgs boson mass is near the electroweak scale. A good measurement that is possible to perform is the Higgs boson's coupling. In the k -framework [15] coupling modifiers are introduced in order

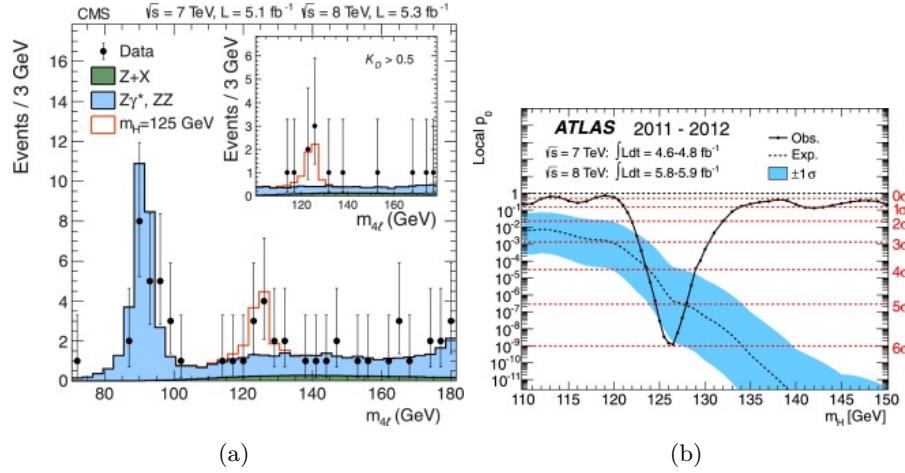


Figure 1.7. (a) Distribution of the four-lepton invariant mass for the $ZZ \rightarrow 4\ell$ analysis of CMS. (b) The observed (solid) p_0 local as a function of m_H in the low mass range in ATLAS detector.

to test for deviations in the couplings of the Higgs boson to other particles. Indeed, the couplings to up and down type fermions can be nonuniversal. Additionally, in these models, it is possible that the Higgs boson will couple differently to fermions (k_F) and vector bosons (k_V), Fig. 1.8. This results so far shows a good agreement among the experimental data and the theory.

1.3 New Scalar Particles

Open Questions

The Standard Model is presently the best description of the subatomic nature, but this theory does not provide a complete picture of the world. Sure enough there are still open questions: the gravity is not described in the model, the explanation of the dark matter is not clear and the neutrino mixing and mass are not well understood. The gravity is one of the four fundamental interactions but it is not described in the SM. It is so different from the three other forces. The purpose to establish a common theory that describe all forces is so difficult. The gravity is described in the Einstein's General Relativity (GR) theory. To combine the SM with the GR it is necessary a quantum theory with a new field associated to gravity as mediator: a spin 2 particle called graviton. Right now there are no experimental evidence of the existence of this kind of particle. An other deficit of the SM regards the dark matter. In fact, from astronomical observations, only the 5% of the matter and energy content of our universe is formed by the ordinary matter (hadrons and leptons), the other 95% is composed by dark matter ($\sim 25\%$) and dark energy ($\sim 70\%$). The SM does not offer good candidates or explanations for the dark matter and dark energy problems although some research at colliders have been done. Concerning the neutrinos, in the SM they were assumed massless. However flavour oscillation implies that they must have non zero mass differences. It is not clear if the small neutrino masses can arise from the same electroweak symmetry breaking mechanism that is in act for

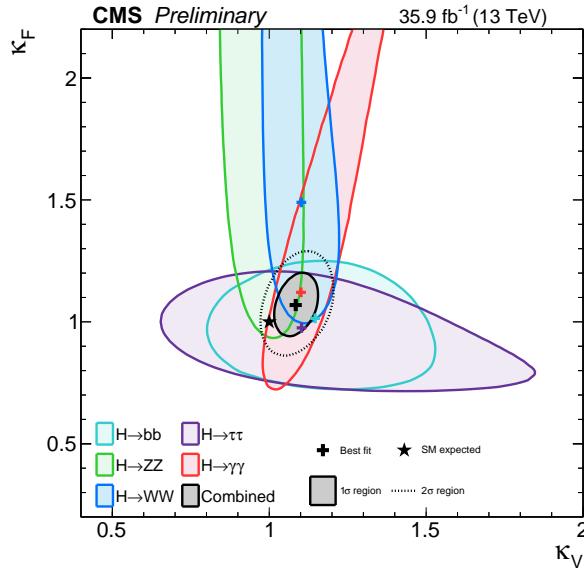


Figure 1.8. The CL regions in the k_F vs k_V parameter space for the model assuming a common scaling of all the vector boson and fermion couplings.

the other SM particles.

These open questions represent deficiencies of the Standard Model. The presence of a hidden sector, defined here to mean extra states that have no SM gauge charge but are charged under some other exotic gauge symmetry, does not necessarily solve any of the problems above. However, in order to identify whether the SM Higgs sector is complete, the searches of additional heavy scalars are performed. They would prove the presence of beyond-the-SM (BSM) physics in the form of a non-minimal Higgs sector [16]. The existence of sibling Higgs boson, denoted X , is motivated in many BSM scenarios, so the research in the full mass range accessible at colliders remains one of the main objectives of the experimental community. This program needs to be continued within the full mass range that is accessible to current and future experiments.

Higgs Singlet Extension

The simplest extension of the SM Higgs sector consist in adding to the complex $SU(2)_L$ doublet Φ a real scalar S which is a singlet under all SM gauge groups. The most general gauge-invariant and renormalisable scalar Lagrangian is,

$$\mathcal{L}_s = (D_\mu \Phi)^\dagger D_\mu \Phi + \partial^\mu S \partial_\mu S - V(\Phi, S), \quad (1.13)$$

where $V(\Phi, S)$ is the scalar potential,

$$V(\Phi, S) = -m^2 \Phi^\dagger \Phi - \mu^2 S^2 + \lambda_1 (\Phi^\dagger \Phi)^2 + \lambda_2 S^4 + \lambda_3 \Phi^\dagger \Phi S^2. \quad (1.14)$$

Here, Z_2 ($S \rightarrow -S$) symmetry is imposed which forbids additional terms in the potential. To determine the condition for $V(\Phi, S)$ to be bounded from below, it is necessary to study its behaviour for large field values. The two vacuum expectation values (VEVs) are defined as,

$$\langle \Phi \rangle \equiv \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}, \quad \langle S \rangle \equiv \frac{x}{\sqrt{2}}, \quad (1.15)$$

with v and x real and non-negative. With this definition of the VEVs, the extrema of V are determined using the following set of equations:

$$\frac{\partial V}{\partial v}(v, x) = v \cdot (-m^2 + \lambda_1 v^2 + \frac{\lambda_2}{2} x^2), \quad (1.16)$$

$$\frac{\partial V}{\partial x}(v, x) = x \cdot (-\mu^2 + \lambda_2 v^2 + \frac{\lambda_2}{2} v^2). \quad (1.17)$$

The physically interesting solutions have $v, x > 0$:

$$v^2 = \frac{\lambda_2 m^2 - \frac{\lambda_3}{2} \mu^2}{\lambda_1 \lambda_2 - \frac{\lambda_3^2}{4}}, \quad (1.18)$$

$$x^2 = \frac{\lambda_1 \mu^2 - \frac{\lambda_3}{2} m^2}{\lambda_1 \lambda_2 - \frac{\lambda_3^2}{4}}. \quad (1.19)$$

For the determination of the extrema we evaluate the Hessian matrix:

$$\mathcal{H}(v, x) \equiv \begin{pmatrix} \frac{\partial^2 V}{\partial v^2} & \frac{\partial^2 V}{\partial v \partial x} \\ \frac{\partial^2 V}{\partial v \partial x} & \frac{\partial^2 V}{\partial x^2} \end{pmatrix} = \begin{pmatrix} 2\lambda_1 v^2 & \lambda_3 v x \\ \lambda_3 v x & 2\lambda_2 x^2 \end{pmatrix}. \quad (1.20)$$

So, the scalar potential $V(\Phi, S)$ is bounded from below if the following conditions are fulfilled,

$$4\lambda_1 \lambda_2 - \lambda_3^2 > 0, \quad (1.21)$$

$$\lambda_1, \lambda_2 > 0, \quad (1.22)$$

where if the first condition is fulfilled, the extremum is a local minimum. The second condition , guarantees that the potential is bounded from below for large field values. The Higgs fields, Φ and S , have non-zero vacuum expectation, denoted by v and x , respectively. Following the unitary-gauge prescription, the Higgs fields is given by,

$$\mathcal{H} \equiv \begin{pmatrix} 0 \\ \frac{h+v}{\sqrt{2}} \end{pmatrix}, \quad S \equiv \frac{h' + x}{\sqrt{2}}. \quad (1.23)$$

Expansion around the minimum leads to the squared mass matrix

$$\mathcal{M}^2 = \begin{pmatrix} 2\lambda_1^2 v^2 & \lambda_3 v x \\ \lambda_3 v x & 2\lambda_2^2 x^2 \end{pmatrix}, \quad (1.24)$$

with the mass eigenvalues

$$m_h^2 = \lambda_1 v^2 + \lambda_2 x^2 - \sqrt{(\lambda_1 v^2 - \lambda_2 x^2)^2 + \lambda_3(xv)^2} \quad , \quad (1.25)$$

$$m_H^2 = \lambda_1 v^2 + \lambda_2 x^2 + \sqrt{(\lambda_1 v^2 - \lambda_2 x^2)^2 + \lambda_3(xv)^2} \quad , \quad (1.26)$$

where h and H are the scalar fields of definite masses m_h and m_H respectively, with $m_h^2 < m_H^2$. The gauge and mass eigenstates are related via the mixing matrix

$$\begin{pmatrix} h \\ H \end{pmatrix} = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} \tilde{h} \\ h' \end{pmatrix} , \quad (1.27)$$

where the mixing angle $-\frac{\pi}{2} \leq \alpha \leq \frac{\pi}{2}$ is given by,

$$\sin 2\alpha = \frac{\lambda_3 xv}{\sqrt{(\lambda_1 v^2 - \lambda_2 x^2)^2 + \lambda_3(xv)^2}} , \quad (1.28)$$

$$\cos 2\alpha = \frac{\lambda_2 x^2 - \lambda_1 v^2}{\sqrt{(\lambda_1 v^2 - \lambda_2 x^2)^2 + \lambda_3(xv)^2}} . \quad (1.29)$$

By the mixing matrix it is clear that the light (heavy) Higgs couplings to SM particles are now suppressed by $\cos \alpha$ ($\sin \alpha$). The heavy Higgs is a new version of the SM Higgs with rescaled couplings to the matter contents and to the gauge fields of the SM. In fact, the only novel channel with respect to the light Higgs case is $H \rightarrow hh$. The corresponding partial decay width, Γ , is given by [17],

$$\Gamma_{H \rightarrow hh} = \frac{|\mu'|^2}{8\pi m_H} \sqrt{1 - \frac{4m_h^2}{m_H^2}} , \quad (1.30)$$

where the coupling strength μ' is,

$$\mu' = -\frac{\sin 2\alpha}{2vx} (\sin \alpha v + \cos \alpha x) \left(m_h^2 + \frac{m_H^2}{2} \right) . \quad (1.31)$$

In collider phenomenology, is important:

- the suppression of the production cross section of the two Higgs states induced by the mixing
- the suppression of the Higgs decay modes to SM particles,

For the high mass scenario, i.e. the case where the heavy Higgs boson is identified with the discovered Higgs state at ~ 125 GeV, $|\sin \alpha| = 1$ corresponds to the complete decoupling of the second Higgs boson and therefore the SM-like scenario.

	Φ_1	Φ_2	u_R	d_R	ℓ_R	Q_L, L_L
Type-I	+	-	-	-	-	+
Type-II	+	-	-	+	+	+
Type-X	+	-	-	-	+	+
Type-Y	+	-	-	+	-	+

Figure 1.9. Variation in charge assignments of the Z_2 symmetry.

Minimal Supersymmetric Standard Model

With the a Higgs boson discovery, possible extensions of the Standard Model has been introduced. One of the simplest ways to extend the scalar sector of the SM is to add one more complex doublet to the model. The resulting two Higgs doublet models (2HDMs) [18] can provide additional CP-violation coming from the scalar sector and can easily originate dark matter candidates. More evolved models with additional field content have a 2HDM like scalar sector, as the Minimal Supersymmetric Standard Model. 2HDMs have a richer particle spectrum with one charged and three neutral scalars. All neutral scalars could in principle be the scalar discovered at the LHC.

Two versions of the 2HDM, one CP-conserving and the other explicitly CP-violating, exist. The Yukawa Lagrangian rises to scalar exchange flavour changing neutral currents (FCNCs) that are constrained by experiments results. A simple way to take in account the FCNCs constrains is to impose a Z_2 symmetry on the scalar doublets ($\Phi_1 \rightarrow -\Phi_1$ and $\Phi_2 \rightarrow -\Phi_2$) and a corresponding symmetry to the quark field.

These arguments lead four Yukawa models: types-I, types-II, types-Y (flipped) and types-X (Lepton Specific) [19, 20]. For example, the Higgs sector of the minimal super-symmetric standard model (MSSM) is the 2HDM with a supersymmetric relation among the parameters of the Higgs sector, whose Yukawa interaction is of type-II, in which only a Higgs doublet couples to up-type quarks and the other couples to down-type quarks and charged leptons. On the other hand, a model model that tries to explain neutrino masses, dark matter corresponds to a type-X Yukawa interaction In this model, the Higgs sector is the two Higgs doublets with extra scalar singlets, in which only a Higgs doublet couples to quarks, and the other couples to leptons. Therefore it is important to experimentally determine the type of the Yukawa interaction in order to select the true model from the different predict by 2HDMs.

The most general Yukawa interaction under the Z_2 symmetry can be written as,

$$\mathcal{L}_{Yukawa}^{2HDM} = -\bar{Q}_L Y_u \tilde{\Phi}_u u_R - \bar{Q}_L Y_d \Phi_d d_R - \bar{L}_L Y_l \Phi_l l_R + h.c., \quad (1.32)$$

where Φ_f , with $f = u, d, l$, is either Φ_1 or Φ_2 . There are four independent Z_2 charge assignments on quarks and charged leptons, Fig 1.9. In the type-I, all quarks and charged leptons obtain their masses from the VEV of Φ_2 . In the type-II, masses of up-type quarks are generated by the VEV of Φ_2 , while those of down-type

quarks and charged leptons are acquired by that of Φ_1 . The Higgs sector of the MSSM is a special 2HDM whose Yukawa interaction is of type-II. The type-X Yukawa interaction (all quarks couple to Φ_2 while charged leptons couple to Φ_1) is used in the model described in [20]. The remaining one is referred to as the type-Y. The most general Higgs potential under the softly broken Z_2 symmetry is given by,

$$\begin{aligned} V^{2HDM} = & m_1^2 \Phi_1^\dagger \Phi_1 + m_2^2 \Phi_2^\dagger \Phi_2 - m_3^2 (\Phi_1^\dagger \Phi_1 + \Phi_2^\dagger \Phi_2) + \frac{\lambda_1}{2} (\Phi_1^\dagger \Phi_1)^2 \\ & + \frac{\lambda_2}{2} (\Phi_2^\dagger \Phi_2)^2 + \lambda_3 (\Phi_1^\dagger \Phi_1 + \Phi_2^\dagger \Phi_2) + \lambda_4 (\Phi_1^\dagger \Phi_2 + \Phi_2^\dagger \Phi_1) \\ & + \lambda_5 [(\Phi_1^\dagger \Phi_2)^2 + (\Phi_2^\dagger \Phi_1)^2], \end{aligned} \quad (1.33)$$

where the parameters m_3^2 and λ_5 are complex, in general but in the following, assuming CP invariance, they are real. The Higgs doublet fields can be parametrized as,

$$\Phi_i = \begin{pmatrix} \omega_i^+ \\ \frac{1}{\sqrt{2}}(v_i + h_i + iz_i) \end{pmatrix}, \quad (1.34)$$

and the mass eigenstates are defined by,

$$\begin{aligned} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} &= R(\alpha) \begin{pmatrix} H \\ h \end{pmatrix}, \\ \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} &= R(\beta) \begin{pmatrix} z \\ A \end{pmatrix}, \\ \begin{pmatrix} \omega_1^+ \\ \omega_2^+ \end{pmatrix} &= R(\beta) \begin{pmatrix} \omega^+ \\ H^+ \end{pmatrix}, \end{aligned} \quad (1.35)$$

where the rotation matrix is given by,

$$R(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}. \quad (1.36)$$

There are five physical Higgs bosons, i.e., two CP-even states h and H , one CP-odd state A , and a pair of charged states H^\pm , and z and ω^\pm are Nambu-Goldstone bosons that are eaten as the longitudinal components of the massive gauge bosons. The eight parameters, m_i^2 ($i = 1, 3$) and λ_j ($j = 1, 5$) in the Higgs sector are replaced by eight physical parameters:

- the VEV $v = \sqrt{v_1^2 + v_2^2} \sim 246$ GeV,
- the mixing angles α and β ($\tan \beta = v_1/V_2$) with $\sin(\beta - \alpha) = 1$,
- boson masses, m_h, m_H, m_A, m_H^\pm ,
- soft breaking mass parameter, $M = m_3/\sqrt{\sin \beta \cos \beta}$.

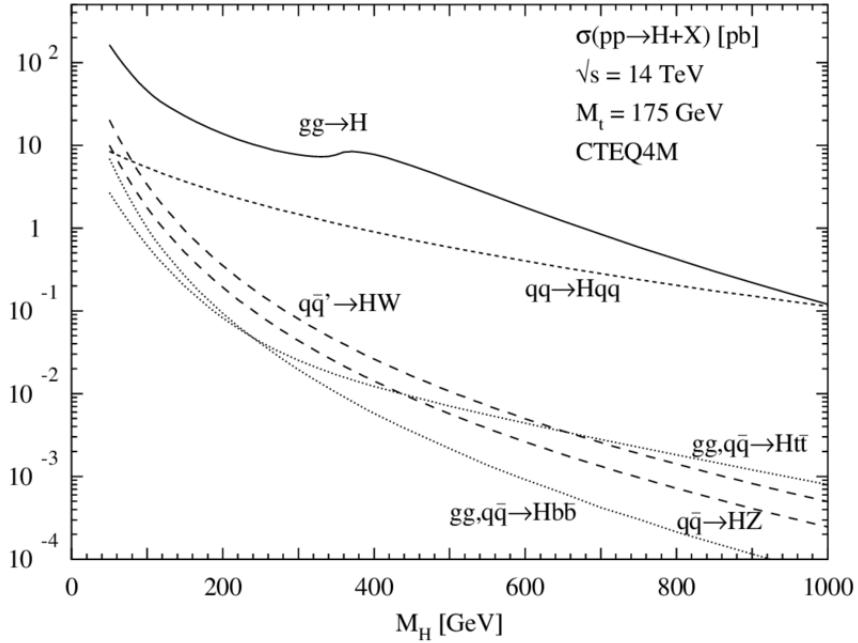


Figure 1.10. Higgs-production cross sections at the LHC for various production mechanisms as a function of the Higgs mass. QCD corrections are included except for Higgs bremsstrahlung [21].

For the successful electroweak symmetry breaking, a combination of quartic coupling constants should satisfy the condition of vacuum stability. In addition it is also take into account bounds from perturbative unitarity to restrict parameters in the Higgs potential. The top and bottom Yukawa coupling constants cannot be taken to o large. The requirement $|Y_{t,b} < \pi|$ at the tree level can provide a milder constraint $0.4 < \tan \beta < 91$.

The wide range of possibilities for Higgs boson mass spectrum hierarchies and branching ratios in 2HDMs yields a diversity of production and decay channels that are relevant for multi-lepton signatures at the LHC. Multi-lepton final states become especially important when the decay of one Higgs scalar to a pair of Higgs scalars or a Higgs scalar and a vector boson is possible.

The WW channel for high mass particle

Let's now concentrate on the $X \rightarrow WW$ channel, that is the final state using in the following for the high mass searches. The two dominant production mechanisms of high mass SM-like Higgs boson are the gluon gluon fusion and Vector Boson Fusion, Fig. 1.10. The first one is the main mechanism for mass values below 1 TeV, above the VBF production mechanism become more and more important as m_X increases. In the search for high mass Higgs boson, in many models, the WW final state, along with ZZ, is the dominant decay channel of X for masses above $2m_Z$ threshold. This fact is evident in Fig. 1.11 (a), where the WW branching ratio (in green) dominates in the high mass region. More detailed results on the

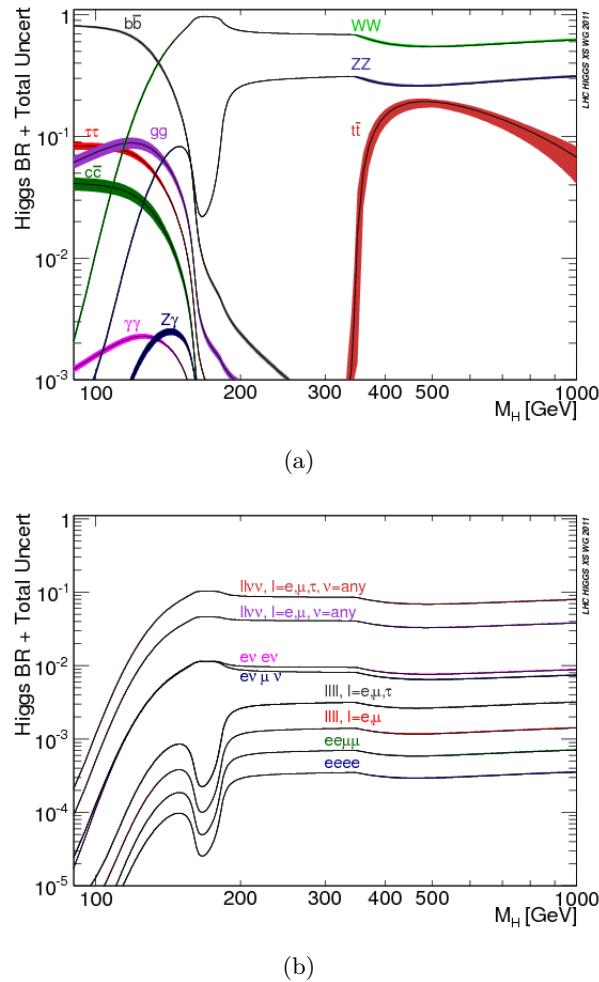


Figure 1.11. (a) Higgs branching ratios and their uncertainties for the full mass range [22]. (b) Higgs branching ratios for the different $H \rightarrow 4\ell$ and $H \rightarrow 2\ell 2\nu$ final states and their uncertainties for the full mass range [22].

decays $H \rightarrow WW$ and $H \rightarrow ZZ$ with the subsequent decay chain are presented in Fig. 1.11 (b). However even if the yield for the decay channels started by the $X \rightarrow WW$ decay is higher, the presence of neutrinos in the final state does not allow to have a complete reconstruction of the decay. This fact makes this channel very challenging. As we have already noticed, gluon gluon fusion cross section is between one and two orders of magnitude larger than that of VBF for a wide range of Higgs masses. Nevertheless, the VBF becomes competitive when the mass approaches to 1 TeV. Furthermore, in case of gluon gluon production mechanism, in addition to the two lepton and two neutrinos, there are rarely one or more jets coming from initial state radiation, Sec. 3.2. The VBF production, providing two more jets (the VBF jets coming from the hadronization quarks from production) to the final state, benefits from a highly reduced background with respect to the gluon gluon production mode, such that even if the VBF production mechanism has a branching ratio smaller than the gluon-gluon fusion, a higher signal-to-noise ratio is expected.

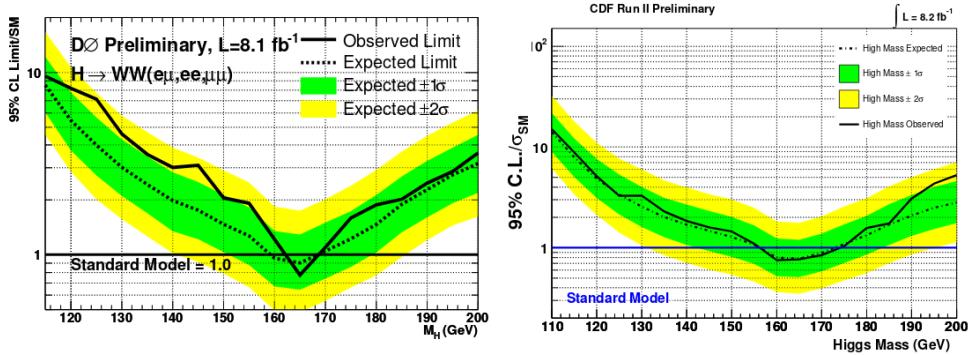


Figure 1.12. Combined limits using the di-lepton channels for D0 (left) and CDF (right).

X → WW searches at colliders

The search of high mass particle with WW final state has been widely performed at experiments at hadron colliders to search for new particles beyond the SM. The resonant WW production has been studied at both the Fermilab Tevatron Collider, and the CERN Large Hadron Collider, with the progressively increasing energy collision and integrated luminosity. Each machine in its time has therefore probed the highest masses of resonances accessible. A review of the different searches performed by D0, CDF, ATLAS and CMS, their techniques, data, results, and limits on new particles decaying to WW are described.

Searches at Tevatron Before the Higgs boson discovery, the mass searches for the high masses SM-like Higgs boson has been performed at the CDF and D0 detector, being the Higgs boson mass a free parameter. These searches result in exclusions of the high mass range $156.5 < m_H < 173.7$ GeV for CDF and $161 < m_H < 170$ GeV for D0 [23]. The high mass searches at CDF and D0 require at least one electron or a muon in the final state in order to suppress the QCD background. Given this requirement all possible decay modes are considered to maximise the signal acceptance. The di-lepton plus missing transvere energy channel requires two electrons or muons (plus neutrinos) of opposite charge in the final state. This represents a small WW decay branching ratio, but a clean signature offering the highest sensitivity of all the high mass channels. CDF and D0 set limits by combining all of the SM high mass channels with up to 8.2 fb^{-1} of integrated luminosity, Fig 1.12.

Searches at LHC After the discovery of the Higgs boson at LHC, the two experiments, ATLAS and CMS, have been focused on high mass searches using the data collected at $\sqrt{s} = 7, 8$ (Run-I) and 13 TeV (Run-II). A search for a heavy Higgs boson in the $H \rightarrow WW$ and $H \rightarrow ZZ$ decay channels has been performed by the CMS experiment based on proton-proton collision data samples corresponding to an integrated luminosity of up to 5.1 fb^{-1} at $\sqrt{s} = 7$ TeV and up to 19.7 fb^{-1} at $\sqrt{s} = 8$ [24]. This analysis is performed in a mass range $145 < m_H < 1000$ GeV and searches for a heavy Higgs boson in the EW singlet extension of the SM. The peculiarity of this analysis is that the full Run-I statistics is considered and

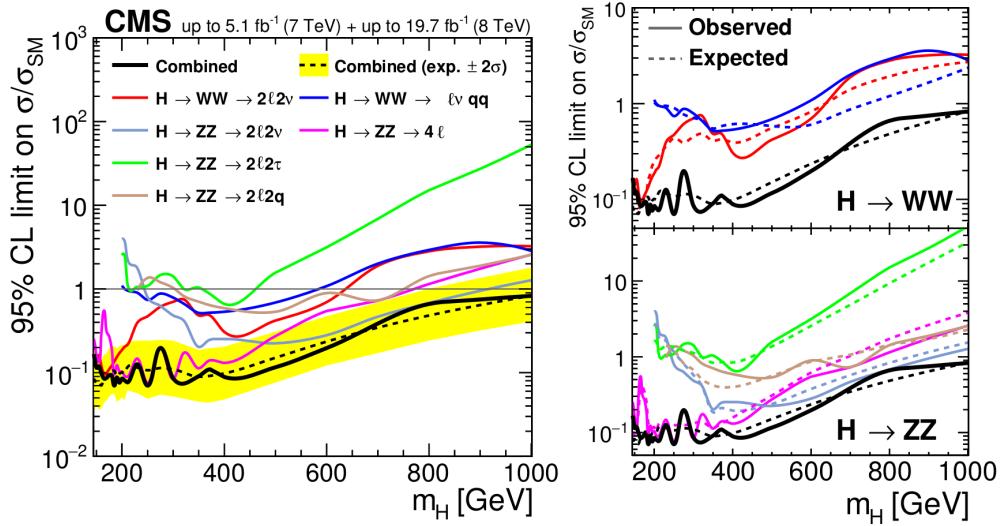


Figure 1.13. Upper limits at the 95% CL for each of the contributing final states and their combination.

the mass range reach 1 TeV for the first time in CMS. In the case of a high Higgs boson decaying into a pair of W bosons, the fully leptonic ($X \rightarrow WW \rightarrow 2\ell 2\nu$) and semileptonic ($X \rightarrow WW \rightarrow \ell\nu qq$) final states are considered in this analysis. Decays containing four charged leptons ($X \rightarrow WW \rightarrow 2\ell 2\ell'$), two charged leptons and two quarks ($X \rightarrow ZZ \rightarrow \ell\nu qq$) and two charged leptons and two neutrinos ($X \rightarrow ZZ \rightarrow 2\ell 2\nu$) are considered. No significant excess over the expected SM background has been observed and exclusion limits have been set. The combined results obtained for a heavy Higgs boson with SM-like couplings for all the different contributing final states are displayed in Fig. 1.13. On the left, the observed 95% CL limit is shown for each final state. On the right, the expected and observed limits are displayed for each of the individual channels as well as the combined result. Using the 13 TeV proton-proton collision data produced at the LHC in 2015, corresponding to an integrated luminosity of 2.3 fb^{-1} , CMS has performed the search for a final state with different flavour leptons ($X \rightarrow WW \rightarrow 2\ell 2\nu$) in the mass range $200 < m_H < 1000 \text{ GeV}$ [25]. The search has been carried out in the 0-jets, 1-jet and VBF categories in order to increase the signal sensitivity to different production mechanisms and maximize the exclusion limits. No significant excess with respect to the SM background expectation has been observed, and the exclusion limits on the cross section times branching ratio to $WW \rightarrow 2\ell 2\nu$ ave been reported, Fig. 1.14 The ATLAS collaboration published the results of a search for a heavy neutral scalar decaying to two W bosons using the datasets collected in 2015 and early 2016 at a centre-of-mass energy $\sqrt{s} = 13 \text{ TeV}$ corresponding to an integrated luminosity of 13.2 fb^{-1} in the mass range between 300 GeV and 3 TeV [26]. In this analysis, categories with one- and at least two-jets are optimised for a vector boson fusion-like signal and the remaining category is quasi-inclusive for a gluon gluon fusion-like signal. The search sensitivity depends on the assumed Higgs boson width. Two different hypotheses are tested: a narrow width approximation, where the width

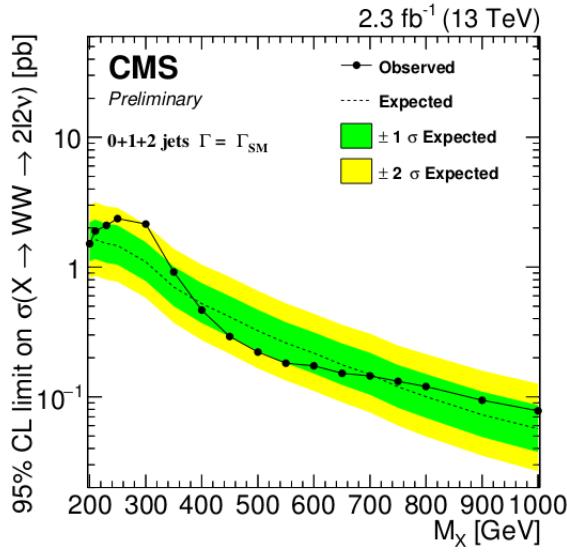


Figure 1.14. Expected and observed exclusion limits at 95% CL on the sum of ggH and VBF cross sections times branching fraction for the combination of the three jet categories as a function of the resonance mass. The black dotted line corresponds to the central value while the yellow and green bands represent the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainties respectively.

of the heavy Higgs boson is smaller than the experimental resolution, and a large width assumption, where widths of 5%, 10%, and 15% of the heavy Higgs boson mass are considered. Upper limits are set on the production cross section times the $H \rightarrow WW$ branching ratio in two scenarios: a high-mass Higgs boson with a narrow width, and with intermediate widths (of 5, 10, 15% of the heavy Higgs boson mass), Fig. 1.15. Values above 4.3 pb (1.4 pb) at $m_H = 300$ GeV (400 GeV) and above 0.051 pb (0.071 pb) at 3 TeV are excluded at 95% CL by the gluon gluon fusion quasi-inclusive NWA (LWA 15%) analysis. For the VBF NWA case, the upper exclusion limit ranges between 1.1 pb at $m_H = 300$ GeV to 0.03 pb at 3 TeV.

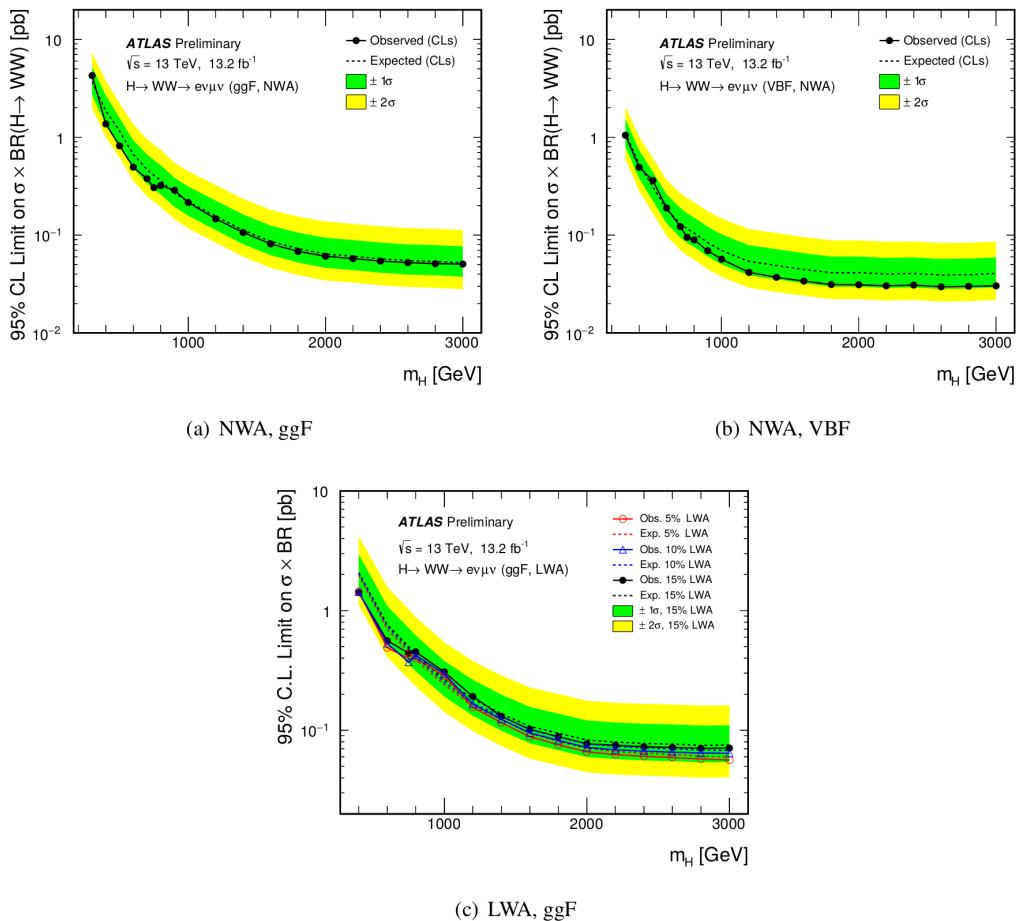


Figure 1.15. 95% CL upper limits on the Higgs production cross section times branching ratio in the analysis, for signals with narrow-width (gluon gluon fusion or VBF) in the top row and the 5%, 10% and 15% width lineshapes (gluon gluon fussion only) in the bottom. The green and yellow bands show the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainties on the expected limit.

Chapter 2

The CMS experiment at LHC

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [27] at CERN, on 2008, is the largest and most powerful hadron collider ever built. Installed in the underground tunnel which housed the Large Electron Positron Collider (LEP), the leptonic accelerator in operation until 2nd November 2000, the LHC accelerator has the shape of a circle with a length of about 27 km and is located at a depth varying between 50 m to 175 m, straddling the Franco-Swiss border near Geneva. It is designed to collide two 7 TeV counter-circulating beams of protons resulting in a center-of-mass energy of 14 TeV, or two beams of heavy ions, in particular lead nuclei at an energy of 2.76 TeV/nucleon in the center-of-mass frame. The transition from a leptonic collider to a hadronic collider entailed the following advantages: first, it has been possible to build a machine that having the same size of the previous one (and therefore accommodated in the same LEP tunnel, substantially reducing the cost and time of construction), could reach a higher energy in the center-of-mass frame. This is due to the much lower amount of energy loss through synchrotron radiation emitted by the accelerated particles, that is proportional to the fourth power of the ratio E/m between their energy and their mass. Secondly, the composite structure of protons compared to the elementary structure of electrons allows LHC to be able to access simultaneously a wider energy spectrum, despite the production of many low energies particles in a complex environment. This feature is particularly important for a machine dedicated to the discovery of “new” physics. Schematic description of the accelerator complex installed at CERN is shown in Fig. 2.1 The acceleration is performed in several stages. The protons source is a Duoplasma- tron: the protons are obtained by removing electrons from a source of hydrogen gas and then sent to the LINAC2, a 36 m long linear accelerator which generates a pulsed beam with an energy of 50 MeV using Radio Frequency Quadrupoles (RFQ) and focusing quadrupole magnets. The beam is subsequently sent to the Proton Synchrotron Booster (PSB), a circular accelerator consisting of four superimposed synchrotron rings with a circumference of about 160 m, which increases the proton energy up to 1.4 GeV. Then, protons are injected into the Proton Synchrotron (PS), a single synchrotron ring with a circumference of about 600 m where the energy is increased to 25 GeV. The sequential combination of these two synchrotrons also allows to

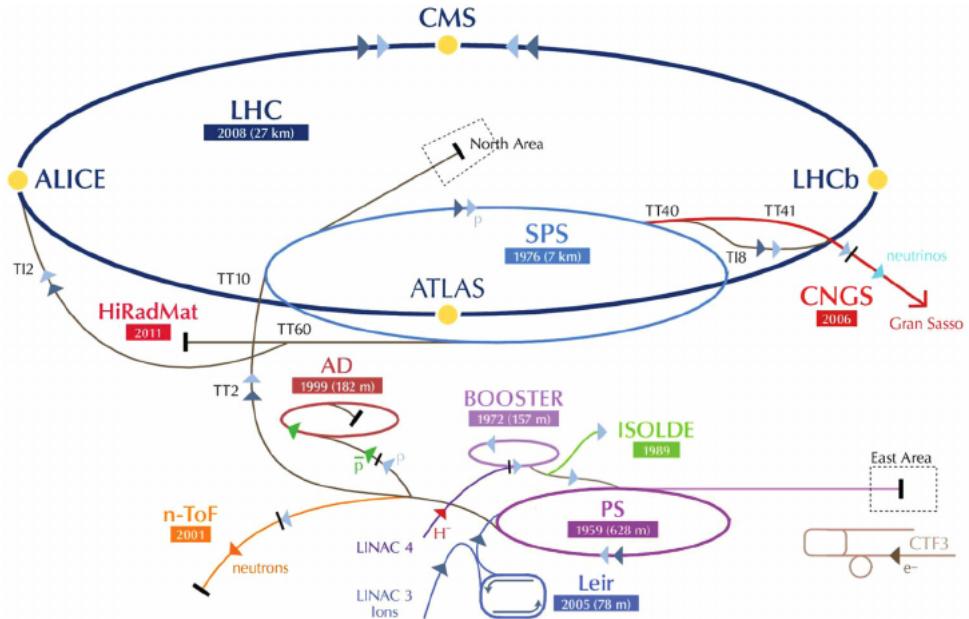


Figure 2.1. Schematic description of the accelerator complex installed at CERN.

create a series of protons bunches interspersed by 25 ns (i.e. at the frequency of 40 MHz) as required for the final correct operation of LHC. The final proton injection stage is the Super Proton Synchrotron (SPS), a synchrotron with a circumference of approximately 7 km where protons reach an energy value of 450 GeV. Subsequently, protons are extracted and injected into the LHC ring via two transmission lines, to generate two beams running in opposite directions in two parallel pipes and which are accelerated up to the energy of interest. The beams collide at four interaction points where the four main experiments: ALICE, ATLAS, CMS and LHCb. Two small experiments, TOTEM and LHCf, which focus on forward particles, have also been built. The 7 TeV per-beam-energy limit on the LHC is not determined by the electric field generated by the radiofrequency cavity but by the magnetic field necessary to maintain the protons in orbit, given the current technology for the superconducting magnets. One of the most important parameter of an accelerator is the instantaneous luminosity \mathcal{L} . For a process having a cross section σ and producing N particles per time unit, the instantaneous luminosity \mathcal{L} is defined by the relation

$$N = \mathcal{L}\sigma \quad (2.1)$$

Then the integrated luminosity L can be defined as

$$L = \int \mathcal{L} dt \quad (2.2)$$

2.2 The Compact Muon Solenoid experiment

The Compact Muon Solenoid (CMS) is a general purpose detector optimized for the proton-proton interactions analysis with the expected energy and luminosity of the LHC particle accelerator design, identifying with precision muons, electrons and

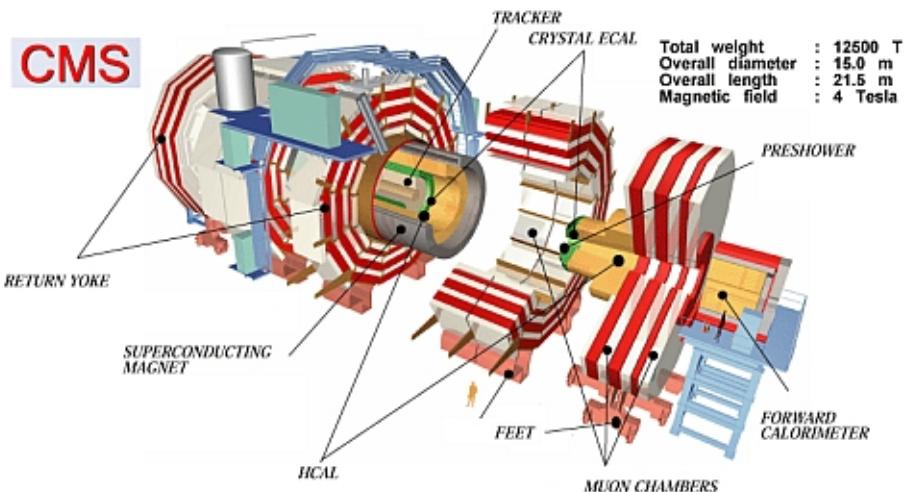


Figure 2.2. A view of the CMS detector with its subdetectors labeled.

photons. It has been designed to investigate a wide range of physics, with the search for the Higgs boson as one of the main highlights. Search for new physics is also an important goal of the experiment, as well as top physics and, of course, Standard Model precision measurements. Although it has the same scientific goals as the ATLAS experiment, it uses different technical solutions and a different magnet-system design. The CMS experiment is one of the largest international scientific collaborations in history, involving more than 4000 people (particle physicists, engineers, technicians, students and support staff) from about 180 universities and institutes in more than 40 countries. The experiment is placed in a cavern 100 m underground in the area called Point 5 (an old LEP access point) near the village of Cessy, in France. The coordinate system used in CMS is a right-handed Cartesian system, having the origin in the nominal beam collision point inside the detector. The x-axis points radially towards the center of the LHC ring, the y-axis is directed upwards along the vertical and the z- axis is oriented along the direction of the beams, along the counter-clockwise direction of the ring if seen from above. The cylindrical symmetry of CMS design and the invariant description of proton-proton physics suggest to define a new coordinate system based on pseudo-angular coordinates, given by the triplet (r, ϕ, η) where r is the distance from the z -axis, ϕ is the azimuth angle measured on the $x - y$ plane starting from the x -axis and η is the pseudorapidity, App. A.

The CMS detector, Fig. 2.2, is 21.6 m long, has a diameter of 15 m and a weight of about 12,500 tons. The constructive element that characterizes the experiment is a solenoidal superconducting magnet which produces an internal constant magnetic field of 3.8 T along the direction of the beams. The CMS detector is designed as a dodecagonal base prism. The central part of the prism, named barrel, contains several layers of detectors with cylindrical symmetry, coaxial with respect to the direction of the beams. A set of detector disks, called endcaps, close the detector at its ends, to ensure its tightness. From the inner region to the outer one, the various components of CMS are:

- Silicon Tracker: it is placed in the region $r < 1.2$ m and $|\eta| < 2.5$. It consists of a silicon pixel vertex detector and a surrounding silicon microstrip detector, with a total active area of about 215 m^2 . It is used to reconstruct charged particle tracks and vertices;
- Electromagnetic Calorimeter (ECAL): it is placed in the region $1.2 \text{ m} < r < 1.8 \text{ m}$ and $|\eta| < 3$. It consists of scintillating crystals of lead tungstate and it is used to measure the trajectory and the energy released by photons and electrons;
- Hadron Calorimeter (HCAL): it is placed in the region $1.8 \text{ m} < r < 2.9 \text{ m}$ and $|\eta| < 5$. It consists of brass layers alternated with plastic scintillators and it is used to measure the direction and the energy released by the hadrons produced in the interactions;
- Superconducting Solenoidal Magnet: it is placed in the region $2.9 \text{ m} < r < 3.8 \text{ m}$ and $|\eta| < 1.5$. It generates an internal uniform magnetic field of 3.8 T along the direction of the beams, necessary to deflect the charged particles in order to allow a measurement of their momentum through the curvature observed in the tracking system. The magnetic field is closed with an iron yoke 21.6 m long with a diameter of 14 m, where a residual magnetic field of 1.8 T is present, in the opposite direction with respect to the 3.8 T field;
- Muon System: it is placed in the region $4 \text{ m} < r < 7.4 \text{ m}$ and $|\eta| < 2.4$. It consists of Drift Tubes (DT) in the barrel region and Cathode Strip Chambers (CSC) in the endcaps. A complementary system of Resistive Plate Chambers (RPC) is used both in the barrel and in the endcaps. This composite tracking system for muons is used to reconstruct tracks released by muons that pass through it. The muons chambers are housed inside the iron structure of the return yoke that encloses the magnetic field.

The Tracker

The silicon tracker is the detector closest to the beams collision point. Its goal is the high resolution reconstruction of the trajectories of charged particles originating from the interaction point and the identification of the position of secondary vertices produced by particles with a short mean life time (in particular hadrons containing the b quark, that decay after few hundreds of μm). The events produced in the proton- proton collisions can be very complex and track reconstruction is an involved pattern recognition problem. Indeed, at the nominal instantaneous luminosity of operation, an average of about 20 pile-up events overlapping to the event of interest are expected, leading to about 1000 tracks to be reconstructed every 25 ns. In order to make the pattern recognition easier, two requirements are fundamental: a low occupancy detector and a large redundancy of the measured points (hits) per track. The first requirement is achieved building a detector with high granularity [3]. The redundancy of the hits is instead achieved having several detecting layers, and is necessary to reduce the ambiguity on the assignment of the hits to a given track. Nevertheless, the amount of tracker material has to be as low as possible, in order

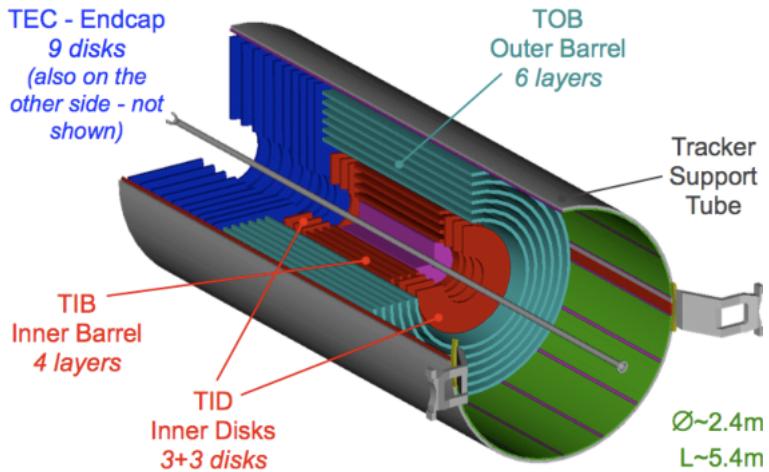


Figure 2.3. Three-dimensional schematic view of the CMS silicon tracker.

to avoid compromising the measurement of the particle trajectory. An excessive amount of material would indeed deteriorate the measurement, mainly because of the increased probability of particle multiple scattering. The outer detectors such as ECAL would be influenced by the material as well, for example because of the increased probability for a photon to convert to an electron-positron pair in the tracker material. For this reasons, the tracker layers are limited in number and thickness. The tracker comprises a large silicon strip detector with a small silicon pixel detector inside it. In the central η region, the pixel tracker consists of three co-axial barrel layers at radii between 4.4 cm and 10.2 cm and the strip tracker consists of ten co-axial barrel layers extending outwards to a radius of 110 cm. Both subdetectors are completed by endcaps on either side of the barrel, each consisting of two disks in the pixel tracker, and three small plus nine large disks in the strip tracker. The endcaps extend the acceptance of the tracker up to $|\eta| < 2.5$. A three-dimensional schematic view of the tracker is shown in Fig. 2.3, while in Fig. 2.4 a pictorial representation of a slice of the tracker is displayed, showing the various layers of the subdetectors. The whole tracker has a cylindrical shape with a length of 5.8 m and a diameter of 2.5 m, with the axis aligned to the beams direction. The average number of hits per track is 12-14, allowing high reconstruction efficiency and low rate of fake tracks.

The Pixel Vertex Detector The pixel vertex detector, is mainly used in CMS as a starting point for the reconstruction of tracks and is essential for the reconstruction of the primary vertex (PV) and any possible secondary vertices. It is placed in the region closest to the collision point, where the particle flow is maximum. It covers the region $|\eta| < 2.5$ and is composed of a central part (barrel) and by two forward parts (endcaps). The barrel is composed of three concentric cylindrical sectors 53 cm long, located at an average distance r of 4.4 cm, 7.3 cm and 10.2 cm. Each half-cylinder is composed of ladders and half ladders that serve as support and cooling structure for the modules of pixels, with each ladder containing 8 modules. In total, the barrel is composed of 768 modules. Each endcap is composed of two disks placed at a distance of 34.5 cm and 46.5 cm from the nominal beams impact

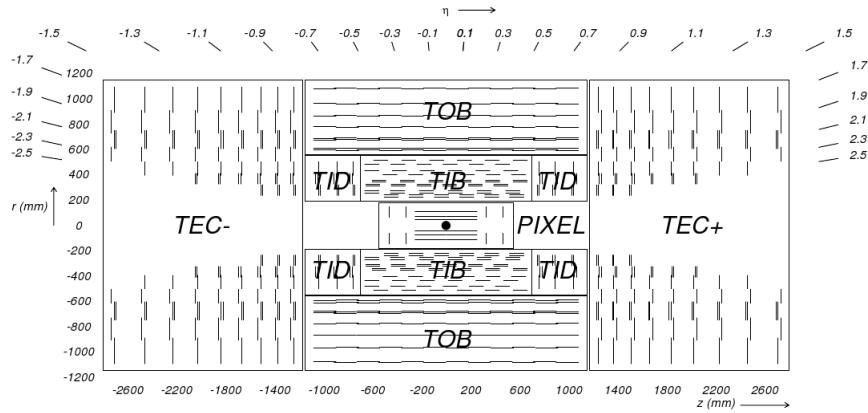


Figure 2.4. Pictorial view of a tracker slice in the r - z plane. Pixel modules are shown in red, single-sided strip modules are depicted as black thin lines and strip stereo modules are shown as blue thick lines.

point. They cover a radius r in a range between 6 cm and 15 cm in such a way that each track included in the detector acceptance passes through at least two layers. Each disk is divided into 24 segments, on each of which 7 modules of different sizes are mounted, for a total of 672 modules on all the endcaps. Each module is composed of several units that contain a highly integrated and segmented silicon sensor with a thickness of $250 \mu\text{m}$. In order to optimize the reconstruction of vertices and the track parameters near the vertex, a set of rectangular pixels with a size of $150 \times 100 \mu\text{m}^2$ are used, with the $100 \mu\text{m}$ side oriented along the $r\phi$ direction in the barrel region and along the z direction in the endcap region. The resolution in the hit reconstruction is about $10\text{-}15 \mu\text{m}$ in the barrel and about $15 \mu\text{m}$ in the endcaps.

The Silicon Microstrip Detector In the region of the detector that is more than 20 cm far from the beam, the flow of particles is sufficiently limited to allow the use of a silicon microstrip detector (Silicon Strip Tracker, SST). Overall, this detector consists of 15400 units (modules), composed of one or two sensors stucked on a support of carbon fiber, together with the readout electronics. In case of a “doubled” sensor, the second detector is rotated with respect to the first one in order to have strips forming an angle of 100 mrad between them. This “stereo” combination, although of lower resolution, is preferable compared to a pixel segmentation because it has a lower number of readout channels. The ambiguities due to the hit recognition are resolved in the process of reconstruction of the entire track. The silicon microstrip tracker is 5.4 m long, extending up to a distance of 1.1 m from the axis of the beams. It consists of a barrel and two endcaps and it is divided into four distinct parts, TIB and TOB, and TID and TEC.

The Electromagnetic Calorimeter (ECAL)

The main function of an electromagnetic calorimeter is to identify electrons and photons and to measure accurately their energy. The electromagnetic calorimeter

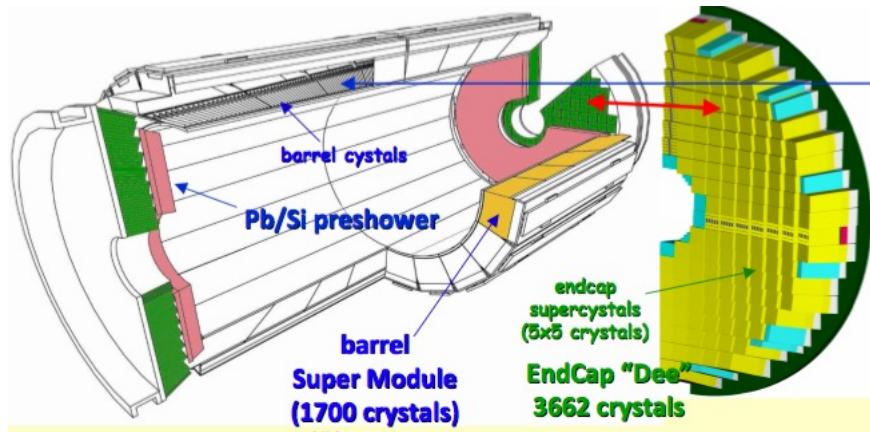


Figure 2.5. Schematic representation of the electromagnetic calorimeter ECAL.

(Fig. 2.5) of CMS (ECAL, Electromagnetic CALorimeter) is a homogeneous calorimeter with cylindrical geometry, whose elements are scintillating crystals of lead tungstate (PbWO_4) with a truncated pyramidal shape. It consists of an ECAL Barrel (EB) with 61200 crystals and two ECAL Endcaps (EE) containing 7324 crystals each one. Crystals are grouped into 5×5 matrices called towers. The barrel has an inner radius of 129 cm, a length of 630 cm and it extends in the region $|\eta| < 1.479$. Crystals in the ECAL barrel have the following dimensions: $22 \times 22 \text{ mm}^2$ at the front face, $26 \times 26 \text{ mm}^2$ at the rear face, and a length of 23 cm, corresponding to $25.8 X_0$. Each submodule, consisting in a 5×2 crystals arrays mounted on a glass fiber structure, forms the elementar unit of EB. The granularity of a single crystal is about 1 grade. To avoid that cracks might align with particle trajectories, the crystal axes are tilted by 3 grade with respect to the direction from the interaction point, both in the η and ϕ direction. Each endcap covers the region $1.479 < |\eta| < 3$ and is formed by two semicircular halves of aluminum called dees. Crystals in endcaps have a length of 22 cm and frontal area equal to $28.6 \times 28.6 \text{ mm}^2$. They are arranged in supercrystals with 5×5 elementary unity. Unlike the crystals in the barrel, arranged in a $\eta - \phi$ symmetry, the endcap crystals are arranged according to a xy geometry. Two preshower detectors are placed in front of the endcaps in order to separate the showers produced by a primary γ from those produced by a primary π_0 . This detector, which covers the region $1.653 < |\eta| < 2.6$, is a sampling calorimeter and it consists of two disks of lead converters that start the electromagnetic shower of the incident photon/electron, alternating with two layers of silicon microstrip detectors in which a measurement of the released energy and the identification of the shower profile are performed. The strips are arranged orthogonally in the two planes, according to a xy configuration. The choice of the PbWO_4 crystals as scintillating material for ECAL is due to several reasons. First, the high-density the short radiation length and the reduced Molière radius 5 ($R_M = 2.2 \text{ cm}$) allow to build a compact and high granularity calorimeter. Furthermore, the 15 ns decay scintillation time allows to collect about 80% of the emitted light during the 25 ns that exist between two consecutive beam interactions in the LHC design performance. Finally, the PbWO_4 crystals have a good intrinsic radiation hardness and therefore they can operate for years in the hard LHC environment,

with a modest deterioration in performance. The main disadvantage of these crystals is the low light yield which makes an internal amplification for the photodetectors necessary. This is achieved through the use of silicon avalanche photodiodes in barrel and single stage photomultipliers in the endcaps, both resistant to the radiation and to the strong magnetic field of CMS. The energy resolution of a homogeneous calorimeter is usually expressed by the sum in quadrature of three terms, according to the formula,

$$\frac{\sigma_e}{E} = \frac{a}{\sqrt{E}} + \frac{b}{E} + c, \quad (2.3)$$

The stochastic term a is the dominant term at low energies: it includes the contribution of statistical fluctuations in the number of photoelectrons generated and collected. The noise term b includes contributions from the electronic noise, both due to the photodetector and to the preamplifier, and from pileup events. The constant term c , dominant at high energies, takes into account several contributions: the stability of the operating conditions (in particular of temperature and voltage), the presence of dead material in front of the crystals and the rear leakage of the electromagnetic shower, the longitudinal non uniformity of the crystal light yield, the intercalibration errors and the radiation damage of the crystals. The ECAL barrel energy resolution for electrons in beam tests has been measured to: $a = 2.8\% \text{ GeV}^{-1/2}$, $b = 12\% \text{ GeV}$, $c 0.3\%$, where the energy E is measured in GeV.

The Hadronic Calorimeter (HCAL)

The hadronic calorimeter HCAL (Hadronic CALorimeter), together with the electromagnetic calorimeter, makes a complete calorimetric system for the jet energy and direction measurement. Furthermore, thanks to its tightness, it can provide a measurement of the features of non-interacting particles, such as neutrinos, by measuring the missing energy deposited in the transverse plane, E_T^{Miss} or MET. The CMS hadronic calorimeter is a hermetic sampling calorimeter that covers the region $|\eta| < 5$. As shown in Fig. 2.7, it is divided into four subdetectors: HB (Barrel Hadronic Calorimeter), located in the barrel region inside the magnet, extending up to pseudorapidities $|\eta| < 1.4$; HE (Endcap Hadronic Calorimeter), situated in the endcaps region inside the magnet, extends in the pseudorapidity region $1.3 < |\eta| < 3$, partially overlapping the HB coverage; HO (Outer Hadronic Calorimeter, also called tail-catcher, placed along the inner wall of the magnetic field return yoke, just outside of the magnet; HF (Forward Hadronic Calorimeter), a sampling calorimeter consisting of quartz fibers sandwiched between iron absorbers, consisting of two units placed in the very forward region ($3 < |\eta| < 5$) outside the magnetic coil. The quartz fibers emit Cherenkov light with the passage of charged particles and this light is detected by radiation resistant photomultipliers. In order to maximize particle containment for a precise missing transverse energy measurement, the amount of absorber material was maximized, reducing therefore the amount of the active material. Since HCAL is mostly placed inside the magnetic coil, a non-magnetic material like brass was chosen as absorber. HB and HE are therefore made with brass absorber layers interleaved with plastic scintillators (wavelength shifters, WLS) coupled to transparent optical fibers, which transmit the light to the HPD (Hybrid Photodiodes) photodetectors.

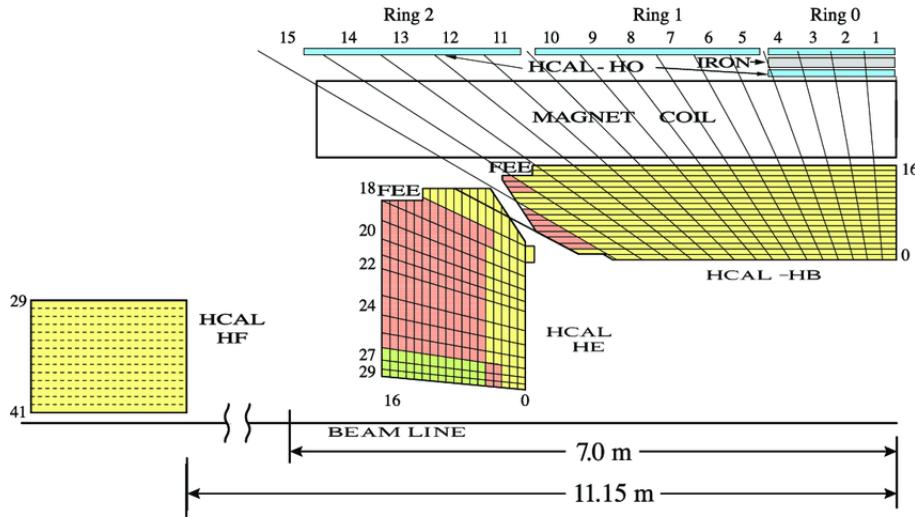


Figure 2.6. A schematic rz view of a quadrant of the CMS hadronic calorimeter HCAL.

The Solenoid Magnet

The CMS magnet, which houses the tracker, the electromagnetic and the hadronic calorimeters, is the biggest superconducting solenoid ever built in the world. The solenoid achieves a magnetic field of 3.8 T in the free bore of 6 m in diameter and 12.5 m in length. The energy stored in the magnet is about 2.6 GJ at full current. The superconductor is made of four Niobium-Titanium layers. In case of a quench, when the magnet loses its superconducting property, the energy is dumped to resistors within 200 ms. The magnet return yoke of the barrel is composed with three sections along the z-axis; each one is split into 4 layers (holding the muon chambers in the gaps). Most of the iron volume is saturated or nearly saturated, and the field in the yoke is about the half (1.8 T) of the field in the central volume.

The Muon chamber

The CMS Muon System is dedicated to the identification and measurement of high p_T muons, in combination with the tracker. Furthermore, it provides a time measurement of the bunch-crossing and also works as trigger for events involving muons. Momentum measurement, in the muon system, is determined by the muon bending angle at the exit of the 3.8 T coil, considering the interaction point as the origin of the muon. Up to p_T values of 200 GeV, the resolution of the muon system is dominated by multiple scattering and the best resolution is rather given by the silicon tracker, , Fig 2.8. The system is placed outside the magnetic coil, embedded in the return yoke, to fully exploit the 1.8 T return flux. It consists of three independent subsystems, as shown in Fig. 2.9: drift tubes (DT), cathode strip chambers (CSC) and resistive plate chambers (RPC). The DT and the CSC provide an excellent spatial resolution for the measurement of charged particle momentum; the RPC are used for trigger issues because of the very good timing. The active parts of the muon system are hosted into stations which are interleaved by the iron layers of the return yoke of the magnet.

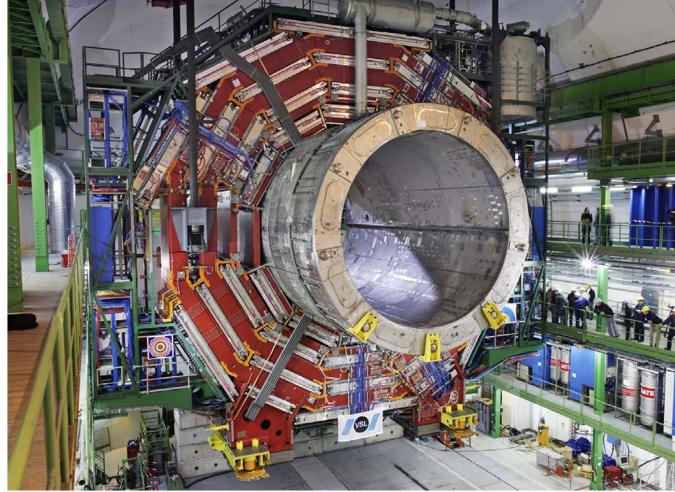


Figure 2.7. Arrival of the magnet in the tunnel on February 28, 2007.

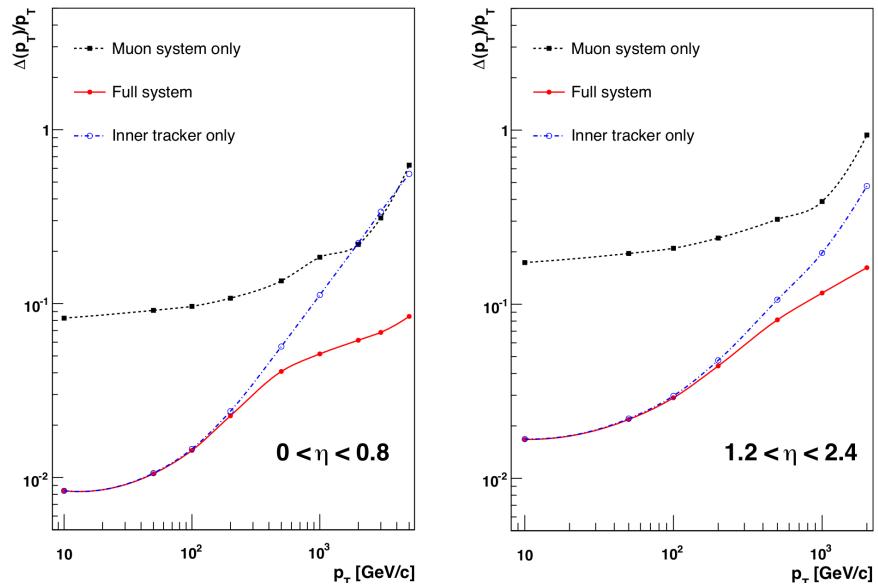


Figure 2.8. Trend resolution for the muon systems. On the left the barrel zone. On the right the endcap.

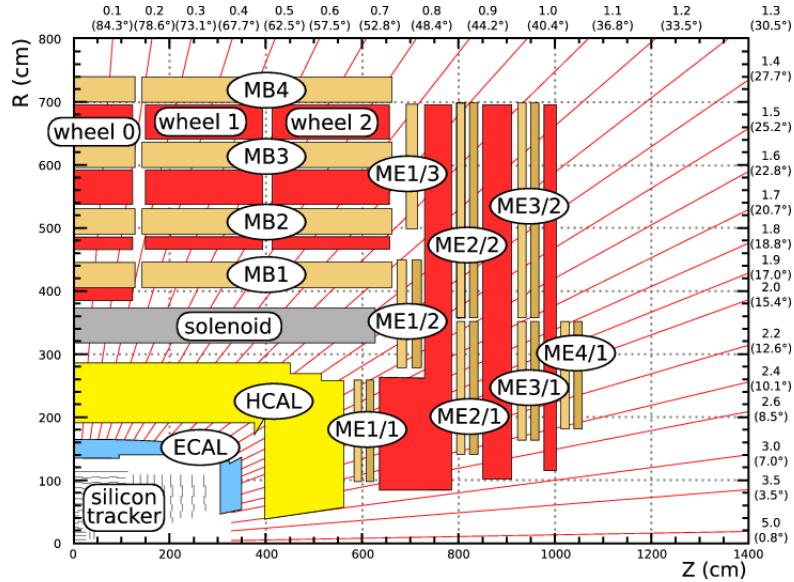


Figure 2.9. Schematic overview of the muon chambers.

Trigger and Data Acquisition

LHC can produce interactions at 40 MHz frequency, but only a small fraction of these events can be written on disk. On one hand the speed at which data can be written to mass storage is limited, on the other hand the vast majority of events produced is not interesting, because it involves low transverse momentum interactions (minimum bias events). Thus, a trigger system is needed to save interesting events at the highest possible rate. The maximum rate of events written on disk is about 800 Hz. CMS has chosen a two-level trigger system, consisting of a Level-1 Trigger (L1) and a High Level Trigger (HLT). Level-1 trigger runs on dedicated processors, and accesses coarse level granularity information from calorimetry and muon system. A L1 Trigger decision has to be taken for each bunch crossing within $3.2 \mu s$. Its task is to reduce the data flow from 40 MHz to about 100 kHz. The High Level Trigger is responsible for reducing the L1 output rate down to a maximum rate of the order of 1 kHz. The HLT code runs on a farm of commercial processors and can access the full granularity information of all the subdetectors.

2.3 Data recoiled and future plans

The first proton beam circulated in the LHC on September 2008, after more than a decade of construction and installation. An incident occurred between two magnets, causing the release of helium into the tunnel and mechanical damage. After that, in March 2010 started the Run-I, a fruitful data taking era that lasted until 2012. It was decided not to operate the LHC at its design parameters, and proton proton collisions took place at a centre of mass energy of 7 TeV and 8 TeV. The amount of recoil data (in CMS) in this period is reported in Fig. 2.10. At the end of 2012, LHC operations halted for two-year due to long shutdown (LS1). In 2015 with centre-of-mass energy of 13 TeV the proton proton collision restarted (Run-II) and

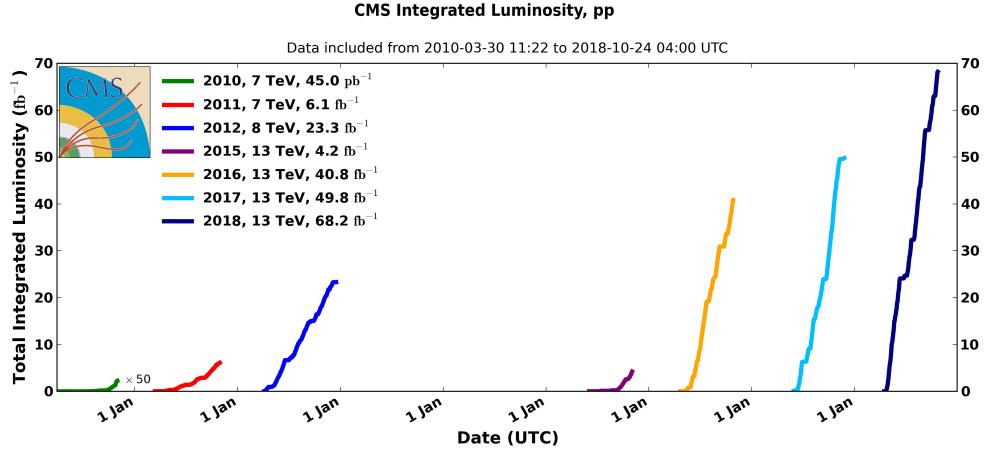


Figure 2.10. Run-I and Run-II integrated luminosity.

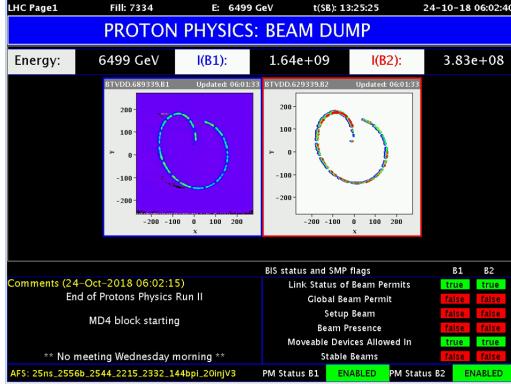


Figure 2.11. Last proton-proton beam dump at the end of Run-II.

in the 2016, LHC was ready to deliver a large dataset to the experiments. The data collected in 2016 are used in the high mass analysis that is the subject of this thesis. Overall, the data correspond to 35.9 fb^{-1} that are data validated for the physics analyses, including the dead times of the experiment. The 2016 LHC operations can be regrouped into several time-periods, labelled with a letter from A to H. In the 2017 and 2018 the data collection has been continued and the total integrated luminosity for RUN-II is around $\sim 150 \text{ fb}^{-1}$. These are a very large amount of data that has been achieved from high energy physics collider. In October 2018 the proton collision has been stopped, Fig. 2.11, and all operation (ion collision) will halt in 2019 for a second long shutdown of ~ 2.5 year (LS2) for the machine and experiments upgrade. The Run-III will start in 2021 with an energy of 14 TeV. After that, the third long shutdown (LS3) starting in 2024 will conclude the Phase I of LHC operations that started back in 2008. The high-luminosity LHC (HL-LHC), or Phase II, starting in ~ 2016 will represent an unprecedented way to study very rare phenomena at the LHC. The machine is expected to deliver, during a decade of operations, an integrated luminosity of about 3000 fb^{-1} , Fig. 2.12.

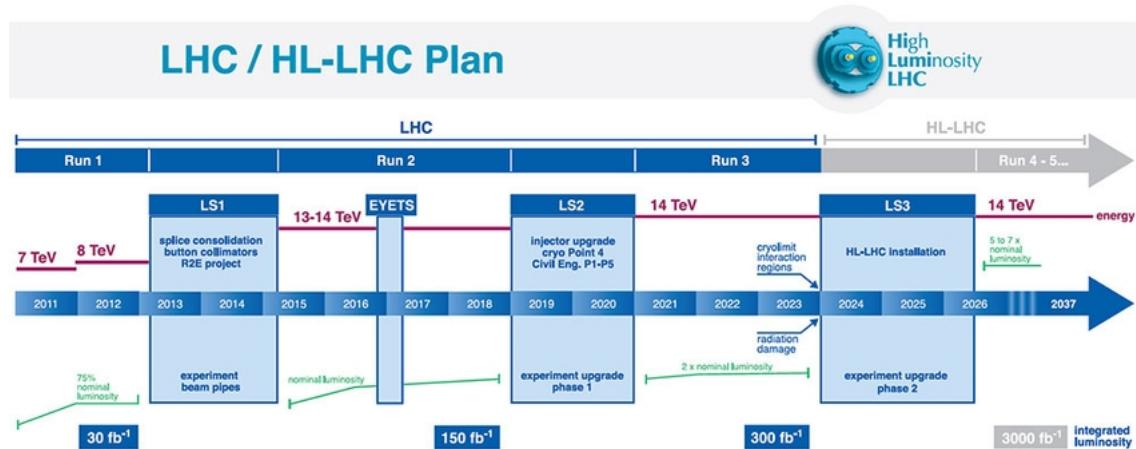


Figure 2.12. Schedule of LHC and HL-LHC operations.

Chapter 3

Monte Carlo Generators

As we have seen in the collisions between high energy protons, hundreds of particles are generally present in the final state. Given the complexity of the events it is necessary to use Monte Carlo generators, i.e. programs that allow to simulate the realistic result of the collisions assuming a certain model for the processes involved. The use of Monte Carlo generators is necessary because it is impossible to predict what happens event-by-event: in fact, in quantum mechanics we can only calculate the probability of having a certain result. The simulation of an event is carried out in successive steps [28, 29], as schematized in Fig. 3.1, thus subdividing the problem into several parts of lower complexity.

The various steps are summarized here:

- Hard process: the incident protons are composed by partons (quarks and gluons) and the hard process consists in a collision between two partons, coming from different hadrons. The matrix element of the process is evaluated perturbatively and often only the lowest perturbative order, called leading order (LO), is calculated.
- Parton shower: the incoming or outgoing partons participating in the hard process can emit gluons: in fact, in analogy with the electromagnetic interaction, a particle with an accelerated color charge can radiate for the bremsstrahlung. The gluons in turn, can produce quark-antiquark pairs thus generating the parton showers. The emission of additional partons takes place mainly in the collinear space respect to the initial parton and progressively with less energies. In the final state there will be a set of partons, called jet, located in the collinear respect to the initial parton. This probabilistic process can be simulated as a Markov process and is implemented in the parton shower algorithms we will discuss later.
- Multiple interactions: in a single collision, it may happen to have more pairs of partons interacting. In this case it is said that there are multiple interactions in addition to the hard process.
- Hadronization: in the evolution of the event the partons are gradually generated with ever lower relative momenta. For momentum values of 1 GeV the confinement forces prevail. At these energy scales, the perturbation theory

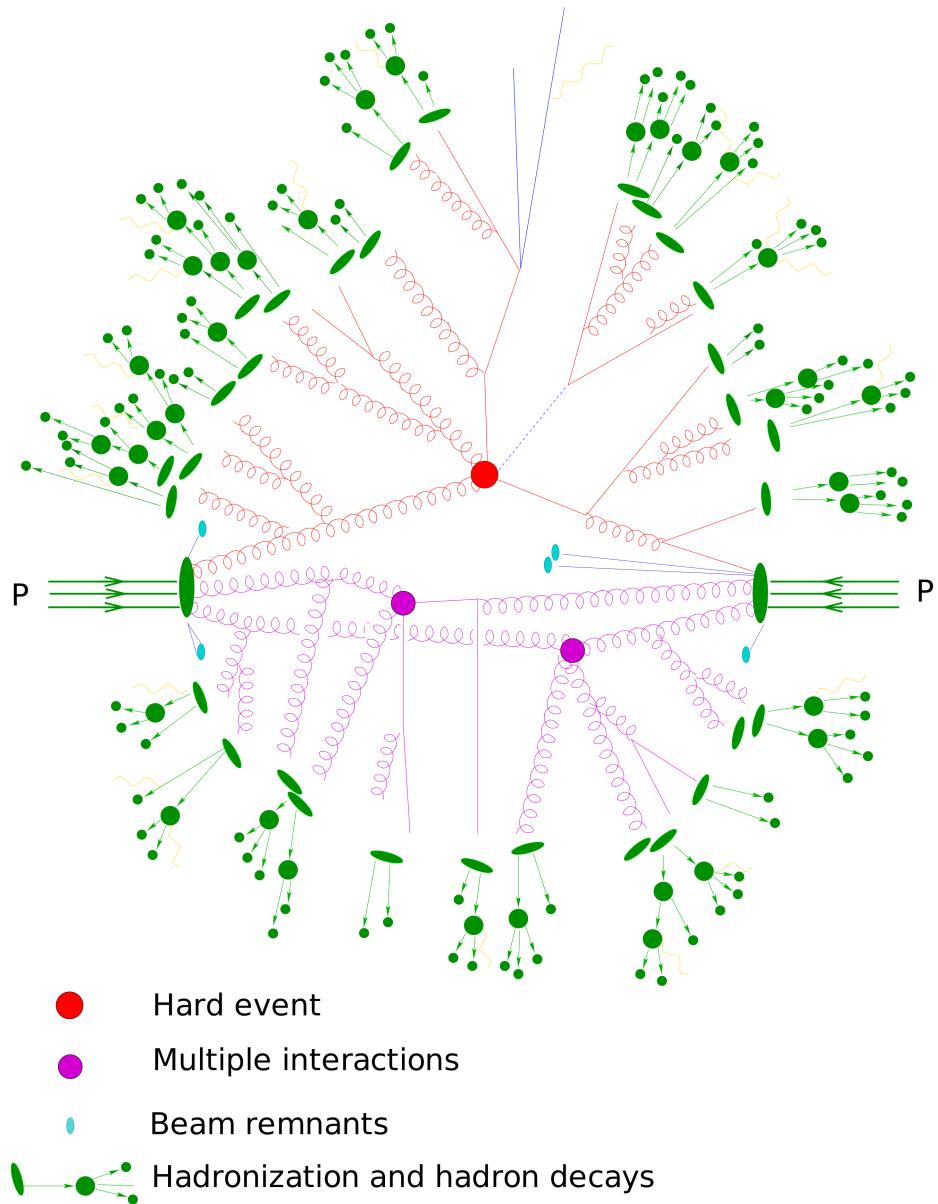


Figure 3.1. Schematic representation of an event generated within an event generator. The partons coming from the protons indicative participate in both the hard process and multiple interactions. Subsequently there is the hadronization.

fails in the description, so we resort to non perturbative models which describe the formation of real hadrons. This hadronization process preserves the jet structure which can therefore be observed experimentally.

- Decaying of unstable particles: many of the particles produced in the primary process are unstable and they decayed unless they interact directly with the detector.

In the Monte Carlo simulation all these steps are considered sequentially: the result of each step is the starting point of the next. At the end, in a single event, there are hundreds of particles each of which has a dozen degrees of freedom (mass, flavor, impulse, average life, spin, vertex production, etc.), so there is a high number of parameters that came into play and must be simulated for each event. The final aim is to provide a realistic description of what happens in high-energy collisions, in order to compare the Monte Carlo model with the experimental data. Schematically, the cross section of the final state is given by,

$$\sigma_{\text{final state}} = \sigma_{\text{hard process}} \mathcal{P}_{\text{tot, hard process} \rightarrow \text{final state}}, \quad (3.1)$$

integrated over the total phase space and summed over all possible final states (e.g. the production of two or more jets). This is the measurable quantity associated with the hard process.

3.1 Hard process

In many processes of interest to LHC high momenta come into play, to produce high mass particles or energetic jet. The simulation of these events is the main goals of the Monte Carlo generators. The cross section for a scattering $ab \rightarrow n$ process is given [29] by,

$$\begin{aligned} \sigma &= \sum_{a,b} \int_0^1 dx_a dx_b \int f_a^{h_1}(x_a, \mu_F) f_b^{h_2}(x_b, \mu_F) d\hat{\sigma}_{ab \rightarrow n}(\mu_F, \mu_R) \\ &= \sum_{a,b} \int_0^1 dx_a dx_b \int d\Phi_n f_a^{h_1}(x_a, \mu_F) f_b^{h_2}(x_b, \mu_F) \\ &\times \frac{1}{2\hat{s}} |\mathcal{M}_{ab \rightarrow n}(\Phi_n, \mu_F, \mu_R)|^2, \end{aligned} \quad (3.2)$$

where

- $f_a^h(x, \mu)$ are the parton density functions (PDF) that depend on the x fraction of the parton a 's energy (Bjorken variable) respect to the h , and on the μ_F factorization scale, which has been introduced in the Eq. XX.
- $\hat{\sigma}_{ab \rightarrow n}$ is the partonic cross section of the process $ab \rightarrow n$. The total differential cross section is given by the product of the corresponding square matrix element, $|\mathcal{M}_{ab \rightarrow n}|^2$, with the incident particle flow $1/(2\hat{s}) = 1/(2x_a x_b s)$, where \sqrt{s} is the energy of the system's center of mass.

- The matrix element $|\mathcal{M}_{ab \rightarrow n}(\Phi_n, \mu_F, \mu_R)|^2$ can be written as the sum on all Feynman diagrams,

$$\mathcal{M}_{ab \rightarrow n} = \sum_i \mathcal{F}_{ab \rightarrow n}^{(i)}. \quad (3.3)$$

- $d\Phi_n$ it is the phase space differential for n particles in the final state.

The phase space will not be all physical space but will contain cuts for two reasons: the first is that the cuts will reflect the geometry and acceptance of the detector; the second because it is necessary put a cut on the transverse impulse of the particles produced in the process to avoid divergences in the calculation of the cross section ¹. In general, the calculation of the matrix element would require the calculation of all the Feynman diagrams which grow in a factorial way (Fig. 3.2) with the number of particles in the final state.

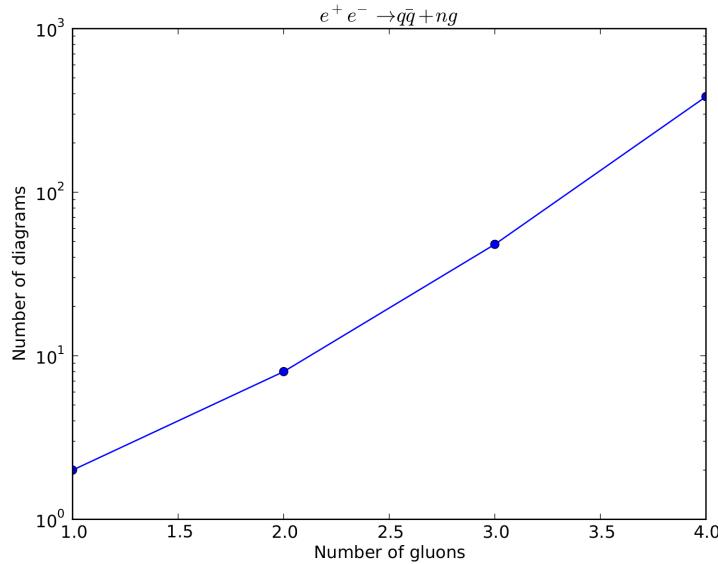


Figure 3.2. Trends in the number of Feynman diagrams as the number n of gluons increases in the process $e^+e^- \rightarrow q\bar{q} + ng$.

Usually the Monte Carlo events generators can compute the matrix element at the leading order for the Standard Model $2 \rightarrow 1$, $2 \rightarrow 2$ and $2 \rightarrow 3$ [?] processes. However, if we stopped at the first perturbative order, we would have only a rough description of the process: in fact, subsequent orders involve important corrections both to the shape of the distributions and to the total cross section. LO is useful for a first study but and it is important to evaluate next-to-leading-order (NLO) ². The cross section calculated at the NLO is composed of three parts: by the LO part or part of Born, by the real and virtual part of the emission corrections (Fig. 3.3),

$$d\sigma^{NLO} = d\tilde{\Phi}_n [\mathcal{B}(\tilde{\Phi}_n) + \alpha_s \mathcal{V}(\tilde{\Phi}_n)] + d\tilde{\Phi}_{n+1} \alpha_s \mathcal{R}(\tilde{\Phi}_{n+1}), \quad (3.4)$$

¹You can imagine having a singularity similar to that one has in scattering classic Coulomb.

²For some particularly important processes, for example $gg \rightarrow H$, the next-next-to-leading-order (NNLO) calculations are even available.

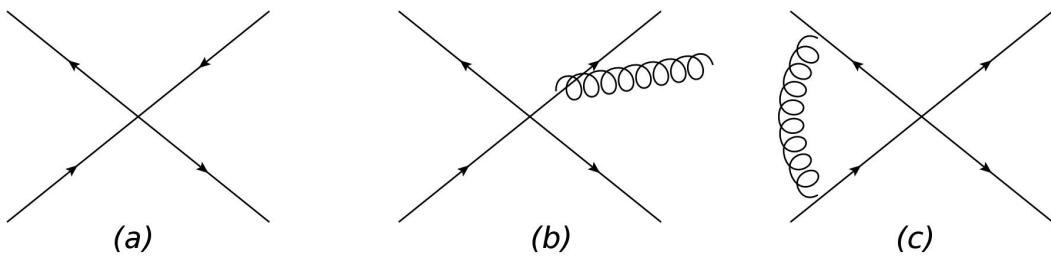


Figure 3.3. Examples of Feynman diagrams (a) Born, (b) real, (c) virtual.

where \mathcal{B} , \mathcal{R} and \mathcal{V} denote the Born, the real and the virtual part respectively. The integral must be made on the n or $n + 1$ final state particles and on the Bjorken variables related to the incident partons. Suppose, in the Born approximation, the process $2 \rightarrow 2$. If you want to go to the next order, NLO, you have to keep the element with an additional parton in the final state, the $2 \rightarrow 3$ process, and virtual correction with a loop in the $2 \rightarrow 2$ process. It should be noted that the cross-section for processes of the type $2 \rightarrow 3$ is divergent when the energy of one of the partons tends to zero (soft divergence) or when two parts are collinear (collinear divergence).

3.2 Parton shower

In a collision between partons a charge of color is accelerated, so there will be bremsstrahlung emission. When studying a process of the type $2 \rightarrow n$, where n represents the number of partons in the final state, the LO matrix elements (called tree-level) will have divergences in the collinear and soft case. In particular, the processes that suffer from this type of divergence are $q \rightarrow qg$, $\bar{q} \rightarrow \bar{q}gq$, $g \rightarrow gg$: the first are similar processes to $e \rightarrow e\gamma$ in QED, while the third is due to the fact that QCD is not an Abelian theory. The process $g \rightarrow q\bar{q}$ does not have this type of divergence. The divergences of the tree-level matrix element can be removed by introducing the virtual corrections into the calculation, but they will be in the next order; these calculations are therefore particularly complex and they are only possible for a limited number of processes. The parton shower [?] algorithms offer an alternative and simple way to eliminate the collinear and soft divergences through:

- an iterative structure that combines the three states suffering from divergences in a single multi-partonic state,
- the introduction of the form factor of Sudakov.

The incoming or outgoing partons, which are far (temporally) from hard process, are called on-shell indeed the module of their four-momenta is equal to the mass at rest. However, closer you get to the interaction, the partons can be in a state called off-shell, where the module of the their four-momenta does not correspond to the mass at rest due to the uncertainty principle ($\Delta E \Delta t \sim \hbar$). For this reason they are able to emit other partons and the energy of emitted partons is higher if they are closer to the scattering. If the emission occurs before the scattering, it is called initial state radiation (ISR), while after the interaction it is called final state radiation (FSR).

Each partition is characterized by a “virtuality scale” Q^2 that corresponds roughly to a shower temporal scale. It is important to stress that different definitions are available for Q^2 ; however regardless of the chosen convention, the Q^2 scale increases as it approaches in the hard process, so then in the ISR, and decreases away, so in the FSR. If we take the FSR, the evolution starts at a Q_{max}^2 scale that is related to the hard process and it ends when a limit scale is reached, Q_0 , which will be on the order of 1 GeV.

The most common choice used is to set $Q^2 = p^2 = E^2 - |\vec{p}|^2$. With this convention in a process of type $a \rightarrow bc$, in FSR case, $Q^2 > 0$, that is of type time-like, and it will decrease until the limit scale Q_0 is reached. The ISR case is more complicated: in this case a and b (supposed off-shell) have p^2 of type space-like, then redefines $Q_i^2 = -m_i^2$ in order to guarantee the increasing order of Q^2 , i.e. $Q_b^2 > Q_a^2$. In contrast, c will not participate in hard process and will have $p^2 > 0$ and therefore its shower will evolve like that of the FSR.

Final State Radiation

In the parton shower approach, the final radiation state is modeled through a series of divisional processes of the type $a \rightarrow bc$. This is evident from the process $q\bar{q}g$, Fig. 3.3 (b), where the first order matrix element corrections correspond to the emission of a gluon. The evolution of the shower is described by two parameters: the fraction of energy carried by one of the two outgoing partons, $z = E_b/E_a$, and the order variable t . As we said, a possible choice for t is the virtuality Q_a^2 of the incoming parton. In the collinear limit the probability of division $d\mathcal{P}_{a \rightarrow bc}$, in z and $t = \ln(Q^2/\Lambda^2)$ is:

$$d\mathcal{P}_{a \rightarrow bc} = \sum_{bc} \frac{\alpha_{abc}}{2\pi} P_{a \rightarrow bc} dt dz, \quad (3.5)$$

where $dt = \frac{dQ^2}{Q^2}$, α_{abc} it is the coupling constant that regulates the division process and $P_{a \rightarrow bc}$ is the kernel splitting; these are universal functions and are valid in the collinear limit:

$$\begin{aligned} P_{q \rightarrow qg} &= \frac{4}{3} \frac{1+z^2}{1-z}, \\ P_{g \rightarrow gg} &= 3 \frac{(1-z)(1-z)^2}{z(1-z)}, \\ P_{g \rightarrow q\bar{q}} &= \frac{n_f}{2} (z^2 + (1-z)^2), \end{aligned} \quad (3.6)$$

where n_f is the quarks flavour number *quark*. However the probability evaluated is larger than the unity indeed it suffers from the same divergences of the matrix element at the LO. The expression 3.5 is evaluated in the collinear approximation. In particular, there are two types of divergences: collinear, due to the dependency of type $1/Q^2$, and soft which corresponds to the limit $z = 1$.

To solve this problem, in the parton shower approach, the probability of dividing t and $t + dt$ is evaluated; this is obtained with the integration of Eq 3.5 over z in the intervals $[z_{min}(t), z_{max}(t)]$:

$$d\mathcal{P}_{a \rightarrow bc} = \left(\sum_{bc} \int_{z_{min}(t')}^{z_{max}(t')} \frac{\alpha_{abc}}{2\pi} P_{a \rightarrow bc} dt dz \right) dt. \quad (3.7)$$

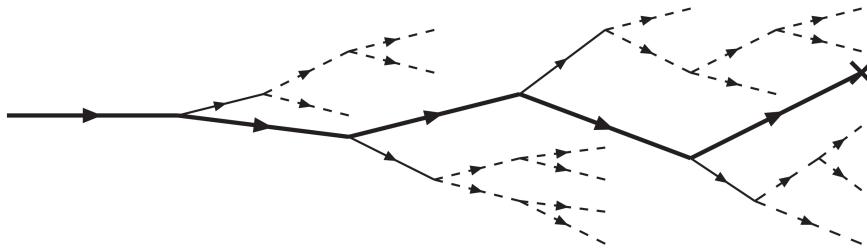


Figure 3.4. Evolution of the initial state. The bold line corresponds to the part that will undergo the hard process (represented by a cross). Thin lines represent the partons that can not recombine, while the dashed lines are fluctuations that may or may not recombine.

As in other physical situations³ the probability of something happening at t is given by the probability that this happens between t and $t + dt$, multiplied by the probability that this has not already occurred between the initial instant t_0 and t . In this case then the division probability at t is:

$$d\mathcal{P}_a^{\text{FSR}}(t) = d\mathcal{P}_a \cdot \exp \left(- \sum_{bc} \int_{t_0}^t dt' \int_{z_{\min}(t')}^{z_{\max}(t')} \frac{\alpha_{abc}}{2\pi} P_{a \rightarrow bc}(z) dz \right), \quad (3.8)$$

where t_0 is the shower starting scale. The exponential term is called Sudakov factor and it represents the probability of non-division. If you want to interpret it in terms of Feynman diagrams, this represents the virtual corrections of LO matrix element. This total process can be combined to have more emissions at different steps: this will result in a partons shower which will be ordered in decreasing Q . Finally, the description given by parton shower is correct if you have collinear jet and it fail in configurations where there are well separated partons.

Initial State Radiation

The initial state radiation evolution is much more complicated than the final state. Indeed the quark and the gluons are emitted and absorbed continuously, inside the incoming proton. The initial state radiation is already present during the hard scattering. The ISR simulation could start from the on-shell parton before the interaction and after could evolve to higher and higher Q^2 scales until the hard process. However, this approach is very inefficient because the interest process is particularly rare and because it has the same probability to happen as in nature. In the event generators a different approach is used: first the hard process is produced and then it we try to rebuild back what may have happened. This procedure is called backward evolution, Fig. 3.4.

It is necessary to evaluate the probability for the process of type $a \rightarrow bc$, that a parton b has been produced by the parton a . For this reason the partonic density function is introduced. This evolves according to the DGLAP [?] equation,

³For example radioactive decay.

$$\frac{df_b(x, t)}{dt} = \sum_{ac} \int_x^1 \frac{dx'}{x'} f_a(x', t) \frac{\alpha_{abc}}{2\pi} P_{a \rightarrow bc} \left(\frac{x}{x'} \right), \quad (3.9)$$

where $f_{a,b}(x, t)$ are the parton PDFs a, b , that has x fraction of the incident and scale proton momenta $t = \ln(Q^2/\Lambda^2)$, instead $P_{a \rightarrow bc}$ is the kernel splitting function. In the backward evolution the probability that the parton b has been generated from a in the interval between t and $t - dt$ is given by:

$$d\mathcal{P}_b(t) = \frac{df_b(x, t)}{f_b(x, t)} = |dt| \sum_{ac} \int \frac{dx'}{x'} \frac{df_a(x', t)}{f_b(x', t)} \frac{\alpha_{abc}}{2\pi} P_{a \rightarrow bc} \left(\frac{x}{x'} \right), \quad (3.10)$$

while the probability of non-division between the scale t_{max} and $t < t_{max}$ is:

$$S_b(x, t, t_{max}) = \exp \left(- \int_t^{t_{max}} dt' \sum_{ac} \int \frac{dx'}{x'} \frac{df_a(x', t')}{f_b(x', t')} \frac{\alpha_{abc}}{2\pi} P_{a \rightarrow bc} \left(\frac{x}{x'} \right) \right), \quad (3.11)$$

Finally then the probability of combining b in a is given in the range between t and $(t - dt)$ from:

$$\begin{aligned} d\mathcal{P}_b^{\text{ISR}}(t) &= - \frac{dS_b(x, t, t_{max})}{dt} dt \\ &= \sum_{ac} \int \frac{dx'}{x'} \frac{df_a(x', t)}{f_b(x', t)} \frac{\alpha_{abc}}{2\pi} P_{a \rightarrow bc} \left(\frac{x}{x'} \right) \cdot S_b(x, t, t_{max}) dt \end{aligned} \quad (3.12)$$

In this case the form factor Sudakov is different respect to FSR as it contains the PDFs. This means that the parton shower results do not depend only on the algorithm but also on the PDFs used.

Resummation

When calculating an observable of the QCD in a perturbative way, the expansion in terms of α_S contains terms of the type $\alpha_S^n L^k$ ($k < 2n$), where $L = \ln(q_{cut}/s)$, being q_{cut} the cut on resolvable emission. When we consider small values of q_{cut} the logarithm of the perturbative expansion becomes large and the perturbative series diverges. The main perturbative order of the expansion is n only if the successive terms of the series are negligible, however this is not guaranteed if there are high value of L . It is therefore necessary to consider the terms that have a high value of the logarithm. The study of these terms is called resummation and is done by putting the terms together in the perturbation series according to their degree of divergence: $\alpha_S^n L^{2n}$ are the leading log (LL) terms, $\alpha_S^n L^{2n-1}$ are the next-to-leading log (NLL) terms, and so on. At the end all α_S orders terms are added. For many processes calculations are available at the NLL. The parton shower reproduces the effects of resuming approximately at the NLL.

Merging among ME and PS

The two different approaches for the matrix element calculation and for the parton shower have advantages and disadvantages. Regarding the ME we have:

- the LO matrix element calculations can be performed exactly in the cases where there are many jet (of the order of six) in the final state,
- a good description of separate partons is performed,
- the perturbative calculations are correct,
- however, the cross section diverges in the collinear and soft case, so an exhaustive description of the internal structure of jet is not possible.

On the other hand the PS:

- it is a universal approach that produces a realistic configuration of the partons,
- the divergences, in the collinear limit, are treated with the introduction of the form factor of Sudakov. So we have an appropriate description of the jet evolution
- however, the method fails when describing separate partons, since the collinear approximation in this case can not be valid.

Clearly the two methods are complementary and their merging is desirable. There are different approaches that combine ME with PS. The main difficulty is to cover the total phase space without overlaps or holes: we want to describe a process in which there are n well separated partons in the final state, using the LO matrix element LO but also including the large logarithms resummation (LL, NLL) which is typical of the PS. A schematic description of the combination for four jet is given in Fig. 3.5. On the horizontal axis are the α_S coupling orders, while on the vertical axis the logarithm. The PS describes the LL ($m = 2n$) and the NLL ($m = 2n - 1$) the green balls (e.g. in the case of $n = 2$, $m = 4, 3$ the two balls in green and marked as “4”). The balls that describe the event with four jet, combining the ME with the PS, are green, the blue and three red marked with the “ 4 ”. The difficulty arises because the ME describes exactly all the balls marked with the “ 4 ”: so if we simply sum up the two approaches we would have double counts of the green spheres called “ 4 ”. The main approaches to merge the ME and PS are:

- parton shower reweight: the basic idea is to start from the process to the lowest order and then re-evaluate the output of the PS as if it had been produced by the ME. This approach does not change the cross section, which remains at the lowest order, but improves the population of the phase space [?, ?].
- CKKW prescription: the phase space is divided into two zones using k_{perp} which is a measure of the cut Q_0^2 : the region in which the jet is produced is filled with the ME, that of evolution with the PS [?, ?].
- The MLM prescription, which is also very widespread, is based on the same principle, but is implemented in a different way.

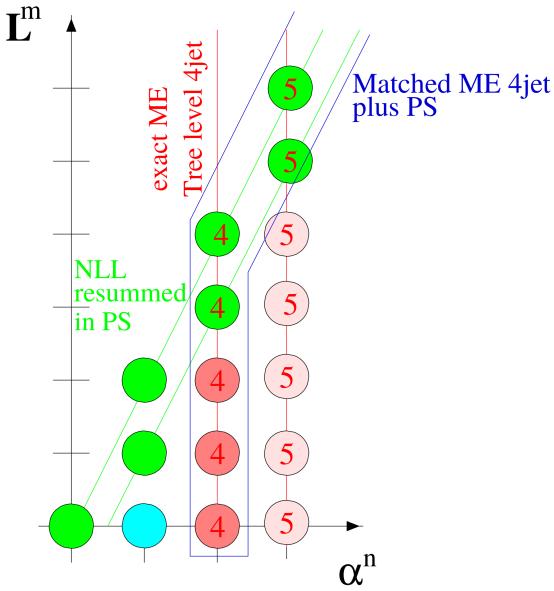


Figure 3.5. Merging among ME and PS.

3.3 Multiple Interaction

Incident protons participating in the interaction are composed by large number of partons (quark and gluons) that can interact independently with each other in addition to the hard process. The total cross section for the QDC process $2 \rightarrow 2$ is dominated by the t process, so the cross section diverges as dp_{perp}^2/p_{perp}^4 for $p_\perp \rightarrow 0$ [?]. So when simulating a real event, in addition to the hard event, characterized by having large transverse transverse momentum, we must also take into account the additional collisions at small p_\perp . If these occur independently then a Poisson distribution is expected, $P_n = \langle n \rangle^n \exp(-\langle n \rangle)/n!$. However, conservation of energy and momenta means that interactions are not effectively independent, thus suppressing the possibility, for $p_{perp} \rightarrow 0$, of having a high number of interactions. It should also be noted that in order to eliminate the divergence it is necessary to introduce a cut-off value of the transverse pulse, below which no collisions are generated.

3.4 Hadronization

In this context, the process of hadronization is a particular model, used in event generators, which describes the transition from the final partonic state to the final hadron state, which is an observable experimental. It is important to underline that this transition is treated in a phenomenological way and not by a rigorous approach. The two most important classes for tuning are the string model and the cluster model. The difference is that the former transforms the partonic systems directly into hadrons, while the second takes an intermediate step where it groups the objects to a scale of ~ 1 GeV.

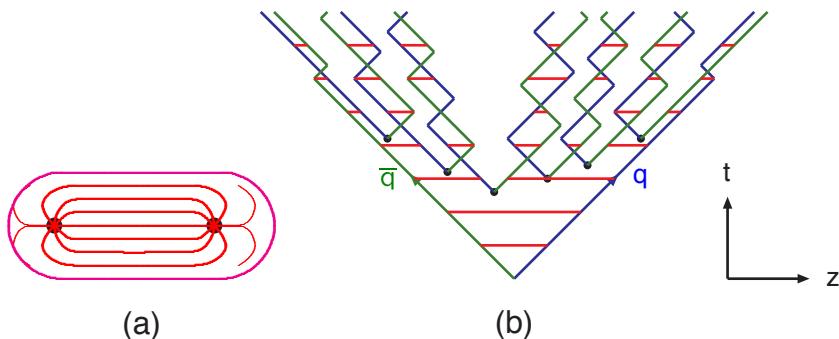


Figure 3.6. (a) The flow tube between a quark and an antiquark moving away. (b) Motion and breaking of a system string.

String Model

The Lund model is the most complete string model: we know from QCD that there is a linear confinement force between the partons that increases with distance. Consider, as an example, the final state in which there are two quarks, $q\bar{q}$. As the partons move away the color flow tube is stretched between q and \bar{q} , Fig. ?? (a). The transverse dimensions of the tube are the typical dimension for the hadrons, therefore about 1 fm. If the tube is assumed to be uniform, the potential increases linearly, $V(r) = \kappa r$, with $\kappa \approx 1$ GeV /fm, string constant. At short distances it would be necessary to introduce an additional Coulomb term, $\sim \frac{\alpha_s}{r}$, however in the Lund model this term is negligible. As the quark and antiquark move away from the interaction vertex, the potential energy accumulated in the string increases until it breaks, giving rise to a pair $q'\bar{q}'$. So the system is divided into two new color singlets $q\bar{q}'$ and $q'\bar{q}$. These two systems will move away repeating the process below. The evolution of the system in space-time is represented in ?? (b). At the end of the process a series of $q_i\bar{q}_i$ pairs are presented, each of which will form a hadron. For now, only the case $q\bar{q}$ has been considered. However, if more partons come from the interaction, the string model becomes more complicated. For an event in which there is an additional gluon, $q\bar{q}g$, the string is stretched between q and g and between g and \bar{q} , Fig. 3.7.

Cluster Model

This hadronization model is based on the pre-confining property of the parton shower: the invariant mass of a single pair of opposite-colored partons is the same at any Q^2 scale. This distribution has its maximum at a mass that is near to the cutoff of parton shower and decreases rapidly to zero, Fig. ?? (a).

In the model, the gluons from parton shower are represented by pairs of color-anticolor lines connected to the vertex. Each color line, near the cutoff, is connected to another colorless line present at the same scale. At this point the contiguous color / anticolor lines are interpreted, in the non-perturbative limit, as quark-antiquark pairs which give rise to mesons, which are observable objects in the final state. This mechanism is represented in ?? (b).

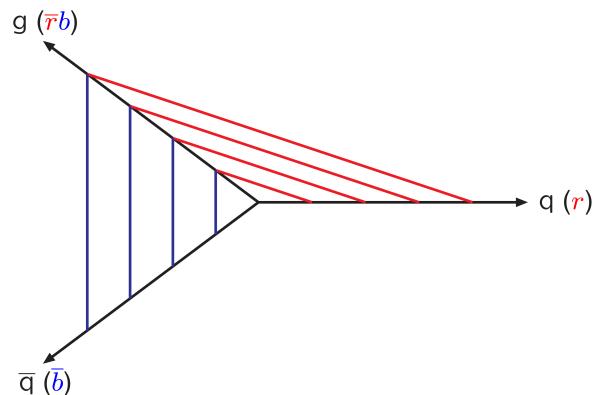


Figure 3.7. Motion of the string in the case $q\bar{q}g$.

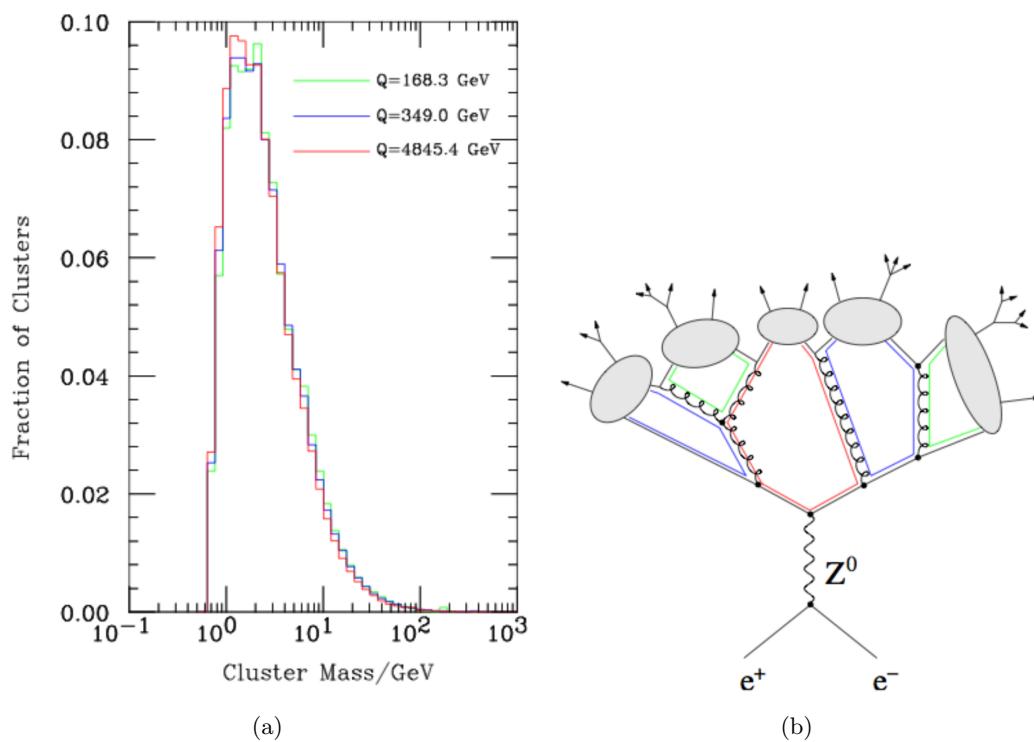


Figure 3.8. (a) Parton shower structure in the cluster model. (b) Invariant mass distribution for singlets.

3.5 Hadronic Decays and Electromagnetic Radiation.

In the hadronization step, unstable hadrons can be produced which decay into other particles. So the final state is the result of the convolution between the hadronization and the decay. The information necessary for the decay of unstable particles is generally taken from the “Particle Data Book” (PDG) [?] which provides the properties (e.g. average life) of a large number of particles. In general, in an event generator, it is necessary to choose which hadrons to include in the simulation and then select the possible decay channels. In addition to hadronic decays, it is also necessary to simulate the emission of electromagnetic radiation. The most common approach adopted is to use algorithms similar to those used to simulate the emission of QCD in parton shower.

3.6 Jets Reconstruction

After the hadronization and the decaying of unstable particles it is possible to estimate the four-momenta of the partons generated in the hard process by the direction and jet energy that are reconstructed starting from the final state particles [?, ?]. The jet reconstruction is made by specific algorithms; these introduce the variable distance, d_{ij} , between two objects (particles or pseudo-jet) defined by,

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2}, \quad (3.13)$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and k_{ti} , y_i and ϕ_i are the transverse momentum, the rapidity and the azimuthal angle of i respectively. R is the radial parameter. The distance between a i and a beam object is also introduced, $d_{iB} = k_{ti}^{2p}$

The algorithms proceed by calculating the minor distance d_{ij} and between all the pairs of particles i, j . For the two particles with a smaller distance, the four-momenta are added. The d_{iB} is evaluated for every i and if it is less than the distance d_{ij} with all other particles j, i , it is considered the jet and it is removed from list of objects present in the event. Finally the distances are recalculated and this whole procedure is repeated until there are no more objects to be added. The value of $p = -1$ defines the algorithm anti- k_t [?], which is the one used, while the free parameter R has been set equal to 0.5.

3.7 Main Monte Carlo generators

In high energy physics, different Monte Carlo generators are available. Each of these has different methods for combining the ME with the PS. A short introduction of MADGRAPH interfaced with PYTHIA, POWHEG and SHERPA is described below.

Madgraph_aMC@NLO

The MADGRAPH [?] approach is very ambitious, in fact the purpose of this generator is to calculate the cross section at the NLO including both real and virtual contributions in the calculation. The hard process is produced by the ME method while

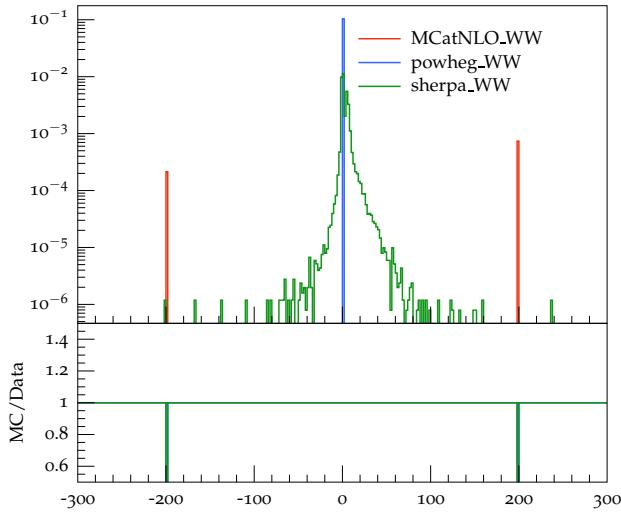


Figure 3.9. Weight distribution, for different Monte Carlo generators, with cross section normalization of 1 fb^{-1} .

the soft emissions are combined with the PS. The first step is to compute the ME NLO corrections for a n parton process, including $n + 1$ partons coming from the real corrections and n from the virtual ones. Next how the parton shower populates the phase space at $n + 1$ parton is evaluated, excluding the Sudakov form factor at this step. To get the true state where $n + 1$ is present, MADGRAPH subtracts the PS expression from the $n + 1$ state of the ME. The expression of the PS without a factor of Sudakov and of the ME are in agreement in the limit soft and collinear, so the singularities are deleted thus obtaining a finite value for the cross section in the case of n and $n + 1$ partons. A technical problem is that in the collinear limit there is no certainty that the ME always overhangs the PS. This problem is solved by introducing a fraction of negative-weight events, Fig. 3.9. Finally, the parton shower is applied, which includes the Sudakov factor and thus allows a finite and correct result to be obtained at the NLL.

POWHEG

The idea behind POWHEG [?] is to generate the hardest radiation first, and then pass the event to the parton shower generator. In parton shower generators, the production, ordered in a transverse pulse, of the harshest radiation is always the first; so POWHEG simply replaces this with issuing to the NLO. In POWHEG events are produced with a positive and constant weight (Fig. 3.9).

PYTHIA8

PYTHIA 8 [?] is a generator that can calculate the ME for processes with two particles or partons in the final state, but above all it generates the parton shower and the subsequent synchronization. The parton shower is ordered in a transverse impulse,

p_T , and the first issue is corrected with the method of the reshuffle. For the use of the atomization, use the Lund model.

SHERPA

SHERPA [?] is a Monte Carlo generator that as PYTHIA8 provides a complete description of hadronic collisions, from the calculation of the matrix element, up to the angle. The parton shower includes both QCD and QED emissions, i.e. photons. It can calculate the ME for the main processes (eg $gg \rightarrow H$) at the NLO and combine the ME with the PS. The code is written completely in C ++ language.

3.8 Monte Carlo sample in High Mass Analysis

Several Monte Carlo generators were used in the searching of high mass particle to simulated the signal and the backgrounds. All processes are generated using the NNPDF3.0 [30, 31] parton distribution functions (PDF) for NLO generators, while the LO version of the same PDF is used for LO generators. All the event generators are interfaced to PYTHIA 8.1 [32] for the showering of partons and hadronization, as well as including a simulation of the underlying event (UE) and multiple interaction (MPI) based on the CUET8PM1 tune [33]. For all processes, the detector response is simulated using a detailed description of the CMS detector, based on the GEANT4 package [34]. The simulated samples are generated with distributions for the number of pileup interactions that are meant to roughly cover, though not exactly match, the conditions expected for the different data-taking periods. In order to factorize these effects, the number of true pileup interactions from the simulation truth (as stored in the PileupInfo collection in the Monte Carlo) is reweighted to match the data. The re-weighting is propagated automatically to both the in-time pile up and the out-of-time one. The pileup histogram for reweighting is calculated using the *pileupCalc* tool as described in [35]. Different sources and calculations are used to obtain the cross sections for the different processes at 13 TeV. All simulated sample are summarized in Tab. 3.1.

Signal

In order to perform the resonance search in a large part of the mass spectrum, several signal samples for the gluon-gluon fusion and the vector boson fusion mechanisms have been generated corresponding to different Higgs boson masses in the range between 200 GeV and 3 TeV. All signal samples have been simulated with POWHEG v2 [36, 37, 38], designed to describe the full NLO properties of these processes. In particular, for Higgs produced via gluon fusion [39], and vector-boson-fusion (VBF) [40], the decay of the Higgs boson into two W boson and subsequently into leptons was done using JHUGen [41]. The signals which correspond to a Higgs boson mass of 125 GeV have been simulated accordingly and are treated as backgrounds in the following analysis, including the associated production with a vector boson (W^+H , W^-H , ZH) [42], and gluon fusion produced ZH ($ggZH$). For associated production processes the Higgs boson decay was done via PYTHIA 8.1 [32]. For Higgs signals, the cross sections used are the ones reported but the LHC Higgs Cross

Process	Dataset Name	$\sigma \times BR$ [pb]
Signal ggH	GluGluHToWWTo2L2NuM200-3000	Various
Signal VBF	VBFHToWWTo2L2NuM200-3000	Various
$t\bar{t} \rightarrow WW$ $b\bar{b} \rightarrow 2l2\nu b\bar{b}$	TTTo2L2Nu_13TeV-powheg	87.31
$q\bar{q} \rightarrow WW \rightarrow 2l2\nu$	WWTo2L2Nu_13TeV-powheg	12.178
$q\bar{q} \rightarrow WW \rightarrow l\nu qq$	WpWmJJ-QCD-noTop_13TeV-powheg	0.59
$gg \rightarrow WW \rightarrow 2l2\nu$	GluGluWWTo2L2Nu_MCFM_13TeV	0.5905
Single top	ST tW top 5f inclusiveDecays ST tW antitop 5f inclusiveDecays	35.85 35.85
Drell-Yan	DYJetsToTauTau_13TeV-amcatnloFXFX-pythia8_ext1 DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 DYJetsToLL_M-50_HT100to200_madgraphMLM-pythia8 DYJetsToLL_M-50_HT200to400_madgraphMLM-pythia8 DYJetsToLL_M-50_HT400to600_madgraphMLM-pythia8 DYJetsToLL_M-50_HT600toInf_madgraphMLM-pythia8	1867 6025.26 147.4 40.99 5.678 2.198
Multibosons	WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8 ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8 WWZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8 WZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	5.5950 3.2210 0.1651 0.05565

Table 3.1. Simulated samples for $t\bar{t}$ and WW production.

Section Working Group [43], computed at NNLO and NNLL QCD and NLO EW for gluon fusion, and at NNLO QCD and NLO EW for the rest of the production modes. The branching fractions are the ones reported in [15].

The WW sample

The WW production, irreducible background for the analysis, was simulated in different ways. POWHEG v2 [44] was used for $q\bar{q}$ induced WW in different decays. The cross section used for normalizing WW processes produced via $q\bar{q}$ was computed at NNLO [?]. Gluon gluon fusion produced WW was generated, with and without Higgs diagrams, using MCFM v7.0 [45]. The cross section used for normalizing $q\bar{q}$ produced WW processes was computed at next-to-next-to-leading order (NNLO) [?]. The leading-order (LO) cross section for ggWW is obtained directly from MCFM. For gluon fusion, the difference between LO and NLO cross sections is significantly big. A scale factor of 1.4 is theoretically calculated [?] and applied to the gg \rightarrow WW background.

In the analysis two different WW Monte Carlo samples are merged: the “ $WW \rightarrow 2l2\nu$ NLO” and the “WW plus 2 jet” LO. The second sample, “WW plus 2 quark”, contains final state with two quarks or a gluon-quark system: only the final state with two quarks interferes with the signal. To avoid double count between the two samples a cut on di-jet mass at gen-level, mjj_{GenLev} , is applied. In particular the sample “ $WW \rightarrow 2l2\nu$ at NLO” is used for $mjj_{GenLev} < 100$ GeV and the “WW plus 2 quark” for $mjj_{GenLev} > 100$ GeV. The distribution for the reco di-jet mass is shown in Fig.3.10. In particular the red distribution corresponds to the “ $WW \rightarrow 2l2\nu$ NLO” sample with a cut of $mjj_{GenLev} < 100$, the blue distribution to “WW plus 2 quark” with $mjj_{GenLev} > 100$ GeV. The sum of the red and blue distributions is shown in black. There is a good agreement between the black distribution and the “ $WW \rightarrow 2l2\nu$ NLO” without any mjj_{GenLev} distribution, in green.

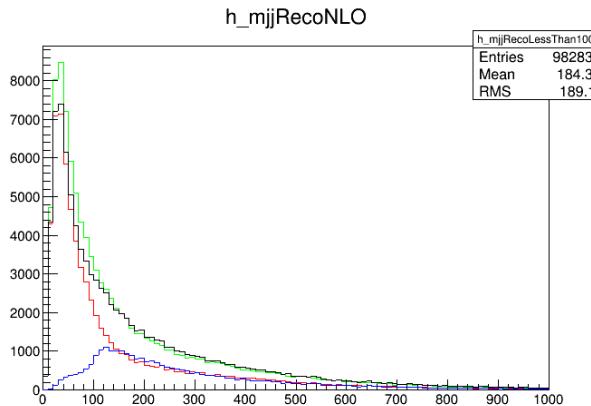


Figure 3.10. Distribution for m_{jj} at RECO level for the merged WW sample.

The Top sample

In order to control the top quark background processes, the analysis is performed in jet bins as described in ???. The jet binning enhances the importance of logarithms of the jet p_T , spoiling the convergence of fixed-order calculations of the $q\bar{q} \rightarrow WW$ process and requiring the use of dedicated resummation techniques for an accurate prediction of differential distributions [46, 47]. Since the p_T of the jets produced in association with the WW system is strongly correlated with its transverse momentum, p_T^{WW} , the simulated $q\bar{q} \rightarrow WW$ events are reweighted to reproduce the p_T^{WW} distribution from the p_T -resummed calculation. A $t\bar{t}$ sample dilepton sample was also generated using POWHEG v2. The cross sections of the different single top processes are estimated by the LHC Top Working group [48] at NLO. The $t\bar{t}$ cross section is also provided by the LHC Top Working group [49], and it is computed at NNLO, with NNLL soft gluon resummation.

The DY sample

For the Drell-Yan backgrounds we use two different sets of samples. For the opposite flavor analysis (Sec 5.7), selecting events with an electron and a muon, a dedicated sample in which only the $Z/\gamma^* \rightarrow \tau\tau \rightarrow e\mu\nu\nu$ decay is simulated. For the same flavor analysis (Sec. 5.8), in which pairs of electrons or muons are selected, a soup of different HT binned DY samples is used. A detailed study about this soup is given below in Sec. 3.8. Drell-Yan production of Z/γ^* is generated using aMADGRAPH [50] and the cross section is scaled using a LO to NNLO k-factor equal to 1.23.

Given the lack of MC statistics in the LO inclusive DY sample the H_T -binned samples are used. This helps increasing the MC statistics especially in the VBF category of the same flavor analysis, which is characterized by large values of H_T . The LO inclusive sample is used for events with $H_T < 100$ GeV and it has been merged to the other samples selecting events with H_T below 100 GeV using the parton level information. The cross sections of those samples have been scaled applying the LO to NNLO k-factor. In Fig. 3.11 the H_T distribution of the sample after the merging is reported, showing a smooth transition between different H_T

samples.

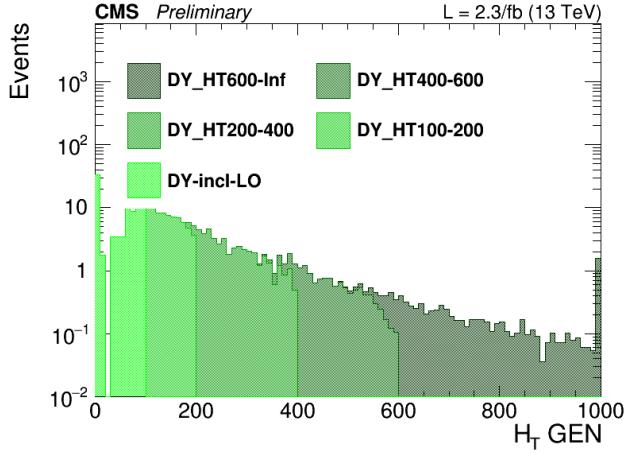


Figure 3.11. H_T distribution for the merged DY sample.

To further check the correct behaviour of the H_T binned samples we compared them to the inclusive LO sample, selecting only the events with a generator level H_T above 100 GeV. The comparison is done in a control region with two same flavor leptons with $p_T > 20$ GeV and $m_{\ell\ell} > 50$ GeV, showing very good agreement between the two samples. The distributions of some variables are shown in Fig. 3.12

To check the differences between the LO inclusive sample and the NLO sample simulated with MC@NLO, the two samples have been compared in a same flavor control region and some variables of interest are shown in Fig. 3.13. The control region is defined requiring two same flavor leptons with $p_T > 20$ GeV and with $m_{\ell\ell} > 50$ GeV.

In Figure 3.14, the effect of this reweighting on a sample enriched in Drell-Yan events is shown. In order to select this sample, events with two electrons with $p_T > 25$ GeV for the leading one and $p_T > 13$ GeV for the trailing one, are selected only if $|m_{\ell\ell} - m_Z| < 10$ GeV.

Other processes

Other multiboson processes, such as WZ, ZZ, and VVV ($V=W/Z$), are generated with amc@NLO and normalized to the cross section obtained at NLO in generation. The cross sections for the remaining processes were directly obtained using the *GenXSecAnalyzer* tool [51] or from the Twiki presented in Ref. [52].

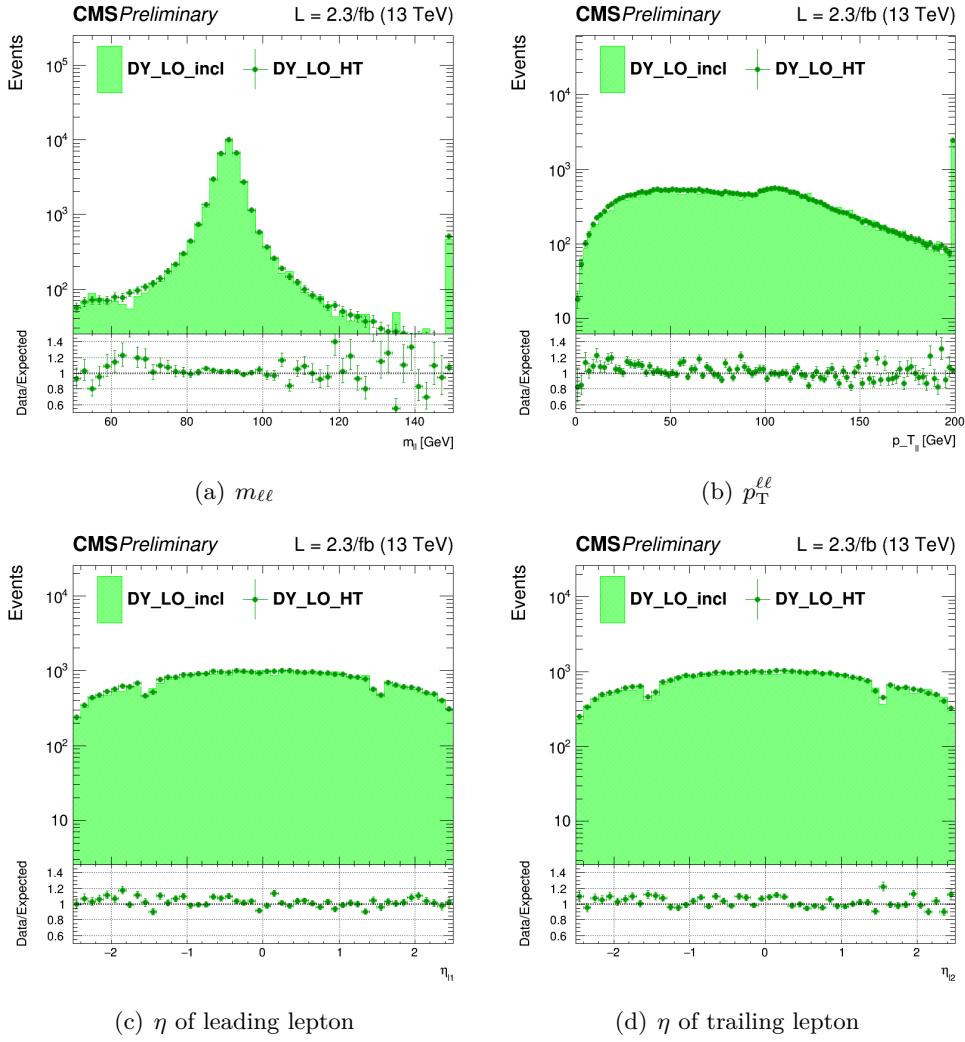


Figure 3.12. Comparison between the inclusive LO DY sample and the H_T binned samples.

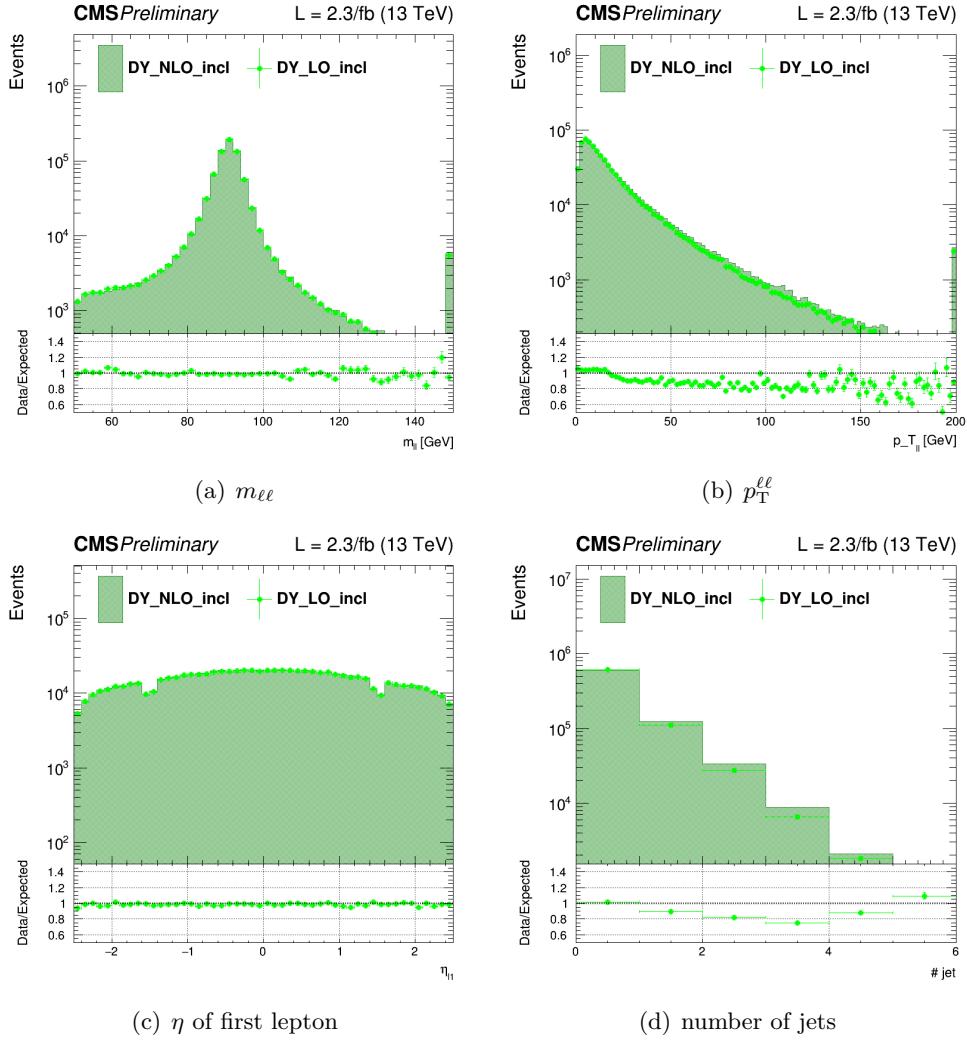


Figure 3.13. Comparison between the LO and NLO DY samples.

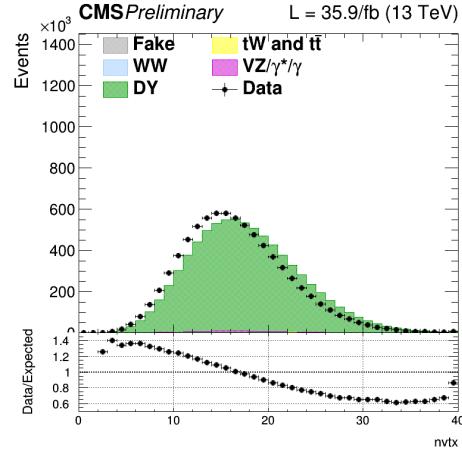


Figure 3.14. Distributions of the number of vertices in a Drell-Yan enriched sample ($Z \rightarrow ee$) in data

Chapter 4

Event Reconstruction

In the search of high mass in W^+W^- channel, leptons (specifically, electrons or muons and neutrinos) and jets are searched for to select signal events. The reconstruction of such objects follow standard algorithms and procedure which have been developed by the experiment. In this chapter the reconstruction of the interested objects are described.

4.1 The Particle Flow

The algorithm for the events reconstruction is called Particle Flow (PF) [53]. Its goal is to reconstruct all stable particles that emerge from a hadron-hadron collision, such as muons, electrons, photons, charged hadrons and neutral hadrons. The PF technique uses the information coming from the CMS sub-detectors and combines them to obtain the type of particle, their direction, the momentum and the energy. The reconstruction of charged particles is allowed thanks to the presence of the silicon tracker that is immersed in the uniform magnetic field of 3.8 T. Thanks to that, a transverse momentum precise measurement could be done down to about 150 MeV. The photon and electron reconstruction is performed by PF technique using the excellent tracking system together with the high granularity of the ECAL, this combination gives a high energy resolution for such particles. The identification of charged-particles, energy clusters in the calorimeters and muon tracks are the first step of the PF. All these information coming from sub-detectors and the information are then connected to each other making blocks of elements which are topologically compatible. Starting by blocks, the candidates particle (PF Candidates) are fully reconstructed and identified as:

- Muons: the combination of a track in the tracker and a track in the muon system rise to a PF muon. The corresponding track is removed from the block after the identification.
- Electrons: a charged-particle track is linked to one or more ECAL clusters (of presents)
- Charged hadrons: PF charged hadrons give rise from the remaining tracks. Tracks can be linked to ECAL and HCAL clusters, and the energy is determined taking into account information from calorimeters.

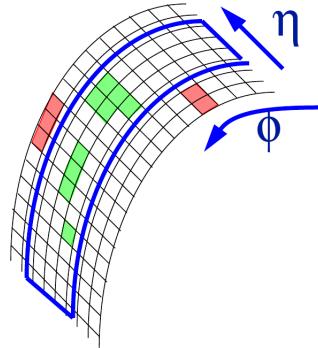


Figure 4.1. Cluster spread in η and ϕ .

- Photons and Neutral hadrons: a PF photons are given by a ECAL clusters not compatible with charged-tracks, instead PF neutral hadrons are given by unaccounted HCAL deposits.

After that, once the list of PF Candidates is defined, the PF jets are clustered. At the end the PF MET is evaluated as the opposite of transverse momentum-vector sum over all reconstructed PF Candidates.

4.2 Lepton reconstruction and identification

Electron reconstruction and identification

The combination of tracker and ECAL information is used for electrons reconstruction. The electrons, emerging from a collision, when interact with the silicon tracker, radiate bremsstrahlung photons that reaches the ECAL with a significant spread in the azimuthal direction ϕ . These energy deposited measured in ECAL are the starting point of the electrons reconstruction algorithm. The energy deposited are associated clusters and in superclusters (clusters of clusters) that take into account the spread in ϕ direction of the bremsstrahlung energy, Fig. 4.1. When superclusters are identified, the reconstruction algorithm tries to match them to track seeds. The track seeds are identified as pairs or triplets of hits in the inner tracker layers. The electrons trajectories are reconstructed using a dedicated modeling that take in account energy loss in the tracker layers via bremsstrahlung radiation. In general, non-Gaussian contributions to the event-by-event fluctuations of the calorimetry and tracking measurements, are introduced due to the bremsstrahlung radiation. A preselection is applied to solve ambiguous cases where several tracks are reconstructed. This is based on matching between the GSF track and the supercluster in η and ϕ . To perform a good resolution electron supercluster must be inside the ECAL acceptance volume, that meaning $\eta < 2.5$ and outside the ECAL barrel-endcap overlap region, $1.4442 < \eta < 1.566$. Several identification variables are used to achieve a good discrimination. These variable are:

- $\Delta\eta_{trk,SC}$ and $\Delta\phi_{trk,SC}$, that measures the spacial matching between the track and the supercluster.

- $\sigma_{in,in}$ that measures the width of the ECAL supercluster in the η direction. It is calorimeter shower shape.
- H/E: is the ratio among the energy deposit in the HCAL tower and the energy of the seed supercluster.
- $|1/E - 1/p|$ the difference of $1/E$ measured in ECAL and the $1/p$ measured in the tracker.
- Number of missing hit
- d_{xy} and d_z , impact parameters respect to the reconstructed primary vertex (PV).
- $\gamma \rightarrow e^+e^-$ veto based on missing hits in the inner layers of the tracker.

Different selections on the previously variables defined different working points. The cuts are different for electrons in the ECAL barrel or in the endcap. These are defined below:

- Tight WP: this corresponds to an average 70% selection efficiency for electrons with $p_T > 20$ GeV. This working point is used where backgrounds are very large. The $X \rightarrow WW$ high mass analysis has large backgrounds like W+jets, where the second lepton is fake. So, in this analysis we use Tight WP for electrons. On the top of that, we apply more cuts to make the electron id selection Trigger safe.
- Medium WP: the average efficiency is about 80% for electrons with $p_T > 20$ GeV. This is also a good starting point for measurements of W and Z cross-sections.
- Loose WP: this working point is used only for very clean final states. The average efficiency is about 90%.
- Veto WP: generally, this is not used for signal selection. However, it is found to be useful for extra lepton veto counting of electrons. The average efficiency is about 95%.

The electrons selected, are also required to pass the isolation criteria that including a pile-up mitigation correction based on the electron effective catchment area. The isolation variable is computed for each electron as,

$$ISO_{EA\ corrected}^{Rel} = [\sum_{ChH} (p_T) + max(0, \sum_{Ph} (p_T)) + \sum_{NH} (p_T) - \rho EA]/p_T^{electron}, \quad (4.1)$$

where ChH is the charged hadrons, Ph is photons, NH is neutral hadrons, ρ is the pile up energy density and A is an effective area. The sums are in a isolated cone of $\Delta R < 0.4$ around the electron direction. The identification and isolation criteria used for the $X \rightarrow WW$ analysis are reported in Tab. 4.1.

Observable	Barrel (EB) cut	Endcap (EE) cut
$ \Delta\eta_{trk,SC} $	0.00308	0.00605
$ \Delta\phi_{trk,SC} $	0.0816	0.0394
$\sigma_{in,in}$	0.011	0.031
H/E	0.060	0.065
$ 1/E - 1/p $	0.013	0.013
Number of missing	1	1
$ d_{xy} $	0.05	1
$ d_z $	1	1
conversion veto	true	true
$ISO_{EA\ corrected}^{Rel}$	0.0588	0.0571

Table 4.1. Electron identification for Tight working point and isolation requirements.

Muon reconstruction and identification

Muons are produced in the interaction point and they can pass all the detector with a negligible energy loss, given a signal in the muon chamber. Their loss of energy is very negligible. Muons are thus reconstructed both in the silicon tracker and in the external muon chambers. The reconstruction starts from the measurements of DT, CSC and RPC sub-detectors. The reconstructed track in the muon spectrometer is called Stand-alone Muon. In the inner silicon tracker the muon track are also reconstructed. Seeds are built using two or three consecutive hits in the pixel and or in the strip detector. The pattern recognition is performed starting from these seeds and proceeding layer by layer, with an iterative technique based on the Kalman Filter technique. At the end of this algorithm, a track fit is performed and the track parameters are updated. The identified tracker track is then combined with a given Stand-alone Muon track in order to construct a global track, which defines a Global Muon. A global fit is performed for each pair of tracks reconstructed in the inner tracker and in the muon system. If more than one track matching the stand-alone track is found, then the one giving the best χ^2 in the global fit is chosen. Exist a second approach that consists in considering all tracker tracks with $p_T > 0.5$ GeV as potential muon candidates. These tracks are extrapolated to the muon system taking into account the magnetic field. If at least one muon segment (a short track stub made of DT or CSC hits) matches the extrapolated tracks, the corresponding tracker track is identified as a Tracker Muon. Quality requirements are applied to ensure a good quality of the reconstruction: the tracker track has to be reconstructed from at least 5 tracker layers with hits;

- at least one hit must be present in the pixel detector;
- least one hit must be present in the muon detector;
- least one muon chamber hit should be included in the Global Muon track fit;
- normalized χ^2 of the Global Muon track fit should be less than 10;
- muon track reconstructed in the tracker must have a distance to the primary vertex (PV) smaller than 2 mm in the transverse plane and smaller than 5

Observable	Value Cut
Is global muon	true
Is PF muon	true
Tracker layers with measurements	> 5
Number of valid pixel hits	> 0
Number of valid muon hits	> 0
Number of matched muon stations	> 1
χ^2 / ndof	< 10
d_{xy} (PV)	< 0.2 cm
d_z (PV)	< 0.5 cm

Table 4.2. Muon identification and isolation requirements..

mm in the longitudinal direction.

The requirements are summarized in Tab.4.2

For the analysis goals, the muons are expected to be isolated, as those produced by W boson decay. Indeed, prompt muons are expected to be isolated in the event, contrary to non-prompt muons that are generally produced within jets and characterized by many nearby particles. Muons coming from W , and so from X , must be isolated and they are requested to pass an isolation criterion which also includes a PU mitigation correction called “ $\Delta\beta$ correction”. The isolation variable is therefore sensitive to the pileup and a correction is needed in order to ensure its independence and robustness on the number of simultaneous interactions. The isolation variable used is,

$$ISO_{\Delta\beta}^{Rel} = [\sum_{ChH} (p_T) + \max(0, \sum_{Ph} (p_T) - 0.5 \times \sum_{ChHUP} (p_T))] / p_T^{electron}, \quad (4.2)$$

where ChH is the charged hadrons, Ph is photons and $ChHUP$ is charged hadrons not coming from the primary vertex. The sum is performed in a cone of 0.4 units in ΔR around the muon. A bias in the muon p_T is introduced by the imperfect knowledge of the magnetic field and the effect of the material distribution. In addition the p_T measurement is sensitive to the alignment of the tracker and muon chambers. To estimate the muon p_T scale and resolution are used different methods. For $p_T < 100$ GeV, the events with two muons arising from the J/Ψ and Z resonance decays are used to correct the p_T scale. Above, so in the high p_T regime, are used cosmic ray muons.

4.3 Jet reconstruction and identification

Jets are the experimental signature of quarks and gluons produced in the hadron collision. They result from the parton hadronization and they play a major role in a hadronic collider where final state with jets have very large cross-section. The technique used for jet reconstruction is the Particle Flow. Using the particle-flow algorithm described in Sec. 4.1, the jets are reconstructed by clustering of the four-momentum vectors of candidates: the information from relevant CMS sub-detectors are used to identify and to reconstruct all visible particles in the event, as muons,

electrons, photons, charged hadrons, and neutral hadrons. The momentum and spatial resolutions of jet identified with PF technique, are greatly improved respect to jet identified with only the calorimeter information. Indeed the use of the tracking detectors allows a better resolution of p_T for the charged particles. The sequential and iterative jet clustering algorithm, that combine four-vectors of input particles, are used to form the jets. A ideal jet clustering algorithm should have the following features:

- infrared safety: infrared singularity should not appear in the perturbative calculation;
- collinear safety: collinear singularities must not appear in the perturbative calculations;
- invariance under boosts: the algorithm should find the same solutions independently by boosts in the longitudinal direction.
- order independence: the algorithm should find the same jets at parton, particle, and detector level;
- straightforward implementation: the algorithm should be straightforward to implement in perturbative calculations.

The jet algorithm should also follow these experimental criteria

- detector independence: the performance of the algorithm should be as independent from the detector;
- minimization of resolution smearing and angle biases: the algorithm should not amplify the inevitable effects of resolution smearing and angle biases;
- stability with luminosity: jet finding should not be strongly affected by pileup;
- efficient use of computing resources: minimum of computer time consumption;
- maximal reconstruction efficiency: the algorithm should efficiently identify all jets;
- ease of calibration and use: algorithm should not obstruct to the reliable calibration and to be straightforward to implement.

Jet definition is not unique, being the parton not a well-defined object, so several approaches for jet clustering are available. Two main algorithms for clustering have been developed. The “conical recombination” where jets are defined as dominant directions of energy flow. The second class is called sequential recombination and it works by defining a distance between pairs of particles, performing subsequent recombinations of the pair of closest particles and stopping when all resulting objects are too far apart. The standard algorithms adopted by CMS are the SISCone in the conical recombination class, and the k_t , Cambridge-Aachen (CA) and anti- k_T algorithms, Sec 3.6.

The jet energy calibration is related, on average, on the energy measured in the detector to the true energy of the corresponding final state particle jet or parton jet. From the clustering of all stable particle, a true particle jet results. The correction is applied as a multiplicative factor C to each component of the raw jet four-momentum vector, p_μ^{raw} , as:

$$p_\mu^{corr} = C \cdot p_\mu^{raw} , \quad (4.3)$$

where correction factor C is composed of the offset correction C_{offset} , the MC calibration factor C_{MC} , and the residual calibrations C_{rel} and C_{abs} for the relative and absolute energy scales, respectively. The various components are applied in sequence:

$$C = C_{offset}(p_\mu^{raw}) \cdot C_{MC}(p'_T, \eta) \cdot C_{rel}(\eta) \cdot C_{abs}(p''_T) , \quad (4.4)$$

where p'_T is the transverse momentum of the jet after applying the offset correction and p''_T is the transverse momentum of the jet after all previous corrections. The details of each component are:

- C_{offset} : an energy offset is produced by the pileup of multiple proton-proton collisions and by electronic noise. The goal of the offset correction is to subtract, on average, the unwanted energy from the jet. For the offset correction estimation, the Jet Area Method has been used. An average p_T -density ρ per unit area is estimated, for each event. The key element for this approach is the jet area A_j . A very large number of infinitely soft four-momentum vectors are artificially added in the event and clustered by the jet algorithm together with the true jet components. The other important quantity for the pileup subtraction is the p_T density ρ , which is calculated with the k_T jet clustering algorithm with a distance parameter $R=0.6$. The quantity ρ is defined on an event-by-event basis as the median of the distribution of the variable p_{Tj}/A_j , where j runs over all jets in the event, and is not sensitive to the presence of hard jets. Therefore, the event-by-event and jet-by-jet offset correction can be defined as,

$$C_{offset} = (p_T^{raw}, A_j, \rho) = 1 - \frac{(\rho - \langle \rho_{UE} \rangle) \cdot A_j}{p_T^{raw}} , \quad (4.5)$$

where $\langle \rho_{UE} \rangle$ is the p_T -density component due to the UE and electronics noise, and is measured in events with exactly one reconstructed primary vertex (no pileup).

- C_{MC} : is based on the simulation and corrects the energy of the reconstructed jets such that it is equal on average to the energy of the generated jets (GenJets). The GenJets reconstruction algorithm is identical to the one applied to the data. The response variable, $R = p_T^{reco}/p_T^{gen}$, in each bin of the GenJet transverse momentum $/p_T^{gen}$, is recorded as jet p_T^{reco} . The average correction in each bin is defined as,

$$C_{MC}(p_T^{reco}) = \frac{1}{\langle R \rangle} , \quad (4.6)$$

- C_{rel} : the goal of the relative jet energy scale correction is to make the jet response flat versus η . This is achieved by employing a Tag and Probe technique, selecting dijet events in data. The size of this residual correction is of the order of 2-3% in the central η region, while it goes up to about 10% in the forward region.
- C_{abs} : the goal of the absolute jet energy scale correction is to make the jet response flat versus p_T . Once a jet has been corrected for η dependence, it is corrected back to particle level.

Each type of correction has uncertainties arising from many different sources. These sources are categorized as: physics modeling in MC, MC modeling of true detector response and potential biases in the methodologies used to estimate the corrections. In CMS more than 16 such sources of uncertainties have been identified. Several are related and can be combined into groups that are relative to the absolute scale, relative scale, extrapolation in p_T , pileup, jet flavor and time stability.

Jets in $X \rightarrow WW$

The jets collection is made by clustering the reconstructed particles known as the Particle Flow Candidates in the CMS software. The reclustering algorithm used is the anti- kt algorithm with a distance parameter equal to 0.4. Pileup mitigation is done with the Charge Hadron Subtraction algorithm which removes charged particles coming from pileup vertices before clustering. The Jet Energy Corrections that are applied to correct the reconstructed jet energy back to the true energy of the final state particles. To reject jets originating from calorimeter or readout electronics noise the loose working point of the PF Jet Identification is applied. This analysis selects a jet by the selection cut as,

$$p_T^{jet} > 30\text{GeV}, |\eta| < 5.0. \quad (4.7)$$

The various distributions of jet after correction are shown at Fig. 4.2

4.4 b-jet identification

To identify the b jets the properties that characterize hadrons containing the quark b. In fact these hadrons have a relatively large mass, around 5 GeV, and a long average life, about 1.5 ps. Given that they can have an impulse of several tens of GeV, the distance that they travel in the detector it is of the order of:

$$\begin{aligned} \delta &\approx t_{medio} v \approx \gamma \tau \beta c \\ &\approx 10 \cdot 1.5 \cdot 10^{-12} \text{s} \cdot 1 \cdot 3 \cdot 10^8 \text{ m/s} \\ &\approx 5\text{mm}. \end{aligned} \quad (4.8)$$

A distance that can be measured since the accuracy of the position of the secondary vertices is $\sim 100 \mu\text{m}$. The CMS collaboration has developed different algorithms to identify the jets coming from quarks b (b-tag algorithms). Each of these algorithms is characterized from a certain efficiency of signal identification (jets from b quark)

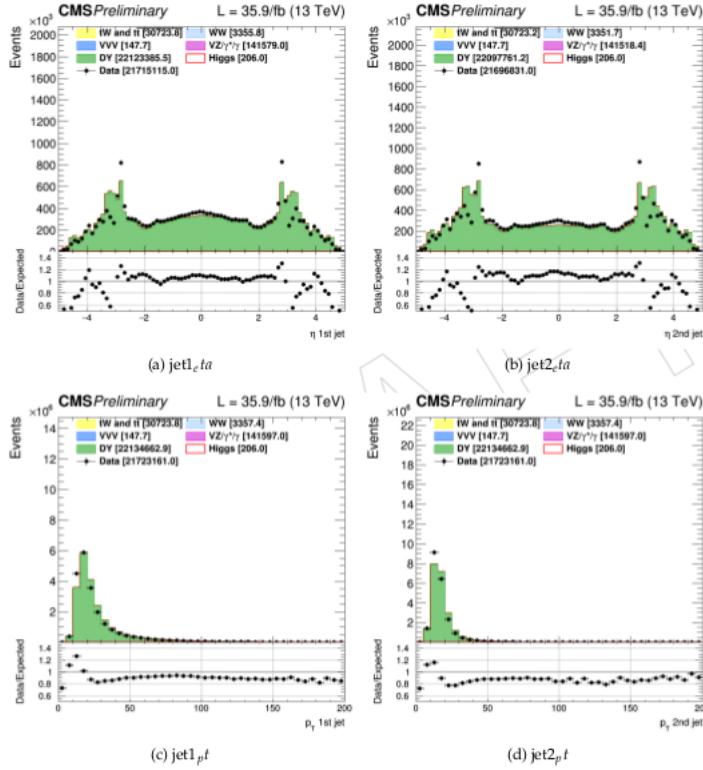


Figure 4.2. The jet kinematic distributions:.

and from the probability of reject the background (jet not from b quark), which in general depend on the transverse impulse, p_T , and from the pseudorapidity of the jet. To build an observable (or discriminator) which distinguishes the jets originated by b from those originating from light quark is used to reconstructed objects such as, traces, vertices and leptons. Some algorithms simple but reliable use a single observable in input, while others more complexes combine different ones to generate the discriminator. Each algorithm produces, in output, a single value of the discriminator for each jet of the event. The first step in identifying of b-jets is the reconstruction of all the jets present in the the event. As mentioned, the CMS collaboration uses for jet reconstruction the anti- k_T algorithm, applied to Particle Flow objects. However the algorithms of identification require a sample with well reconstructed and high purity traces. Therefore further cuts to the traces of the jet are applied. First of all, to reduce the number of traces reconstructed incorrectly, a transverse pulse is required greater than 1 GeV and at least two signals in the silicon pixel detector. Then, at least eight signals per track must be present with a fit that has a $\chi^2/\text{d.o.f.} < 5$, where d.o.f. it is the number of degrees of freedom of the fit. It also applies also a selection on the impact parameters: this is used to increase the fraction of well reconstructed traces and to reduce the contamination due to long life particles like neutral kaons. The transverse distance d_{xy} and longitudinal d_z between the trace and the primary vertex are required to be smaller than 0.2 cm and 17 cm respectively. In Fig. 4.3 it is was a schematic representation of the parameters used in the identification of the b jets.

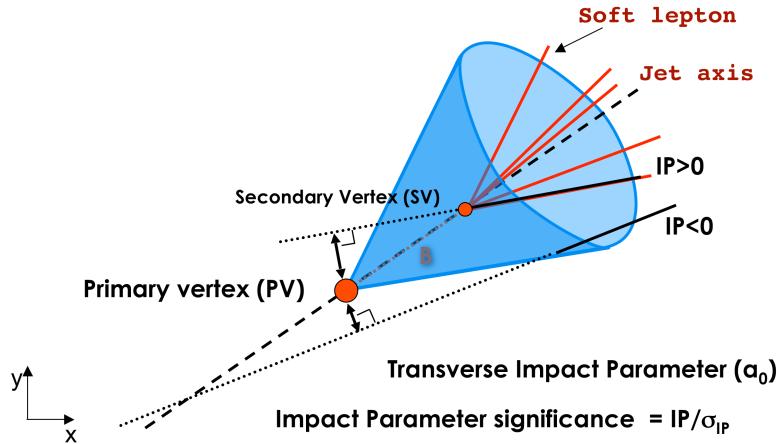


Figure 4.3. Cluster spread in η and ϕ .

Impact Parameter algorithm

The impact parameter (IP) of a track respect to the primary vertex, can be used to distinguish the hadrons b decay respect to background tracks (prompt). The IP is calculated in three dimensions thanks to the excellent resolution of the pixel detector. The impact parameter has the same sign of the product scalar between the direction of the jet and the distance of the primary vertex from the nearest point of the trace. The traces produced by particles due to the decay of a hadron with a long average life they have a positive value of IP, while for the traces originated from the primary vertex (prompt) the IP will be both negative and positive. The variable used as observable is the significance of the parameter of pact, S_{IP} , defined as the relationship between IP and its estimated uncertainty. There significance of the impact parameter can therefore be used as a discriminant between b and non-b jets. The Track Counting (TC) algorithm orders in significance IP traces of jets, from the highest to lowest. There are two different versions of this algorithm that they are distinguished on the basis of which significance value is taken into consideration: Track Counting High Efficiency (TCHE) uses the S_{IP} of the second track , while the Track Counting High Purity (TCHP) use the third. A natural extension of the TC algorithms is the combination of the several tracks impact points. This is implemented in the Jet Probability (JP) and Jet algorithms B Probability (JBP). The former uses an estimator for the likelihood he evaluates which may be the probability that all traces associated with a jet come from from the vertex of primary interaction. The estimator of likelihood, P_{jet} , is defined such as,

$$P_{jet} = \Pi \cdot \sum_{i=0}^{N-1} \frac{(-\ln \Pi)^i}{i!} \quad \text{con} \quad \Pi = \prod_{i=0}^N \max(P_i, 0.005), \quad (4.9)$$

where N is the number of traces considered and P the probability that the track is originated in the primary vertex. Instead the JBP algorithm gives more weight to the tracks (up to a maximum of four) than they are connected to a jet from b; these are identified as they have a high level value of the significance of IP.

Secondary Vertex identification

The presence of a secondary vertex in the event and related variables can be used to discriminate jets coming from a quark b respect to the others. The main variables associated with the secondary vertex are the distance, the flight direction with respect to the primary vertex, the mass and the energy of the particles associated to the traces secondary. To identify a secondary vertex it is required that:

- shares less than 65% of its traces associated with the primary vertex and the significance of the radial distance between the two vertices is beyond 3σ ;
- the flight distance of each candidate is in a cone with $\Delta R < 0.5$ around to the direction of the jet;
- the secondary candidates who have a distance over per 2.5 cm above the primary vertex, both a mass compatible with that of K_0 or that in any case exceeds 6.5 GeV.

The Simple Secondary Vertex (SSV) algorithm uses as a discriminating variable the significance of the flight distance, given by the distance ratio of flight with its uncertainty. Similar to the TC algorithm, there are two versions of SSV: High Efficiency (SSVHE) which uses vertices to which they are associated at least two tracks, and High Purity (SSVHP) that requires at least three tracks. There are more complicated algorithms that consider the secondary vertices together with the information on the average life. This makes it possible to construct a discriminant even when there is no secondary summit, increasing efficiency in terms of to SSV; two of these algorithms are the Combined Secondary Vertex (CSV) and the Combined Secondary Vertex Version 2 + Inclusive Vertex Finder (CSVV2IVF). The difference between the two is that the CSV algorithm uses to reconstruct the vertices share information from jets while the CSVV2IVF algorithm uses all the tracks present in the event. Moreover, in general, there are some variables that are not very related to each other and that they can be used as good discriminators that are:

- the category of vertices;
- the significance of the flight distance in the transverse plane;
- the mass of the vertex;
- the number of vertex tracks;
- the ratio between the energy carried by a track and the other jet tracks;
- the pseudorapidity of the track;
- the number of traces in the jet.

b-tag performance in $X \rightarrow WW$ analysis

In order to assess which tagger and working point are performing better, we have calculated the signal significance for different taggers and working points in the

Tagger (WP)	Significance		
	0 jet	1 jet	0+1 jet
CMVA (loose)	7.31	4.86	8.76
CMVA (medium)	7.39	4.52	8.66
CMVA (tight)	7.35	4.16	8.44
CSV (loose)	7.12	4.64	8.47
CSV (medium)	7.37	4.47	8.62
CSV (tight)	7.36	4.15	8.45
Deep CSV (loose)	7.19	4.80	8.61
Deep CSV (medium)	7.38	4.79	8.80
Deep CSV (tight)	7.39	4.25	8.52

Figure 4.4. Signal significance for different taggers and working points in the 0 and 1 jet categories. The significance for the combination of the 0 and 1 jet categories is shown as well. Only the systematic uncertainties associated to the b tagging scale factors are taken into account for the significance evaluation.

0 and 1 jet categories, which are the most sensitive to the dominant gluon fusion production mode. The significance has been computed running the full analysis but including only the systematic uncertainties associated to the b tagging scale factors. The results are shown in Tab. 4.4. The results show that the usage of CMVA (loose WP) or Deep CSV (medium WP) leads to a comparable signal significance in the combined 0+1 jet category. The CMVA tagger with loose WP has been found to be the best option for the analysis, given the good performance and the nice agreement that has been observed between data and MC, Fig. 4.5 and Fig. 4.6.

4.5 The Missing Transverse Energy

The longitudinal momentum (along the beam axis) in the collision is not known, so the measurement of the total missing energy is impossible. However the initial transverse momentum, carried by the incoming partons, is zero, so in the final state of the collision, for the conservation of the momentum components, the sum of momenta of all the particles must be zero. If a missing transverse momentum, \vec{p}_T^{miss} , in the transverse plane is present, it is the evidence of invisible particles, such as neutrinos particles predicted by some BSM models. The \vec{p}_T^{miss} is defined as,

$$\vec{p}_T^{\text{miss}} = - \sum_{\text{PF } Obj_s} \vec{p}_T^{\text{PF } Obj}, \quad (4.10)$$

where the sum extends over all the PF objects. The \vec{p}_T^{miss} is the negative vectorial sum of transverse momenta of all reconstructed PF objects, and its module is called missing transverse energy, $E_T^{\text{miss}} = |\vec{p}_T^{\text{miss}}|$. As inefficiencies of the tracking algorithm, minimal thresholds in the calorimeter energy estimation, and nonlinearities of the energy response of the calorimeters for hadronic particles can introduce a bias in the \vec{p}_T^{miss} determination, a correction is applied by propagating to the \vec{p}_T^{miss} sum the

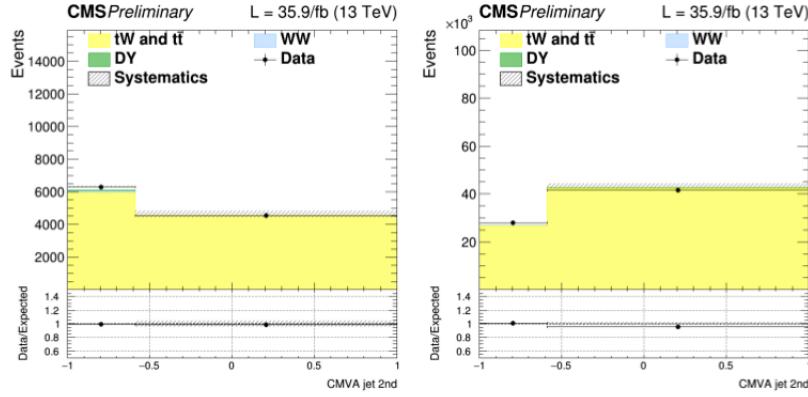


Figure 4.5. CMVA discriminator for jets above 30 GeV (a) and between 20 and 30 GeV (b) in a top enriched control region, after applying the scale factors. The top background normalization is scaled to match data. The systematics band comprises only the uncertainties related to the b tagging scale factors.

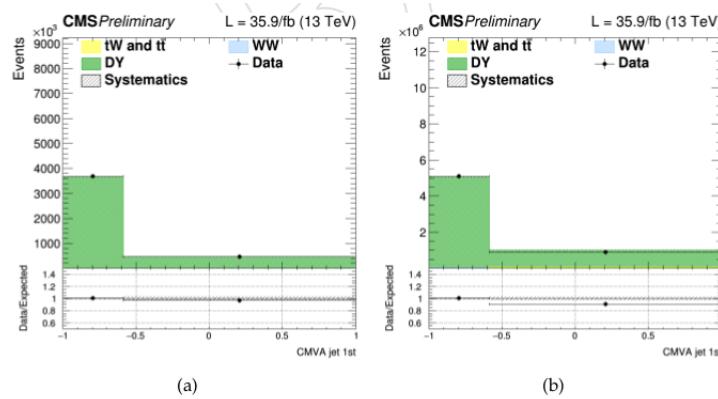


Figure 4.6. CMVA discriminator for jets between 20 and 30 GeV (a) and above 30 GeV (b) in the Z enriched control region. The normalization of the DY background is scaled to match data. The systematics band comprises only the uncertainties related to the b tagging scale factors..

jet energy corrections, in according as,

$$\vec{p}_T^{miss\ correction} = \vec{p}_T^{miss} - \sum_{Jets} (\vec{p}_T^{Corr} - \vec{p}_T), \quad (4.11)$$

where the superscript JEC refers to corrected jets. To estimate the E_T^{miss} systematic uncertainty we have considered the following sources:

- jet p_T ;
- jet resolution;
- muon p_T ;
- electron p_T ;
- unclustered energy.

The E_T^{miss} has been computated varying each of these uncertainties and added in quadrature the differences with respect to the nominal value. For the ϕ uncertainty has been taken as the largest ϕ variation. The distributions of various E_T^{miss} variables, in the electron-muon channel, are shown in Fig. 4.7

4.6 Fake Lepton Background Estimation

Lepton fake rates are measured as a function of the lepton p_T and η , in a single lepton triggered sample. The test of the method and systematic errors are described below. In the analysis the primary source of background from misidentification is W+Jets. QCD multi-jet and hadronic top backgrounds are also present at much smaller level. Events in which W bosons are produced in association with jets give rise to background to WW events when a jet is misidentified as a lepton. These events contain a real lepton and real missing energy from the W decay. With the jet misidentified as a lepton, the W+Jets events have two identified leptons, missing energy, and no other significant event characteristics. As a result, the W+Jets events cannot be readily suppressed by event selection. This background is particularly important at low p_T . The estimation of the fake lepton contribution is based on the “fakeable objet” data-driven method and provides a measurement of the yield and the kinematic distributions of fake background. It is a general technique, applicable to any physics analysis in which particle level selection criteria are used to suppress background. The method can be used with any number of final state particles and is independent of the event selection. The fundamental idea of the fakeable objet method is simple: select a control sample of events enriched in the background being estimated, and then use an extrapolation factor to relate these events to the background in the signal region. The method is data-driven provided the control sample is selected in data, and the extrapolation factor is measured with data. For background arising from particle misidentification, the extrapolation is done in particle identification space (p_T and η) of the lepton. The control sample is defined using a looser particle selection criteria that are chosen such that the rate of misidentification is increased. The extrapolation factor relates background misidentified with this criteria, to background misidentified as passing the full particle selection of the signal region.

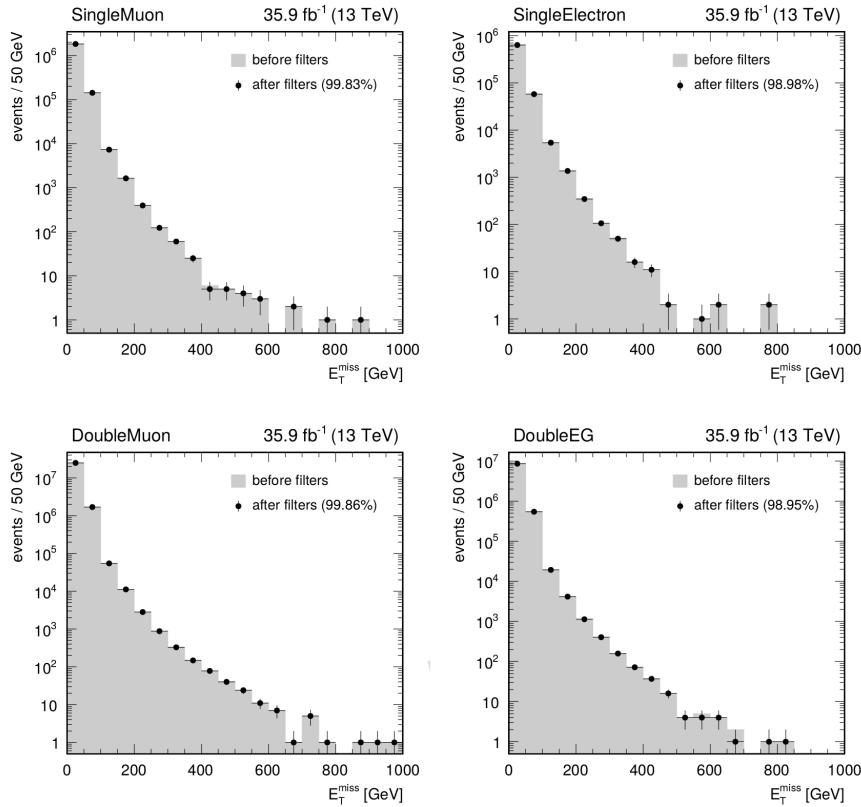


Figure 4.7. The E_T^{miss} filters efficiencies for 35.9 fb^{-1} of 2016 data. The common requirements for the distributions before filters (solid histogram) and after filters (black points) are the trigger and two tight leptons.

Chapter 5

High mass resonances searching

5.1 The X to WW search: summary

As described in Sec. 1.3, the research of new resonance X is one of the mail goals of LHC. With a energy achieved of 13 TeV in the center of mass and the data collected in 2016, it is possible to lead searches in a vast range of mass. One of the main final state channel in a couple of W^+W^- bosons. The ATLAS experiment has been done this kind of searches using the early 2016 statistic, 13.2 fb^{-1} and the results are shown in Fig. 1.15. In the following is reported the $X \rightarrow W^+W^- \rightarrow 2\ell 2\nu$ research with CMS experiment using 2016 data. With the full 2016 statistic, approximately $\sim 36 \text{ fb}^{-1}$ is possible to investigate a wide range of masses and, if will be no evidence of high mass signal, is possible to set tight upper limits on the possible cross section.

The main production mode for the high mass Higgs boson like particle over the all mass spectrum is the gluon-gluon fusion process. However the gluon gluon fusion cross section decreases with m_H but the VBF/gluon fusion cross section ratio increases with the mass, making the VBF production mechanism more and more important as m_H approaches to high values.

The signal samples are interpreted in terms of the EWK singlet and MSSM models described in Sec 1.3. The Higgs boson width and lineshape is reweighted at generator level according to the parameters defined in the model. The interference effects between the signal produced via gluon gluon fusion, the WW background emerging from two gluons and SM Higgs boson, that are expected to change the shape of the signal distribution, have been fully taken into account. A similar treatment is also applied for the interference between VBF high mass signal produced via VBF, the WW plus two quarks background (emerging with the same initial state) and the SM Higgs generated with VBF production mechanism. In general, the interference becomes more and more important as the mass of X increase and it is study in detail, Sec 5.3. Finally, the interference between the $W^+W^- \rightarrow 2\ell 2\nu$ and $ZZ \rightarrow 2\ell 2\nu$ is negligible due to the different phase space characteristic of these processes. In the SM there are some processes that have the same or similar final state respect to the signal. These processes are called background and affecting the signal signature. To estimate and to control the backgrounds in the analysis, control regions are defined for the most important backgrounds. A fully description of backgrounds events are

reported in Sec. 5.5.

The analysis strategy for the high mass search in the $X \rightarrow W^+W^- \rightarrow 2\ell 2\nu$ final state is similar to the high mass analysis performed with 2015 data (2.3 fb^{-1}) [25], but has several improvements. Study on VBF interference, on discriminating variables and a new categorization of the events have been performed. Now indeed, the events are divided according the flavour in the final state:

- opposite-flavour final state, $e^\pm\mu^\mp$,
- same-flavour final state, e^+e^- and $\mu^+\mu^-$.

In the opposite-flavour final state four different jets-categories are defined: the 0-jet, the 1-jet, the 2-jet and the VBF, Sec. 5.7. In the same-flavour final state only the VBF category is considered. Indeed, only the VBF selection cuts are sufficiently tight to reduce the overwhelming Z plus jets background to a manageable level, Sec. 5.8.

5.2 Discriminating variable

This analysis is target as shape analysis, meaning that after applying selection cuts, the events are not simply counted, but rather the data are fitted in a histogram of a discriminating variable with the sum of signal and background templates, and, at the end, the signal yield is extracted. In principle, the variable with the best discriminating power would be the invariant mass of the four lepton system, however it is impossible to reconstruct due to the presence of neutrinos. Usually, in the Higgs boson to $WW \rightarrow 2\ell 2\nu$, the variable used in the analysis is the transverse mass, m_T^H , defined as,

$$m_T^H = \sqrt{2p_T^{\ell\ell}E_T^{\text{miss}}(1 - \cos\Delta\phi(\ell\ell, \vec{p}_T^{\text{miss}}))} \quad (5.1)$$

where $\Delta\phi(\ell\ell, \vec{p}_T^{\text{miss}})$ is the azimuthal angle between the dilepton momentum and \vec{p}_T^{miss} . However m_T^H but also and also the di-lepton mass, $m_{\ell\ell}$, are not very sensitive to different signal mass hypothesis. For this reason a new variable that allows a better sensitivity to different resonance mass hypotheses has been searched. The visible transverse mass, m_T^I , has been introduced, defined as the visible mass,

$$m_T^I = \sqrt{(p_{\ell\ell} + E_T^{\text{miss}})^2 - (\vec{p}_{\ell\ell} + \vec{p}_T^{\text{miss}})^2}. \quad (5.2)$$

This variable is defined as the invariant mass of the four momentum resulting from the sum of the two leptons four-momenta and the missing four-momentum. The distribution of the variables defined above are shown in Fig. 5.1, where it is visible the better power of m_T^I in discriminating different mass hypotheses respect the other variable as m_T^H or $m_{\ell\ell}$. In addition, the usage of this variable also provides a good discriminating power between signal and background.

5.3 Signal interpretation: EW singlet, 2HDM and MSSM

The signal is interpreted in terms of the electroweak singlet model, in 2HDM and finally in MSSM model. The theory part of the models are described in Sec. 1.3.

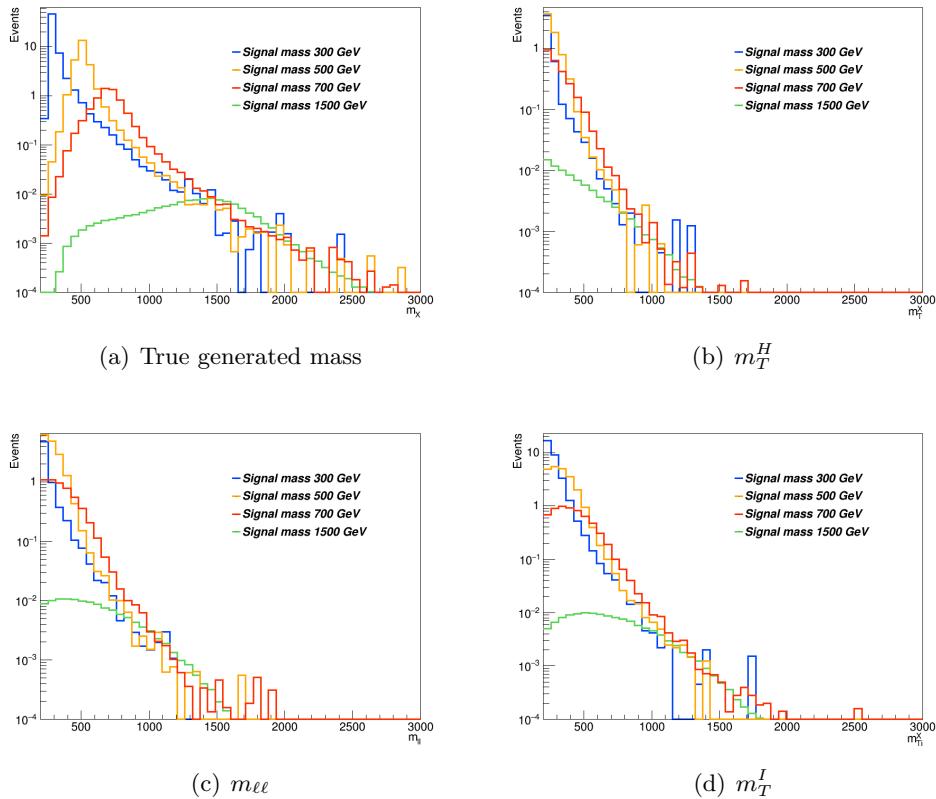


Figure 5.1. Distributions of the generated mass (no possible reconstruction), m_T^H , $m_{\ell\ell}$ and m_T^I variables for different X mass hypothesis. It is clear that the most discriminating variable is m_T^I .

Electroweak singlet model

The EW singlet represents a scalar mixing among the high mass particle and the Higgs boson. This model relies on two parameters: the scale factor of the couplings of the high mass resonance with respect to the SM, C' , and the branching fraction of the electroweak singlet to non-SM decays modes, BR_{new} . The electroweak singlet signal strength, μ' and the modified width, Γ' , are related with the parameters in the model by the following equations:

$$\mu' = C'^2 \cdot (1 - BR_{\text{new}}) \quad (5.3)$$

$$\Gamma' = \Gamma_{\text{SM}} \cdot \frac{C'^2}{1 - BR_{\text{new}}} \quad (5.4)$$

The high mass signal samples for different mass hypothesis have been reweighted according to this model. At the moment only the $BR_{\text{new}} = 0$ hypothesis has been investigated while we tested different C' values. In Fig. 5.2 are shown the $m_{\ell\ell}$ and m_T templates corresponding to a high mass boson of 700 GeV for three different C' values: $C' = 1$, corresponding to the SM Higgs decay width, $C' = 0.5$, corresponding to $\Gamma' = 2.5 \cdot 10^{-2} \Gamma_{\text{SM}}$, and $C' = 0.1$, corresponding to $\Gamma' = 10^{-2} \Gamma_{\text{SM}}$. A value of $BR_{\text{new}} = 0$ is considered in all cases. We note that the signal shape is not very sensitive to different C' values.

2HDM and MSSM models

The 2HDM is a well motivated extension of the SM. It contains two Higgs doublets, from which a total of five Higgs bosons are predicted: Two CP-even bosons h and H , a CP-odd boson A and two charged bosons H^\pm . In most theories, h exhibits the features of the SM Higgs boson, while H is a CP-even Higgs boson at a higher mass. The 2HDM comprises many free parameters. Two of these are of particluar interest:

- $\tan \beta$: The ratio $\frac{v_u}{v_d}$ of the vacuum expectation values of the two Higgs doublets.
- α : The mixing angle of the two scalar Higgs bosons h and H .

The quantity $\cos(\beta - \alpha)$ is also of interest, as the coupling of the heavy scalar Higgs boson H to two vector bosons is proportional to this factor. In the decoupling limit, which occurs at $\cos(\beta - \alpha) = 0$, all couplings become SM-like. A 2HDM of type-2 is considered in this study. Here up-type quarks couple to one doublet, while down-type quarks and leptons couple to the other doublet.

The MSSM is a type-2 2HDM. On tree level only two parameters are left free. By convention, these parameters are chosen to be $\tan \beta$ and m_A , the mass of the pseudoscalar Higgs boson. The exclusion limits can be set in a two-dimensional plane as a function of these two parameters. Due to higher order diagrams additional free parameters occur. Benchmark scenarios are then used in order to constrain these parameters. Here two MSSM scenarios are used: the m_h^{mod+} scenario and the hMSSM scenario [54].

The necessary model predictions for these scenarios are provided by the LHC

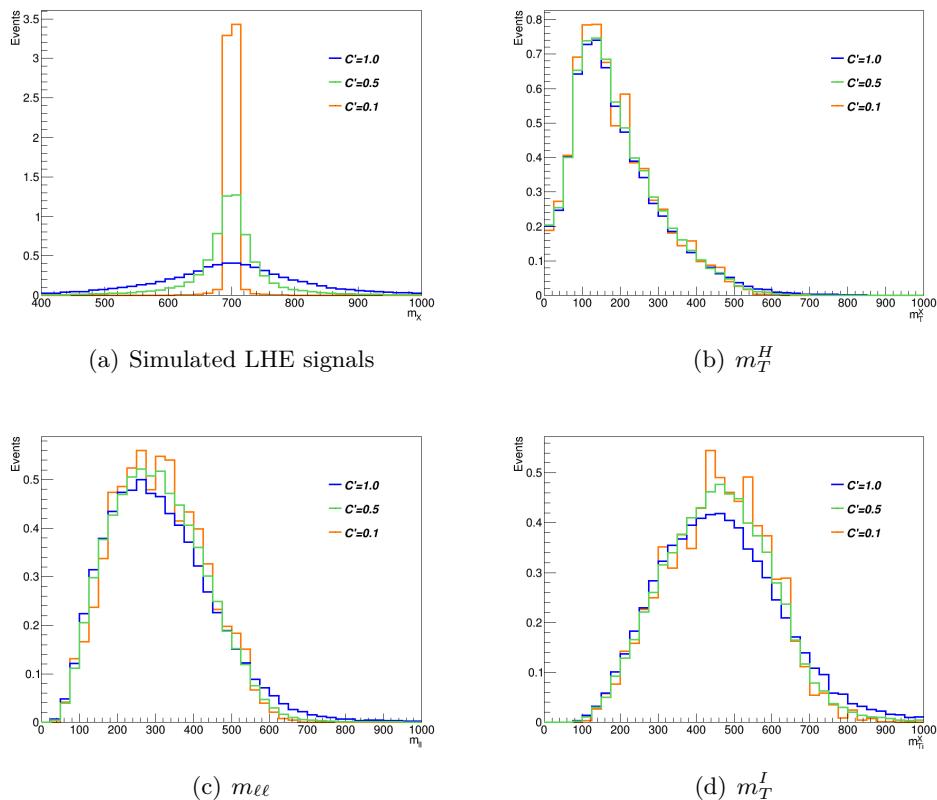


Figure 5.2. Distributions of the signals, the m_T^H , the $m_{\ell\ell}$ and the m_T^I variables at generator level for different values of C' , without any selection.

Higgs Cross Section Working Group [55]. For both MSSM scenarios the ggF cross sections have been computed with SusHi (v.1.4.1)[56]. These cross sections include NLO supersymmetric QCD corrections and NNLO QCD corrections for the top quark contribution in the effective theory of a heavy top quark, as well as electroweak effects by light quarks. The masses of the Higgs bosons, their mixing, the branching fractions and the effective Yukawa couplings in the m_h^{mod+} scenario are all calculated with FeynHiggs (v.2.10.2)[57, 58, 59, 60, 61]. For the hMSSM scenario the branching fractions are obtained from HDECAY (v.6.40)[62, 63]. The results for general 2HDM are obtained using the ggF cross sections computed with SusHi (v.1.5.0) and the branching fractions from 2HDMC (v.1.7.0)[64]. The VBF cross sections are calculated using an approximation. The BSM Higgs production cross sections for VBF, which are provided for different masses by the LHC Higgs Cross Section Working Group [65], are taken and multiplied by $\cos^2(\beta - \alpha)$, resulting in VBF cross sections for a heavy CP-even Higgs boson.

The exclusion limits obtained for the MSSM scenarios are displayed in the m_A - $\tan\beta$ plane. A fine grid is chosen in this plane, and for each point of this grid a maximum likelihood fit is performed after the m_A and/or $\tan\beta$ dependent values of the model, such as cross sections and masses of the Higgs bosons are calculated. These fits are done using the asymptotic method. Performing a maximum likelihood fit in this manner is equivalent to a hypothesis test, where the signal hypothesis is tested against the SM-and-background hypothesis. The signal hypothesis for a combination of m_A and $\tan\beta$ is excluded at 95 % confidence level. In the two-dimensional plane this limit is determined from interpolation between the points of the grid. The limits in the more general 2HDM are obtained in the same way, although a different parameter is chosen in place of m_A .

5.4 Study of the Interference effects

When a resonance X , with a non negligible width is considered, it is important to take into account also the interference effects both with the WW background , with same initial and final state, and with the Higgs boson off-shell tail. In this analysis the interference effects between the new signal X produced in gluon-gluon fusion and in vector-boson-fusion is taken into account. The effect of the various interference terms are shown in 5.3 and 5.4 for the two different production mechanism, gluon-gluon fusion and vector-boson fusion. The contribution of the interference of high mass resonances X with the WW background and with the Higgs boson have opposite sign and partially cancel out. This cancellation effect is different for different resonance masses. The interference contribution is thus non negligible and is included in the fit, Sec 6.1.

5.5 Main Background processes

Inside the SM that are several processes that have the same or a similar final state of the signal, called background processes. The most important background processes contributing to this final state are non resonant $q\bar{q} \rightarrow W^+W^-$, the top production

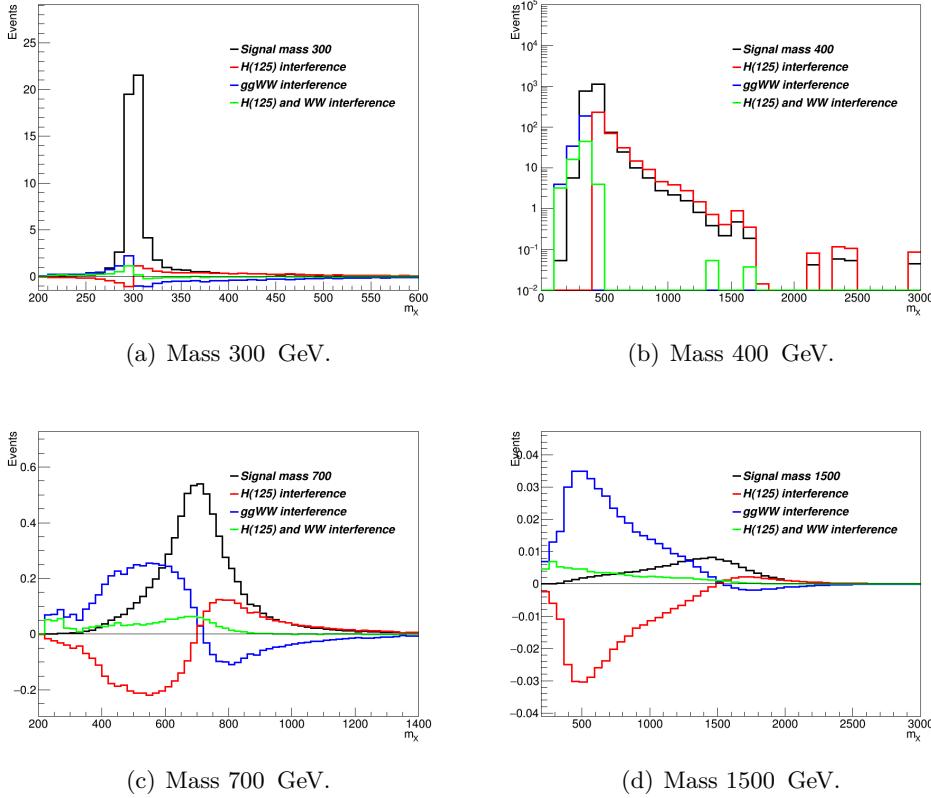


Figure 5.3. Distribution of for the X mass resonance, produced via gluon-gluon fusion for different masses. In black the high mass signal. In red the interference between the high mass signal and the Higgs boson. In blue the interference between the high mass signal and the background. In green the total interference i.e. high mass signal, Higgs bison and background.

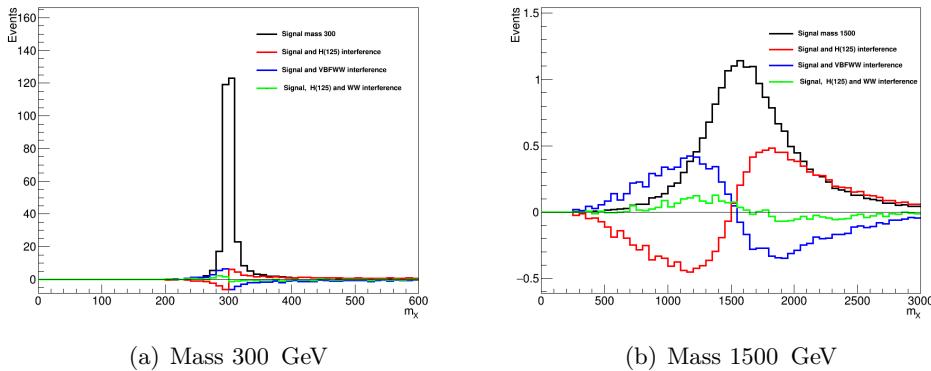
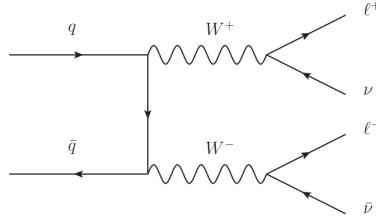


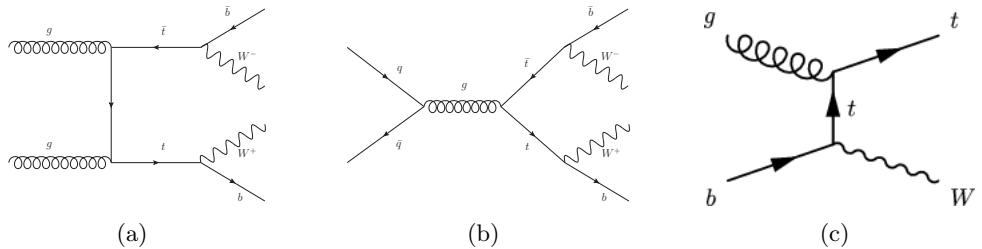
Figure 5.4. Distribution of for the X mass resonance, produced via vector-boson-fusion fusion for different masses. In black the high mass signal. In red the interference between the high mass signal and the Higgs boson. In blue the interference between the high mass signal and the background. In green the total interference i.e. high mass signal, Higgs bison and background.

($t\bar{t}$ and single-top) and the Drell-Yan process. Other background the W and the Z bosons. All these processes have been simulated with Monte Carlo generators and the simulation details have been discussed in Sec 3.8. Following, the description of the main different background processes:

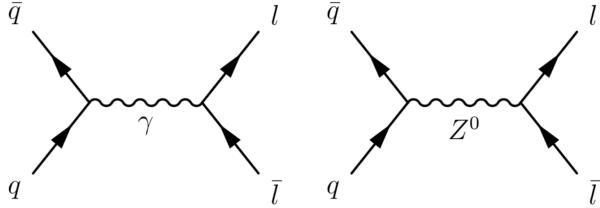
- *Non-resonant WW ($q\bar{q} \rightarrow W^+W^-$)*: this background is characterized by a final state identical to the signal, however the lepton kinematics for signal and $q\bar{q} \rightarrow W^+W^-$ processes is rather different. For the signal process, the W bosons originate from a spin-0 particle decay and their spins must therefore be antiparallel, implying that the charged leptons produced in their decays appear preferentially in the same hemisphere [66]. In contrast, there is no preferential spin direction in the background case. For this reason the azimuthal angle difference between the two leptons is on average smaller for signal than for background, resulting in a smaller dilepton invariant mass in the former case. The Feynman diagram of the process is reported below.



- *Top ($t\bar{t}$ and single-top)*: the $t\bar{t}$ events can give a signal-like signature if the decay chain is $t \rightarrow Wb$ and subsequently $W \otimes \ell\nu$. In such a case, in fact, there are two leptons and two jets (from the hadronization of the b quark) in the final state. This process is especially important when the signal is produced via VBS or when the signal is produced with jets coming from initial or final state radiation. The single-top instead is characterized by the presence of W boson and a top quark. Following, some examples of Feynman diagrams for the top background. The $t\bar{t}$ in (a) and (b) diagrams, the single-top in (c).



- *Drell-Yan*: the Drell-Yan process is defined as the annihilation of a quark-antiquark pair into a lepton-antilepton pair. his process is described at leading order by the two Feynman diagrams drawn. These two amplitudes are proportional to the fine structure constant $\alpha \sim 1/137$. This kind of background is particular important for the same flavour final state of the signal having two leptons of same flavour.



- *W+jet*: this background is characterized by a W boson, decaying in $\ell\nu$ in addition to a jet. A fake lepton arise from the misidentified jet and lead the same final state of the signal (two leptons and missing-transverse-energy).
- *Other*: other background processes involved the Z bosons, such as $WZ/W\gamma^*$, ZZ^* with a misidentified lepton and $Z\gamma$ with γ conversion.

The main background processes the WW production and the top production are estimated using data. Instrumental backgrounds arising from non-prompt leptons in $W+jets$ production and mis-measurement of E_T^{miss} in Drell-Yan events are also estimated from data. The contribution from $W\gamma^*$ is estimated partly from data. The contribution of other sub-dominant backgrounds is obtained directly from simulated samples. The different data-driven background estimations are explained in the following sections. More precisely top and Drell-Yan backgrounds normalizations have been extracted directly from data-simulation comparison in specific control regions enriched in either one or the other background separately for the different events categories, using the rateParam feature of the combine package [67].

5.6 Data sample and Triggers

Data in CMS

Data recorded in proton proton collisions at 13 TeV during all 2016 was used in the analysis, with a total integrated luminosity of 35.9 fb^{-1} . The data has been reprocessed in the reprocessing campaign characterized by the submission date *03Feb2017* in CMS. In Table 5.1 the different data streams used are presented. All runs are taken at 25 ns and recorded in seven different periods.

Triggers

In the high mass analysis final states with 2 leptons are studied, therefore a combination of both single and double lepton triggers are used. For the electrons A

Data Taking Era	Stream
Run2016C	SingleMuon DoubleMuon SingleElectron DoubleEG MuonEG
Run2016D	SingleMuon DoubleMuon SingleElectron DoubleEG MuonEG
Run2016E	SingleMuon DoubleMuon SingleElectron DoubleEG MuonEG
Run2016F	SingleMuon DoubleMuon SingleElectron DoubleEG MuonEG
Run2016G	SingleMuon DoubleMuon SingleElectron DoubleEG MuonEG
Run2016H	SingleMuon DoubleMuon SingleElectron DoubleEG MuonEG

Table 5.1. Data samples used in the analysis. The total integrated luminosity corresponds to 35.9 fb^{-1} .

Dataset	HLT path
SingleElectron	HLT_Ele45_WP Loose_Gsf_v* HLT_Ele27_eta2p1_WP Loose_Gsf_v*
SingleMuon	HLT_IsoMu22_v* HLT_IsoTkMu22_v*
MuonEG	HLT_Mu8_TrkIsoVVL_Ele17_CaloIdL_TrackIdL_IsoVL_v* HLT_Mu17_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v*
DoubleMuon	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_v* HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_v*
DoubleEG	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*

Table 5.2. HLT paths used in the analysis.

combination of triggers is necessary to increase the statistics and to adjust the p_T threshold requirement. Efficiencies are measured using the Tag and Probe method described in App. B. For the trigger efficiencies calculation, it is applied the complete identification/isolation criteria on the leptons. The results for the trigger efficiencies for electron triggers using Tag and Probe method are more 95%. Also for the calculation of the muon trigger efficiency the tag-and-probe method has been used. The triggers used in the analysis are summarized in Table 5.2 In Figure 5.5 the trigger efficiency for a gluon fusion signal with mass 300 GeV is shown for electrons (left) and muons (right). An average trigger efficiency greater than 99% is found, as shown in Figure 5.6. The triggers used in the analysis are summarized in Table 5.2.

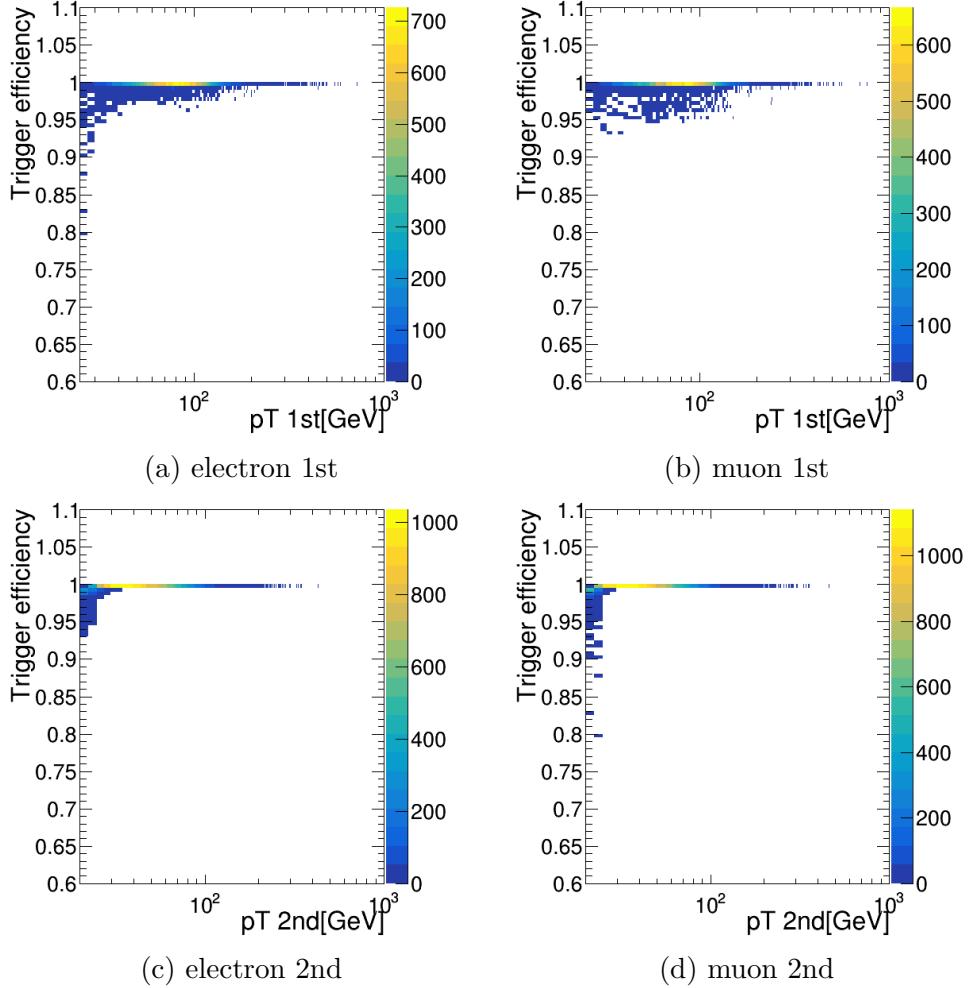


Figure 5.5. Trigger efficiency per event as a function of the lepton p_T for electrons (a) where leading lepton is an electron, and (c) where trailing lepton is an electron, and muons (b) where leading lepton is a muon, and (d) where trailing lepton is a muon, for a gluon fusion 300 GeV MC sample. In this plots the other lepton not shown is integrated. An average trigger efficiency greater than 99% is found.

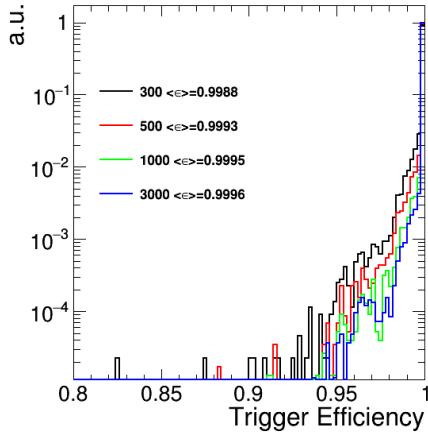


Figure 5.6. Trigger efficiency distribution for four MC samples corresponding to masses of 300, 500, 1000 and 3000 GeV. An average trigger efficiency greater than 99% is found.

5.7 Opposite Flavor final state

General overview

The goal of the analysis which will be described in the following is the search for a high mass Higgs boson decaying into WW with opposite-flavour final state $W^+W^- \rightarrow \mu^\pm e^\mp 2\nu$. The analysis take in account the gluon gluon fusion and the VBF production mechanism.

The events are subdivided in signal and control regions. The signal events are expected to fill mainly to first one. Instead the second one are expected to be fill up by backgrounds events and they are used to normalized the backgrounds shape.

To avoid a possible bias of the experimenter in the process of developing the analysis strategy and optimizing the selection requirements, firstly a blind analysis has been performed by defining a background-only sideband control region and only looking at data events falling in such a region. Precisely, the control region is defined according to selection criteria that exclude the signal events. Data in the signal region, i.e. were not examined until the event selection criteria were settled, only relying on Monte Carlo simulations for what concerns signal. In this way, the analysis strategy has been decided looking at the simulated events. Only when data distributions in signal-depleted regions confirmed the robustness of the analysis, data in signal region were inspected. For these reasons, the agreement between data and Monte Carlo was checked on a sideband region.

Signal region

The events are requested to pass single or double lepton triggers, and exactly one electron and one muon are requested to be reconstructed in the event. One of the two leptons is requested to have a p_T greater than 25 GeV, the other is requested to have p_T greater than 20 GeV and both leptons are requested to be well identified and

isolated, to reject non-prompt leptons and leptons coming from QCD sources. To suppress background processes with three or more leptons in the final state, such as ZZ, WZ, Z γ , W γ or triboson production, no additional identified and isolated lepton with $p_T > 10$ GeV should be reconstructed. The low dilepton invariant mass region dominated by QCD production of leptons is not considered in the analysis and $m_{\ell\ell}$ is requested to be higher than 50 GeV to reduce the SM Higgs boson ($m_H=125$ GeV) contamination. A moderate cut on the missing-transverse-energy is applied ($\text{MET} > 20$ GeV) due to the presence of neutrinos in the final state. Since a high mass signal is searched a cut on $m_T^I > 100$ GeV is applied. A cut on the transverse momentum ($p_T^{\ell\ell} > 30$ GeV) and on the $m_T^H > 60$ GeV are applied against $DY \rightarrow \tau\tau$ background. Finally, against the top background, all jets above 20 GeV are requested not to be identified as b-jets according to the cMVAv2 tagger, loose WP. The full selection, defined as the “WW opposite-flavour selection”, is summarized:

- Two isolated leptons with different charge and flavor ($\mu^\pm e^\mp$);
- p_T of the leading lepton > 25 GeV;
- p_T of the trailing lepton > 20 GeV;
- Third lepton veto: veto events if a third lepton with $p_T > 10$ GeV;
- $m_{\ell\ell} > 50$ GeV, to reduce H(125) contamination;
- MET > 20 GeV;
- $m_T^I > 100$ GeV;
- $p_T^{\ell\ell} > 30$ GeV;
- $m_T^H > 60$ GeV;
- no b-tagged (cMVAv2 loose WP) jets with $p_T > 20$ GeV;

Events passing the “WW opposite-flavour selection” are categorized according to the jet multiplicity, counting jets above 30 GeV, to enhance the sensitivity, especially against the top background.

- **0 jet**, no jets are required in the event;
- **1 jet**, exactly 1 jet is required in the event;
- **2 jet**, exactly 2 jets are required in the event and in addition the condition $\Delta\eta_{jj} < 3.5$ **or** $m_{jj} < 500$ GeV;
- **VBF**, exactly 2 jets are required in the event and in addition the condition $\Delta\eta_{jj} > 3.5$ **and** $m_{jj} > 500$ GeV;

where the 2 jet and VBF regions are mutually exclusive by construction. To extract high mass boson signals in these four categories, the strategy is followed: the m_T^I distribution is fitted as the sum of signal and background templates. Different binnings have been chosen for the m_T^I distributions in the different categories. The binning was chosen to have at least 10 top Monte Carlo events in each bin of the template. The chosen bins are:

- **0/1/2 jet**, [100,150,200,250,300,350,400,450,500,550,600,650,700,750,800,900,1000,2000]
- **VBF**, [100,150,200,250,300,350,400,500,700,1000,2000]

where the first number represents the lower edge of the first bin while the other numbers represent the upper edges. The last bin is an overflow bin. The m_T^I distributions for the signal regions are presented in the four categories in Figs. 5.7.

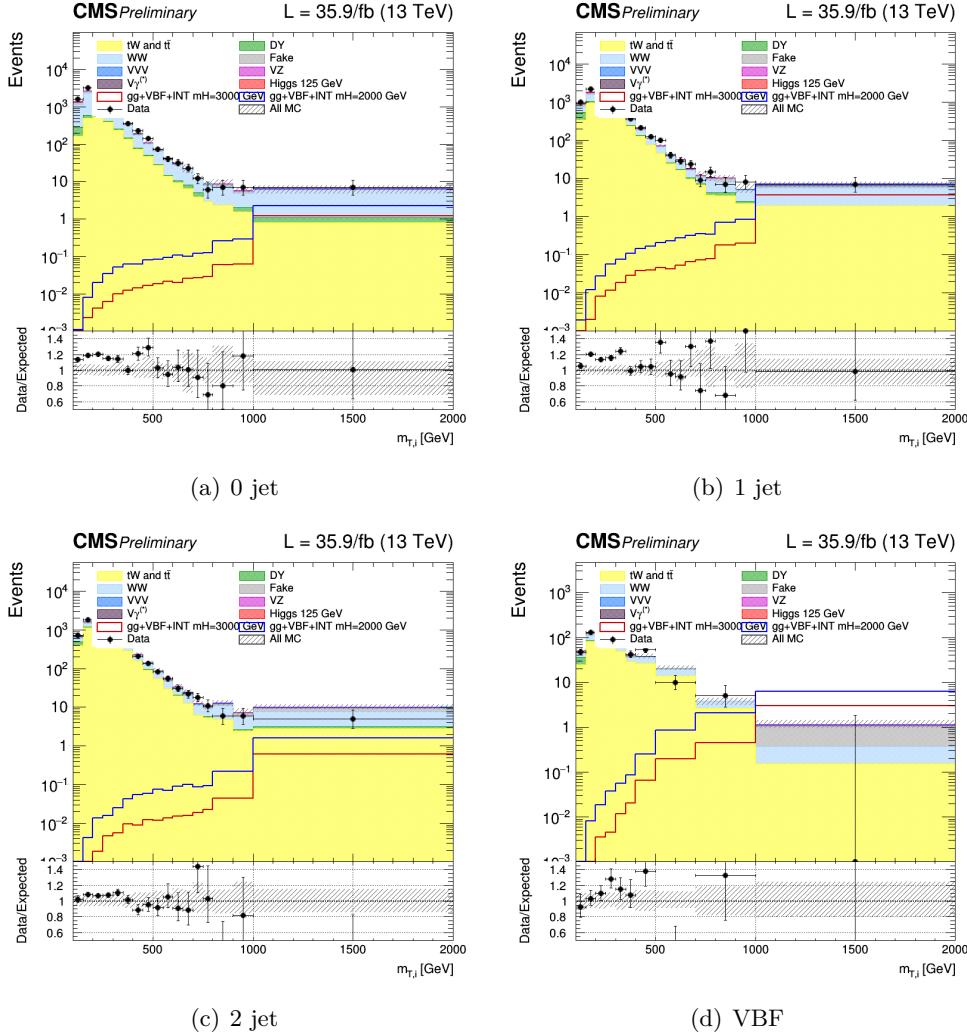


Figure 5.7. Unblinding distributions m_T^I in the signal region for 0, 1, 2 and VBF categories. The signal hypothesis corresponding to m_X of 2 and 3 TeV.

Drell-Yan $\tau\tau$ control region

To normalize the Drell-Yan $\tau\tau$ background to the data, control regions have been defined, as close as possible to the signal region, but enriched in $Z \rightarrow \tau^+\tau^-$. In particular, the “WW OF selection” is used with inverted m_T^H cut, i.e. $m_T^H < 60$. In addition a cut on the invariant mass of the two leptons $50 \text{ GeV} < m_{\ell\ell} < 80 \text{ GeV}$ is requested to exclude possible contribution from non-prompt leptons (low limit) and from $t\bar{t}$ (high limit). For each signal category, a corresponding Drell-Yan $\tau\tau$ control regions is defined. We thus have 4 total Drell-Yan $\tau\tau$ control regions, for 0 jets, 1 jets, 2 jets and VBF. The control plots for several variables in a Drell-Yan enriched phase space for the four jets categories are shown in Figs. 5.8, 5.9, 5.10, 5.11. In general there is a good agreement between data and MC.

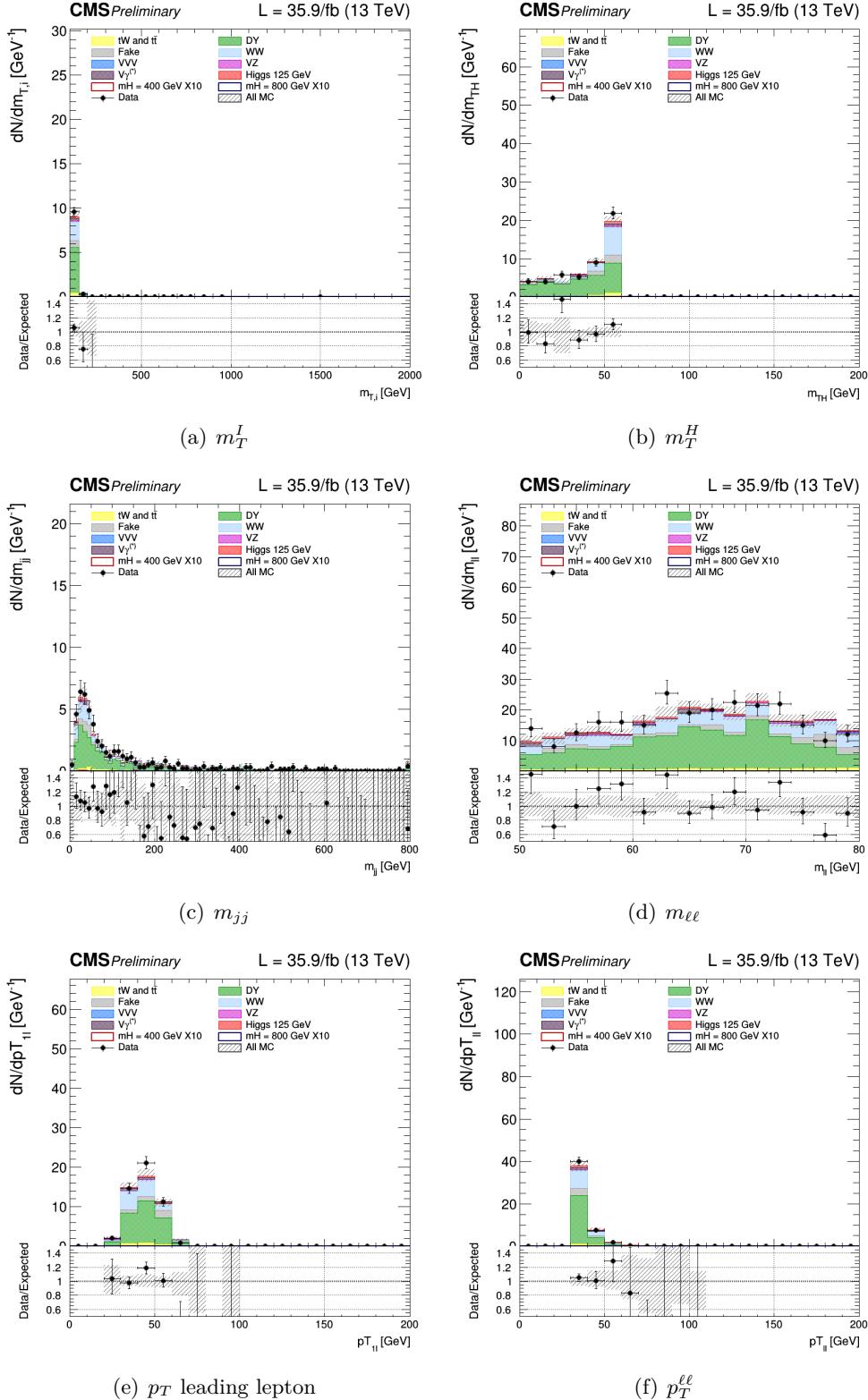


Figure 5.8. Control plots for several variables in a Drell-Yan enriched phase space for events with 0 jet.

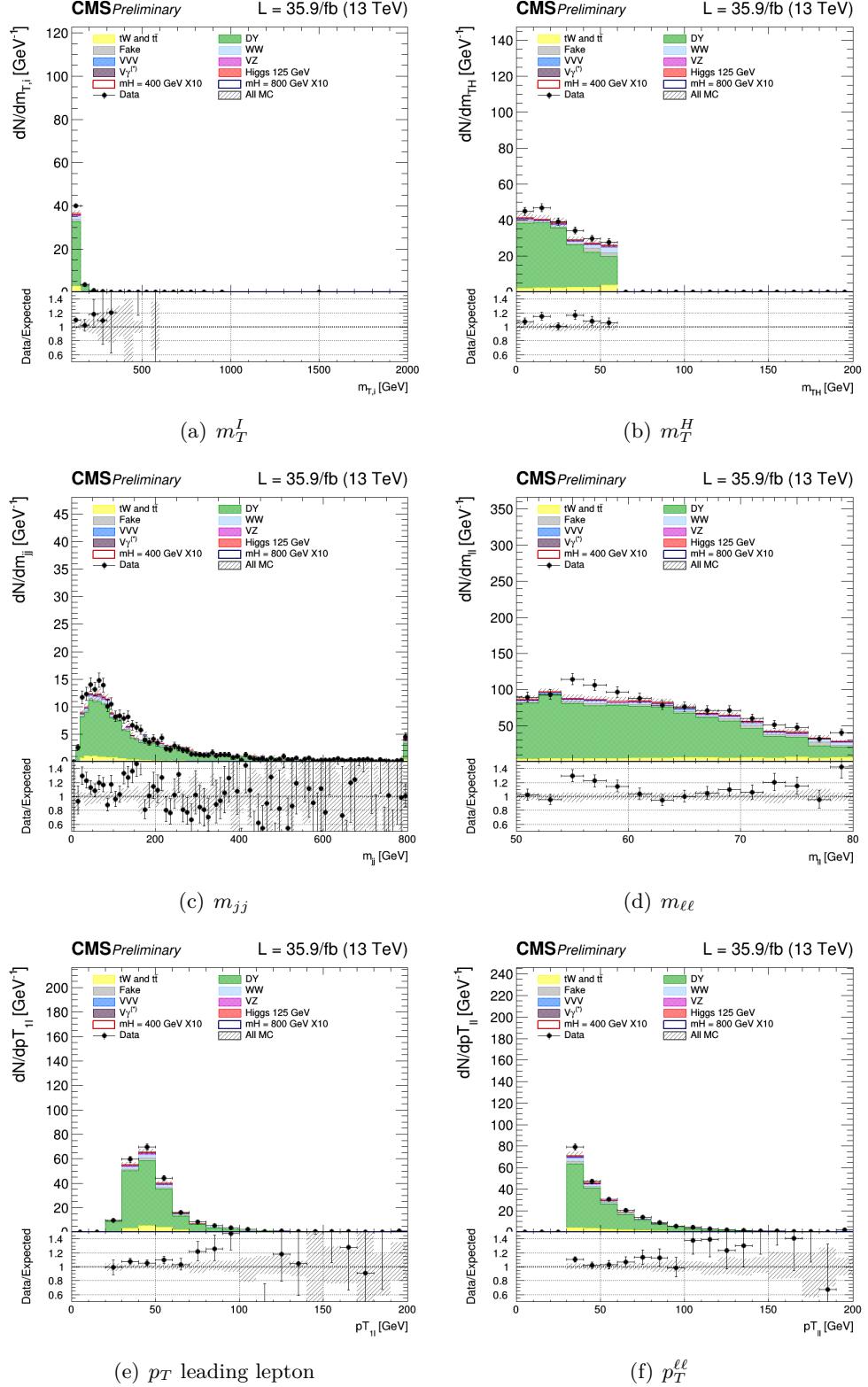


Figure 5.9. Control plots for several variables in a Drell-Yan enriched phase space for events with 1 jet.

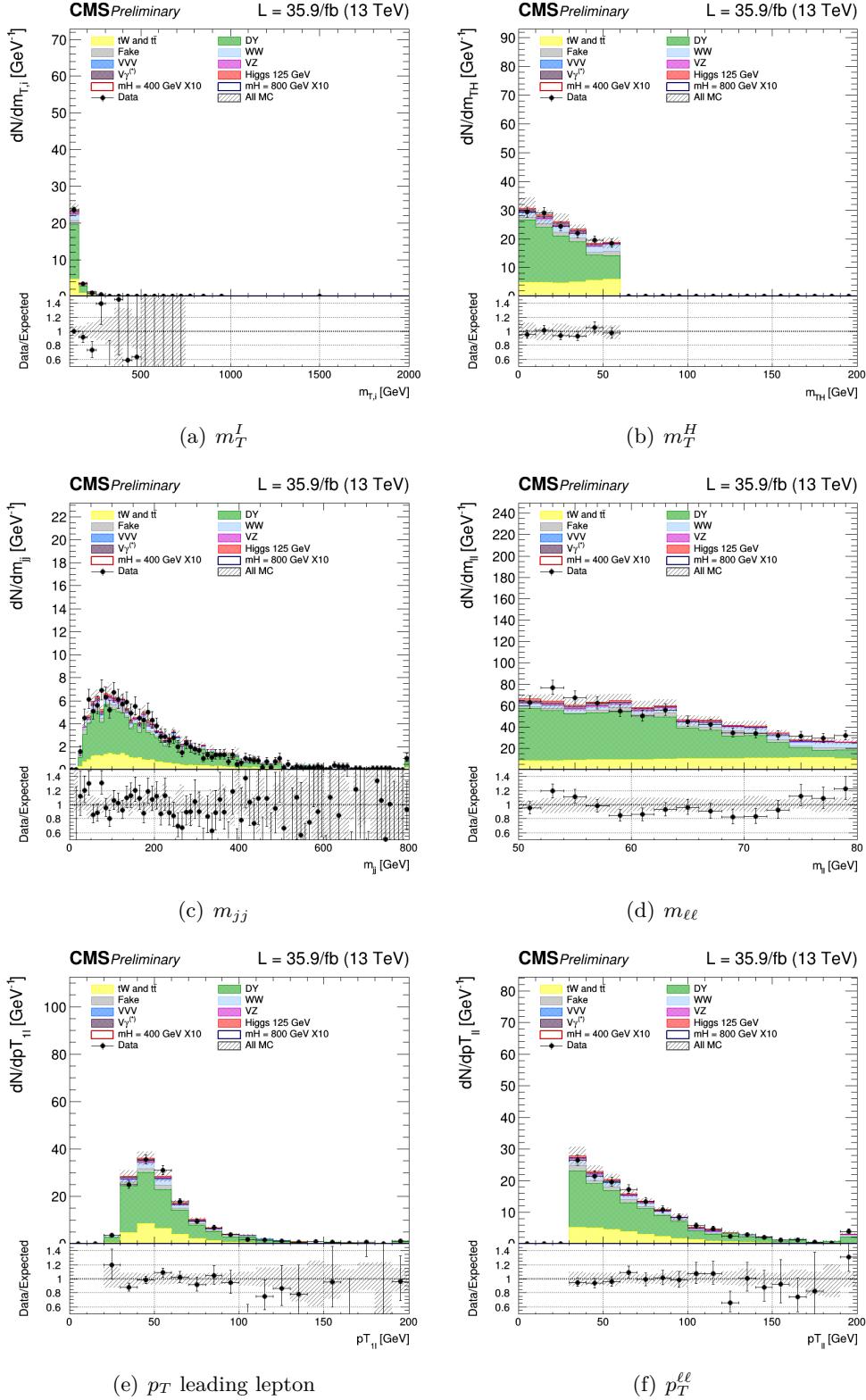


Figure 5.10. Control plots for several variables in a Drell-Yan enriched phase space for events with 2 jet.

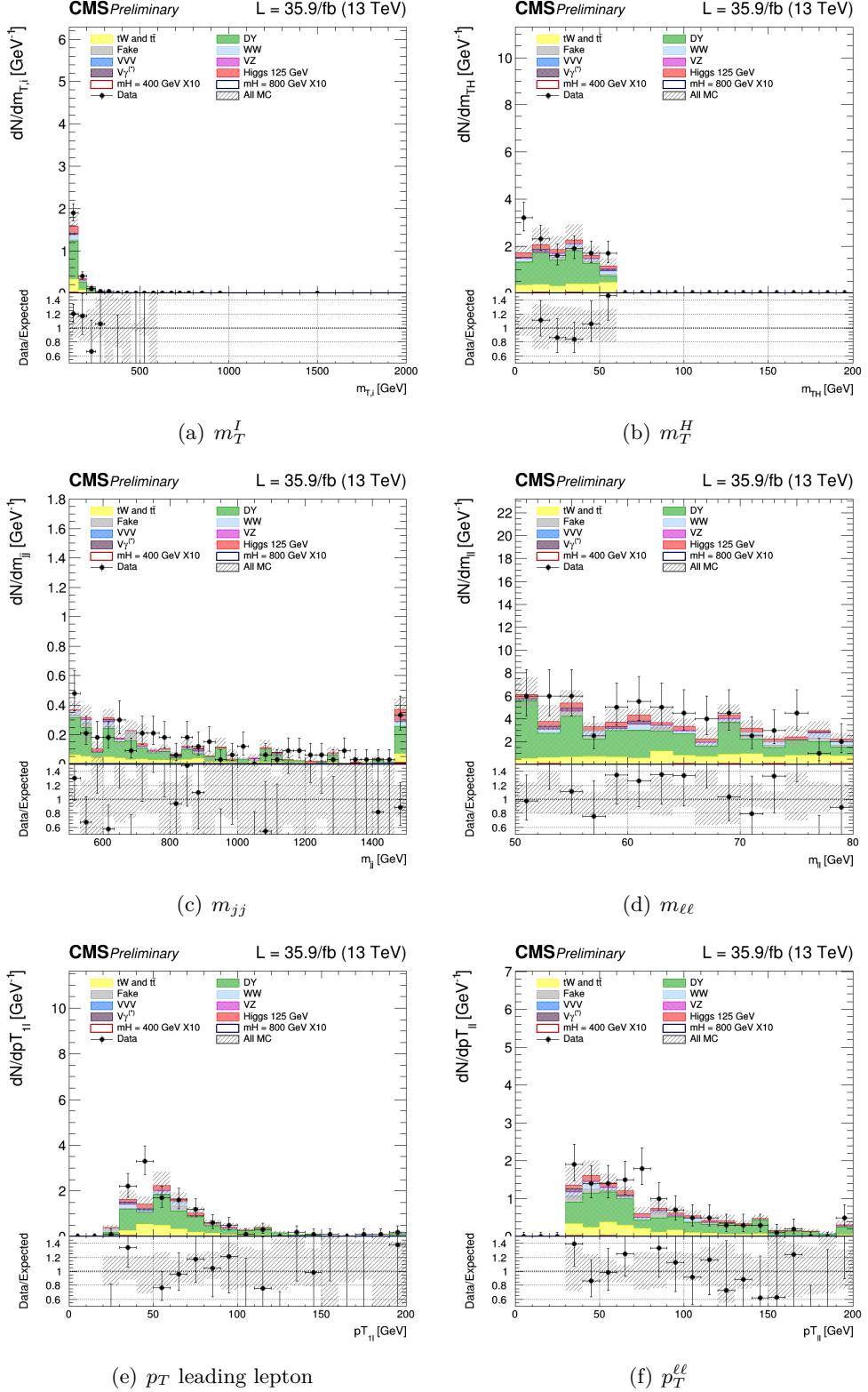


Figure 5.11. Control plots for several variables in a Drell-Yan enriched phase space for events for VBF.

Top control region

Similarly to the Drell-Yan $\tau\tau$ case, control regions are defined for the Top background, and they are used to normalize the top background to data. The “WW OF selection” is used with inversion of the veto on b-jets. In particular the following conditions are imposed to select a top enriched control region for each of the 4 signal regions:

- **0 jet**, at least one b-tagged jet with $20 < p_T < 30$ GeV is required;
- **1 jet**, exactly one b-tagged jet with p_T above 30 GeV is required;
- **2 jet**, exactly 2 jets with at least one of them b-tagged and in addition the condition $\Delta\eta_{jj} < 3.5$ **or** $m_{jj} < 500$ GeV;
- **VBF**, exactly 2 jets with at least one of them b-tagged and in addition the condition $\Delta\eta_{jj} > 3.5$ **and** $m_{jj} > 500$ GeV.

A jet is considered b-tagged if its cMVAv2 score is above the threshold defining the loose working point. The control plots for several variables in a top enriched phase space for events are shown in the Fig. Figs. 5.12, 5.13, 5.14, 5.15. The last bin in the distribution is the overflow.

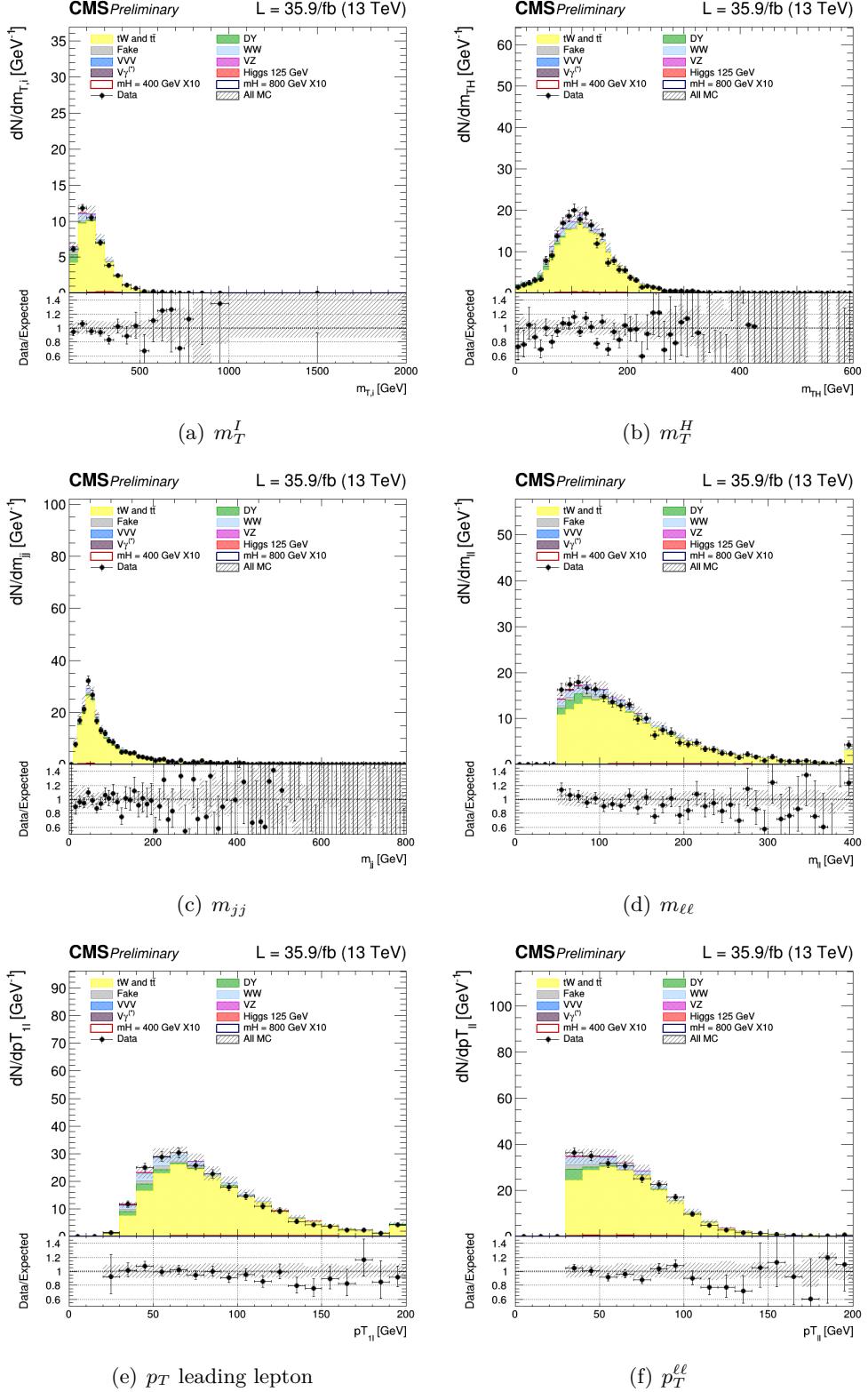


Figure 5.12. Control plots for several variables in the Top enriched phase space for events with 0 jet.

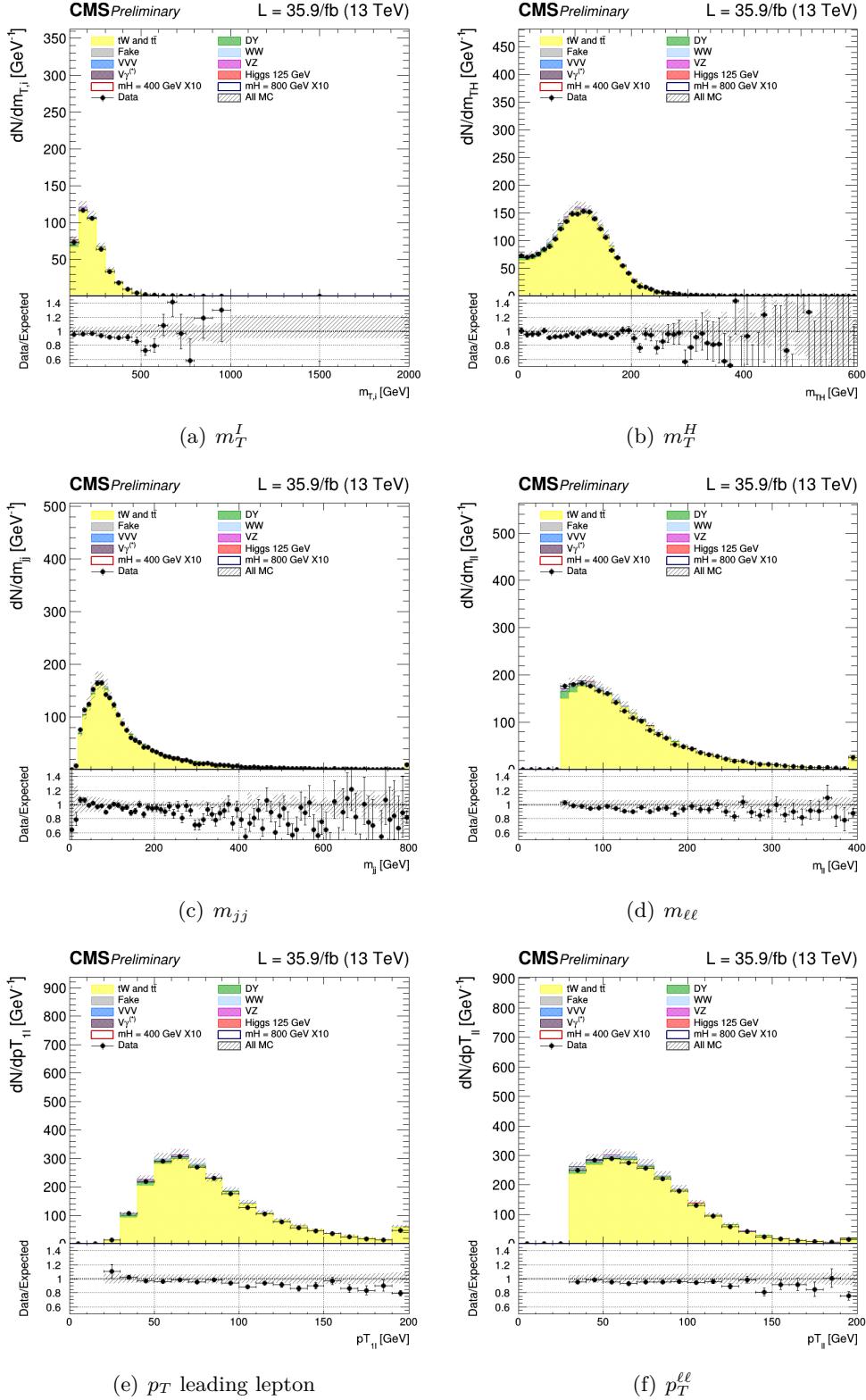


Figure 5.13. Control plots for several variables in the Top enriched phase space for events with 1 jet.

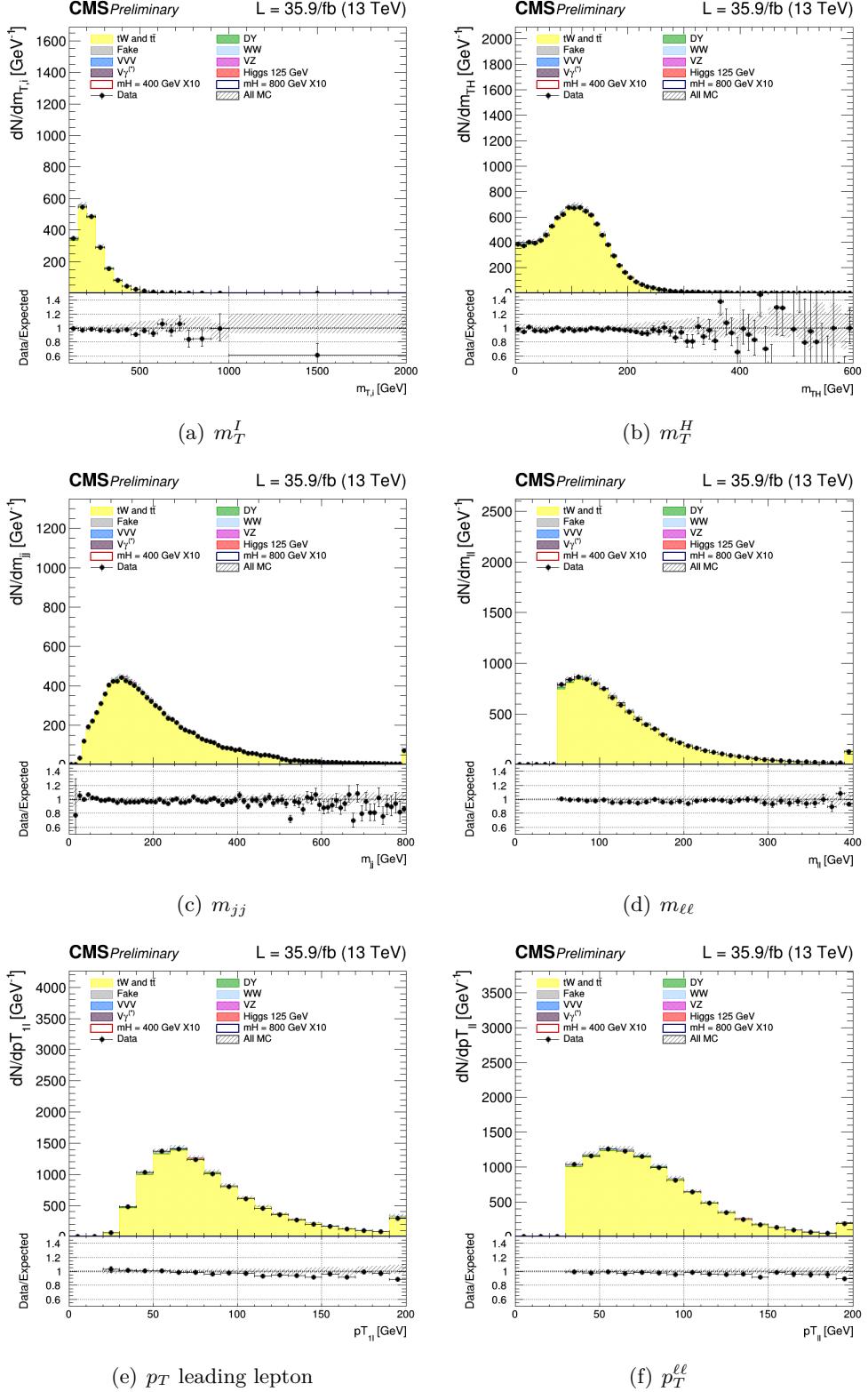


Figure 5.14. Control plots for several variables in the Top enriched phase space for events with 2 jet.

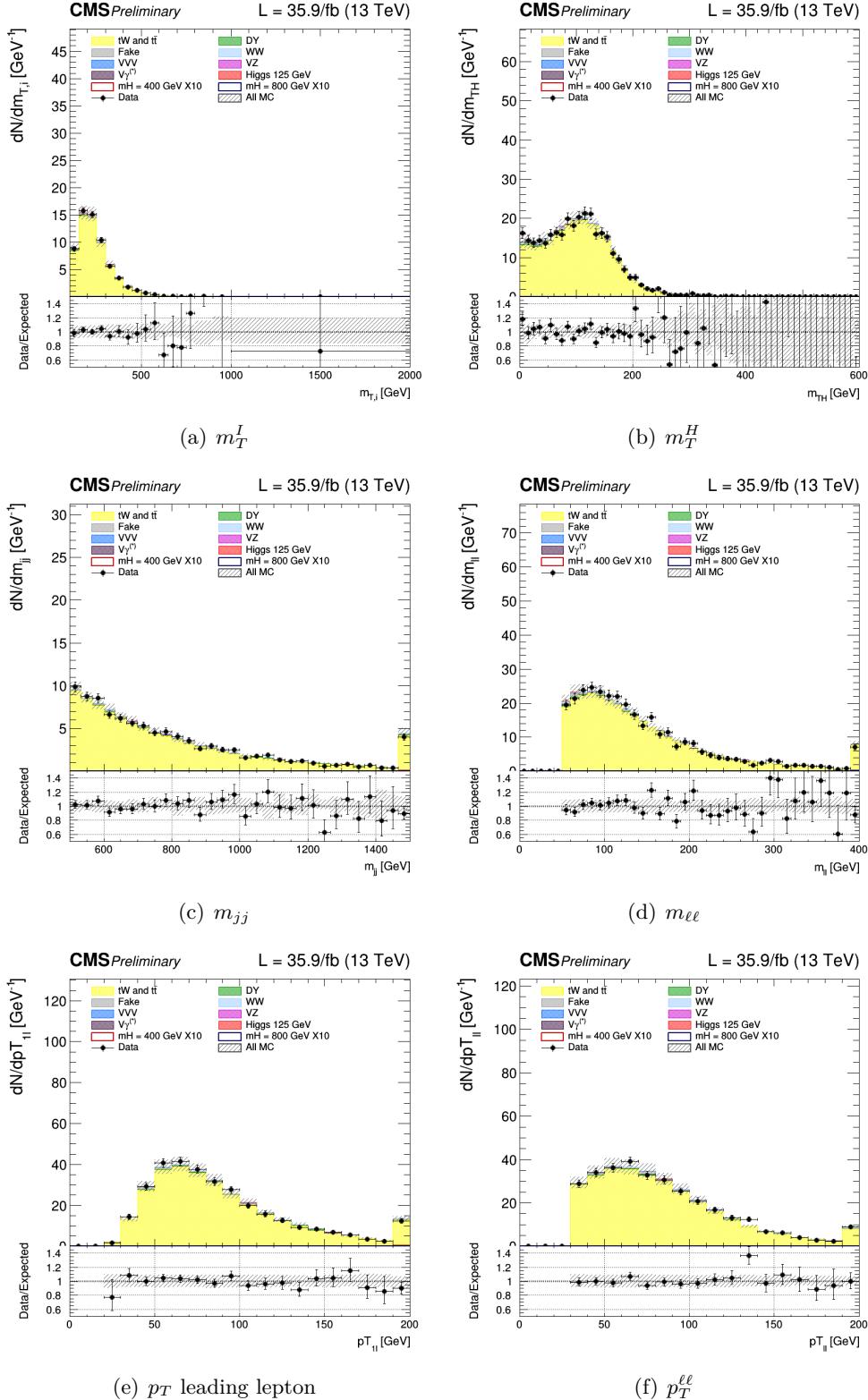


Figure 5.15. Control plots for several variables in the Top enriched phase space for events in VBF region.

5.8 Same Flavor final state

The analysis if the same-flavour final state $W^+W^- \rightarrow \mu^\pm\mu^\mp 2\nu$ and $W^+W^- \rightarrow e^\pm e^\mp 2\nu$ is described.

Signal region

Events are requested to pass single or double lepton triggers and all the physics objects definitions are the same as in the OF analysis. The final state consists of two well identified electrons or two muons with $p_T > 20$ GeV, opposite charge, and large missing transverse energy from the undetected neutrinos.

In addition to the backgrounds described for the OF final state, the background from $DY \rightarrow \mu^+\mu^-$ and $DY \rightarrow e^+e^-$ is very large in this final state. Indeed, due to this very large background, the SF analysis only targets the VBF topology, where the DY background is suppressed by the tight jet requirements. In addition, an invariant mass of the two leptons larger than 120 GeV is requested. The full selection, defined as the “WW same flavour selection”, is :

- Two isolated leptons with same flavor and opposite charge ($\mu^\pm\mu^\mp$ and $e^\pm e^\mp$);
- p_T of the leading and trailing lepton > 20 GeV;
- Third lepton veto: veto events if a third lepton with $p_T > 10$ GeV;
- $m_{\ell\ell} > 120$ GeV
- $p_T^{\ell\ell} > 30$ GeV;
- MET > 50 GeV;
- $m_T^I > 100$ GeV;
- At least 2 jets non b-tagged (according to cMVAv2 loose WP) with $p_T > 30$ GeV.
- $\Delta\eta_{jj} > 3.5$;
- $m_{jj} > 500$ GeV;;

Similarly to the opposite-flavour analysis, the signal is extracted from a template fit of the m_T^I distribution. The m_T^I distributions has the following binning:

- **VBF**, [100,150,200,250,300,350,400,450,500,600,700,1000];

where the first number represents the lower edge of the first bin while the other numbers represent the upper edges. The last bin is an overflow bin. The binning has been chosen in order to have at least 10 expected Top-backgrounds event and at least 10 expected Drell-Yan events in each bin of the template.

The distributions for the signal region of m_T^I variable is shown in Fig. 5.7

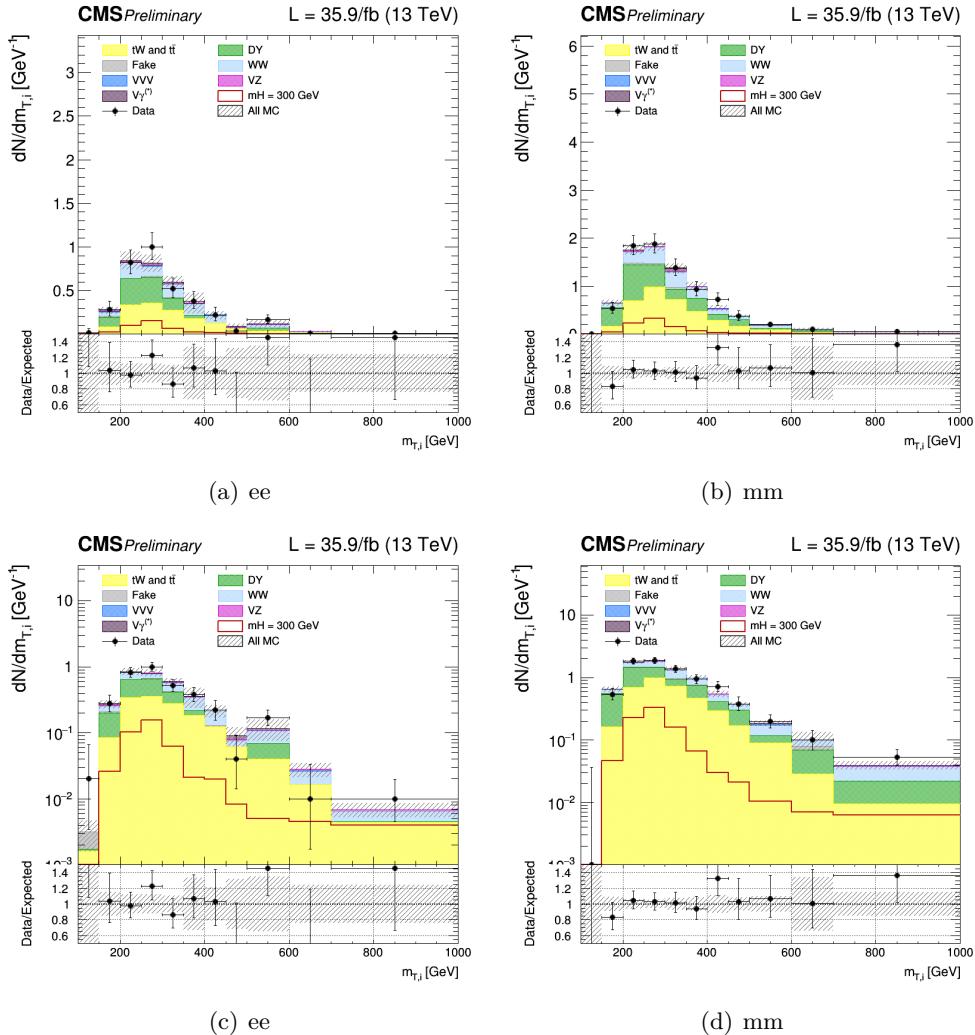


Figure 5.16. Unblinding distributions m_T^I in the signal region for ee and $\mu\mu$ categories in linear and log scale. The signal hypothesis corresponding to m_X of 300 GeV.

Drell-Yan control region

The main background for the SF analysis is the DY. A control region has been defined, as close as possible to the signal one to be used for the normalization of the DY background, separately for electrons and muons. The control region is defined by the “WW same flavour selection”, except for the $m_{\ell\ell}$ requirement which is changed to $70 \text{ GeV} < m_{\ell\ell} < 120 \text{ GeV}$ to include the Z boson. The missing transverse energy distribution in the data shows discrepancies respect to Monte Carlo simulation in ee and $\mu\mu$ Drell-Yan control regions. A correction is applied reweighting all the simulated samples with a weight per event which depends on the MET value. The weight is evaluated as the ratio between data, one subtracted all backgrounds except the DY, and the Drell-Yan itself, in each bins of the distribution, separately for ee and $\mu\mu$ categories. The weight is assumed to be linear as function of the MET value.

This kind of reweighting allows to correct for shape differences between data and MC, , Fig. 5.17. The control plots for several variables in a Drell-Yan enriched phase space for the ee and $\mu\mu$ are shown in Figs. 5.19 for the dielectron case and Figs. 5.20 for the dimuon case. In general there is a good agreement between data and MC.

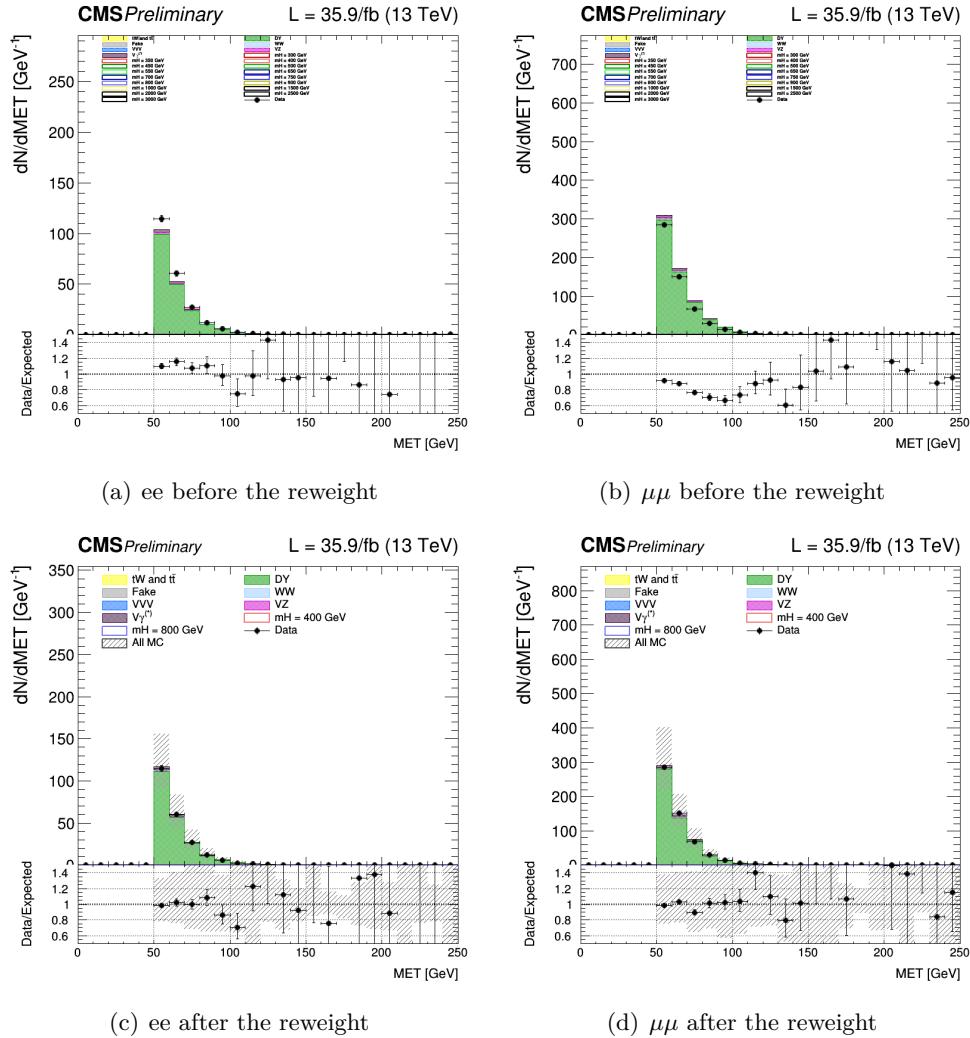


Figure 5.17. MET control plots for Drell-Yan fot ee categories in *a* and for $\mu\mu$ in *b* before the reweight. In *c* and *d* the same distribution after the correction.

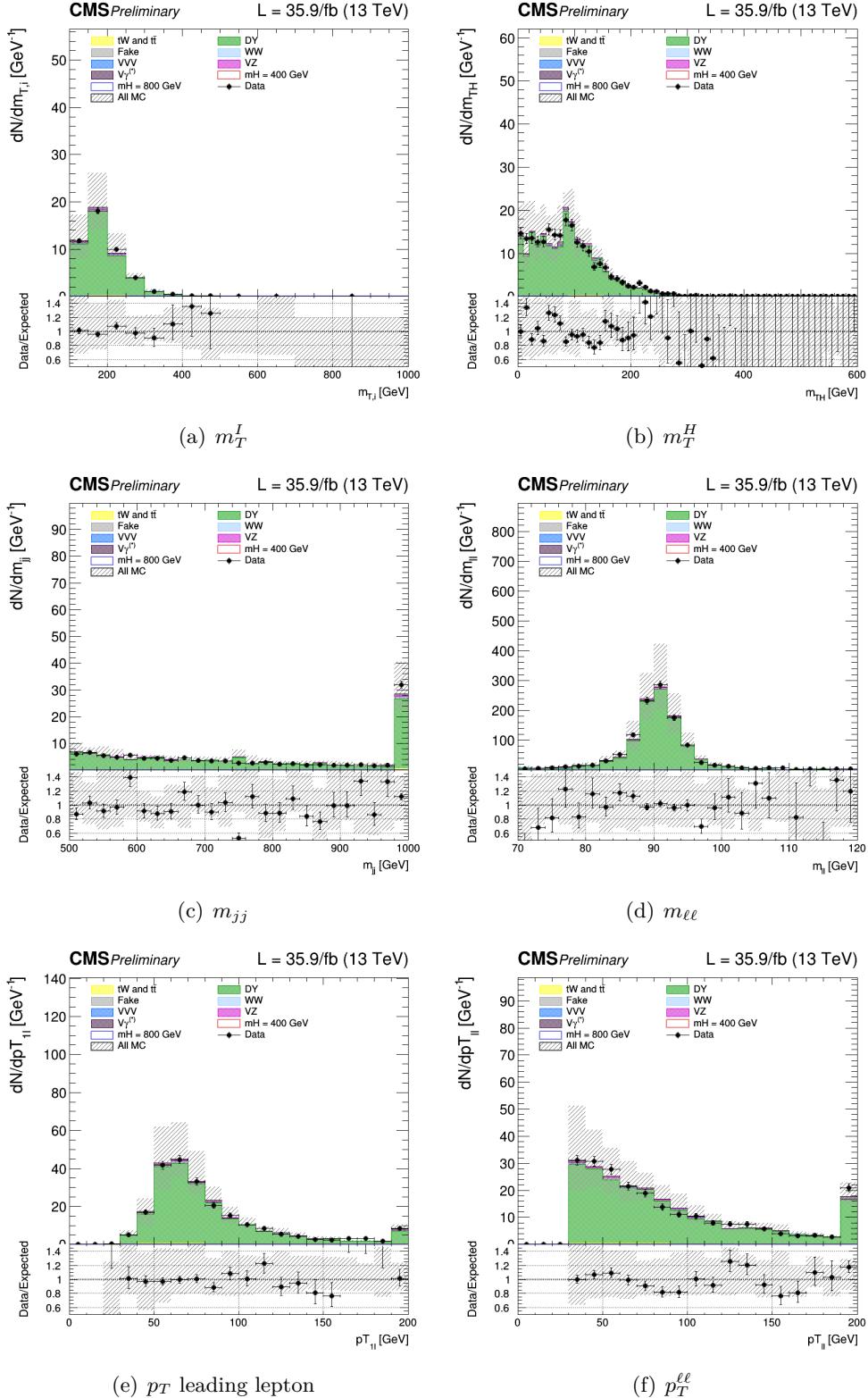


Figure 5.18. Control plots for several variables in a Drell-Yan enriched phase space for ee.

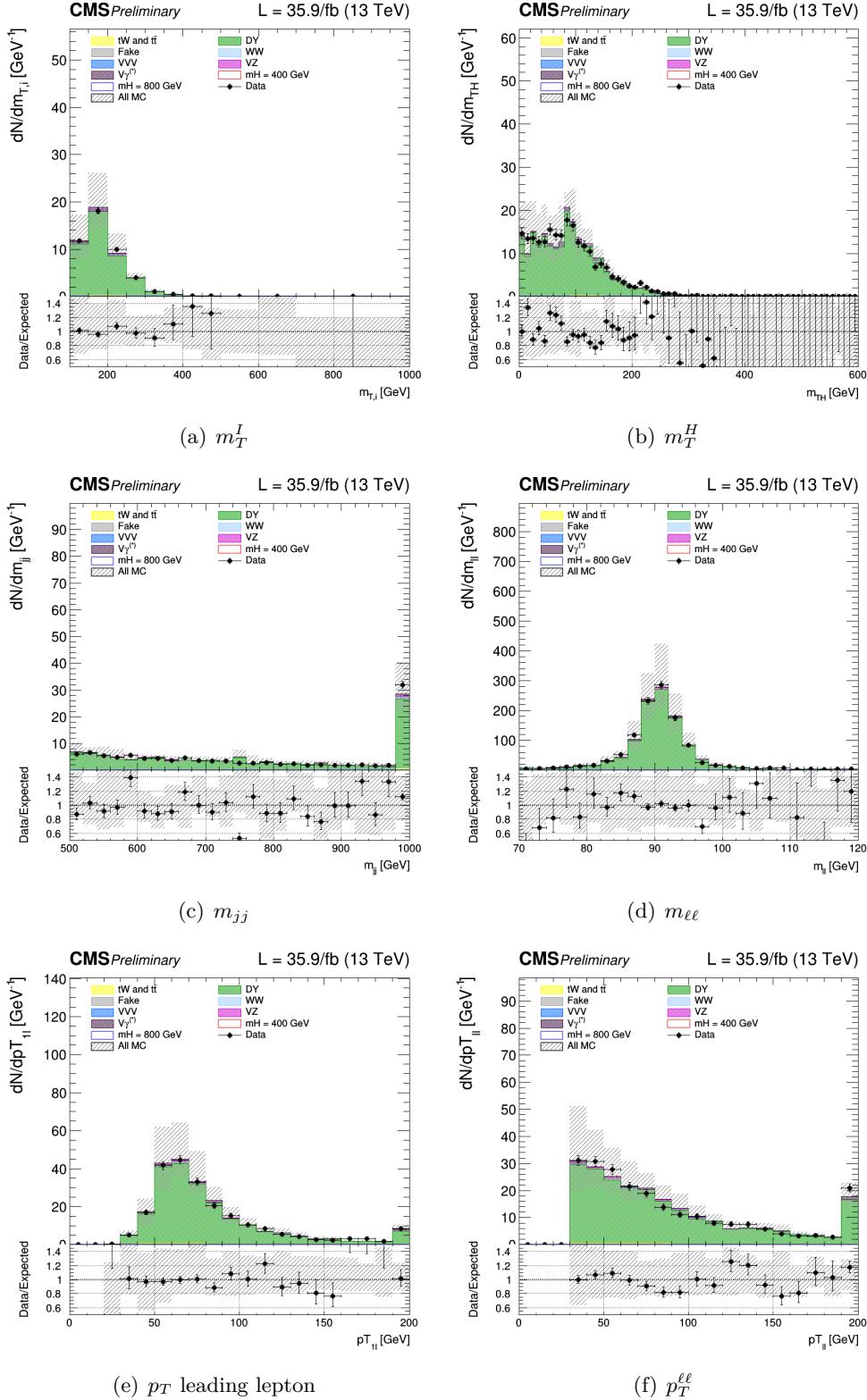


Figure 5.19. Control plots for several variables in a Drell-Yan enriched phase space for ee.

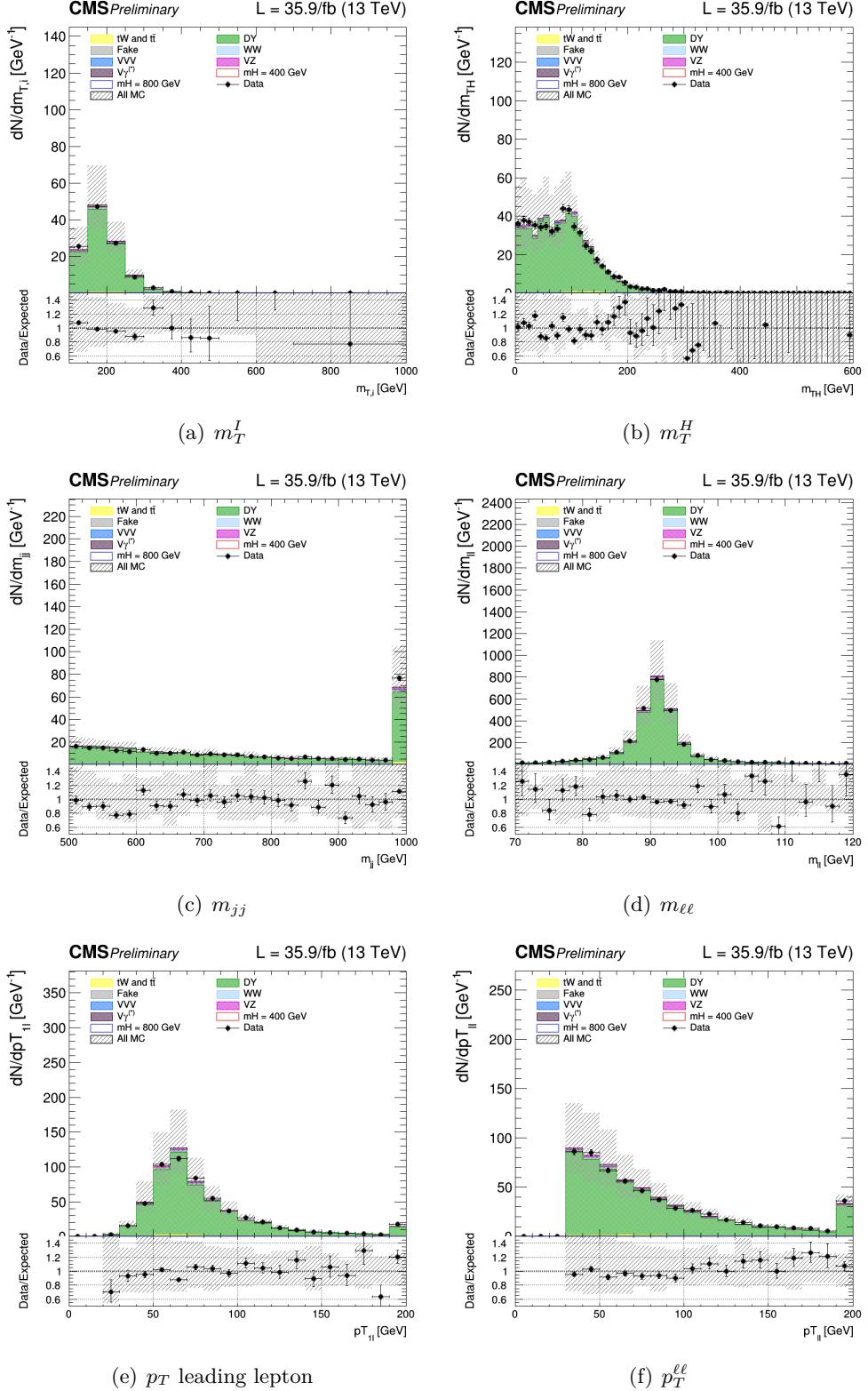


Figure 5.20. Control plots for several variables in a Drell-Yan enriched phase space for $\mu\mu$.

Top control region

A top-enriched control region is defined to normalize the top backgrounds, separately for electrons and muons. The “WW SF selection” is required with the inversion of the b-tagging requirement, i.e. the two jets are both requested to be b-tagged according to cMVAv2 loose WP. The control plots for several variables in a top enriched phase space for events are shown in the Figs. 5.21 for the dielectron case and 5.22 for the dimuon case. Good agreement is observed between data and MC.

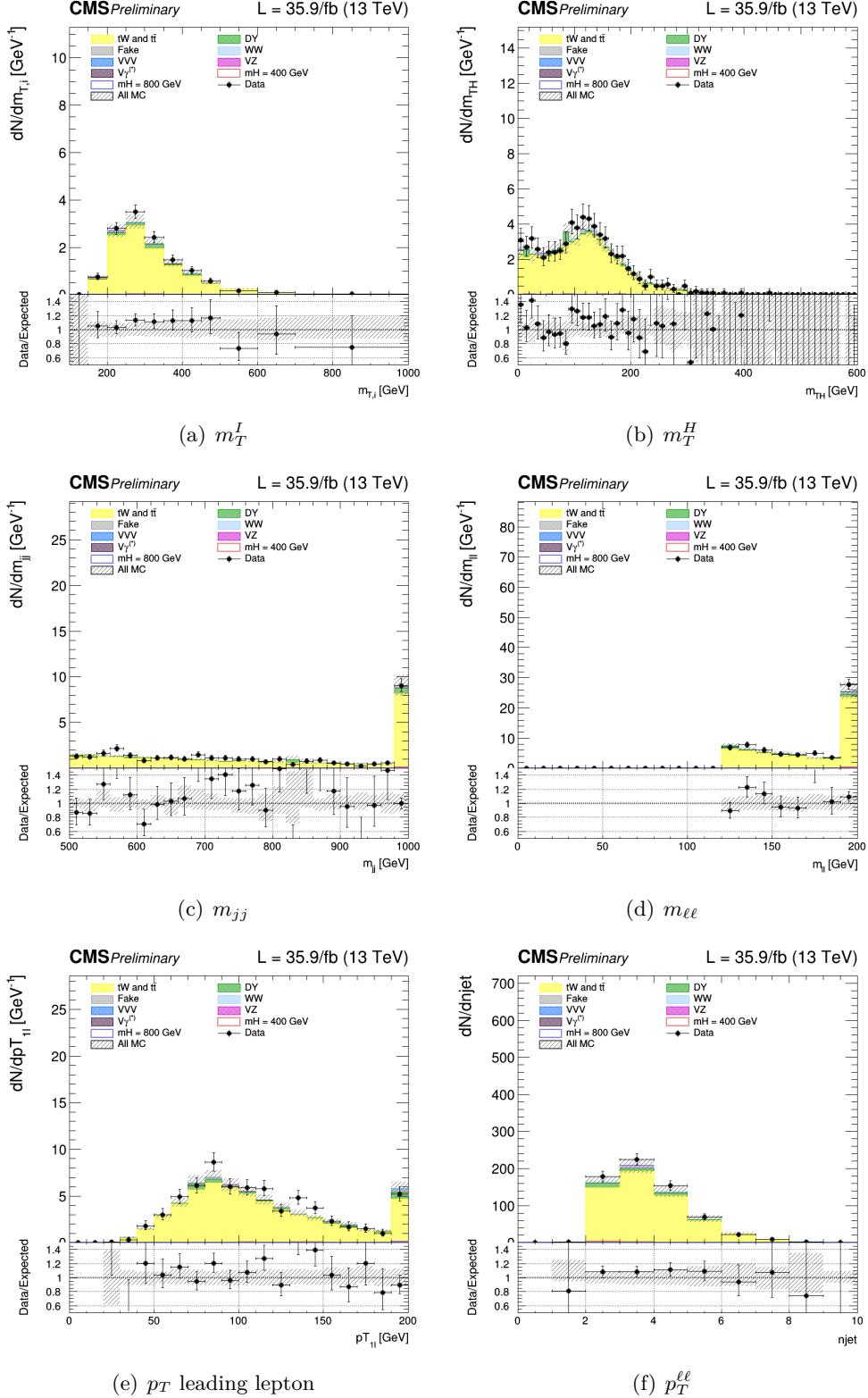


Figure 5.21. Control plots for several variables in a Top enriched phase space for ee.

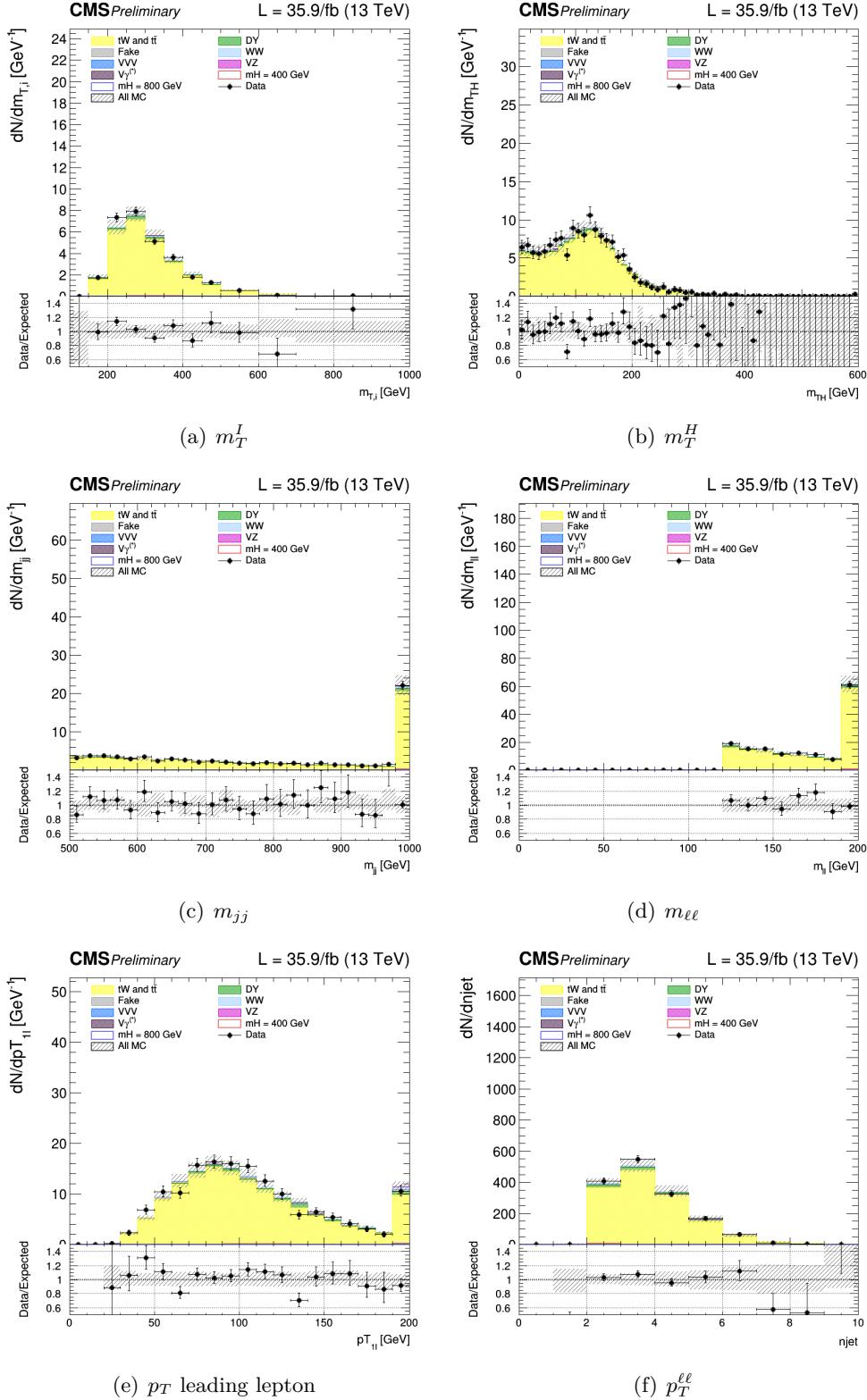


Figure 5.22. Control plots for several variables in a Top enriched phase space for $\mu\mu$.

5.9 Systematic uncertainties

Systematic uncertainties are introduced as nuisance parameters in the fit and can affect the normalization and the shape of the different contributions.

Statistical uncertainties from MC simulated events are taken into account. Systematic uncertainties are represented by individual nuisance parameters with log-normal or shape-based distributions. The uncertainties affect the overall normalization of the signal and backgrounds as well as the shape of the predictions across the distribution of the observables. Correlations between systematic uncertainties in different categories and final states are taken into account. Systematic uncertainties play an especially important role in this analysis where no strong mass peak is expected due to the presence of undetected neutrinos in the final state. Below we describe in detail sources and quantities of systematics in this analysis and their effects on the signal and background processes. A list of the most important background uncertainties is given below.

Background normalization uncertainties

One of the most important sources of systematic uncertainty is the normalization of the backgrounds that are estimated on data control samples whenever is possible. The signal extraction is performed subtracting the estimated backgrounds to the event counts in data. The amount of uncertainty depends on the considered background:

- jet-induced background: normalization and kinematic shapes are derived from a data control region and both normalization and shape systematic uncertainties are considered. A conservative 30% uncertainty on the fake rate is assumed correlated across the different analysis regions. The contribution to the uncertainty in the signal region due to the limited electron statistics in the background enriched control regions is about 10%, while the contribution due to the limited muon statistics 3%.
- WW background: The normalization of the WW background is performed independently in each jet multiplicity via the rateParam feature of combine. A WW electroweak (VBS) sample is used in addition to the standard WW sample in the phase spaces with at least two jets, where its contribution becomes non negligible. The uncertainty in the cross section for this process is evaluated using the variations of the renormalization and factorization QCD scales, as well as the PDF variations, and amounts to 11%.
- $t\bar{t}$ and tW backgrounds: Top events are estimated with b-tagging in data control regions. The two top background enriched control regions are defined as additional categories in the fit while the kinematic shapes are taken from the simulation corrected for the b-tagging discriminant scale factors. The top normalization is correlated between the top control region and the Higgs signal categories separately in the different jet multiplicities, and these normalizations are left unconstrained using the rateParam feature of combine.
- Drell-Yan background: The Drell-Yan background enters the different flavor analysis via the leptonic decays of the τ leptons from $Z\gamma^* \rightarrow \tau\tau$. In the

different flavor analyses the normalization of these background is controlled via the rateParam feature of combine and with a dedicated control region in each jet bin category.

- $W\gamma^*$ background: The kinematic shape of this background is predicted by simulation, normalized to its data-driven estimate, and constrained within the respective uncertainty, which is 25%.
- WZ : The kinematic shapes of this backgrounds are predicted by simulation and normalized to their theoretical predictions in the different and same flavour analysis.
- $Z\gamma^*$: The kinematic shapes of this backgrounds are predicted by simulation and normalized to their theoretical predictions in the different and same flavor analysis.
- ZZ: The kinematic shapes of this backgrounds are predicted by simulation and normalized to their theoretical predictions in the different and same flavor analysis.

Experimental uncertainties

Effects from experimental uncertainties are studied by applying a scaling and/or smearing of certain variables of the physics objects, followed by a subsequent recalculation of all the correlated variables. This is done for MC simulation, to account for possible systematic mismeasurements of the data. All experimental sources except luminosity are treated both as normalization and shape uncertainties. For background with a data-driven normalization estimation, the shape uncertainty is considered only. The following experimental systematic sources have been taken into account.

- Luminosity: The uncertainty determined by the CMS luminosity monitoring is 2.3% for 13 TeV data.
- Lepton trigger systematics: Lepton trigger systematics are of the order of less than 1%. These uncertainties are computed by varying the tag selection as well as the Z window in the tag and probe method used to compute the corresponding scale factors.
- Lepton reconstruction and identification efficiency: The lepton reconstruction and identification efficiencies are measured with the tag and probe method in data. To correct for the difference in the lepton identification efficiencies between data and MC, data/MC scale factors dependent on p_T and η are applied to the MC. The resulting uncertainty in the signal region is 1% for electrons and 2% for muon.
- Muon momentum and electron energy scale: Uncertainties on both the scale and resolution individually amount to 0.6-1% for electrons and 0.2% for muons.

- MET miss modelling: The MET miss measurement is affected by the possible mismeasurement of individual particles addressed above, as well as the additional contributions from the pile-up interactions. The effect of the missing transverse momentum resolution on the event selection is studied by propagating each component of the MET uncertainty to the absolute value and direction of MET.
- Jet energy scale (JES) uncertainties: We estimate this uncertainty applying the official jet uncertainties on the JES and compute the variation of the selection efficiency. JES uncertainty affects the rates in the signal region at the level of 10%.
- b-jet misidentification modelling: The uncertainties on the selection of non-b jets is taken into account by looking at the b-jet misidentification efficiency. The uncertainties on these scale factors need to be taken into account, and are of the order of a few percent.

Theoretical uncertainties

- PDF and higher-order corrections (renormalization and factorization scale): PDF uncertainties and the missing knowledge on higher-order corrections, evaluated by means of scale variation, directly affect the cross section, as well as the acceptance of a simulated process. The uncertainties that arise from using different PDF sets were obtained by reweighting events with different PDF sets.
- Underlying event and parton shower modelling: The underlying event (UE) and parton shower (PS) modelling uncertainties are estimated by comparing samples interfaced with different parton showers (Pythia vs Herwig) and UE tunes
- Single top tW and tt ratio: The ratio between the single top and top pair cross section is varied by the uncertainty on the ratio between their cross sections, calculated considering scale variations,
- A QCD and PDF scales for the signal samples at different masses. The uncertainties are taken from the Yellow Report 3 and the same values are used both for the large width hypothesis and for different values of C' . The effect of QCD and PDF scale uncertainties on the analysis selection has also been taken into account.
- The categorization of events based on jet multiplicity introduces additional uncertainties related to higher order corrections. These uncertainties are associated to the ggH production mode and are evaluated independently following the recipe described in [68] and are 5.6% for the 0-jet and 13% for the 1-jet and 20% for the 2-jet and VBF categories.

The top background shape is estimated from simulation and corrected using a data driven b-tagging scale factor. The normalization is measured in a top quark

enriched control region obtained inverting the b-veto requirement of the signal region. Three control regions are defined, one for each jet bin category. A nuisance parameter is added to take into account the effect of the parton shower uncertainty on the top background.

The DY background shape is also estimated from simulation and analogously to the Top background, the DY normalization is measured with a data driven technique in three control regions enriched in DY events.

A dedicated nuisances for MET reweighting in DY control region is introduced in SF analysis. It is evaluated separately for ee and $\mu\mu$ categories. The uncertainty is quoted as maximum and minimum best-fit lines of the linear fit.

Chapter 6

Results and Interpretation

6.1 Statistical interpretation

In the research of high mass Higgs boson, processes that have been predicted but not yet seen are searched . Given that no excesses over the SM expectation are seen in the mass spectra, the upper limits on the cross sections are computed.

The Bayesian and the classical frequentist [69], with a number of modifications, are two statistical approaches commonly used in high energy physics for characterising the absence of a signal. Both methods allow one to quantify the level of incompatibility of data with a signal hypothesis, which is expressed as a confidence level (C.L.) [70]. For excluding a signal the C.L. 95% is a common choice. The C.L. probabilistic interpretation is used when stating the non-existence of a signal is not straightforward and the subject of a vast body of literature as in the high mass analysis. The procedure used to establish the upper limits calculation is based on frequentist test using a likelihood ratio as a test statistic. In addition to the parameter of interest such as the cross section of the signal, the signal and the background models contain a nuisance parameters whose values are not taken in account as known *a priori* but rather must be fitted from the data [71]. In the following the frequentist approach is described. The expected high mass signal event yields will be generically denoted as s and the backgrounds as b .

The frequentist approach is built to discriminate signal from background events. The most powerful statistic test, in accordance to the Neyman-Pearson lemma [69], is the likelihoods ratio $\lambda(\mu)$,

$$\lambda(\mu) = \frac{\mathcal{L}(\text{data}|\mu s + b)}{\mathcal{L}(\text{data}|b)} \quad (6.1)$$

where, \mathcal{L} is the likelihood function from the product of Poisson probabilities and μ is the strength of the signal process (the case $\mu = 0$ corresponds to background only hypothesis, $\mu = 1$ the nominal signal hypothesis). One can see that $0 \leq \lambda(\mu) \leq 1$, μ near 1 is evidence of good agreement among data and the hypothesized μ value. It is convenient, for numerical reason, to use the test statistic q_μ defined as,

$$q_\mu = -2 \ln \lambda(\mu) \quad (6.2)$$

where high value of q_μ correspond to more likely incompatibility between data and μ , i.e. background only hypothesis.

Using the statistic test q_μ , is possible to quantify the level of disagreement between the data and the hypothesis, p -value, defined as,

$$p_\mu = \int_{q_\mu, obs}^{\infty} f(q_\mu | \mu) dq_\mu \quad (6.3)$$

where $q_{\mu, obs}$ is the value of statistic test q_μ observed from the data and $f(q_\mu | \mu)$ is the pdf of q_μ under the assumption of the signal strength μ .

The systematic uncertainties on signal $s(\theta)$ and background $b(\theta)$ rates are introduced in test statistic. The test statistic then would take the following form:

$$q_\mu = \frac{\mathcal{L}(data | \mu, \hat{\theta}_\mu)}{\mathcal{L}(data | 0, \hat{\theta}_0)}, \quad (6.4)$$

where $\hat{\theta}_\mu$ and $\hat{\theta}_0$ are maximum likelihood estimators for the signal+background hypothesis (with the signal strength factor μ) and for the background-only hypothesis ($\mu = 0$). The profile likelihood test statistic is introduced to prevent negative signal as,

$$\tilde{q}_\mu = \frac{\mathcal{L}(data | \mu, \hat{\theta}_\mu)}{\mathcal{L}(data | \hat{\mu}, \hat{\theta})}, \quad 0 \leq \hat{\mu} \leq \mu, \quad (6.5)$$

where $\hat{\mu}$ and $\hat{\theta}$ gives the global maximum of the likelihood. The constrain $0 \leq \hat{\mu}$ is due to a positive signal rate, while the $\hat{\mu} \leq \mu$ is imposed by hand in order to guarantee a one-sided confidence interval.

At this point is useful to evaluate the observed statistic test \tilde{q}_μ^{obs} and the nuisance parameters $\hat{\theta}_0^{obs}$, $\hat{\theta}_\mu^{obs}$ that escribing the experimentally observed data for the background-only and signal+background hypotheses, respectively. With this in mind, the pdf of the test statistic in constructed by generating toy MC pseudo-data for both the background-only and signal+background hypotheses, $f(\tilde{q}_\mu | \mu, \hat{\theta}_\mu^{obs})$ and $f(\tilde{q}_\mu | \mu, \hat{\theta}_0^{obs})$. The corresponding p -value for the signal+background and background-only hypotheses, p_μ and p_b are given by:

$$p_\mu = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs} | signal + background) = \int_{q_\mu, obs}^{\infty} f(\tilde{q}_\mu | \mu, \hat{\theta}_\mu^{obs}) d\tilde{q}_\mu \quad (6.6)$$

$$1 - p_b = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs} | background - only) = \int_{q_0, obs}^{\infty} f(\tilde{q}_\mu | 0, \hat{\theta}_0^{obs}) d\tilde{q}_\mu. \quad (6.7)$$

The $CL_s(\mu)$ is given by the ratio,

$$CL_s(\mu) = \frac{p_\mu}{1 - p_b} \quad (6.8)$$

To quote the 95% of confidence level upper limits on μ , μ is adjust until reaches $CL_S=0.05$. For the background-only hypothesis, the expected median upper-limit and $\pm 1\sigma$ and $\pm 2\sigma$ bands are generated with a large set of background-only pseudo-data. The CL_S is evaluated for each of them. Then, one can build a cumulative probability distribution of results by starting integration from the side corresponding

to low event yield. The point at which the cumulative probability distribution crosses the quantile of 50% is the median expected value. The $\pm 1\sigma$ (68%) band is defined by the crossings of the 16% and 84% quantiles. Crossings at 2.5% and 97.5% define the $\pm 2\sigma$ (95%) band.

In the high mass analysis, the interference contribution is not negligible, as described in 5.4, and it is included as part of the signal. In particular during the fit the interference term is scaled by $\sqrt{\mu}$. However, to prevent possible negative probability distribution function of the interference, during the fit the signal yield is computated as,

$$Yield = \sqrt{\mu} \times (S + B + I) + (\mu - \sqrt{\mu}) \times (S) + (1 - \sqrt{\mu}) \times (B) \quad (6.9)$$

where S is the signal, B the background and I the interference.

6.2 Limits and results

Electroweak singlet

The final binned fit is performed using the m_T^I histogram for all signals and the number of events for the backgrounds. For the oppiste-flavour and same-flavour analysis, for every categories and for every mass point from 200 GeV up to 3 TeV the significance and the 95% CL upper exclusion limit are calculated. The expected final limit from the combination of the OF and SF analysis are shown in Fig. 6.1. This limit represent a considerable improvement respect to the high mass search done with 2015 data and the expected limits is compatible with the ATLAS results for a similar analysis, Sec 1.3 .

2HDM results

In Fig. 6.2 the exclusion limits are shown for the m_h^{mod+} scenario on the left and the hMSSM scenario on the right. The dashed line marks the limit, while the green area shows the side of the limit that is excluded. The bands surrounding the limit indicate the $\pm 1, 2\sigma$ contours. For both scenarios the region at low values of m_A and $\tan \beta$ is excluded. These results complement well with the exclusion limit given by the MSSM $H \rightarrow \tau\tau$ analysis, where the sensitivity is lower for low m_A and $\tan \beta$.

In Fig. 6.3 and 6.4 the exclusion limits are shown for a 2HDM. The limits in 6.3 for both a type-1 and type-2 2HDM is displayed in a $\cos(\beta - \alpha)$ - $\tan \beta$ plane, in which the neutral heavy Higgs boson masses are $m_H = m_A = 200, 300, 500$ GeV and the convention $\sin(\beta - \alpha) > 0$ is used. The plots in Fig. 6.4 show the limit in the m_H - $\tan \beta$ plane. Here it is again assumed that $m_H = m_A$ and $\sin(\beta - \alpha) > 0$, but here the relationship between β and α is $\cos(\beta - \alpha) = 0.1$. The exclusion limits seen here are larger compared to those produced in the similar analysis by ATLAS. A possible reason may be the choice of the discriminating variable m_T^I or the different categorization.

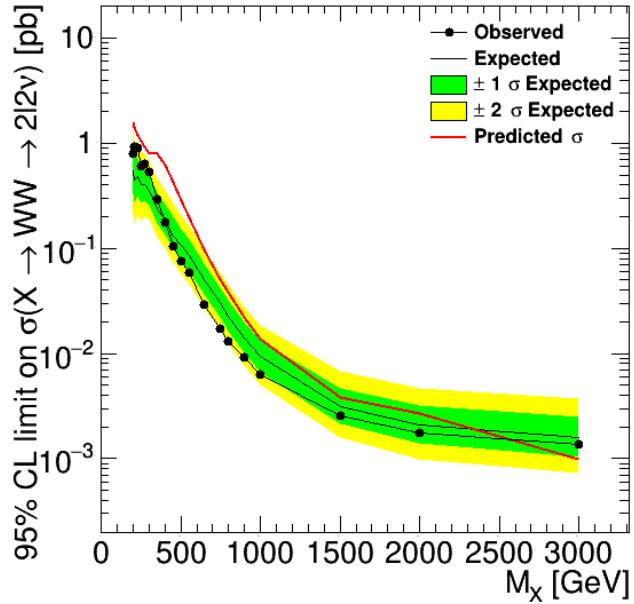


Figure 6.1. 95% CL exclusion limits, on the production ggH and VBF cross section times branching fraction as a function of the mass for the combination of the two analysis OF and SF, in the full mass range. The red line represent the predicted cross-section for EW high mass bosons.

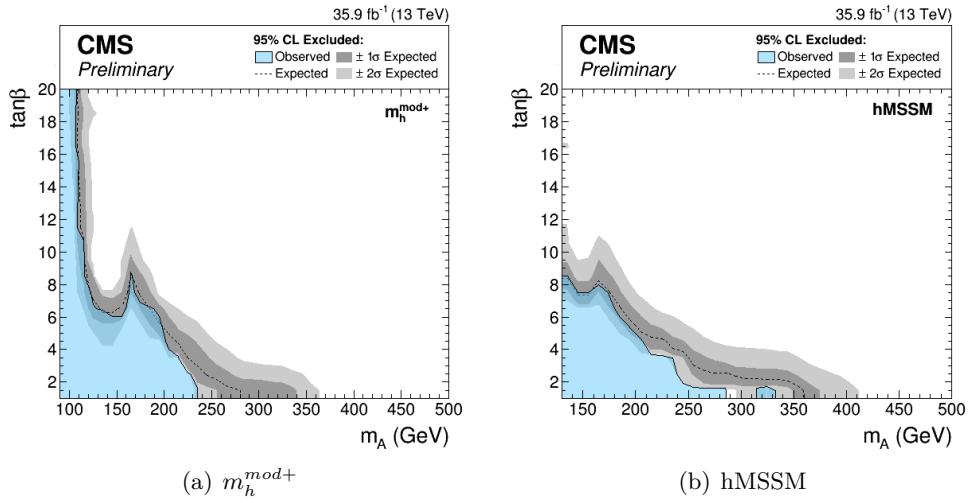


Figure 6.2. 95% CL exclusion limits on the MSSM m_h^{mod+} scenario (left) and the hMSSM scenario (right).

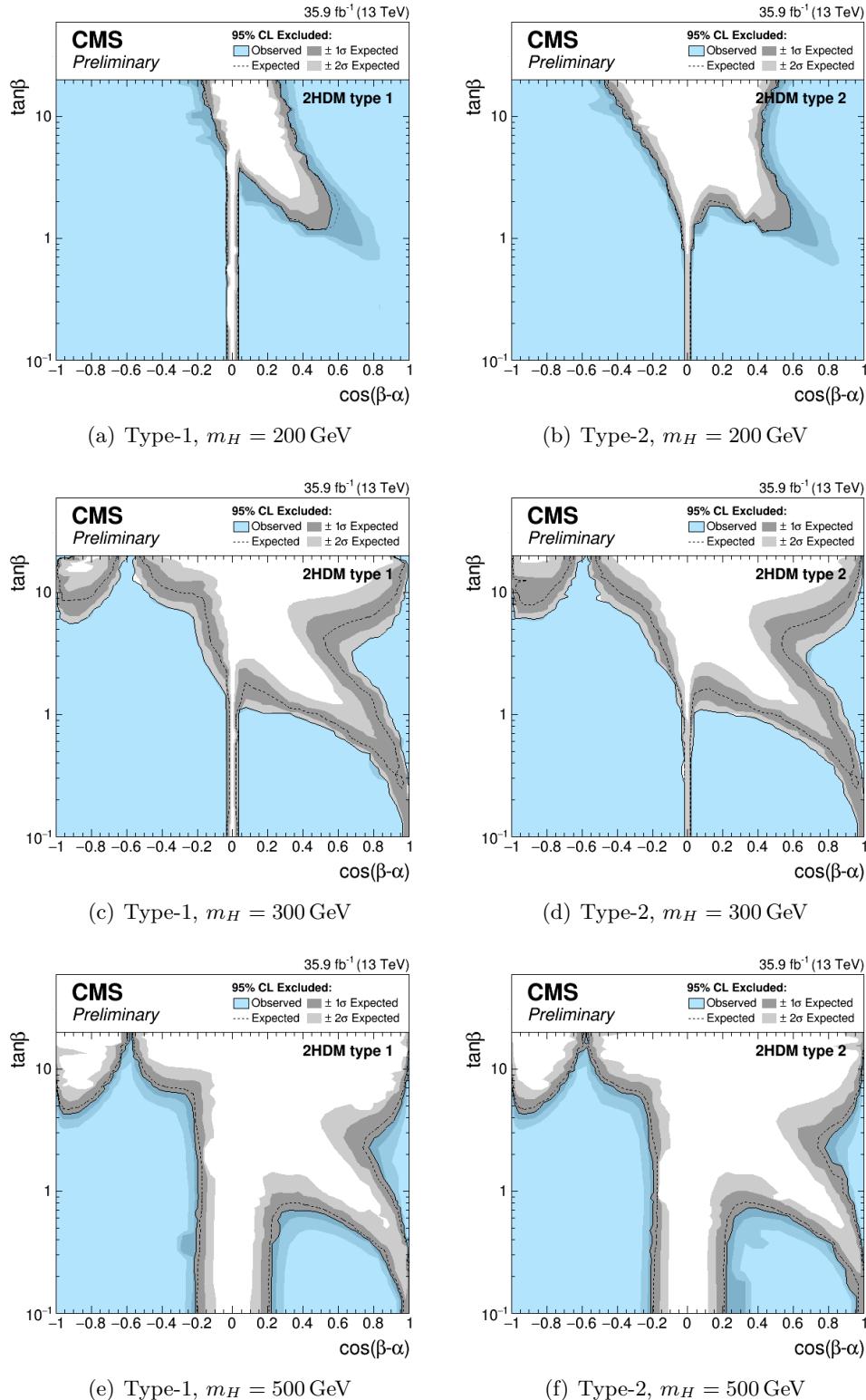


Figure 6.3. 95% CL exclusion limits on a 2HDM with $\cos(\beta - \alpha)$ on the x-axis. Limits are shown for a type-1 and type-2 2HDM for different masses $m_H = 200, 300, 500$ GeV.

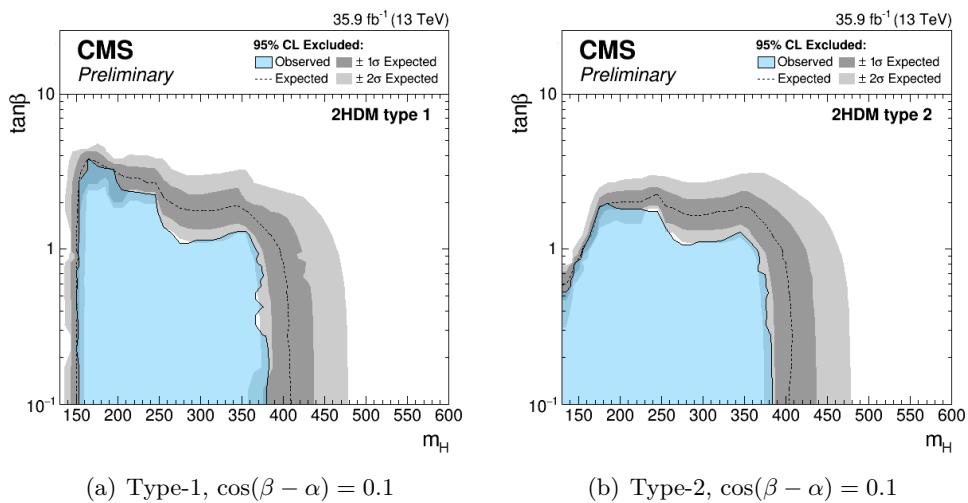


Figure 6.4. 95% CL exclusion limits on a 2HDM with m_H on the x-axis. Limits are shown for a type-1 and type-2 2HDM for $\cos(\beta - \alpha) = 0.1$.

Appendix A

Boost-invariant variables

Since the center-of-mass of the parton-parton scattering is normally boosted in the beam direction with respect to that of the two incoming hadrons, it is therefore useful to classify the final state in terms of variables that are invariant under Lorentz transformations in that direction. For this purpose, the variables rapidity y , transverse momentum p_T , and azimuth angle ϕ are introduced. In term of these variables, the four-momentum of a particle of mass m may be written as,

$$p^\mu = (E, p_x, p_y, p_z) \quad (\text{A.1})$$

where where p_x , p_y and p_z are the Cartesian coordinates of the momentum \vec{p} . The rapidity y is defined by the relation,

$$y = \frac{1}{2} \ln\left(\frac{E + p_z}{E - p_z}\right) \quad (\text{A.2})$$

and is not invariant for relativistic transformations but since it transforms according to the law $y' = y + \beta$, (where β is the relative velocity between two frames), the rapidity differences Δy are boost invariant. In the approximation of ultra-relativistic particles rapidity can be approximated by the pseudorapidity η , defined as:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \sim y \quad (\text{A.3})$$

where θ is the angle between the particle and the beam direction, and is therefore directly measurable in the detector.

Appendix B

Electron Efficiencies from Tag and Probe Method

One of the well established data-driven approach for measuring the particle efficiencies is the so called Tag and Probe method. The Tag and probe method uses a known mass resonance (e.g. J/Ψ , Z) to select particles of the desired type, and probe the efficiency of a particular selection criterion on these particles. In general the “tag” is an object that passes a set of very tight selection criteria designed to isolate the required particle type. Tags are often referred to as a “golden” electrons or muons and the fake rate for passing tag selection criteria should be very small. A generic set of the desired particle type (i.e. with potentially very loose selection criteria) known as “probes” is selected by pairing these objects with tags such that the invariant mass of the combination is consistent with the mass of the resonance. Combinatorial backgrounds may be eliminated through any of a variety of background subtraction methods such as fitting, or sideband subtraction. The definition of the probe objects depend on the specifics of the selection criterion being examined. The simple expression to get the efficiency as a function of p_T and η is given below:

$$\epsilon = \frac{N_{Pass}^{Probes}}{N_{Pass}^{Probes} + N_{Fail}^{Probes}} \quad (B.1)$$

This method is used here to get the Identification efficiency of electrons. In this case, the Tag is a well identified and isolated electron which also made to pass an electron Trigger to increase the purity. Once the Tag electron is selected then we look for another object which is reconstructed as electron. After passing some kinematical selection, we form the total probe selection. The invariant mass of the pair is then reconstructed from Tag and Probe electron and a Z boson mass window is imposed. After that, we ask the Probe to pass the Id working point to be checked and compute the efficiency. This is done for both data and MC. Once we get the efficiency for both data and MC then we compute the scale factors, which are the ratio of efficiencies of Data and MC. These scale factors are used then in the analysis to scale the MC to correct the difference in efficiencies between data and MC. The Pile-Up reweighting is also applied on MC during the computation of efficiencies. A MC truth matching has also been applied in case of the computation using simulation. There are basically two methods to estimate the efficiencies. One

is the Counting Method and the other is the Fitting Method. The counting method is used when there is less background. We are computing efficiencies with loose Z mass window and hence there is a possibility of background contamination under the Z peak. Hence, proper handling of background becomes important. In order to take into account the effect of background, the Fitting method has been used to estimate the efficiency.

Bibliography

- [1] F. Halzen and Alan D. Martin. *QUARKS AND LEPTONS: AN INTRODUCTORY COURSE IN MODERN PARTICLE PHYSICS*. 1984.
- [2] Sheldon L. Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22(4):579 – 588, 1961.
- [3] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967.
- [4] Abdus Salam. Weak and Electromagnetic Interactions. *Conf. Proc.*, C680519:367–377, 1968.
- [5] A. B. Balantekin and W. C. Haxton. Neutrino Oscillations. *Prog. Part. Nucl. Phys.*, 71:150–161, 2013.
- [6] Particle Data Group, J. Beringer, et al. Review of Particle Physics. *Phys. Rev. D*, 86:010001, 2012.
- [7] Luigi Di Lella and Carlo Rubbia. *The Discovery of the W and Z Particles*, pages 137–163.
- [8] et alt. Abe. Evidence for top quark production in $p^- p$ collisions at $\sqrt{s} = 1.8$ tev. *Phys. Rev. Lett.*, 73:225–231, Jul 1994.
- [9] S. Abachi et al. Observation of the top quark. *Phys. Rev. Lett.*, 74:2632–2637, 1995.
- [10] P.W. Higgs. Broken symmetries, massless particles and gauge fields. *Physics Letters*, 12(2):132 – 133, 1964.
- [11] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.
- [12] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29, 2012.
- [13] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett.*, B716:30–61, 2012.

- [14] Timo van Ritbergen and Robin G. Stuart. On the precise determination of the Fermi coupling constant from the muon lifetime. *Nucl. Phys.*, B564:343–390, 2000.
- [15] J R Andersen et al. Handbook of LHC Higgs Cross Sections: 3. Higgs Properties. 2013.
- [16] Tania Robens and Tim Stefaniak. Status of the Higgs Singlet Extension of the Standard Model after LHC Run 1. *Eur. Phys. J.*, C75:104, 2015.
- [17] Robert M. Schabinger and James D. Wells. A Minimal spontaneously broken hidden sector and its impact on Higgs boson physics at the large hadron collider. *Phys. Rev.*, D72:093007, 2005.
- [18] A. Barroso, P. M. Ferreira, Rui Santos, Marc Sher, and Joao P. Silva. 2HDM at the LHC - the story so far. 2013.
- [19] V. Barger, J. L. Hewett, and R. J. N. Phillips. New constraints on the charged higgs sector in two-higgs-doublet models. *Phys. Rev. D*, 41:3421–3441, Jun 1990.
- [20] Mayumi Aoki, Shinya Kanemura, Koji Tsumura, and Kei Yagyu. Models of yukawa interaction in the two higgs doublet model, and their collider phenomenology. *Phys. Rev. D*, 80:015017, Jul 2009.
- [21] Abdelhak Djouadi. Higgs physics at future colliders: Recent theoretical developments. *Pramana*, 62(2):191–206, Feb 2004.
- [22] A. Denner, S. Heinemeyer, I. Puljak, D. Rebuzzi, and M. Spira. Standard Model Higgs-Boson Branching Ratios with Uncertainties. *Eur. Phys. J.*, C71:1753, 2011.
- [23] Konstantinos A. Petridis. High Mass Standard Model Higgs searches at the Tevatron. *EPJ Web Conf.*, 28:07002, 2012.
- [24] Vardan Khachatryan et al. Search for a Higgs boson in the mass range from 145 to 1000 GeV decaying to a pair of W or Z bosons. *JHEP*, 10:144, 2015.
- [25] Search for high mass Higgs to WW with fully leptonic decays using 2015 data. Technical Report CMS-PAS-HIG-16-023, CERN, Geneva, 2016.
- [26] Search for a high-mass Higgs boson decaying to a pair of W bosons in pp collisions at $\sqrt{s}=13$ TeV with the ATLAS detector. Technical Report ATLAS-CONF-2016-074, CERN, Geneva, Aug 2016.
- [27] Thomas Sven Pettersson and P Lefèvre. The Large Hadron Collider: conceptual design. Technical Report CERN-AC-95-05-LHC, Oct 1995.
- [28] T. Sjostrand. Monte Carlo generators. 2006.
- [29] A. Buckley et al. General-purpose event generators for LHC physics. *Phys. Rept.*, 504:145–233, 2011.

- [30] Richard D. Ball, Valerio Bertone, Stefano Carrazza, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Nathan P. Hartland, and Juan Rojo. Parton distributions with QED corrections. *Nucl. Phys.*, B877:290–320, 2013.
- [31] Richard D. Ball, Valerio Bertone, Francesco Cerutti, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Jose I. Latorre, Juan Rojo, and Maria Ubiali. Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO. *Nucl. Phys.*, B855:153–221, 2012.
- [32] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008.
- [33] Vardan Khachatryan et al. Event generator tunes obtained from underlying event and multiparton scattering measurements. 2015.
- [34] S. Agostinelli et al. GEANT4—a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250, 2003.
- [35] 2015 pileup json files. https://twiki.cern.ch/twiki/bin/view/CMS/PileupJSONFileforData#2015_Pileup_JSON_Files.
- [36] Paolo Nason. A New method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004.
- [37] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 11:070, 2007.
- [38] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010.
- [39] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. NLO Higgs boson production via gluon fusion matched with shower in POWHEG. *JHEP*, 04:002, 2009.
- [40] Paolo Nason and Carlo Oleari. NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG. *JHEP*, 02:037, 2010.
- [41] A. V. Gritsan *et. al.* S. Bolognesi, Y. Gao. Jhugen. <http://www.pha.jhu.edu/spin/>.
- [42] Gionata Luisoni, Paolo Nason, Carlo Oleari, and Francesco Tramontano. $HW^\pm/HZ + 0$ and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO. *JHEP*, 10:083, 2013.
- [43] Sm higgs production cross sections at $\sqrt{s} = 13\text{--}14$ tev. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt1314TeV>.
- [44] Tom Melia, Paolo Nason, Raoul Rontsch, and Giulia Zanderighi. W+W-, WZ and ZZ production in the POWHEG BOX. *JHEP*, 11:078, 2011.

- [45] John M. Campbell, R. Keith Ellis, and Ciaran Williams. Bounding the Higgs width at the LHC: Complementary results from $H \rightarrow WW$. *Phys. Rev.*, D89(5):053011, 2014.
- [46] Patrick Meade, Harikrishnan Ramani, and Mao Zeng. Transverse momentum resummation effects in W^+W^- measurements. *Phys. Rev.*, D90(11):114006, 2014.
- [47] Prerit Jaiswal and Takemichi Okui. Explanation of the WW excess at the LHC by jet-veto resummation. *Phys. Rev.*, D90(7):073009, 2014.
- [48] Nlo single-top channel cross sections. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec>.
- [49] Nnlo+nnll top-quark-pair cross sections. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO>.
- [50] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.
- [51] How to compute cross sections with the genxsecanalyzer. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/HowToGenXSecAnalyzer>.
- [52] Summary table of samples produced for the 1 billion campaign, with 25ns bunch-crossing. <https://twiki.cern.ch/twiki/bin/view/CMS/SummaryTable1G25ns>.
- [53] Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. Technical Report CMS-PAS-PFT-09-001, CERN, Geneva, Apr 2009.
- [54] Stefania Gori, Ian-Woo Kim, Nausheen R. Shah, and Kathryn M. Zurek. Closing the Wedge: Search Strategies for Extended Higgs Sectors with Heavy Flavor Final States. *Phys. Rev. D*, 93(WSU-HEP-1508. WSU-HEP-1508):075038. 30 p, Feb 2016. Comments: 30 pages, 10 figures.
- [55] Lhc hxswg for bsm higgs (wg3). <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG3>.
- [56] Robert V. Harlander, Stefan Liebler, and Hendrik Mantler. SusHi: A program for the calculation of Higgs production in gluon fusion and bottom-quark annihilation in the Standard Model and the MSSM. *Comput. Phys. Commun.*, 184:1605–1617, 2013.
- [57] S. Heinemeyer, W. Hollik, and G. Weiglein. FeynHiggs: A Program for the calculation of the masses of the neutral CP even Higgs bosons in the MSSM. *Comput. Phys. Commun.*, 124:76–89, 2000.
- [58] S. Heinemeyer, W. Hollik, and G. Weiglein. The Masses of the neutral CP - even Higgs bosons in the MSSM: Accurate analysis at the two loop level. *Eur. Phys. J.*, C9:343–366, 1999.

- [59] G. Degrassi, S. Heinemeyer, W. Hollik, P. Slavich, and G. Weiglein. Towards high precision predictions for the MSSM Higgs sector. *Eur. Phys. J.*, C28:133–143, 2003.
- [60] M. Frank, T. Hahn, S. Heinemeyer, W. Hollik, H. Rzehak, and G. Weiglein. The Higgs Boson Masses and Mixings of the Complex MSSM in the Feynman-Diagrammatic Approach. *JHEP*, 02:047, 2007.
- [61] T. Hahn, S. Heinemeyer, W. Hollik, H. Rzehak, and G. Weiglein. High-Precision Predictions for the Light CP -Even Higgs Boson Mass of the Minimal Supersymmetric Standard Model. *Phys. Rev. Lett.*, 112(14):141801, 2014.
- [62] A. Djouadi, J. Kalinowski, and M. Spira. HDECAY: A Program for Higgs boson decays in the standard model and its supersymmetric extension. *Comput. Phys. Commun.*, 108:56–74, 1998.
- [63] A. Djouadi, M. M. Muhlleitner, and M. Spira. Decays of supersymmetric particles: The Program SUSY-HIT (SUspect-SdecaY-Hdecay-InTerface). *Acta Phys. Polon.*, B38:635–644, 2007.
- [64] Johan Rathsman and Oscar Stal. 2HDMC - A Two Higgs Doublet Model Calculator. *PoS*, CHARGED2010:034, 2010.
- [65] Bsm higgs production cross sections at $\sqrt{s} = 13$ tev. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBSMAt13TeV>.
- [66] John Ellis and Dae Sung Hwang. Does the ‘Higgs’ have Spin Zero? *JHEP*, 09:071, 2012.
- [67]
- [68] Radja Boughezal, Xiaohui Liu, Frank Petriello, Frank J. Tackmann, and Jonathan R. Walsh. Combining Resummed Higgs Predictions Across Jet Bins. *Phys. Rev.*, D89(7):074044, 2014.
- [69] G. Cowan. Statistical data analysis. 1998.
- [70] Procedure for the LHC Higgs boson search combination in Summer 2011. (CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11), Aug 2011.
- [71] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J.*, C71:1554, 2011. [Erratum: Eur. Phys. J.C73,2501(2013)].